EPA/841-B-96-004
September 1997

MONITORING GUIDANCE FOR DETERMINING THE EFFECTIVENESS OF NONPOINT SOURCE CONTROLS

FINAL

U.S. Environmental Protection Agency
Nonpoint Source Control Branch
401 M Street, S.W.
Washington, DC 20460

September 1997

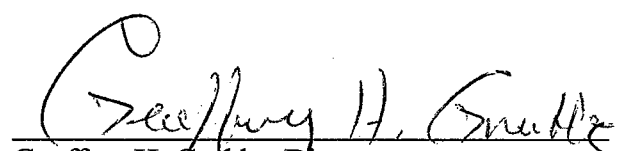
FOREWORD

The diffuse nature of nonpoint sources (e.g., agriculture, forestry, urban areas) and the variety of pollutants generated by them create a challenge for their effective control. Although progress has been made in the protection and enhancement of water quality, much work is still needed to identify nonpoint source management strategies that are both effective and economically achievable under a wide range of conditions. Monitoring will play an important role in this effort.

This nonpoint source monitoring and evaluation guide is written for use by both those who monitor and those who evaluate and fund monitoring proposals. For example, the federal, State, and Tribal agencies that support monitoring activities might use the guide to assess the technical merits of proposed monitoring and evaluation plans. These same agencies, university personnel, and others that carry out the monitoring and evaluation might use this guide to formulate their plans.

This guidance addresses the design of water quality monitoring programs to assess both impacts from nonpoint source pollution and the effectiveness of control practices and management measures. There are diverse opinions regarding the most effective way to design a monitoring program. Since each situation is different and may need a unique monitoring approach, this guidance presents the theory and information needed to design monitoring programs tailored to particular situations.

Readers are encouraged to consult, as well, the additional resources listed in this document. For example, companion documents, *Techniques for Tracking, Evaluating, and Reporting the Implementation of Nonpoint Source Control Measures: I. Agriculture, II. Forestry, and III. Urban* (USEPA, 1997), present guidance on determining the extent of and trends in implementation of nonpoint source control practices and management measures. In addition to these resources, readers are urged to contact monitoring and quality assurance experts in academia and at the local, State, Tribal, and federal levels for assistance in developing monitoring plans.



Geoffrey H. Grubbs, Director
Assessment and Watershed Protection Division

ACKNOWLEDGMENTS

Mr. Steven A. Dressing of the Assessment and Watershed Protection Division (AWPD) in USEPA's Office of Water was the principal author of the 1988 draft document and managed the development of the final guidance. Contractual support for the final document was provided by Tetra Tech, Inc. under EPA contract 68-C3-0303. Dr. Jon Harcum wrote much of the text for Chapters 2 and 4. Drs. Michael Barbour, Mohammed Lahlou, and Sam Stribling and Mr. Sam Pett prepared Chapter 3, and Dr. Esther Peters and Mr. Pett prepared Chapter 5. Mr. George Townsend coordinated Tetra Tech's efforts, and Ms. Marti Martin edited the document.

Special thanks go to the following reviewers for their insights, technical assistance, and critical review of various drafts of the guidance:

Donald W. Meals of the Vermont Agency of Natural Resources
Dr. R. Peter Richards of Heidelberg College, Ohio
Thomas E. Davenport of USEPA, Region 5
Patricia L. Lietman of US Geological Survey, Harrisburg, Pennsylvania
Gretchen Hayslip of USEPA, Region 10
Dr. Jean D. Spooner of North Carolina State University
Dr. Deanna L. Osmond of North Carolina State University
Dr. Benno P. Warkentin of Oregon State University
Jeanne Goodman of South Dakota Dept. of Environment & Natural Resources
Steven W. Coffey of North Carolina State University
Martin Brossman of USEPA, Headquarters
Donald Martin of USEPA, Region 10
Charles Kanetsky of USEPA, Region 3
Chris Faulkner of USEPA, Headquarters
Keith R. Seiders of Washington Department of Ecology
Dr. John C. Clausen of the University of Connecticut
William H. Clark of Idaho Division of Environmental Quality
Dr. Ed Liu of USEPA, Region 9
Dr. Paul D. Robillard of Penn State University
Dr. Richard P. Maas of the University of North Carolina-Asheville
Hoke Howard of USEPA, Region 4
Annie McLeod of South Carolina Dept. of Health and Environmental Control
Ralph Reznick of Michigan Dept. of Environmental Quality
Robert Alpern of New York City Department of Environmental Protection
Dr. James L. Plafkin of USEPA, Headquarters (posthumously)

and many others who reviewed earlier drafts over the past several years. Much appreciation is due to the many water quality experts whose field and theoretical work has laid the foundation for this and related monitoring guides.

CONTENTS

PAGE

List of Tables v

List of Figures vii

Credits ix

Glossary xi

1. Overview of the Nonpoint Source Problem
   - 1.1 Definition of a Nonpoint Source 1-1
   - 1.2 Extent of Nonpoint Source Problems in the United States 1-1
   - 1.3 Effects of Nonpoint Source Pollutants 1-4
   - 1.4 Major Categories of Nonpoint Source Pollution 1-6
     - 1.4.1 Agriculture 1-6
     - 1.4.2 Urban Sources 1-7
     - 1.4.3 Removal of Streamside Vegetation 1-7
     - 1.4.4 Hydromodification 1-8
     - 1.4.5 Mining 1-9
     - 1.4.6 Forest Harvesting 1-10
     - 1.4.7 Construction 1-10
     - 1.4.8 Marinas 1-11
   - 1.5 Water Resource Considerations 1-12
     - 1.5.1 Rivers and Streams 1-12
     - 1.5.2 Lakes, Reservoirs, and Ponds 1-13
     - 1.5.3 Estuaries 1-16
     - 1.5.4 Open Coastal Waters 1-18
     - 1.5.5 Ground Water 1-19
   - 1.6 Climate 1-20
   - 1.7 Soils, Geology, and Topography 1-21

2. Developing a Monitoring Plan
   - 2.1 Introduction 2-1
   - 2.2 Monitoring Objectives 2-4
     - 2.2.1 Monitoring Objective Category: Problem Definition 2-5
     - 2.2.2 Monitoring Objective Category: Model Development 2-6
     - 2.2.3 Monitoring Objective Category: Evaluation 2-6
     - 2.2.4 Monitoring Objective Category: Conduct Research 2-6
   - 2.3 Data Analysis and Presentation Plans 2-7
   - 2.4 Variable Selection 2-7
     - 2.4.1 Physical and Chemical Water Quality Data 2-10
     - 2.4.2 Biological Data 2-10
     - 2.4.3 Precipitation Data 2-10

2.4.4 Land Use Data 2-11
2.4.5 Topographic Data 2-12
2.4.6 Soil Characteristic Data 2-12
2.5 Program Design 2-12
2.5.1 Probabilistic Designs 2-14
2.5.2 Targeted Site Location Study Designs 2-21
2.6 Example Program Design 2-25
2.7 Roles and Responsibilities 2-27
2.8 Quality Assurance Project Planning 2-29
2.9 Chemical and Physical Monitoring 2-29
2.10 Recommended References 2-30

3. Biological Monitoring of Aquatic Communities
3.1 Introduction 3-1
3.1.1 Rationale and Strengths of Biological Assessment 3-2
3.1.2 Limitations of Biological Assessment 3-3
3.2 Habitat Assessment 3-4
3.3 Overview of Biological Assessment Approaches 3-6
3.3.1 Screening-Level or Reconnaissance Bioassessment 3-6
3.3.2 Paired-Site Approach 3-7
3.3.3 Composited Reference Site Bioassessment 3-10
3.4 Reference Sites and Conditions 3-11
3.5 Rapid Bioassessment Protocols 3-16
3.6 The Multimetric Approach for Biological Assessment 3-18
3.7 Sampling Considerations 3-20
3.7.1 Benthic Macroinvertebrate Sampling 3-21
3.7.2 Fish Sampling 3-24
3.8 Biomonitoring Program Design 3-26
3.8.1 Process of Randomized Sampling Site Selection 3-29
3.8.2 Targeted Site Selection 3-32
3.8.3 Integrated Network Design 3-32
3.9 Monitoring Trends in Biological Conditions 3-34
3.10 Overview of Some State Programs 3-38

4. Data Analysis
4.1 Introduction 4-1
4.1.1 Estimation and Hypothesis Testing 4-1
4.1.2 Characteristics of Environmental Data 4-4
4.1.3 Recommendations for Selecting Statistical Methods 4-4
4.1.4 Data Stratification 4-8
4.1.5 Recommended Reading List and Available Software 4-9
4.2 Summary (Descriptive) Statistics 4-10
4.2.1 Point Estimation 4-10
4.2.2 Interval Estimation 4-18
4.3 Graphical Data Display 4-20

4.4 Evaluation of Test Assumptions 4-25
4.4.1 Tests of Normality 4-26
4.4.2 Tests of Equal Variance 4-31
4.4.3 Tests of Randomness 4-33
4.5 Evaluation of One or Two Independent Random Samples 4-33
4.5.1 Tests for One Sample or Paired Data 4-34
4.5.2 Two-sample Tests 4-47
4.5.3 Magnitude of Differences 4-51
4.6 Comparison of More Than Two Independent Random Samples 4-52
4.6.1 One-Factor Comparisons 4-53
4.6.2 Two-Factor Comparisons 4-58
4.6.3 Matched Data 4-61
4.6.4 Multiple Comparisons 4-63
4.7 Regression Techniques 4-64
4.7.1 Overview 4-64
4.7.2 Simple Linear Regression 4-65
4.7.3 Nonlinear Regression and Transformations 4-74
4.7.4 Multiple Regression 4-75
4.7.5 Multivariate Regression 4-78
4.8 Analysis of Covariance 4-79
4.9 Evaluation of Time Series 4-85
4.9.1 Monotonic Trends 4-86
4.9.2 Correlation Coefficients 4-90
4.10 Multivariate Analysis 4-91
4.10.1 Canonical Correlation 4-92
4.10.2 Cluster Analysis 4-93
4.10.3 Principal Components and Factor Analysis 4-93
4.10.4 Discriminant Analysis 4-94
4.11 Extreme Events 4-94
4.11.1 Rainfall Analyses 4-95
4.11.2 Design Flows 4-101
4.11.3 Frequency of Extreme Events 4-105

5. Quality Assurance and Quality Control
5.1 Introduction 5-1
5.1.1 Definitions of Quality Assurance and Quality Control 5-1
5.1.2 Importance of QA/QC Programs 5-1
5.1.3 EPA Quality Policy 5-2
5.2 Data Quality Objectives 5-3
5.2.1 The Data Quality Objectives Process 5-4
5.2.2 Data Quality Objectives and the QA/QC Program 5-7
5.3 Elements of a Quality Assurance Project Plan 5-8
5.3.1 Group A: Project Management 5-8
5.3.2 Group B: Measurements and Acquisition 5-14
5.3.3 Group C: Assessment/Oversight 5-17

5.3.4 Group D: Data Validation and Usability 5-18
5.4 Field Operations 5-19
5.4.1 Field Design 5-19
5.4.2 Sampling Site Selection 5-21
5.4.3 Sampling Equipment 5-22
5.4.4 Sample Collection 5-22
5.4.5 Sample Handling and Transport 5-23
5.4.6 Safety and Training 5-23
5.5 Laboratory Operations 5-24
5.5.1 General Laboratory QA and QC 5-24
5.5.2 Instrumentation and Materials for Laboratory Operations 5-24
5.5.3 Analytical Methods 5-26
5.5.4 Method Validation 5-26
5.5.5 Training and Safety 5-26
5.5.6 Procedural Checks and Audits 5-26
5.6 Data and Reports 5-26
5.6.1 Generation of New Data 5-27
5.6.2 Uses of Historical Data 5-27
5.6.3 Documentation and Record Keeping 5-28
5.6.4 Report Preparation 5-28

References R-1

Index I-1

Appendices
A. Review of Available Monitoring Guidances A-1
B. Data Sources B-1
C. Example Monitoring Programs C-1
D. Statistical Tables D-1

LIST OF TABLES

TABLE	TITLE	PAGE
1-1	Sources of nonpoint source pollution and their contribution to the impairment of water quality in the United States	1-5
2-1	General characteristics of monitoring types	2-4
2-2	Applications of six sampling designs to estimate means and totals	2-21
3-1	General strengths and limitations of biological monitoring and assessment	3-5
3-2	Five tiers of the rapid bioassessment protocols	3-18
3-3	Scoring criteria for the core metrics as determined by the 25th percentile of the metric values from the Middle Rockies-Central Ecoregion, Wyoming	3-23
3-4	Scoring criteria for the metrics as determined by the 25th percentile of the metric values for the two aggregated subecoregions for Florida streams	3-24
3-5	Comparison of probabilistic and targeted monitoring designs	3-28
3-6	Waterbody stratification hierarchy	3-30
3-7	Summary of the primary technical issues related to biological monitoring for nonpoint source evaluations	3-39
3-8	Selected biomonitoring program components, Delaware DNREC	3-40
3-9	Selected biomonitoring program components, Florida DEP	3-41
3-10	Selected biomonitoring program components, Montana DHES	3-42
3-11	Selected biomonitoring program components, North Dakota DEH	3-43
3-12	Selected biomonitoring program components, Vermont DEC	3-44
3-13	Fish IBI metrics used in various regions of North America	3-46
3-14	Examples of metric suites used for analysis of macroinvertebrate assemblages	3-49
4-1	Errors in hypothesis testing	4-2
4-2	Methods for characterizing data	4-5
4-3	Methods for routine data analysis	4-6
4-4	Total nitrogen runoff concentrations for a single storm event in Florida	4-13
4-5	Total nitrogen runoff concentrations for a single storm event in Florida and example calculations for the EMC	4-14
4-6	Raw data by time period	4-15
4-7	Loadings rate data	4-15
4-8	Calculation of plotting position for the sulfate data from Station 16 in Figure 4-8	4-27
4-9	Table of skewness test for normality for sample sizes less than 150	4-29
4-10	Selected summary statistics for the sulfate data from Station 16 in Figure 4-8...	4-29
4-11	Values of kurtosis test for normality for small sample sizes	4-30
4-12	Example analysis of the Sharpiro-Wilk *W* test using the sulfate data from Station 16 in Figure 4-8	4-32
4-13	Highland Silver Lake TSS data for site 1	4-38
4-14	Evaluation of power using the postimplementation TSS data	4-43
4-15	Nonparametric evaluation of postimplementation data using the Wilcoxon Signed Ranks test	4-45

List of Tables

4-16 Sign test for comparing paired BOD_5 concentrations 4-46
4-17 Summary of parametric tests used to evaluate difference between means 4-49
4-18 Nonparametric evaluation of postimplementation data using the Mann-Whitney test 4-51
4-19 ANOVA notation 4-54
4-20 Common one-way ANOVA output format 4-55
4-21 Trout population from streams in the coastal plain region 4-55
4-22 One-way ANOVA of stream trout data from the coastal plain region using stream as the treatment 4-56
4-23 Rank of trout population data from streams in the coastal plain region 4-57
4-24 Common two-way ANOVA output format 4-59
4-25 Stream trout population 4-60
4-26 Two-way ANOVA of trout population data using an interaction term 4-60
4-27 Common two-way ANOVA without replication output format 4-62
4-28 Assumptions necessary for the purposes of linear regression 4-66
4-29 Runoff sampler calibration data 4-67
4-30 Regression analysis of runoff sampler calibration data 4-68
4-31 Common ANOVA output format for linear regression 4-72
4-32 ANOVA for regression of treatment watershed runoff on control watershed runoff during calibration 4-82
4-33 ANOVA for regression of treatment watershed runoff on control watershed runoff during treatment 4-83
4-34 ANCOVA for comparing regression lines 4-83
4-35 ANCOVA for comparing regression lines from calibration and treatment (hand calibrations) 4-84
4-36 ANCOVA for comparing regression lines from calibration and treatment (computerized software) 4-84
4-37 Annual total rainfall for 21 years 4-88
4-38 Analysis of rainfall data using Mann-Kendall τ test 4-89
4-39 Analysis of rainfall data using Spearman's rho 4-91
4-40 Theoretical log-probability frequency factors 4-96
4-41 Linearized rainfall frequency variate for equation 4-109 4-96
4-42 Linearized rainfall duration variate for equation 4-109 4-96

5-1 Common QA and QC activities 5-2
5-2 Elements required in an EPA Quality Assurance Project Plan 5-9
5-3 Checklist of items that should be considered in the field operations section of a QA/QC program 5-20
5-4 Checklist of items that should be considered in the laboratory operations section of a QA/QC program 5-25

LIST OF FIGURES

FIGURE	TITLE	PAGE
1-1	Waterbody types affected by nonpoint sources of pollution, by state	1-2
1-2	Leading nonpoint sources of pollution that impair rivers and streams	1-2
1-3	Leading nonpoint sources of pollution that impair lakes and reservoirs	1-3
1-4	Leading nonpoint sources of pollution that impair estuaries	1-3
1-5	Leading nonpoint sources of pollution that impair ocean shorelines	1-3
1-6	Leading nonpoint sources of pollution that impair Great Lakes shoreline miles	1-4
1-7	Vertical sediment concentration and flow velocity distribution in a typical stream cross section	1-13
1-8	Schematic diagram of stream vertical showing relative position of sediment load terms	1-14
1-9	Important differences between lakes and reservoirs	1-15
1-10	Hydraulic residence time, assuming inflow = outflow	1-15
1-11	A cross-sectional view of a thermally stratified lake in mid-summer	1-16
1-12	Phytoplankton chlorophyll *a* concentration in Chautaugua Lake's northern basin and and southern basin, 1977	1-16
1-13	Mixing of salt water and fresh water in an estuary	1-17
1-14	Chesapeake Bay salinity levels over time and space	1-17
1-15	Estuarine drainage area versus fluvial drainage area	1-18
1-16	Nitrate concentration versus depth below water table	1-19
1-17	Comparison of water movement from irrigation furrows into two different soil types	1-21
2-1	Development of a monitoring project	2-2
2-2	Expectations report outline	2-8
2-3	Simple random sampling for silviculture	2-14
2-4	Stratified random sampling for silviculture	2-16
2-5	Systematic sampling for silviculture	2-18
2-6	Cluster sampling for silviculture	2-19
2-7	Nested paired and paired watershed study designs	2-22
2-8	Map of the Rock Creek Rural Clean Water Program study area	2-26
2-9	St. Albans Bay watershed sampling locations	2-28
3-1	Sample calculations of biological metrics	3-8
3-2	The process for metric selection and validation and development of reference conditions	3-12
3-3	Approach to establishing reference conditions	3-17
3-4	Selection and application of the different tiers of RBP depend on monitoring objectives	3-19
3-5	Organizational structure of attributes that can serve as metrics	3-21
3-6	Areas in which various fish IBI metrics have been used	3-22
3-7	Some trends that might be observed during the course of a biological monitoring program	3-35
3-8	Sample power analysis of a bioassessment method	3-36
4-1	Comparison of α and β	4-3
4-2	Precipitation, runoff, total nitrogen, and total phosphorus from a single storm event in	

Florida 4-12
4-3 Comparison of several theoretical distributions 4-17
4-4 Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a time series plot 4-21
4-5 Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a histogram 4-22
4-6 Stem and leaf plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware 4-22
4-7 Boxplots of dissolved oxygen concentrations by month from 1980 through 1989 for the Delaware River at Reedy Island, Delaware 4-23
4-8 Boxplot of sulfate concentrations from 1993 and 1994 for the Rio Grande near El Paso, Texas 4-24
4-9 Bivariate scatter plot of total suspended solids and flow at 36th Street storm sewer in Denver, Colorado 4-24
4-10 Time series plot of dissolved orthophosphate from 1989 through 1994 for portions of the Delaware River 4-25
4-11 Probability plot of sulfate data from Station 16 in Figure 4-8 4-28
4-12 Preimplementation data set 4-36
4-13 Postimplementation data set 4-37
4-14 Log-transformed preimplementation data set 4-39
4-15 Log-transformed postimplementation data set 4-40
4-16 One- and two-sided *t* test for post-BMP mean TSS concentration 4-42
4-17 Evaluation of power using the log-transformed postimplementation TSS data 4-43
4-18 Split versus flow rate 4-67
4-19 Plot of residuals versus predicted values 4-69
4-20 Plot of split residuals 4-70
4-21 Plot of split versus flow rate with confidence limits for mean response and individual estimates 4-74
4-22 Comparison of regression analyses using raw and log-transformed data 4-76
4-23 Comparison of regression equations for data from two periods 4-80
4-24 Storm runoff calibration and treatment periods in Vermont 4-81
4-25 One-hour rainfall to be expected at a return period of 2 years 4-97
4-26 24-hour rainfall to be expected at a return period of 2 years 4-98
4-27 One-hour rainfall to be expected at a return period of 100 years 4-99
4-28 24-hour rainfall to be expected at a return period of 100 years 4-100

5-1 Sample organization chart for a quality assurance project plan 5-10
5-2 Sample quality assurance objectives 5-12
5-3 Sample custody chart 5-15

CREDITS

Figure 2-2, Tables 4-2, 4-3	Modified and reprinted from R.C. Ward, J.C. Loftis, and G.B. McBride; *Design of water quality monitoring systems*; 1990; pages 82-83, page 92, and page 97, respectively; with permission from Van Nostrand Reinhold Company, New York.
Figure 5-3	Modified and reprinted with permission from A. Cross-Smiecinski and L.D. Stetzenback, *Quality planning for the life science researcher: Meeting quality assurance requirements*. Copyright CRC Press, Boca Raton, Florida. ©1994.
Table 2-2	Modified and reprinted from R.O. Gilbert, *Statistical methods for environmental pollution monitoring*, 1987, page 20, with permission from Van Nostrand Reinhold Company, New York.
Table 4-9	Modified from Dr. E.S. Pearson, *Tables for Statisticians and Biometricians*, vol. 1, Tables 34B and 34C, with permission from the *Biometrika* Trustees.
Tables 4-11, D9	Modified and/or reprinted from *Biometrika Tables for Statisticians*, 3rd ed., vol. I, Table 34, Part A and Table 15, respectively, 1966, with permission from the *Biometrika* Trustees.
Table 4-28	Modified and reprinted from D.R. Helsel and R.M. Hirsch, *Statistical Methods in Water Resources*, 1995, page 225, with kind permission from Elsevier Science - NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.
Table 4-40	Modified and reprinted from V.T. Chow, *The log-probability law and its engineering applications*, 1954, Separate No. 536, 80, with permission from the American Society of Civil Engineering.
Tables 4-41, 4-42	Modified and reprinted from *Monthly Weather Review*, 1962, pp. 87-88, L.L. Weiss: A general relation between frequency and duration of precipitation, with permission from the American Meteorological Society.
Tables D4, D5	Reprinted from *Biometrika*, 1965, vol. 52, pages 591-611, S.S. Shapiro and M.B. Wilk, An analysis of variance test for normality (complete samples), with permission from the *Biometrika* Trustees.
Table D7	Reprinted from *Journal of the American Statistical Association*, 1965, vol. 60, pp. 320-333, D.B. Owen: The power of Student's *t* test, with permission from the American Statistical Association.
Table D10	Reprinted from Glasser and Winter, 1961, with permission from the *Biometrika* Trustees.

Table D11

Reprinted by permission of John Wiley & Sons, Inc., New York, from W.J. Conover, *Practical Nonparametric Statistics*, Copyright © (1980 John Wiley & Sons, Inc.), page 447.

GLOSSARY

Atomic absorption spectrophotometry: A method to determine the elemental composition of a substance by vaporizing the sample and measuring at specific wavelengths the amount of radiation absorbed, which is proportional to the concentration of the element in the sample.

Bathymetry: The measurement of depths of water in oceans, seas, and lakes; also, information derived from such measurements.

Biomagnify: The process by which chemical pollutants become increasingly concentrated in animal tissue as the pollutants are passed up the food chain.

Biological oxygen demand (BOD): the amount of oxygen required by aerobic organisms to carry out oxidative metabolism in water containing organic matter, such as sewage.

Chemical oxygen demand (COD): a measure of the quantity of oxidizable components present in water.

Coliform bacteria: Bacteria present in mammalian feces, used as an indicator of the presence of human feces, bacteria, viruses, and pathogens in the water column.

Congener: An organism that is a member of the same genus as another animal or plant; a chemical substance that is related in some way to another.

Coriolis effect: The deflection relative to the earth's surface of an object that is moving on or above the earth, due to the action of the Coriolis force. An object that is moving horizontally above the earth's surface in the Northern Hemisphere tends to show a rightward deflection, and one in the Southern Hemisphere tends to show a leftward deflection.

Dissolved oxygen: The concentration of free molecular oxygen in the water column.

Diurnal: Active primarily during daylight hours.

Estuarine drainage area: The land and water component of a watershed that drains directly into estuarine waters.

Eutrophic: Rich in nutrients and hence having excessive plant growth, which removes oxygen from the water.

Eutrophication: The process of becoming eutrophic.

Fluvial drainage area: land and freshwater portions of watersheds upstream of estuarine drainage areas.

GIS (geographic information system): A computer system specialized for storage, manipulation, and presentation of geographical information, such as topography, political subdivisions, geology, vegetation, flood plains, etc. *[Note: Do not confuse GIS with GPS, global positioning system.]*

Habitat alteration: Changes in a habitat that make it less suitable for the organisms inhabiting it, create conditions favorable to invasion by species not present prior to the changes, or limit its ecosystem function.

Hydraulic residence time (HRT): amount of time necessary to fill a lake, or average amount of time water entering a lake stays in the system, calculated as lake volume/flow rate

Hydromodification: Alteration of the hydrologic characteristics of coastal and noncoastal waters, which in turn could cause degradation of water resources.

Impervious: The characteristic of a surface that prevents or retards the entry of water into or through it and causes water to run off the surface.

Ion chromatography: A technique for separating components from a mixture using ionic attraction by placing the mixture in a mobile phase that is passed over a stationary phase.

Macroinvertebrate: Invertebrate organisms that can be seen with the naked eye.

Macrophyte: Plants visible to the naked eye.

Nonpoint source: Generally, any unconfined and diffuse source of contamination, such as stormwater or snowmelt runoff, or atmospheric pollution. Legally, a nonpoint source of water pollution is any source of water pollution that does not meet the legal definition of "point source" in section 502(14) of the Clean Water Act.

Oligotrophic: Poorly nourished; describes a lake with low plant productivity and high transparency.

Onsite disposal system (OSDS): Sewage disposal system designed to treat wastewater at a particular site. Septic tank systems are common OSDS.

Oxygen-demanding substance: A substance whose decomposition in water uses oxygen.

Pervious: The characteristic of a surface that allows the entry of water into or through it and causes little water to run off the surface.

Phytoplankter: An organism of phytoplankton—microscopic algae and microbes that float freely in open water of lakes and oceans.

Point source: Any discernable, confined or discrete conveyance (pipe, ditch, channel, tunnel, conduit, well, discrete fissure, container, rolling stock, concentrated animal feeding operation, or vessel or other floating craft) from which pollutants are or may be discharged.

Pollution (water): The presence of excessive amounts of dredged spoil, solid waste, incinerator residue, sewage, garbage, sewage sludge, munitions, chemical wastes, biological materials, radioactive materials, heat, wrecked or discarded equipment, rock, sand, cellar dirt, and industrial, municipal, and agricultural waste in water.

Riparian areas: Vegetated ecosystems along a waterbody through which energy, materials, and water pass. Riparian areas characteristically have a high water table and are subject to periodic flooding and influence from the adjacent waterbody.

SAV (submerged aquatic vegetation): Macrophytes that are rooted and grow beneath the water surface.

Sluice gate: A gate for regulating or stopping flow in a conduit or passage where surplus water is carried off

Substrate: The substance, base, or nutrient on which an organism lives and grows, or the surface to which a fixed organism is attached.

Suspended sediment: The very fine soil particles that remain in suspension in water for a considerable period of time.

Turbidity: A cloudy condition in water due to suspended silt or organic matter.

Vadose zone: The subsurface zone in soil that contains air or gases generally under atmospheric pressure, between the land surface and the zone of saturation.

Weir: A device for measuring or regulating the flow of water.

1. OVERVIEW OF THE NONPOINT SOURCE PROBLEM

1.1 DEFINITION OF A NONPOINT SOURCE

Nonpoint sources of water pollution are both diffuse in nature and difficult to define. Nonpoint source pollution can generally be defined as the pollution of waters caused by rainfall or snowmelt moving over and through the ground. As water moves over or through the soil, it picks up and carries away natural pollutants and pollutants resulting from human activity, finally depositing them into lakes, rivers, wetlands, coastal waters, and ground waters. Habitat alteration (such as the removal of riparian vegetation) and hydrologic modification (such as damming a river or installing bridge supports across the mouth of a bay) can cause adverse effects on the biological and physical integrity of surface waters and are also treated as nonpoint sources of pollution. Atmospheric deposition, the deposition of airborne pollutants onto the land and into waterbodies, is also considered to be nonpoint source pollution. At the federal level, the term *nonpoint source* is defined to mean any source of water pollution that does not meet the legal definition of *point source* in section 502(14) of the Clean Water Act:

> The term "point source" means any discernible, confined and discrete conveyance, including but not limited to any pipe, ditch, channel, tunnel, conduit, well, discrete fissure, container, rolling stock, concentrated animal feeding operation, or vessel or other floating craft, from which pollutants are or may be discharged. This term does not include agricultural storm water discharges and return flows from irrigated agriculture.

The distinction between nonpoint sources and diffuse point sources is sometimes unclear. Although diffuse runoff is usually treated as nonpoint source pollution, runoff that enters and is discharged from conveyances, such as those described above, is treated as a point source discharge and hence is subject to the federal permit requirements under section 402 of the Clean Water Act.

1.2 EXTENT OF NONPOINT SOURCE PROBLEMS IN THE UNITED STATES

Over the last two decades, significant achievements have been made nationally in the protection and enhancement of water quality. Much of this progress, however, has resulted from controlling point sources of pollution. Although some state, tribal, and local nonpoint source management programs have been developed and are being implemented, pollutant loads from nonpoint sources present continuing problems for achieving water quality goals and maintaining designated uses in many parts of the United States.

Data provided by state water quality officials and contained in the *National Water Quality Inventory 1994 Report to Congress* (USEPA, 1995), referred to as the "1994 305(b) report," indicate that nonpoint sources negatively affect rivers and streams in 49 of the 52 states and territories that reported data; lakes, reservoirs, and ponds in 41 of the 52 states and territories that reported data; estuaries and coastal waters in 20 of the 26 coastal states and territories that reported data; and the Great Lakes in 4 of the 8 Great Lakes states (Figure 1-1). The categories of nonpoint source pollution affecting these waterbodies include agriculture, atmospheric deposition, channelization, construction, contaminated sediment, contaminated ground water, flow regulation, forest harvesting (silviculture), ground water loading, highway maintenance/runoff, hydrologic and habitat modification, in-place contamination, land development, land disposal, marinas, onsite disposal systems, recreational activities, removal of riparian vegetation, resource extraction, shoreline modification, streambank destabilization, and unspecified or other nonpoint source pollution.

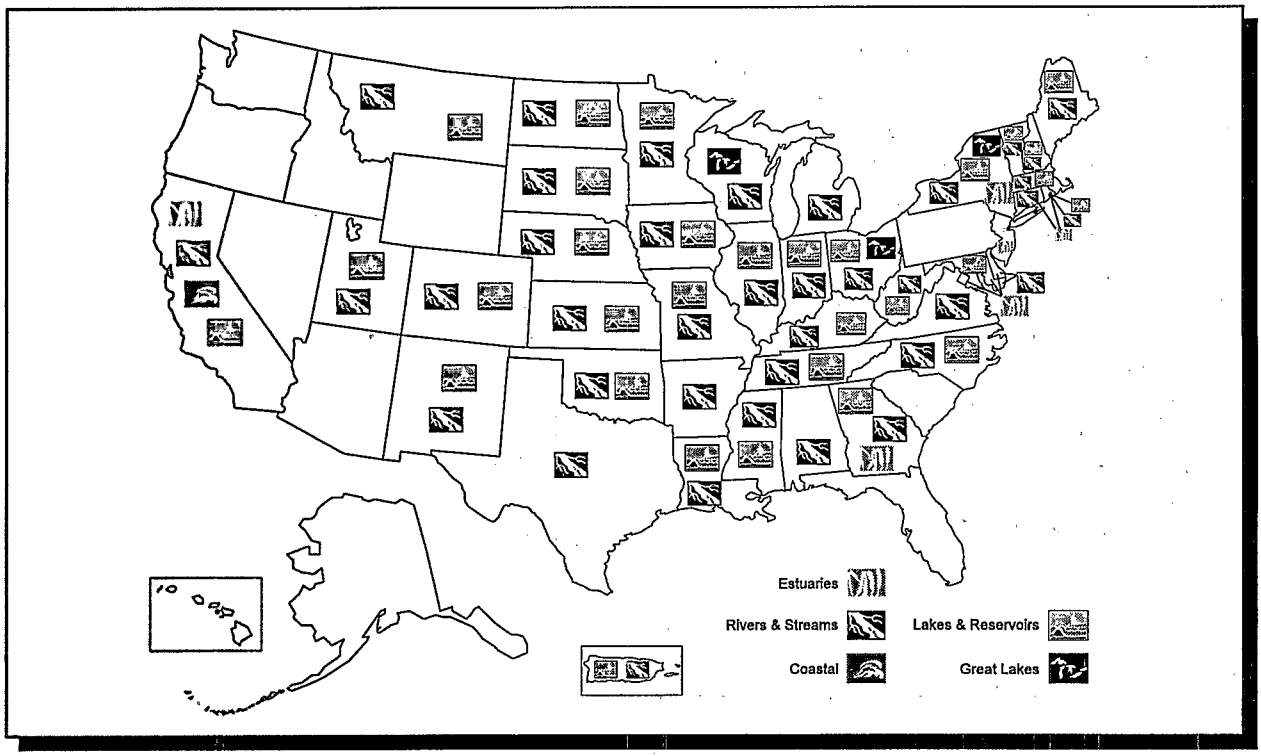


Figure 1-1. Waterbody types affected by nonpoint sources of pollution, by state (USEPA, 1995).

According to the 1994 305(b) report, agriculture is the primary source of pollution affecting rivers and streams. Forty-six states and territories list it as a major source of water quality impairment. Sixty percent of impaired river and stream miles are reported to be negatively affected by agricultural sources of pollution. Other sources that affect rivers and streams include natural sources[1] (reported as affecting 19% of impaired river miles), municipal point sources (17%), hydrologic and habitat modification (17%), urban runoff/storm sewers (12%), resource extraction (11%), removal of riparian vegetation (10%), forest harvesting (9%), industrial point sources (7%), unspecified or other nonpoint source pollution (7%), stream bank destabilization (6%), channelization (5%), petroleum activities (5%), construction (5%), land disposal (5%), recreational activities (3%), flow regulation (3%), onsite disposal systems (2%), highway maintenance and runoff (2%), and land development (2%) (Figure 1-2).

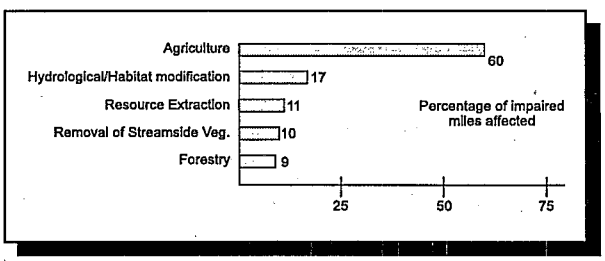


Figure 1-2. Leading sources of nonpoint pollution that impair rivers and streams (USEPA, 1995).

Agriculture is also considered the most significant pollution source affecting lakes, reservoirs, and ponds. Twenty-seven states list agriculture as a major source of impairment to the water quality of lakes, reservoirs, and ponds. Fifty percent of

[1]Natural sources refer to a variety of naturally occurring water quality problems, including natural deposits of salts, nutrients, and metals in soils that leach into surface and ground waters; warm-weather and dry-weather conditions that raise water temperatures, depress dissolved oxygen concentrations, and dry up shallow waterbodies; and low-flow conditions and tannic acids from decaying leaves that lower pH and dissolved oxygen concentrations in swamps that drain into streams (USEPA, 1995).

impaired lake, reservoir, and pond surface acres are reported to be negatively affected by agriculture, followed by municipal point sources (which affect 19% of impaired river miles), urban runoff/storm sewers (18%), unspecified and other nonpoint source pollution (15%), natural sources (14%), hydrologic and habitat modification (12%), industrial point sources (11%), land disposal (11%), construction (9%), flow regulation (7%), highway maintenance and runoff (6%), contaminated sediment (6%), atmospheric deposition (6%), onsite disposal systems (5%), forest harvesting (5%), resource extraction (4%), shoreline modification (3%), land development (3%), recreational activities (3%), and spills (2%) (Figure 1-3).

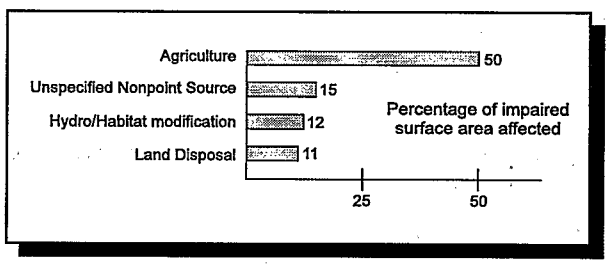


Figure 1-3. Leading sources of nonpoint pollution that impair lakes, reservoirs, and ponds (USEPA, 1995).

The water quality of estuaries is impaired more by urban runoff/storm sewers than by any other pollution source. Twelve states list urban runoff/storm sewers as a major source of impairment to estuarine water quality. Forty-six percent of impaired estuarine waters are reported to be negatively affected by urban runoff/storm sewers, followed by municipal point sources (39%), agriculture (34%), natural sources (30%), industrial point sources (27%), petroleum activities (13%), construction (13%), land disposal (13%), upstream sources (11%), unspecified and other nonpoint source pollution (10%), spills (8%), combined sewer outfalls (5%), resource extraction (5%), contaminated sediment (4%), marinas (3%), onsite disposal systems (3%), wastewater lagoons (3%), forest harvesting (5%), atmospheric deposition (2%), and recreational activities (2%) (Figure 1-4).

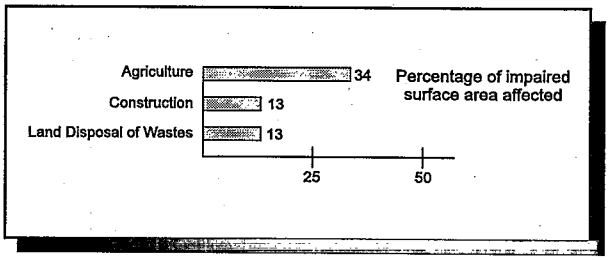


Figure 1-4. Leading sources of nonpoint pollution that impair estuaries (USEPA, 1995).

Ocean shoreline waters are impaired more by urban runoff/storm sewers than by other source of pollution. Two states and Puerto Rico list urban runoff/storm sewers as a major source of impairment to ocean shoreline water quality. Forty-eight percent of impaired ocean shoreline waters are reported to be negatively affected by urban runoff/storm sewers, followed by industrial point sources (34%), natural sources (25%), land disposal (25%), onsite disposal systems (23%), agriculture (20%), unspecified and other nonpoint source pollution (19%), combined sewer outfalls (11%), recreational activities (11%), municipal point sources (7%), atmospheric deposition (3%), spills (3%), ground water loading (3%), and land development (2%) (Figure 1-5).

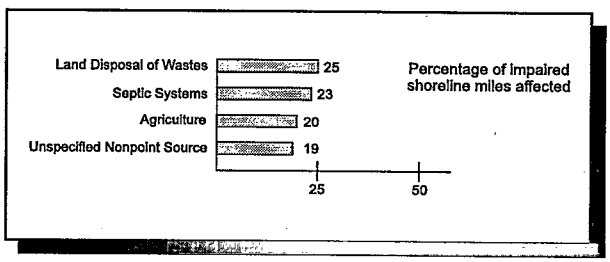


Figure 1-5. Leading sources of nonpoint pollution that impair ocean shorelines (USEPA, 1995).

Four of the eight Great Lakes states list nonpoint source pollution as negatively affecting the quality of their Great Lakes shoreline miles, with atmospheric deposition as the most damaging source of pollution to the lakes. Three states list atmospheric deposition as a source of impairment to Great Lakes shoreline miles, though none list it as a major source of impairment. Twenty-one percent of

impaired Great Lakes shoreline miles are reported to be negatively affected by atmospheric deposition, followed by discontinued discharges (20%), contaminated sediment (15%), land disposal (9%), unspecified and other nonpoint source pollution (6%), agriculture (4%), urban runoff/storm sewers (4%), industrial point sources (4%), municipal point sources (4%), combined sewer outfalls (3%), onsite disposal systems (2%), spills and illegal dumping (2%), streambank destabilization (1%), construction (1%), in-place contamination (1%), contaminated ground water (<1%), highway maintenance and runoff (<1%), and hydrologic and habitat modification (<1%) (Figure 1-6).

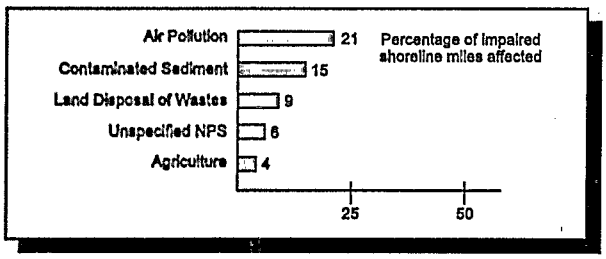


Figure 1-6. Leading sources of nonpoint pollution that impair Great Lakes shoreline miles (USEPA, 1995).

Table 1-1 summarizes the information on the impairment of the Nation's water quality by nonpoint sources of pollution.

1.3 Effects of Nonpoint Source Pollutants

Nonpoint sources can generate both conventional pollutants (e.g., bacteria, oxygen-demanding substances) and toxic pollutants (e.g., pesticides, petroleum products), just as point sources do. Even though nonpoint sources can contribute many of the same kinds of pollutants as point sources, however, these pollutants are usually generated in different volumes, combinations, and concentrations.

Pollutants from nonpoint sources are mobilized primarily during rainstorms or snowmelt. Consequently, waterborne nonpoint source pollution is generated irregularly, in contrast to the more continuous discharges of point sources of pollution. However, the adverse impacts of NPS pollution downstream from its source, or on downgradient waterbodies, can effectively be continuous under some circumstances. For example, sediment-laden runoff that is not completely flushed out of a surface water prior to a subsequent storm can combine with runoff from that storm to create a continuous adverse impact; toxic pollutants carried in runoff and deposited in sediment can exert a continuous adverse impact long after a rainstorm; physical alterations to a stream course caused by runoff can have a permanent and continuous effect on the watercourse; and the chemical and physical changes caused by NPS pollution can have a continuous adverse impact on resident biota. Hence, the noncontinuous generation of NPS pollution does not necessarily translate into noncontinuous impacts on receiving waterbodies.

Sediment, nutrients, pathogens, salts, toxic substances, petroleum products, and pesticides are the pollutants contributed to surface and ground waters by various nonpoint sources. Each of these pollutants, as well as habitat alteration and hydrologic modification, can have adverse effects on aquatic systems and, in some cases, on human health.

- Nitrogen and phosphorus are contained in commercial fertilizers and manure. The addition of excessive amounts of these nutrients to marine and freshwater systems, where nitrogen and phosphorus, respectively, are generally limiting to plant growth, can lead to accelerated eutrophication.

- Waste from livestock and pets also contains bacteria that contaminate swimming, drinking, and shellfishing waters, as well as oxygen-demanding substances that deplete dissolved oxygen levels in aquatic systems. Suspended sediment generated by construction, overgrazing, logging, and other activities in riparian areas, along with that carried in runoff

Table 1-1. Sources of nonpoint pollution and their contribution to the impairment of water quality in the United States (USEPA, 1995).

		Rivers & Streams	Lakes, Ponds & Reservoirs	Estuaries	Ocean Shoreline	Great Lakes Shoreline
Agriculture	Major[a]	36	28	4	1	1
	Length/Area Percent[b]	134,557 miles 60%	3,349,585 acres 50%	3,321 mi^2 34%	74 miles 20%	226 miles 4%
Land Disposal	Major	18	10	4	1	1
	Length/Area Percent	10,360 miles 5%	712,890 acres 11%	1,217 mi^2 13%	92 miles 25%	458 miles 9%
Onsite Disposal Systems	Major	13	6	3	0	1
	Length/Area Percent	5,428 miles 2%	335,702 acres 5%	271 mi^2 3%	87 miles 23%	96 miles 2%
Construction[c]	Major	13	13	2		1
	Length/Area Percent	10,365 miles 5%	624,901 acres 9%	1,253 mi^2 13%		48 miles 1%
Recreational Activities	Major	7	4	1	0	
	Length/Area Percent	7,796 miles 3%	189,828 acres 3%	180 mi^2 2%	40 miles 11%	
Atmospheric Deposition	Major		5	0	0	0
	Length/Area Percent		369,348 acres 6%	236 mi^2 2%	12 miles 3%	1,068 miles 21%
Hydrologic and Habitat Modification	Major	28	13			0
	Length/Area Percent	37,080 miles 17%	832,152 acres 12%			11 miles < 1%
Resource Extraction	Major	28	14	2		
	Length/Area Percent	24,059 miles 11%	236,999 acres 4%	514 mi^2 5%		
Silviculture	Major	13	9	1		
	Length/Area Percent	20,315 miles 9%	307,366 acres 5%	235 mi^2 2%		
Contaminated Sediment	Major		8	1		2
	Length/Area Percent		381,183 acres 6%	395 mi^2 4%		749 miles 15%
Unspecified and Other Nonpoint Source Pollution	Major	12	6	2	1	2
	Length/Area Percent	16,318 miles 7%	988,714 acres 15%	991 mi^2 10%	72 miles 19%	296 miles 6%
Total Impared[d]		224,236 miles	6,682,200 acres	9,700 mi^2	374 miles	5,077 miles

Source: USEPA, 1995.

[a] Number of states and/or territories in which the source is reported as a major source of water quality impairment.

[b] Length or area of waterbody type impaired by the source, and percent of total impaired length or area of waterbody type (reported at bottom of column) that is impaired by the particular source.

[c] Does not include road construction and maintenance.

[d] Figures in columns might not add to total at bottom if multiple sources impair the same stretch or area of surface water.

from cropland, highways, and bridges, reduces sunlight to aquatic plants, smothers fish spawning areas, and clogs filter feeders and fish gills.

- Salts from irrigation water concentrate at the soil surface through evapotranspiration. Salts used on roads accumulate along the edges of roads and are often carried via storm sewer systems to surface waters. Salts cause the soil structure to break down, decrease water infiltration, and decrease the productivity of cropland, and they can be toxic to plants at high concentrations.

- Some pesticides are persistent in aquatic systems and biomagnify in animal tissue (primarily fish tissue) as they are passed up through the food chain. Biomagnification has detrimental physiological effects in animals and negative human health impacts. Herbicides that are toxic to aquatic plants remove a food source for many aquatic animals, as well as the protective cover that aquatic vegetation offers to many organisms.

- Finally, the trampling of stream bottoms by livestock and equipment; stream bank erosion caused by grazing, logging, and construction; conversion of natural habitats to agricultural, urban, and other land uses; flow regulation; and activities in riparian areas can reduce the available habitat for aquatic species, increase erosion, and create flow regimes that are detrimental to aquatic life.

1.4 Major Categories of Nonpoint Source Pollution

1.4.1 Agriculture

Agriculture is the leading source of impairment to the Nation's rivers, affecting 60 percent of the impaired river miles in the United States, according to the 1994 305(b) report (USEPA, 1995). Agriculture was also reported as a source of impairment to 50 percent of impaired lake, reservoir, and pond acres; 34 percent of impaired estuary square miles; 20 percent of impaired ocean shoreline miles; and 4 percent of impaired Great Lakes shoreline miles. Wetland loss and wetland degradation were attributed to agriculture by 10 states and 8 states, respectively (USEPA, 1995).

The primary agricultural nonpoint source pollutants are nutrients, sediment, animal wastes, salts, and agricultural chemicals. Direct impacts on habitats are also associated with agriculture. Nitrogen and phosphorus are the two major nutrients from agricultural land that degrade water quality. Nutrients are applied to agricultural land in several different forms and come from various sources, including commercial fertilizer, manure from animal production facilities, municipal and industrial treatment plant sludge and/or effluent applied to agricultural lands, legumes and crop residues, irrigation water, and atmospheric deposition.

Greatly increased loadings of sediment to runoff and surface waters can result from land disturbance and clearing for agricultural operations and from stream bank erosion due to increased instream flows. Sediment loss and runoff are especially high if it rains or if high winds occur while the soil is being disturbed or soon afterward.

Animal waste includes the fecal and urinary wastes of livestock and poultry; process water; and the feed, bedding, litter, and soil from confined animal facilities. Runoff water and process wastewater from confined animal facilities can contain oxygen-demanding substances; nitrogen, phosphorus, and other nutrients; organic solids; salts; bacteria, viruses, and other microorganisms; and sediment.

Large amounts of salt can be added to agricultural soils by irrigation water that has a natural base load of dissolved mineral salts, regardless of whether the water is supplied by ground water or surface water sources. Irrigation water is consumed by plants and lost to the atmosphere by evaporation, and the salts in the water remain on and become concentrated in

the soil. Salt accumulation leads to soil dispersion, soil compaction, and possible toxicity to plants and soil fauna.

Agricultural chemicals—including pesticides, herbicides, fungicides, and their degradation products—can enter ground and surface waters in solution, in emulsion, or bound to soil colloids. Some types of agricultural chemicals are resistant to degradation and can persist and accumulate in aquatic ecosystems. Normal application to agricultural fields is a major source of pesticide contamination of surface water and ground water. Other sources are atmospheric deposition; drift during application; misuse; and spills, leaks, and discharges associated with pesticide storage, handling, and disposal.

Impacts on habitats and adjacent surface waters result from planting crops too close to surface waters and from livestock grazing. Riparian vegetation and its pollutant buffering capacity are lost when crops are planted too close to surface waters. Livestock grazing can cause loss of cover vegetation on pasturelands, resulting in erosion, loss of plant diversity on pasturelands, and adverse impacts on stream courses and surface waters. If allowed access to streams, cattle can trample riparian vegetation and disturb stream bank soils, leading to bank erosion, and can alter riparian vegetation species composition through selective grazing. Grazing animals also add fecal contamination to streams and ponds.

1.4.2 Urban Sources

Urban runoff and pollutants carried in storm sewers reportedly impair 12 percent of the Nation's impaired river miles; 18 percent of impaired lake, reservoir, and pond acres; 46 percent of impaired estuary square miles; 48 percent of impaired ocean shoreline miles, and 4 percent of impaired Great Lakes shoreline miles (USEPA, 1995). Wetland degradation is attributed to pollution from urban runoff/storm sewers by six states (USEPA, 1995).

The major pollutants in runoff from urban areas are sediment, nutrients, oxygen-demanding substances, road salts, heavy metals, petroleum hydrocarbons, pathogenic bacteria, viruses, and toxic chemicals. These are generated directly from the use of insecticides, road salts, and fertilizers, and indirectly from automobile exhaust, oil drippings from trucks and cars, brake lining wear, and various urban activities (USEPA, 1977).

During urbanization, pervious, vegetated ground is converted to impervious, unvegetated land. Land imperviousness in urban areas—as rooftops, roads, parking lots, and sidewalks—can range from 35 percent or lower in lightly urbanized areas to nearly 100 percent in heavily urbanized areas. Increases in pollutant loadings generated from human activities are associated with urbanization, and imperviousness results in increased stormwater runoff volumes and altered hydrology in urban areas. Urban runoff carries these increased pollutant loadings to surface waters, typically without treatment.

Imperviousness results in large volumes of stormwater runoff delivered to surface waters much more quickly than normal, which can result in scouring of stream banks and streambeds and increased sediment loadings to surface waters. Combined with the increased runoff velocities that occur during spring snowmelts and rain-on-snow events in urbanized watersheds, floods often occur more frequently and with greater severity in urbanized areas (Buttle and Xu, 1988). Major snowmelt events can produce peak flows with as much as 20 times the volume of baseflows in urban areas (Pitt and McLean, 1992).

1.4.3 Removal of Streamside Vegetation

Removal of streamside vegetation is reported to be a leading source of impairment to rivers and streams and was reported to affect 10 percent of impaired river and stream miles in the 1994 305(b) report (USEPA, 1995). Somewhere between 70 and 90 percent of natural riparian ecosystems in the United

States have been lost to human activities (Windell, 1983, cited in USEPA, 1991).

Losses of riparian vegetation are attributed to conversion to farmland, drainage for agriculture, forest harvesting, channelization, damming, creation of impoundments, irrigation diversions, ground water pumping, and overgrazing (Brinson et al., 1981).

The biological communities in streams depend on inputs of energy from outside sources. The primary source of energy and nutrients in small, low-order streams is organic debris (e.g., leaf litter) deposited from riparian vegetation. When riparian vegetation is removed, this source of energy and nutrients is eliminated or reduced. Stretches of streams and rivers are left with sunlight as the only source of energy and largely devoid of nutrient inputs. Other essential inputs to rivers and streams, such as woody debris—which provides microhabitats for fish and invertebrates, are also lost when streamside vegetation is removed (USEPA, 1991).

Riparian habitats, regardless of regional location, have many characteristics important to surrounding communities. They have a high rate of energy, nutrient, and species exchange; they are highly productive; they provide a unique microclimate with respect to upslope conditions; they have high edge-to-area ratios (similar to ecotone areas); and they support diverse faunal assemblages that are often unique within the local environment (USEPA, 1991). Loss of riparian vegetation therefore has negative effects on surrounding biotic communities.

Riparian vegetation also has an enormous capacity to store water. When it is removed, the natural hydroperiods of streams and rivers are altered and the loss of the buffering effects of water released by riparian vegetation during low flow periods and water stored by riparian vegetation during periods of flooding can cause severe stress to aquatic plant and animal communities. Riparian vegetation protects stream banks from erosion due to flowing water, and this protection is also lost when the vegetation is removed. Increases in erosion, turbidity, and sedimentation usually result (Brinson et al., 1981).

Riparian vegetation also removes sediment as water passes through it, rebuilds floodplains, provides shelter for aquatic animals and wildlife under overhanging banks, provides food to aquatic and terrestrial wildlife, buffers water temperatures, and improves water quality for downstream users (USDOI, 1991). Degraded water quality, increased severity of flooding, loss of wildlife, increased stream temperatures, and increased expense to purify water for public uses are therefore some of the consequences of the removal of riparian vegetation.

1.4.4 Hydromodification

Hydromodification and habitat alteration are reported to be a source of impairment to 17 percent of impaired river and stream miles and 12 percent of impaired lake, reservoir, and pond acres, and a source of wetland degradation in five states and wetland loss in one state (USEPA, 1995).

Hydromodification includes channelization or channel modification and flow alteration. Channel modification is river and stream channel engineering undertaken for the purpose of flood control, navigation, drainage improvement, and reduction of channel migration potential (Brookes, 1990). Straightening, widening, deepening, or relocating existing stream channels; excavation of borrow pits, canals, underwater mining, and other practices that change the depth, width, or location of waterways or embayments in coastal areas; and clearing or snagging operations are examples of channel modification. Channel modification typically results in more uniform channel cross sections, steeper stream gradients, and reduced average pool depths.

Flow alteration describes a category of hydromodification activities that result in either an increase or a decrease in the usual supply of fresh

water to a stream, river, or estuary. Flow alterations include diversions, withdrawals, and impoundments. In rivers and streams, flow alteration can also result from transportation embankments, tide gates, sluice gates, weirs, and the installation of undersized culverts. Levees and dikes are also flow alteration structures.

Channel modification can deprive wetlands and estuarine shorelines of enriching sediment; change the ability of natural systems to both absorb hydraulic energy and filter pollutants from surface waters; increase transport of suspended sediment to coastal and near-coastal waters during high-flow events; increase instream water temperature; and accelerate the discharge of pollutants (Sherwood et al., 1990). Hydromodification often diminishes the suitability of instream and riparian habitat for fish and wildlife through reduced flushing, lowered dissolved oxygen levels, saltwater intrusion, interruption of the life cycles of aquatic organisms, and loss of streamside vegetation.

1.4.5 Mining

Mining, or resource extraction, is reported as a source of impairment to 11 percent of impaired river and stream miles; 4 percent of impaired lake, pond, and reservoir acres; and 5 percent of impaired estuary square miles. It also accounts for wetland loss in two states and wetland degradation in one state (USEPA, 1995).

Numerous pollutants are released from coal and ore mining. Acid mine drainage from coal mining contains sulfates, acidity (low pH), heavy metals, ferric hydroxide or "yellow boy," and silt (USEPA/USDOI, 1995; Zielinski, n.d.). The heavy metals released from mining activities include silver, arsenic, copper, cadmium, mercury, lead, antimony, and zinc (Horowitz et al., 1993).

Ore mining, both past and present, is a significant source of mercury contamination (Leigh, 1994). Mercury was used to separate gold and silver from ore and is contained in waste piles from the amalgamation process (Oak Ridge National Laboratory, 1993). It is estimated that 5.5×10^9 g of metallic mercury was released into the Carson River Drainage Basin during processing of the Comstock Ore at Virginia City, Nevada, in the 1800s. The mill no longer stands, but mercury-contaminated tailings were left behind to create a long-term, significant source of mercury contamination of soil and air (Gustin et al., 1995). Mercury was also used in eastern mining. Gold mining in the Georgia piedmont from 1829 to 1940 left mercury-contaminated alluvium. Mercury concentrations in historical alluvium have been found to exceed background by as much as two orders of magnitude near the core of the mining district (Leigh, 1994).

Public health can be threatened by contaminants released from ore mining. Exposure pathways for this contamination include ingestion of fish and waterfowl, as well as ingestion, inhalation, and direct contact with contaminated soil, sediment, and surface water. Potential exposure is also possible through ingestion of crops irrigated with contaminated surface water or grown in contaminated soil (Oak Ridge National Laboratory, 1993).

Coal mining creates significant acidity problems. The results of a study done to characterize the causes of acidity in lakes and streams in the United States show that 26 percent of streams were acidified primarily by acid mine drainage and 47 percent by atmospheric deposition (Kaufmann et al., 1992).

Three types of mines are created for coal extraction. *Drift or slope mines* are driven into valley walls to expose coal. *Shaft mines* are driven perpendicular to the ground. These mines must be pumped continuously to extract infiltrating water, and when abandoned they fill with water. *Surface mining* extracts coal from the surface after overlying soil and rock have been removed. Surface mines leach metals and acids as seeps or springs, and they can have flows of up to 500 gallons per minute (SCRIP,

1996). Abandoned, self-draining underground mines and coal cleaning refuse piles are the worst potential sources of acid mine drainage. Contamination from them can continue for 800 to 3,000 years as all of the exposed acidic materials from the mines slowly leach pollutants to ground waters (USDOI, n.d.; Zielinski, n.d.). Acid mine drainage from surface mines is also a problem but is more controllable (USDOI, n.d.). Of the acidification caused by acid mine drainage, an estimated 40 percent is from active surface and underground mines and 60 percent is from abandoned mines (Zielinski, n.d.).

The effects of acid mine drainage can be devastating. In severely affected streams, ferric hydroxide blankets stream bottoms, smothering eggs, and covers gills and body surfaces (Zielinski, n.d.). Once a stream's acid-neutralizing capacity has been depleted by acidity entering the stream, the acidity begins to alter the biota. Fish are absent from streams with a pH less than 4.5, and vascular plants are lacking in streams with a pH less than 4 (Zielinski, n.d.).

Approximately 11,990 miles of streams are reported to be degraded by acid mine drainage in Pennsylvania, West Virginia, Ohio, Kentucky, Maryland, Indiana, Illinois, Oklahoma, Iowa, Missouri, Kansas, Tennessee, Virginia, Alabama, and Georgia (USDOI, n.d.; Zielinski, n.d.). Eighty percent of these (10,507 miles) are in Pennsylvania, West Virginia, Ohio, Kentucky, Tennessee, Maryland, and Alabama (Zielinski, n.d.). The worst acid mine drainage pollution is in Pennsylvania and West Virginia and a few areas of southeastern Ohio (USDOI, n.d.). Pennsylvania alone has 7,800 abandoned or inactive underground mines below the water table; one billion gallons is the estimated daily influx of acid mine drainage to surface waters from these mines (Zielinski, n.d.).

Most water quality problems associated with mining, however, are considered to be point source problems and are regulated under state and federal NPDES permits (USEPA, 1978a).

1.4.6 Forest Harvesting

Forest harvesting is reported as a source of impairment to 9 percent of impaired river and stream miles; 5 percent of impaired lake, pond, and reservoir acres; and 5 percent of impaired estuarine square miles. Two states attribute wetland degradation to forest harvesting and three states attribute wetland loss to forest harvesting (USEPA, 1995). On federal lands, such as national forests, many water quality problems can be attributed to the effects of timber harvesting and related activities (Whitman, 1989).

Forest harvesting operations can degrade water quality in several ways in waterbodies that receive drainage from forest lands. Sediment, organic debris, nutrients, and silvicultural chemicals are pollutants associated with forest harvesting operations. Construction of forest roads and yarding areas, as well as log dragging during harvesting, can accelerate erosion and sediment deposition in streams, which fouls instream habitats. Removal of overstory riparian shade can increase stream water temperatures; harvesting operations can leave slash and other organic debris to accumulate in waterbodies, which can deplete dissolved oxygen and alter instream habitats. Fertilizer applications can add excessive nutrients to aquatic habitats and accelerate eutrophication. Pesticide applications can increase organic and inorganic chemical concentrations in waterbodies, which can lead to adverse wildlife and habitat impacts (Brown, 1985).

1.4.7 Construction

Construction is reported as a source of impairment to 5 percent of impaired river and stream miles; 9 percent of impaired lake, pond, and reservoir acres; 13 percent of impaired estuarine square miles; and 1 percent of impaired Great Lakes shoreline miles. Two states attribute wetland degradation to construction and four states attribute wetland loss to construction (USEPA, 1995).

Many potential pollutants are associated with construction activities. These include sediment; pesticides (insecticides, fungicides, herbicides, and rodenticides); fertilizers used for vegetative stabilization; petrochemicals (oils, gasoline, and asphalt degreasers); construction chemicals such as concrete products, sealers, and paints; paper; wood; garbage; and sanitary wastes (Washington State Department of Ecology, 1991). The variety of pollutants present and the severity of their effects depend on the nature of the construction activity, the physical characteristics of the construction site, and the proximity of surface waters to the nonpoint pollutant source.

Runoff from construction sites is by far the largest source of sediment in urban areas under development (York County Soil and Water Conservation District, 1990). Soil erosion accounts for over 90 percent of sediment losses by tonnage in urbanizing areas, where most construction activities occur (Canning, 1988). Uncontrolled construction site sediment loads have been reported to be on the order of 35 to 45 tons per acre per year (Novotny and Chesters, 1989; Yorke and Herb, 1976, 1978). Loadings from undisturbed woodlands are typically less than 1 ton per year (Leopold, 1968).

Petroleum products used during construction include fuels and lubricants for vehicles, power tools, and general equipment maintenance. Asphalt paving also can be particularly harmful since it releases various oils for a considerable time period after application.

Solid waste on construction sites includes trees and shrubs removed during land clearing and structure installation, wood and paper from packaging and building materials, scrap metal, sanitary wastes, rubber, plastic, glass, and masonry and asphalt products.

Chemical pollutants, such as paints, acids for cleaning masonry surfaces, cleaning solvents, asphalt products, soil additives used for stabilization, pollutants in wash water from concrete mixers, and concrete-curing compounds, can also be used on construction sites and carried in runoff.

1.4.8 Marinas

Marinas are reported as a source of impairment to 3 percent of impaired estuarine square miles. Puerto Rico reports port construction as a source of wetland degradation, and the construction of wharves, piers, and bulkheads is reported as a source of wetland loss by two states (USEPA, 1995).

Marinas are located right at the water's edge, so there is often no buffering of the release of pollutants from them to waterways. Consequently, the concentrations of pollutants in marina waters and sediment, and in the tissues of organisms living in or near marinas, can be elevated.

The primary pollutants associated with marinas are sewage discharged from boats, which contains high concentrations of fecal coliform bacteria and organics; metals; and petroleum hydrocarbons. These pollutants enter the water in marinas through discharges and spills from boats and docks, and stormwater runoff from marina uplands. The concentration of dissolved oxygen in marina basins can be lowered by inadequate flushing and the decomposition of organics, such as those in sewage and fish offal.

Marina or port construction can negatively affect the ecology of an area; effects include loss of habitat and alterations to local hydrodynamics. Protective measures like bulkheads and jetties are built near marinas to prevent damage to boats and shoreline structures, and marinas and ports are areas of concentration of boat traffic. Both the attenuation of waves by in-water structures and the creation of waves by boat passage affect shoreline processes, which can increase turbidity, resuspend pollutants in sediment, and increase shoreline erosion (USFWS, 1982).

Studies have shown that boats can be a source of fecal coliform bacteria in estuaries with high boat densities and poor flushing (Fisher et al., 1987; Gaines and Solow, 1990; Milliken and Lee, 1990; NCDEM, 1990; Sawyer and Golding, 1990; Seabloom et al., 1989). Fecal coliform levels in marinas and mooring fields become most elevated during periods of high boat occupancy and usage, such as holiday weekends.

Metals and metal-containing compounds are contained in fuel additives, antifoulant paints, ballast, and other marina structures. Arsenic is used in paint pigments, pesticides, and wood preservatives. Zinc anodes are used to deter corrosion of metal hulls and engine parts. Copper and tin are used as biocides in antifoulant paints. Other metals (iron, chrome, etc.) are used in the construction of marinas and boats. These metals are released to marina waters through spillage, incomplete fuel combustion, wear on boat hulls and marina structures, and boat bilge discharges (NCDEM, 1991). Elevated levels of copper, zinc, cadmium, chromium, lead, tin, and PCBs have been found in oysters, other bivalves, and algae in some marinas (CARWQCB, 1989; Marcus and Stokes, 1985; McMahon, 1989; NCDEM, 1991; Nixon et al., 1973; SCDHEC, 1987; Wendt et al., 1990; Young et al., 1979).

1.5 Water Resource Considerations

Before a monitoring plan that will provide sufficient information for meeting monitoring objectives can be developed, the water resource to be monitored must be understood. Each type of water resource—rivers and streams; lakes, reservoirs, and ponds; estuaries; open coastal waters; and ground waters—possesses unique hydrologic and biological features that must be taken into consideration, and a monitoring program must be structured to either adapt to those features or avoid them.

All water resource types exhibit both temporal (long- and short-term) and spatial (small- and large-scale) variability. Placement of monitoring stations and timing of sampling are affected by these variabilities. For instance, suspended sediment concentrations vary across the width and length of reservoirs; salinity concentrations in estuaries vary vertically and temporally as they are affected by relatively light fresh water flowing over heavier salt water; and ground water quality varies with soil type and geozone. The monitoring guidance provided in this document is appropriate for temporal variability of minutes to a few years.

1.5.1 Rivers and Streams

Generally, streams are of two types, intermittent and perennial. Clearly, sampling cannot be done in intermittent streams when they do not have flow, and year-to-year variations in precipitation affect the duration of their flows, their pollutant loads, and their water quality. Variability in perennial streams and rivers is also affected by seasonal variations in precipitation, including snowfall, reservoir discharge management, and irrigation management. The highest concentrations of suspended sediment and nutrients often occur during spring runoff, winter thaws, or rainstorms.

Other features of streams and rivers that affect monitoring program design include, but are not limited to:

- Lateral spatial variability is most important in streams. Velocity varies vertically and horizontally in streams and affects pollutant concentrations at a given location (Figure 1-7) (USDA-NRCS, 1996).

- Tributary mixing affects lateral variability. Mixing below tributary junctions might be incomplete, with tributary flow primarily following one bank. Meanders produce increased velocities at the outside bank and reduced velocities at the inside bank (USDA-NRCS, 1996).

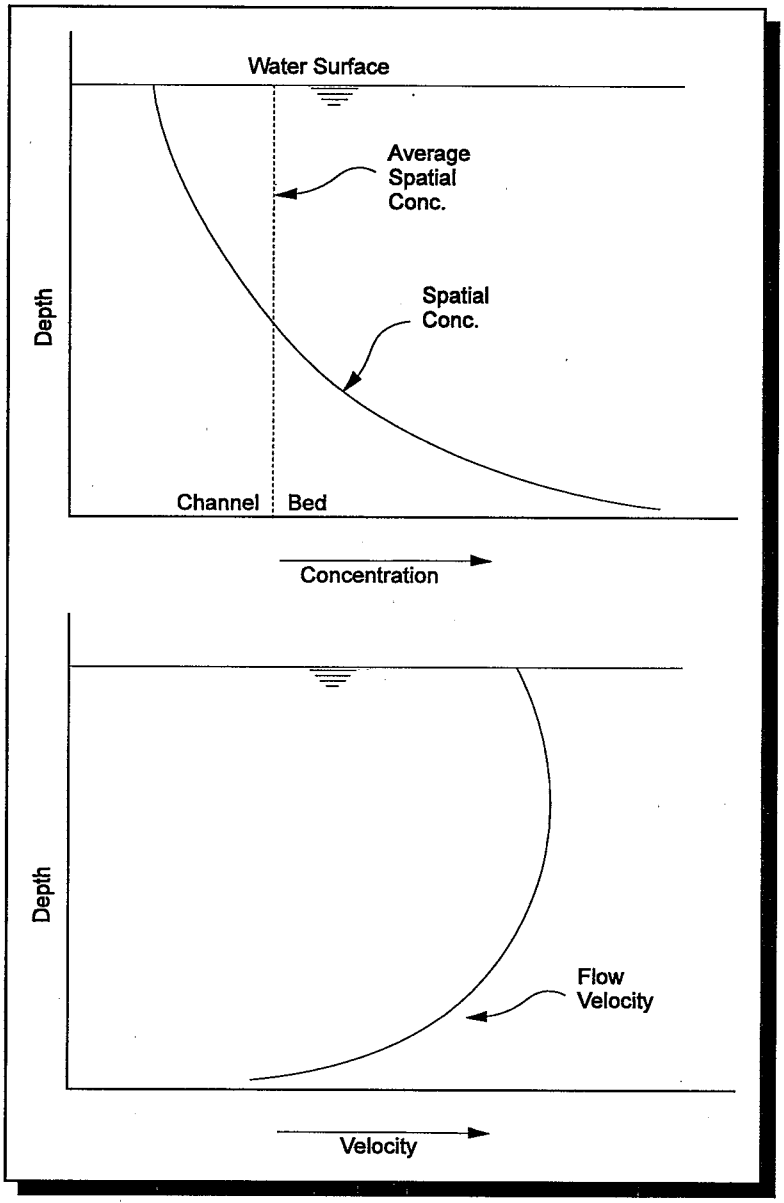


Figure 1-7. Vertical sediment concentration and flow velocity distribution in a typical stream cross section (Brakensiek et al., 1979).

- The complexity of currents at obstructions makes them poor monitoring sites (USDA-NRCS, 1996).

- Vertical variability is particularly important during runoff and in slow-moving streams because suspended solids, dissolved oxygen, and algal productivity can vary substantially with depth (Figure 1-8).

- Toxic contaminants in bed sediment vary laterally and vertically.

- Biological communities vary with type of bed substrate, water temperature, and amount and type of aquatic and riparian vegetation.

Also, when designing a stream or river monitoring program, the effects of tributary flows must be considered. Such flows can add pollutant loads, dilute pollutant loads, and create lateral gradients. Segmentation of a stream into fairly homogeneous segments prior to monitoring might be necessary or prudent. One to several monitoring stations might be necessary in each segment (Coffey et al., 1993). When dividing a stream into homogeneous segments, both land use and drainage area should be considered, since both affect the quantity and quality of flows.

1.5.2 Lakes, Reservoirs, and Ponds

Shape is an important factor that affects spatial variability in lakes, reservoirs, and ponds. Lakes and ponds commonly have simple, roundish shapes, whereas reservoirs usually have complex, dendritic shapes (Figure 1-9). Some lakes and reservoirs are well mixed and homogeneous, while others are stratified and heterogeneous. Stratification in lakes and reservoirs depends on depth and hydraulic residence time (HRT) (Figure 1-10). Deeper lakes and reservoirs are more likely to be strongly stratified, while shallow lakes and reservoirs are usually more uniform vertically.

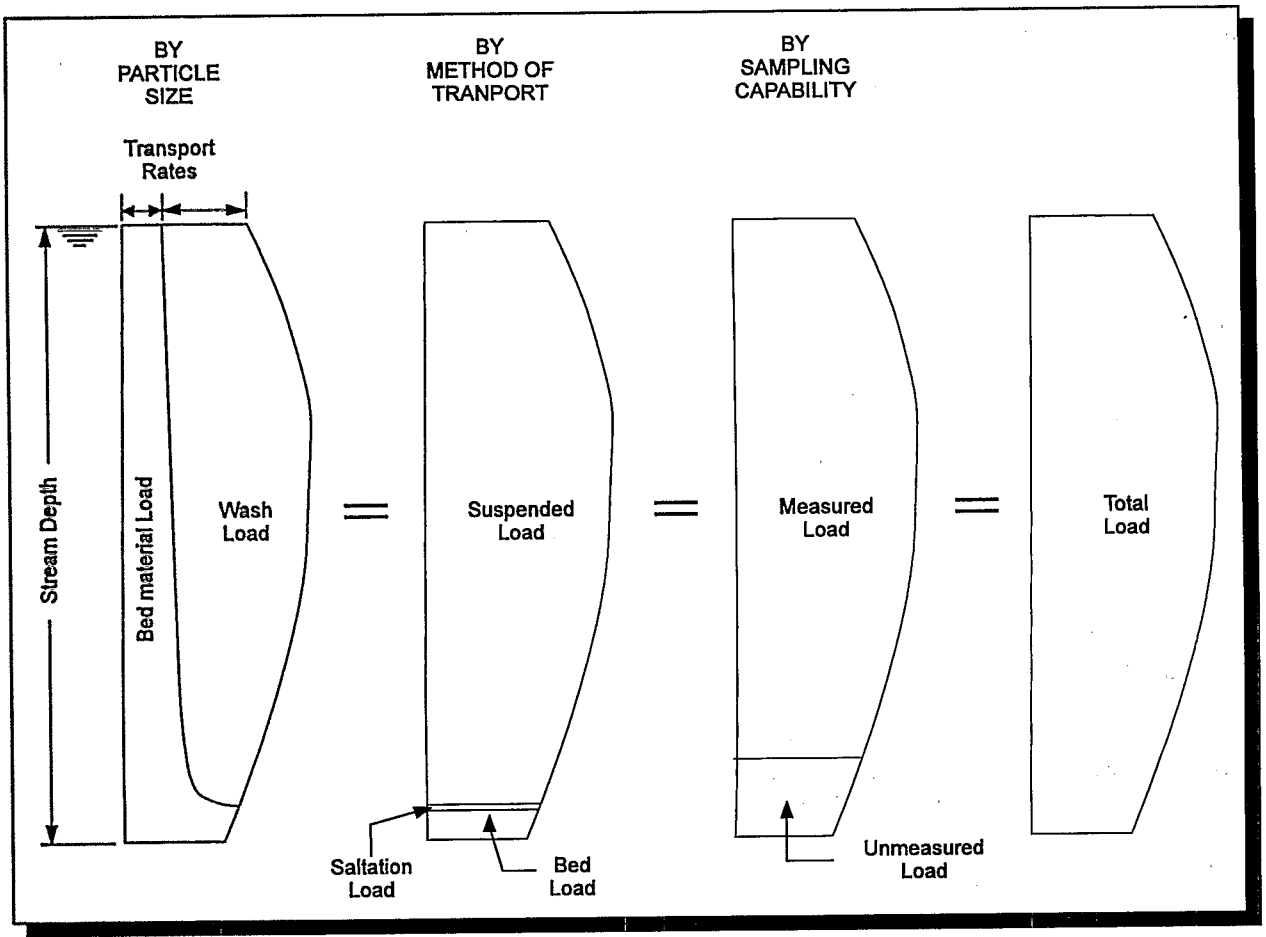


Figure 1-8. Schematic diagram of stream vertical showing relative position of sediment load terms (Brakenslek et al., 1979).

Stratification creates vertical variability in temperature, dissolved oxygen, and nutrient concentrations (Figure 1-11). Stratified lakes and reservoirs can also exhibit little vertical variability in DO and nutrient concentrations, depending on their productivity. Deep, eutrophic lakes and reservoirs exhibit more vertical variability than deep, oligotrophic lakes and reservoirs (USEPA, 1990a).

The implication of HRT for monitoring and the water quality effects of pollution control is that there can be a delay between changes in inflow water quality and a noticeable effect on lake water quality. The length of the delay depends on the HRT of the lake or reservoir and other water quality factors (e.g., biota, sediment, existing water chemistry) (USEPA, 1990a).

Features of lakes, reservoirs, and ponds to keep in mind when designing a monitoring program for them include, but are not limited to, the following:

- Lakes and reservoirs with HRTs of days or weeks might respond to seasonal pollutant loads, whereas lakes and reservoirs with much longer HRTs might not respond so quickly to seasonal loads (USEPA, 1990a).

- The distribution and concentrations of water quality parameters in individual lakes and reservoirs vary seasonally (Figure I-11).

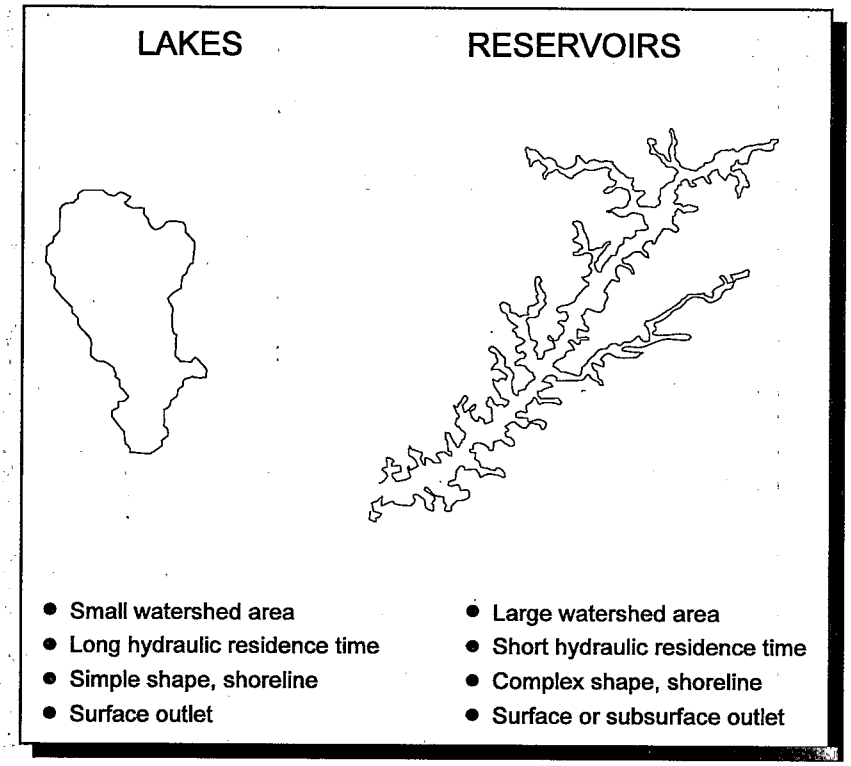


Figure 1-9. Important differences between lakes and reservoirs.

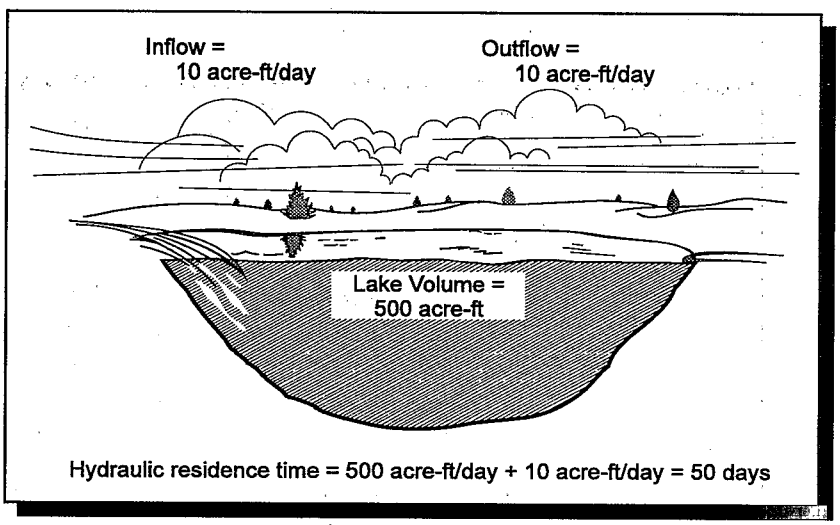


Figure 1-10. Hydraulic residence time, assuming inflow = outflow (After USEPA, 1990).

- Separate lakes and reservoirs in the same geographic area do not necessarily undergo seasonal changes at the same time.
- Pollutants are generally not distributed uniformily throughout lakes and reservoirs due to inflow points and circulation patterns.
- Pollutants such as phosphorus can have residence times in lakes and reservoirs very different from the HRTs of the lakes and reservoirs in which they are found. Some lakes and reservoirs do not respond to reductions in phosphorus loads because of phosphorus contained in lake sediment, which is released when the phosphorus concentration in the water column decreases and sequestered when it increases.
- Short-term variability is an inherent characteristic of most still (lentic) waterbodies. Dissolved oxygen, pH, and temperature can vary considerably over the course of a day.
- Small lakes and reservoirs can respond rapidly to the addition of runoff, which has implications if lake water quality is to be correlated with land treatment or stream water quality.
- The lateral variability of chlorophyll *a* concentrations can vary based on water depth and the diurnal migrations of phytoplankters (Davenport and Kelly, 1984a).

In summary, some important lake and reservoir characteristics and processes that must be

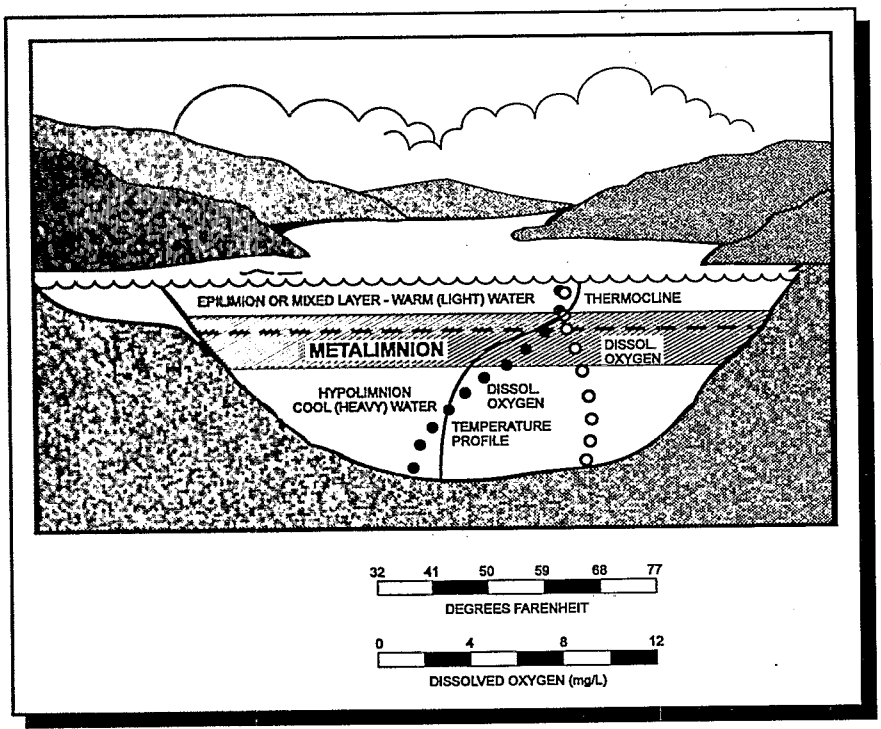


Figure 1-11. A cross-sectional view of a thermally stratified lake in mid-summer. The water temperature profile (curved solid line) illustrates how rapidly the water temperature decreases in the metalimnion compared to the nearly uniform temperatures in the epilimnion and hypolimnion. Open circles represent the dissolved oxygen (DO) profile in an unproductive (oligotrophic) lake: the DO concentration increases slightly in the hypolimnion because oxygen solubility is greater in colder water. Solid circles represent the DO profile in a productive (eutrophic) lake in which the rate of organic matter decomposition is sufficient to deplete the DO content of the hypolimnion (USEPA, 1990).

considered when designing a monitoring program include productivity, depth, stratification, seasonality, HRT, and the locations and sources of inflows. Since sampling locations must accurately represent lake or reservoir conditions, monitoring a round, simply shaped lake might require only a single sampling station, whereas monitoring a dendritic reservoir might require numerous stations to reflect its spatial variability accurately (USEPA, 1990a).

To simplify both sample collection and data interpretation, it can be useful to monitor within the strata of stratified lakes and reservoirs, and to achieve some monitoring objectives, it could be necessary to monitor during periods of peak stratification. Finally, biological monitoring programs must be tailored to the diurnal variations in lakes and reservoirs.

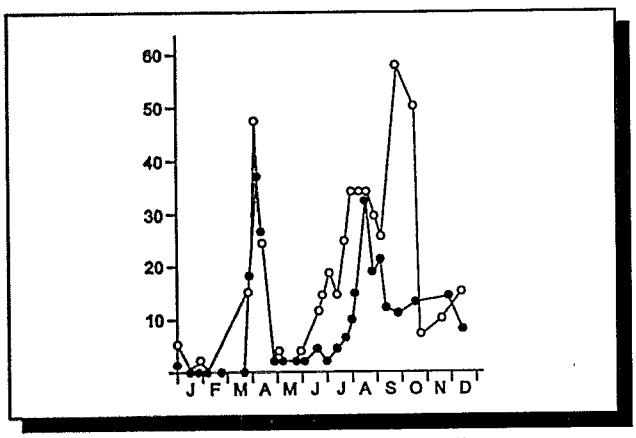


Figure 1-12. Phytoplankton chlorophyll *a* concentration in Chautaugua Lake's northern basin and southern basin, 1977 (Storch, 1986).

1.5.3. Estuaries

The major difference between estuaries and freshwater bodies is in the mixing of fresh water with salt water, and the influence of tides on spatial and temporal variability in estuaries. Incoming tides affect estuaries by pushing salt watershoreward while fresh water is entering(Figure 1-13). Fresh water is lighter, so it flows over the top of salt water, while the force of the tide forces the salt water shoreward and under the inflowing fresh water. Outgoing tides pull the entire water mass oceanward, and the freshwater input fills the gap left by the receding submerged salt water.

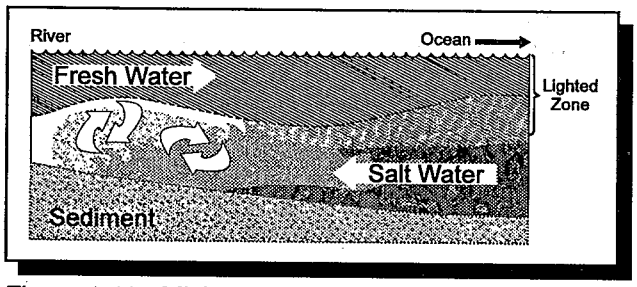


Figure 1-13. Mixing of salt water and fresh water in an estuary (Chesapeake Bay Program, 1995).

These processes affect daily and seasonal salinity distributions (Figure 1-14).

Features to consider when designing a monitoring plan for estuaries include, but are not limited to, the following:

- The volume of an estuary is an important factor in determining its ability to dilute pollutants (NOAA, 1990).
- Short-term variability is related to tidal cycles, which affect the mixing of fresh and salt waters and the position of the fresh water-salt water interface.
- The size of the estuarine drainage area (EDA; Figure 1-15) of an estuary relative to watershed size determines the overall impact of pollutant inputs from the EDA on the estuary (NOAA, 1990).
- In estuaries with large fluvial drainage areas (FDA; Figure 1-15), pollutants added from sources in the EDA might have less overall impact on estuarine water quality than in estuaries with small FDAs (NOAA, 1990).
- Freshwater inflow is a major determinant of the physical, chemical, and biological characteristics of most estuaries. It affects theconcentration and retention of pollutants, the

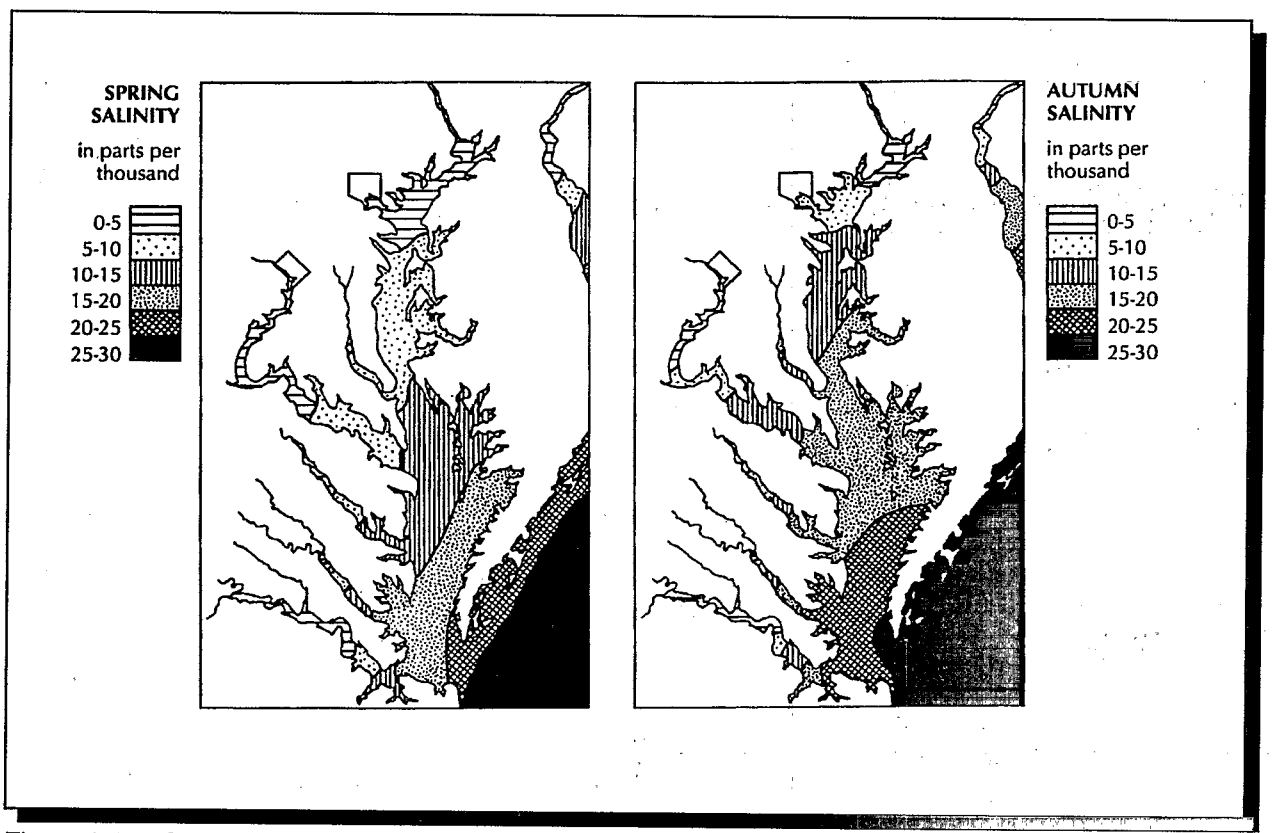


Figure 1-14. Chesapeake Bay salinity levels over time and space (Chesapeake Bay Program, 1995).

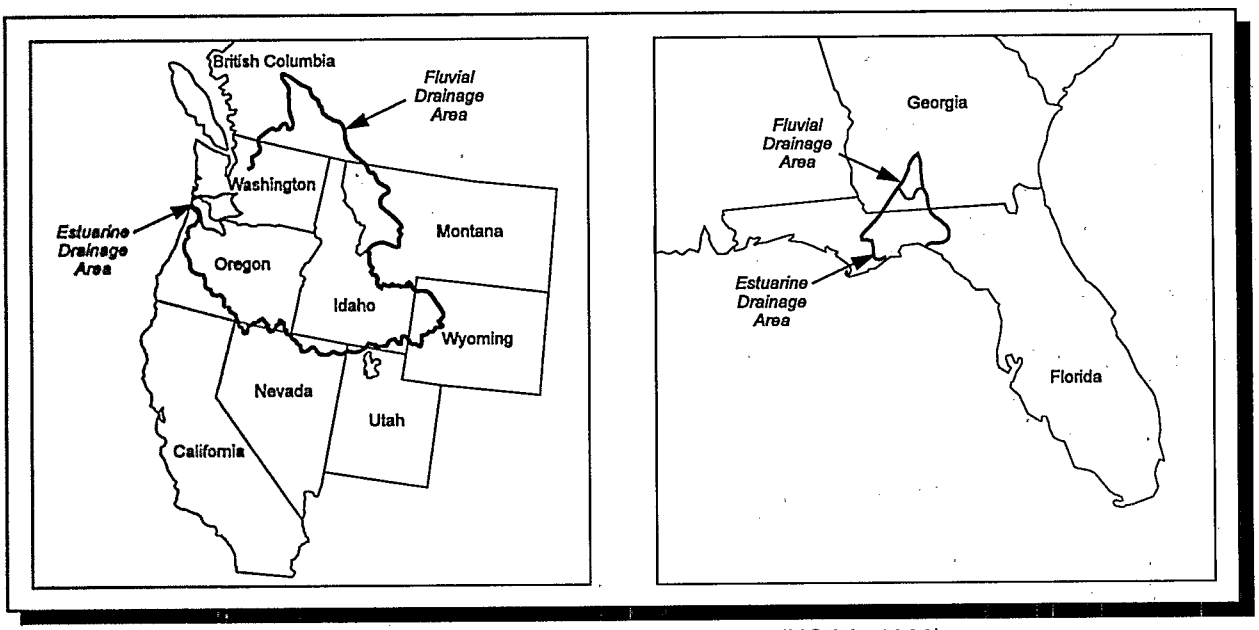


Figure 1-15. Estuarine drainage area versus fluvial drainage area (NOAA, 1990).

distribution of salinity, and the stratification of fresh water and salt water in an estuary (NOAA, 1990).

- Temperature profiles vary seasonally in estuaries.
- Freshwater input to estuaries varies seasonally, and spatial variability in estuaries is affected by the location of freshwater inflows.
- Due to spring runoff, salinity in estuaries is generally higher in fall and lower in spring (Chesapeake Bay Program, 1995).
- The earth's rotation (Coriolis effect), barometric pressure, and bathymetry (submerged sills and banks, islands) affect circulation and spatial variability in estuaries. For instance, Puget Sound contains numerous islands that affect circulation within it.

Stellwagen Bank, a submerged sand bank located across the boundary of Massachusetts Bay and the Atlantic Ocean, has a strong effect on circulation within the bay and within neighboring Cape Cod Bay.

In summary, the most important factors that determine the characteristics of individual estuaries are the sizes of the EDA and the FDA, water surface area, water volume, tidal range, salinity regime, and freshwater inflow. Also, an estuary might contain subestuaries—portions of a large estuary having definable subbasin drainage areas and constituting a significant percentage of either freshwater inflow or water surface area (NOAA, 1990). For instance, San Francisco Bay has very separate northern and southern reaches. The southern reach has a longer HRT, less inflow, and more sewage input, which give it characteristics very different from those of the northern reach and would require a different monitoring program design.

1.5.4 Open Coastal Waters

The major difference between open coastal waters and estuaries is that open coastal waters are not directly influenced by freshwater inflows. To design a monitoring program for open coastal

waters, knowledge of local salinity and circulation patterns is necessary. This helps to identify relatively discrete units of coastal water for monitoring purposes. In open coastal waters, it is particularly important to identify such discrete segments or units from which to sample in order to be able to track conditions over time. An example of a discrete segment or unit of open coastal water is a semienclosed embayment.

Features to consider when designing a monitoring program for open coastal waters include, but are not limited to, the following:

- Monitoring should focus on units for which there is reasonable likelihood that changes in water quality will result from BMP implementation.
- Consider segment or unit size and circulation patterns to estimate the likelihood of water quality improvements from pollution reduction efforts.
- Open coastal waters exhibit gradients in salinity, temperature, and water chemistry both spatially and temporally.
- Surface salinity varies with amount of rainfall input and evaporation (Tchernia, 1980).

1.5.5 Ground Water

Ground water is monitored less than surface waters because of the complexity of ground water systems and the difficulty of obtaining samples. However, ground water monitoring is important because of contamination of drinking water supplies and the

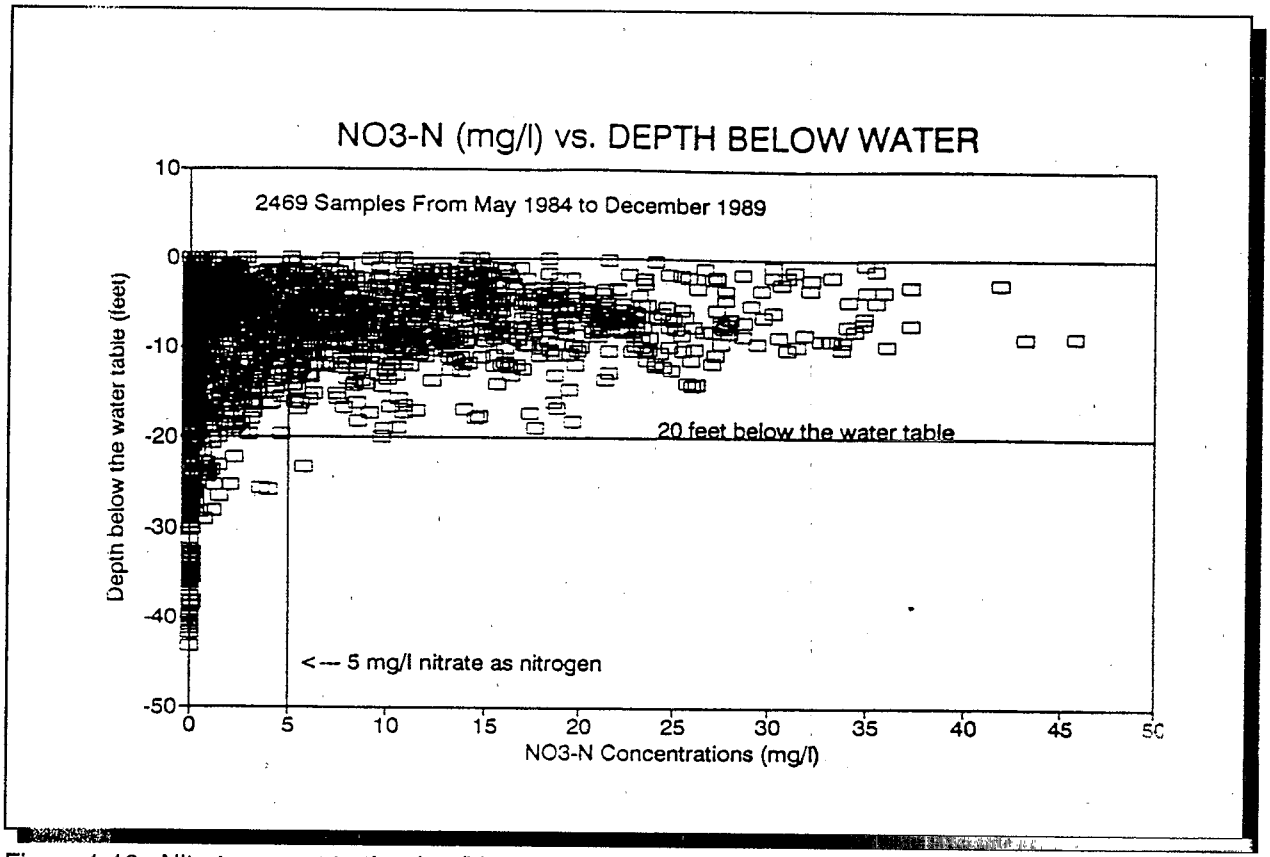


Figure 1-16. Nitrate concentration (mg/L) versus depth below water table (Goodman, 1991).

interplay between surface water quality and ground water quality. In 1990, ground water was the source of drinking water for over half of the Nation's population and for 95 percent of the population in rural areas (USEPA, 1995).The purposes of ground water monitoring include the following:

- To determine the ground water component of a hydrologic/chemical budget for a surface waterbody.
- To document the impact of a polluting activity.
- To identify background water quality.
- To identify trends and variations in water quality.
- To determine the effectiveness of BMPs.

Ground water monitoring often requires a two-stage approach. The first stage consists of a hydrogeologic survey to determine ground water surface elevations and flow directions. This survey requires numerous sampling locations. The second stage is an investigation of water quality, with stations selected based on the results of the first stage and monitoring objectives (Bishoff et al., 1995; Goodman et al., 1996; USDA-NRCS, 1996). More than with surface waters, site-specific information is absolutely necessary to design a ground water monitoring program. Water quality in an aquifer can vary considerably with depth and location (Figure 1-16).

Features to consider when designing a ground water monitoring program include, but are not limited to, the following:

- Local soils and geology.
- The direction of ground water flow.
- The type of ground water system. There are two general types of aquifers, confined and unconfined. Unconfined (water table) aquifers are in direct contact with the atmosphere through the soil. Confined (artesian) aquifers are separated from the atmosphere by an impermeable layer (USDA-NRCS, 1996).
- Selection of well locations depends on the variability of the aquifer's water quality and is complicated by the presence of confining beds, multiple aquifer systems, effects of pollutant density on pollutant transport, and changes in permeability.
- Spatial and temporal variabilities in aquifers cannot be generalized. Some respond to precipitation quickly, whereas others respond slowly.
- Sampling depth and depth to aquifer are important variables to consider in determining initial sampling frequency.

1.6 CLIMATE

Climate introduces elements of temporal variability into a monitoring program's design. When designing a monitoring program, the characteristics of seasonal variability, measurable as both the degree to which seasons are distinguishable (e.g., the difference between winter and summer in Alabama versus the difference in Maine), and the severity of the seasons (e.g., winter in Minnesota versus winter in South Carolina), must be considered. These characteristics determine when and for what amount of time specific variables can be monitored. For some monitoring objectives, year-round monitoring might be necessary, and in cold climates this could call for some means of heating if automated sample collection is to be used (USDA-NRCS, 1996).

Climate also affects the quantity, timing, and intensity of precipitation, the likelihood of catastrophic events (e.g., hurricanes, floods, droughts) that could interrupt sampling, seasonal variations in biological activity, and seasonal

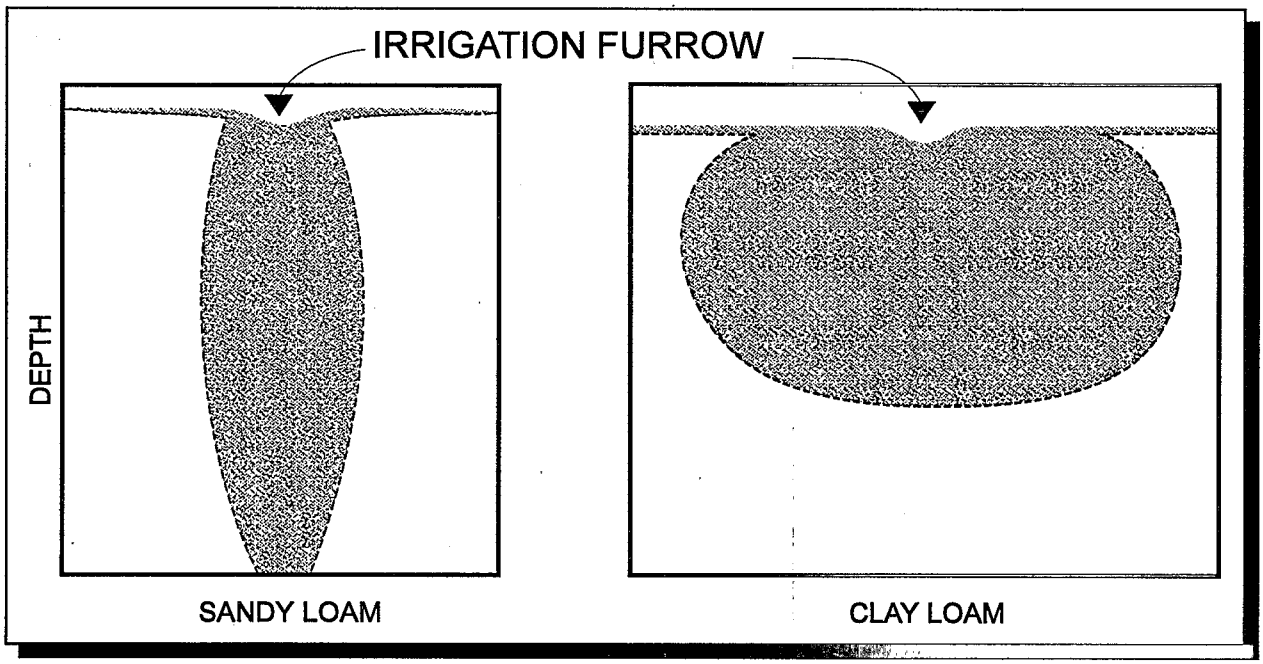


Figure 1-17. Comparison of water movement from irrigation furrows into two different soil types. The zone of infiltration represents water movement after 24 hours (after Brady, 1984).

variations in water quality parameters (e.g., addition of organic matter to streams and rivers from overhanging vegetation). These regional and local features must be factored into monitoring program design.

A few examples of climatic factors that affect monitoring program design follow (MacDonald et al., 1991):

- Variability in precipitation is inversely proportional to average annual precipitation. Drier areas tend to have more year-to-year variation in precipitation.

- Areas with more rainfall tend to have lower concentrations of nutrients and other dissolved ions.

- Climate affects the weathering rate, erosion rate, and vegetation type and the productivity of aquatic biota.

1.7 Soils, Geology, and Topography

Soils, geology, and topography are local or regional features that must be considered in monitoring program design (MacDonald et al., 1991). The permeability, depth, and porosity of soil and bedrock affect background levels of nutrients and dissolved ions in ground water and surface waters. The texture, depth, and permeability of soils also influence the quantities of fertilizers, pesticides, and herbicides that are leached into surface waters and the amount of leaching that occurs immediately after application and during runoff events (Figure 1-17). Slope is an important factor that must be considered when designing a monitoring program. Slope affects the rate and duration of runoff from a watershed, rate of erosion, depth of soil (steep slopes often have less soil overlying the bedrock), and stream characteristics. Slope also affects the likelihood of landslides and debris flow, erosional processes, and weathering rates.

2. DEVELOPING A MONITORING PLAN

2.1 INTRODUCTION

Since the relationship between public health and water quality began to influence legislation in the early 1900s, water quality management and its related information needs have evolved considerably. Today, the Intergovernmental Task Force on Monitoring Water Quality (ITFM, 1995b) defines water quality monitoring as an integrated activity for evaluating the physical, chemical, and biological character of water in relation to human health, ecological conditions, and designated water uses. Water quality monitoring for nonpoint sources of pollution includes the important element of relating the physical, chemical, and biological characteristics of receiving waters to land use characteristics. Without current information, water quality and the effects of land-based activities on water quality cannot be assessed, effective management and remediation programs cannot be implemented, and program success cannot be evaluated.

The most fundamental step in the development of a monitoring plan is to define the goals and objectives, or purpose, of the monitoring program. In the past, numerous monitoring programs did not document this aspect of the design process and the resulting data collection efforts led to little useful information for decision making (GAO, 1986; MacDonald et al., 1991; National Research Council, 1986; Ward et al., 1990). As a result, the identification of monitoring goals is the first component of the design framework outlined by the ITFM (1995b). In general, monitoring goals are broad statements such as "to measure improvements in Elephant Butte Reservoir" or "to verify nutrient load reductions into the Chesapeake Bay." Designing a monitoring plan also includes selecting sampling variables, a sampling strategy, station locations, data analysis techniques, the length of the monitoring program, and the overall level of effort to be invested. Figure 2-1 presents one approach for developing a monitoring plan.

Monitoring programs can be grouped according to the following general purposes or expectations (ITFM, 1995b; MacDonald et al., 1991):

- Describing status and trends
- Describing and ranking existing and emerging problems
- Designing management and regulatory programs
- Evaluating program effectiveness
- Responding to emergencies
- Describing the implementation of best management practices
- Validating a proposed water quality model
- Performing research

The remainder of the design framework outlined by the ITFM (1995b) includes coordination and collaboration, design, implementation, interpretation, evaluation of the monitoring program, and communication. Numerous guidance documents have been developed, or are in development, to assist resource managers in developing and implementing monitoring programs that address all aspects of the ITFM's design framework. Appendix A presents a review of more than 40 monitoring guidances for both point and nonpoint source pollution. These guidances discuss virtually every aspect of nonpoint source pollution monitoring, including monitoring program design and objectives, sample types and sampling methods, chemical and physical water quality variables, biological monitoring, data analysis and management, and quality assurance and quality control.

Once the monitoring goals have been established, existing data and constraints should be considered. A thorough review of literature pertaining to water quality studies previously conducted in the geographic region of interest

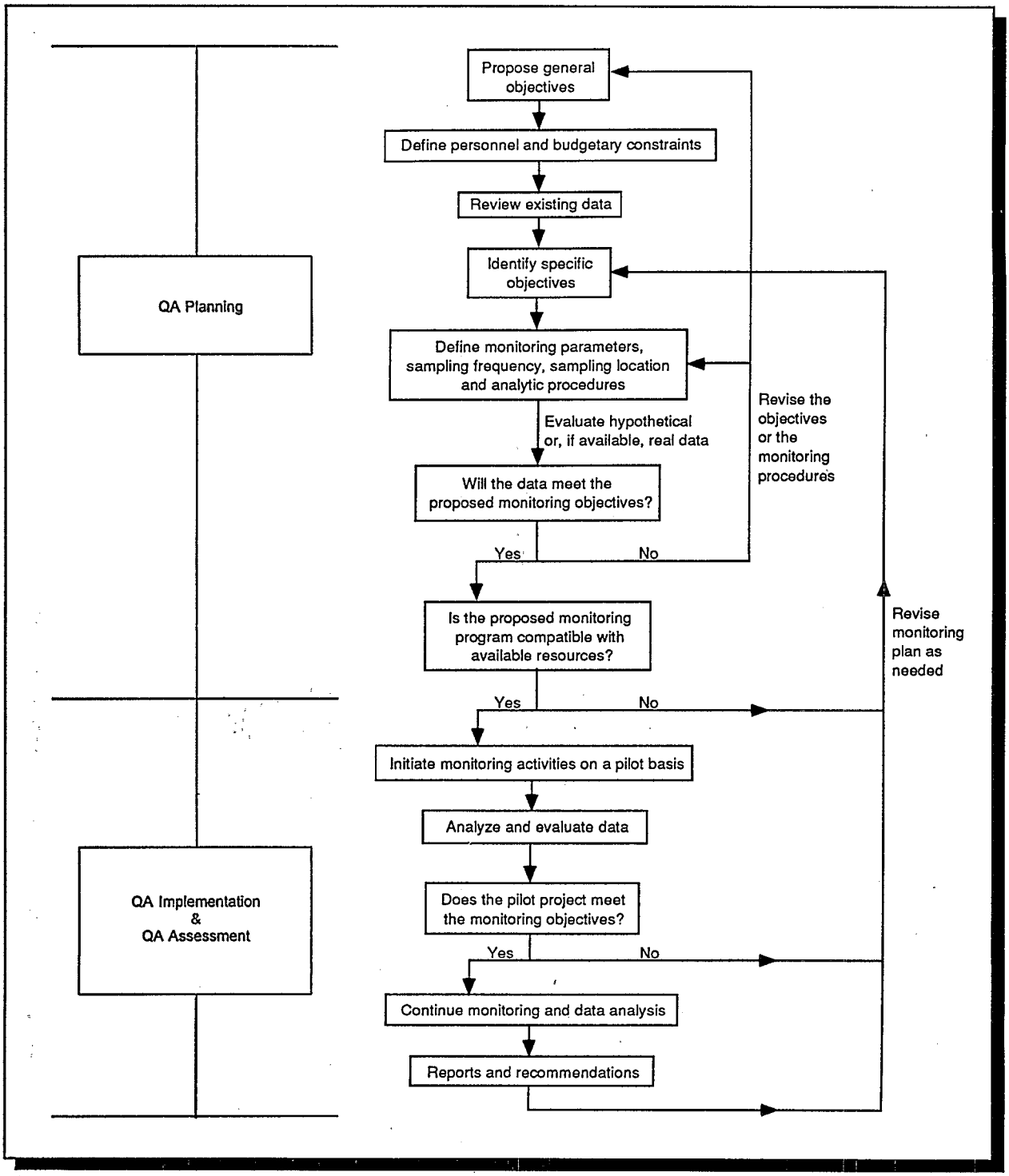


Figure 2-1. Development of a monitoring project (after MacDonald et al., 1991).

should be completed before starting a new study. The review should help determine whether existing data provide sufficient information to address the monitoring goals and what data gaps exist.

Identification of project constraints should address financial, staffing, and temporal elements. Clear and detailed information should be obtained on the time frame within which management decisions need to be made, the amounts and types of data that must be collected, the level of effort required to collect the necessary data, and the equipment and personnel needed to conduct the monitoring. From this information it can be determined whether available personnel and budget are sufficient to implement or expand the monitoring program.

As with monitoring program design, the level of monitoring that will be conducted is largely determined when goals and objectives are set for a monitoring program, although there is some flexibility for achieving most monitoring objectives. Table 2-1 provides a summary of general characteristics of various types of monitoring.

The overall scale of a monitoring program has two components—a temporal scale and a geographic scale. The temporal scale is the amount of time required to accomplish the program objectives. It can vary from an afternoon to many years. The geographic scale can also vary from quite small, such as plots along a single stream reach, to very large, such as an entire river basin. The temporal and geographic scales, like a program's design and monitoring level, are primarily determined by the program's objectives. Hence, unspecific or unclear monitoring objectives present a barrier to selecting the appropriate temporal and geographic scales.

If the main objective is to determine the current biological condition of a stream, sampling at a few stations in a stream reach over 1 or 2 days might suffice. Similarly, if the monitoring objective is to determine the presence or absence of a nonpoint source impact, a synoptic survey might be conducted in a few select locations. If the objective is to determine the effectiveness of a nutrient management program for reducing nutrient inputs to a downstream lake, however, monitoring a subwatershed for 5 years or longer might be necessary. If the objective is to calibrate or verify a model, more intensive sampling might be necessary.

Depending on the objectives of the monitoring program, it might be necessary to monitor only the waterbody with the water quality problem or it might be necessary to include areas that have contributed to the problem in the past, areas containing suspected sources of the problem, or a combination of these areas. A monitoring program conducted on a watershed scale must include a decision about a watershed's size. The effective size of a watershed is influenced by drainage patterns, stream order, stream permanence, climate, number of landowners in the area, homogeneity of land uses, watershed geology, and geomorphology. Each factor is important because each has an influence on stream characteristics, although no direct relationship exists.

There is no formula for determining appropriate geographic and temporal scales for any particular monitoring program. Rather, once the objectives of the monitoring program have been determined, a combined analysis of them and any background information on the water quality problem being addressed should make it clear what overall monitoring scale is necessary to reach the objectives.

Other factors that should be considered to determine appropriate temporal and geographic

Table 2-1. General characteristics of monitoring types .

Type of Monitoring	Number and Type of Water Quality Parameters	Frequency of Measurements	Duration of Monitoring	Intensity of Data Analysis
Trend	Usually water column	Low	Long	Low to moderate
Baseline	Variable	Low	Short to medium	Low to moderate
Implementation	None	Variable	Duration of project	Low
Effectiveness	Near activity	Medium to high	Usually short to medium	Medium
Project	Variable	Medium to high	Greater than project duration	Medium
Validation	Few	High	Usually medium to long	High
Compliance	Few	Variable	Dependent on project	Moderate to high

Source: MacDonald et al., 1991.

scales include the type of water resource being monitored and the complexity of the nonpoint source problem. Some of the constraints mentioned earlier, such as the availability of resources (staff and money) and the time frame within which managers require monitoring information, will also contribute to determination of the scales of the monitoring program.

2.2 *Monitoring Objectives*

Identifying and concisely stating the monitoring objectives are critical steps in the development of a monitoring program. Unlike monitoring goals, monitoring objectives are more specific statements that can be used to complete the monitoring design process including scale, variable selection, methods, and sample size (Plafkin et al., 1989; USDA-NRCS, 1996). Monitoring program objectives must be detailed enough to allow the designer to define precisely what data will be gathered and how the resulting information will be used. Vague or inaccurate statements of objectives lead to program designs that provide too little or too much data, thereby failing to meet management needs or costing too much.

Monitoring programs can be implemented for one or many reasons. The more common types of monitoring program objectives are summarized below. The emphasis of this guidance is on

evaluation monitoring, but information contained herein might also be used to address other types of monitoring. The reader is cautioned that even though two different monitoring programs might share some objective listed below, their designs can be radically different.

2.2.1 Monitoring Objective Category: Problem Definition

(1) Determine whether an impairment exists

Meeting this objective involves an investigation of key parameters to determine the general condition of a habitat or water quality. Measurements of individual pollutants in waterbodies are often taken to determine whether violations of water quality standards are occurring. Biological monitoring is also useful when evaluating whether designated uses are supported. Monitoring associated with this type of objective might reveal that a suspected problem is more complicated or serious than originally thought and that more intensive monitoring studies will be necessary.

(2) Determine the extent of the impairment

Even if a problem is known to exist, the geographic and temporal extent of the problem might not be known. Does the problem affect a stream reach, or does the problem extend to the downstream lake? Some pollution sources are emitted only during certain parts of the year or in association with certain events, such as storms, or might be a problem only during a particular time of the year, such as fish spawning season. Determining the geographic and temporal aspects of a pollution problem will help focus management on BMP systems that will have the most benefit.

(3) Determine the causes and sources of impairment

Monitoring might be required to determine the cause of an environmental problem, such as degraded fish habitat or an algal bloom. Determining the pollution's source is often more difficult than determining its presence because there are often many potential sources whose influences overlap. When conducting monitoring for this purpose, it is important to monitor the appropriate water quality characteristics and account for climatic factors to establish a cause-and-effect relationship, even though it might be difficult to prove.

Point and nonpoint sources often affect the same waterbody, and monitoring might also be required to determine the contribution and relative importance of each to water quality impairment. It might also be necessary to determine which areas are the most critical in causing waterbody impairment. For instance, a high erosion rate on land far from a receiving waterbody might have a lower pollution-causing potential than an area with a lower erosion rate near to a receiving waterbody. Factors such as the timing of pollutant contributions relative to the hydrologic cycle of the waterbody and the ecology of the biological communities must be factored into the analysis. In addition, the distance of pollutant sources from receiving waters, the fate and transport of pollutants from different sources, the magnitude of pollutant contributions from each source, and the distance to the impaired resource of concern (as distinguished from distance to a point of entry into a receiving waterbody, which might be some distance from the actual impairment) should be considered. This type of information can often be used in developing load allocations for nonpoint pollution sources and wasteload allocations for point sources, although extensive monitoring might be required.

2.2.2 *Monitoring Objective Category: Model Development*

(1) Calibrate models

Model calibration is the first stage of testing a model and tuning it to a set of field data. Field data are necessary to guide the modeler in choosing the empirical coefficients in a model before the model can be used to predict the effect of management techniques or activities.

(2) Validate models

Model validation involves the testing of a model using a second set of field data. In most cases, the second set of field data should represent an independent data set that extends the range of conditions for which the model is valid. If an independent data set is not available, a set of randomly selected data should be used for validation. Once a model has been validated, it can be used to assist managers with management decisions within the range of the calibration and validation data sets.

2.2.3 *Monitoring Objective Category: Evaluation (emphasis of this guidance)*

(1) Measure the effectiveness of best management practice (BMP) systems

Individual BMPs or groups of BMPs are monitored to determine the extent of pollution control. Monitoring for individual BMPs can typically be conducted at a plot or field scale, whereas monitoring for BMP systems is usually conducted on a watershed scale because the combined effect of a few or several BMPs is being investigated. Studies of some individual practices can be conducted in a relatively short time (less than 5 years), while others might take longer. Evaluation of BMP systems is typically conducted over a long term (more than 5 years) because BMP implementation can take years to affect water quality. This type of monitoring is difficult due to the presence of pollutant reserves in soil and sediments, the effect of many land uses within a study area, the variety of approaches that landowners use to implement similar systems of BMPs, and the need to track land management as well as water quality and climatic variables.

(2) Analyze trends

The objective here is to answer the question, "Is water quality changing over time?" Baseline monitoring is part of trend analysis because establishing a baseline is essential to analyzing trends. However, baseline monitoring is generally thought of as determining a condition prior to pollutant entry or prior to a change in waterbody condition, whether beneficial or detrimental. Controlling for influencing factors such as climate is necessary if baseline monitoring is to be used as a reference point for trend analysis and management decisions. The ability to relate water quality changes to changes in land management depends on the quality and quantity of data collected on land management practices.

2.2.4 *Monitoring Objective Category: Conduct Research*

Research monitoring is done to address specific research questions. Research monitoring is usually conducted on a plot scale, is well controlled, and is limited to a very specific question. Monitoring and data analysis techniques for research and for other types of monitoring are often very similar, and the difference between them is often one of objective rather than approach. A critical examination of articles about relevant and well-conducted research projects in which monitoring is a key element can provide excellent guidance for the design of a monitoring program.

2.3 Data Analysis and Presentation Plans

Ward et al. (1990) point out that one of the most important and difficult tasks is to identify what information is to be produced by the monitoring effort. It is particularly critical to ensure that policy makers and other stakeholders know the type of information that a monitoring program can produce and that realistic monitoring program expectations are developed. Ward et al. (1990) identify key steps to ensure that realistic expectations are placed on the monitoring program and the associated data analysis:

- Perform a thorough review of the legal basis for the management effort and define the resulting "implications" for monitoring.
- Review the administrative structure and procedures developed from the law in order to define the information expectations of the management staff.
- Review the ability of the monitoring program to supply information.
- Formulate an information expectations report for the monitoring system.
- Present the information expectations report to all users of the information.
- Develop consensus as to an agreeable formulation of information expectations and related monitoring system design criteria.

This process is typically performed as an iterative process that involves the technical staff and the decision makers who developed the monitoring objectives. To develop an information expectations report, the data analyst might need to have formal meetings, develop questionnaires, and conduct interviews to learn what the managers need. In some cases this iterative process might require modifying or redesigning the monitoring program. The data analyst should remember that complete consensus might not be possible.

When developing an information expectations report, the presentation of results should be selected depending on the audience reviewing the information and the objectives of the monitoring program. How quickly must information be presented to information users? To what kind of information and how much information do the decision makers respond favorably? At a minimum, the data analyst should prepare example report formats to be approved by the decision makers, keeping in mind that "a picture is worth a thousand words." In all cases, the goal should be to present clear and accurate information that is not subject to misinterpretation. Ward et al. (1990) present an example outline (Figure 2-2) of what might be considered in an expectations report. (The data analyst should modify this outline to suit individual needs.)

2.4 Variable Selection

In these days of increasing monitoring and evaluation needs and relatively small monitoring and evaluation budgets, it is extremely important for program managers to design efficient monitoring and evaluation programs. The variables selected for a monitoring program should be tied directly to the monitoring objectives. It is often the case that some variables in addition to those of prime interest are monitored because they are relatively cheap to monitor and might provide some useful information for purposes not yet outlined. This is generally reasonable, but the technical staff should (1) anticipate these undefined purposes so that the extra variables are monitored in a manner that yields useful information (e.g., support statistical analyses) and (2) make sure the extra cost associated with monitoring additional variables does not preclude necessary expansions

Expectations Report Outline

- Evolution of Water Quality Management Program
  - Geographical/Hydrological Setting
  - Water Quality Problems
  - Water Quality Laws
  - Management Program Structure
  - Management Procedures
- Information "Expected" by Management Program
  - Implications of the Law Establishing the Program
  - Legal Goals
  - Management Powers and Functions
  - Monitoring Requirements Directly Stated
  - Information Needs of Management Operations
  - Water Quality Criteria
  - Water Quality Standards
  - Permits
  - Compliance
  - Enforcement
  - Construction Loans
  - Planning
  - Water Quality Assessment
- Ability of Monitoring Systems to Produce Water Quality Information
  - Narrative Information
  - Numerical Information-Data
  - Graphical Information
  - Statistical Information
  - Average Conditions
  - Changing Conditions
  - Extreme Conditions
  - Water Quality Indices
- Suggested Information Expectations for Monitoring System
  - Management Information Goal(s)
  - Definition of Water Quality
  - Monitoring System Goal(s)
  - Information Product of Monitoring System
  - Narrative
  - Graphical
  - Statistical
- Resulting Monitoring Network Design Criteria
  - Variable Selection
  - Site Selection
  - Sampling Frequency Determination

Figure 2-2. Expectations report outline (Ward et al., 1990).

or extensions of the monitoring and evaluation program for the variables of prime interest.

In many instances the water quality problem will directly indicate what variables should be monitored. For example, a dissolved oxygen problem would strongly suggest monitoring of dissolved oxygen. (Typically, biochemical oxygen demand, sediment oxygen demand, temperature, and nutrients would be monitored as well.) Or, if the goal is to assess the impact of nonpoint source controls in terms of standards violations, then the variables selected should be those required for the analysis of standards violations.

In some cases, it might be more beneficial to use surrogate measures instead of the variables mentioned in the monitoring goals and objectives. In these cases, objectives for the surrogates that are consistent with the overall monitoring and evaluation goals should be established. The key to using surrogate measures is to be certain that a reliable relationship exists between the true measure and the surrogate measure. For example, if the objective is to monitor the condition of salmon spawning areas, surrogate measures are necessary because the condition of salmon spawning areas is a composite of many factors. Good surrogate variables would be stream bank undercut, embeddedness, and vegetative overhang (Platts et al., 1983). The corresponding surrogate goals could be to reduce cobble embeddedness and to increase vegetative overhang to appropriate levels for salmon spawning. The monitoring goals would then be to document changes in cobble embeddedness and vegetative overhang.

Poor surrogate selection results when a known relationship between the monitoring goals and objectives and the chosen surrogate measures does not exist. For example, a poor surrogate for estimators of sediment delivery to water resources is the unqualified use of erosion rates. Without the existence of a known relationship between these two measures (i.e., sediment delivery ratio), the surrogate will produce misleading results.

Variable selection should also reflect the nonpoint source data analysis and presentation plan. For example, if the plan involves data normalization or grouping prior to data analysis, the variable list should include those variables used to normalize and/or group the data. Some analyses might require discrete observations, whereas others might use continuous data. All monitoring sites should be characterized sufficiently for meaningful data interpretation, including georeferencing. For surface water sites the relevant information may include waterbody name, river reach number and milepoint, location, prevailing winds, shading, bottom sediment, elevation, slope, stream width and depth, drainage area, upstream land use, lake depth, and more. In the case of ground water monitoring, this information includes the aquifer tapped by a well, the depth of the well, the type of well construction, and the well elevation (USGS, 1977). Water level measurements should be included in all ground water studies.

Since there are numerous variables to choose from but monitoring budgets are limited, some method to prioritize variable selection is often necessary. When available, existing data should be used to guide variable selection. Further discussion on variable selection, prioritization, and optimization are provided by USDA-NRCS (1996), MacDonald et al. (1991), and Sherwani and Moreau (1975). In some cases, optimal variable selection is not possible, perhaps due to lack of local data. In such cases, the researcher might need to rely on professional judgment and the review of monitoring programs of similar nature and scope.

Some data requirements for nonpoint source monitoring and evaluation efforts can be met using nationally available data sources. Appendix

B describes some of these data sources and includes information for those interested in accessing the data. Other data sources are available to nonpoint source professionals as well, and state, regional, or local sources of data in particular should be investigated. State agriculture, forestry, and other environmental agencies; counties; municipalities; and state and local health departments are likely sources of water quality, health-related, and land use data and information. Regional planning commissions, local universities, and environmental consultants might also be able to provide data. The sources summarized in Appendix B focus on the major data sources made available to EPA or known to reviewers of this document. The remainder of this section summarizes key data that would normally be considered in a nonpoint source monitoring program.

2.4.1 Physical and Chemical Water Quality Data

Physical and chemical water quality data are essential to almost all nonpoint source monitoring and evaluation efforts, due to the relationships between flow and pollutant characteristics. For example, it might be necessary to establish watershed water budgets so that the location and magnitude of nonpoint sources or background sources can be determined. In other cases, the extent of the floodplain might prove critical to assessments of BMP control needs. Important physical and chemical water quality variables to monitor include flow (streams), temperature, transparency, suspended sediment, sedimentation rate, dissolved oxygen, pH, conductivity, alkalinity/acid neutralizing capacity (lakes), and nutrients. Other factors, such as cobble embeddedness, woody debris, and salinity, might be important depending on type of water body and monitoring goals.

2.4.2 Biological Data

Biological data can be very useful for evaluating water resource impairment due to nonpoint source impacts because aquatic organisms integrate the exposure to various nonpoint sources over time. Measures of biological communities integrate the effects of different pollutant stressors—excess nutrients, toxic chemicals, increased temperature, excessive sediment loading, and others—and thus provide an overall measure of the aggregate impact of the stressors. Monitoring changes in aquatic communities over time can serve as a measure of improvement due to BMPs. The biological survey approach used depends on waterbody type, i.e., stream, river, lake, wetland, or estuary. Important biological parameters to monitor include bacteria, algal biomass, macrophyte biomass and location, macroinvertebrates, and fish populations.

2.4.3 Precipitation Data

Precipitation data, including total rainfall, rainfall intensity, storm interval, and storm duration, have proven to be key to successful interpretation of nonpoint source data in the Nationwide Urban Runoff Program (NURP), Model Implementation Program (MIP), and Rural Clean Water Program (RCWP) studies. By combining precipitation data with pollutant loading evaluations, it has been found that a few storms can account for a large proportion of the total annual pollutant load. Johengen and Beeton (1992) found that, in the Saline Valley RCWP, a few storms accounted for more than 50 percent of the annual loading. Interestingly, they found that initial estimates of suspended solids and phosphorus loadings were only 20 and 50 percent of loadings estimated by adjusting for daily precipitation. The project-mandated weekly sampling had missed the loading spikes that lasted for only a few days.

Research has shown that average annual soil loss can be estimated using only a few site-specific

factors, among which is a rainfall-runoff erosivity factor (R). The other factors used to estimate soil loss are soil erodibility, topography, and land use and management. The Universal Soil Loss Equation (USLE) has been revised and is now known as the Revised USLE (RUSLE), based on research by Renard et al. (1991) and Wischmeier and Smith (1978). The rainfall-runoff erosivity factor found in the RUSLE is also used in several nonpoint source models, including the Agricultural Nonpoint Source Pollution Model (AGNPS) (Young et al., 1985). The Water Erosion Prediction Project (WEPP) Hillslope Profile version erosion model is a "new generation" soil erosion model that can be run both as a continuous simulation model and on a single-storm basis. The model requires a large number of data on management practices, which might be difficult to obtain (Singh and Fiorentino, 1996). A procedure derived from the NURP program uses storm frequency and other factors to determine recurrence intervals for instream pollutant concentrations resulting from urban nonpoint source pollution (USEPA, 1984b).

2.4.4 Land Use Data

Landuse data include information on treatments applied to land, current and historical use of the land, spatial and temporal information on land use activities, and changes in land use made before and during a project. Data on these elements are important for evaluating correlations between land surface activities and water quality. Establishing a correlation between a change in water quality and a change in land treatment must be based on both the detection of a water quality trend and detailed information on changes in land use or management, and it requires rigorous statistical analysis (Goodman, 1991; Meals, 1991, 1992). Land treatment can be linked to water quality impacts at the field, subwatershed, watershed, or project level. In general, the larger the drainage area, the harder it is to associate land treatment and water quality. Subwatershed monitoring is the most effective means for demonstrating water quality improvements from a system of BMPs because at this scale the confounding effects of external factors, other polllutant sources, and other BMPs or BMP systems are minimized (Coffey et al., 1993).

Two key points must be considered in nonpoint source monitoring with respect to linking water quality and land treatment. First, weather and season are important confounding influences on nonpoint source activities because they strongly influence the types of land-based activities that can occur, and hence the timing and quantity of runoff from treated lands and the consequential water quality effects. Second, spatial variation must be considered. The location of land treatments relative to surface waters is likely to vary from year to year, and this adds variation to the effect of land treatment on water quality (Meals, 1991).

Correlations between water quality and land treatment can be made much more easily if land use and land treatment monitoring are considered as part of monitoring design in a project's preliminary stages. It is also very important to control for the effects of hydrologic variation. Paired regression is an effective method to control for background variability and is recommended (Meals, 1991, 1992).

Geographic information systems (GIS) are effective management tools for land use data (Meals, 1991). They allow for tracking and manipulating spatial land use data and remarkably improve the ease of visual inspection and comprehension of the data. Data for GIS are available from a variety of sources, including state agencies, GIS user groups, GIS vendors, universities, consultants, conferences, and numerous publications dedicated to GIS topics (Griffin, 1995).

2.4.5 Topographic Data

Topographic data are also required for many nonpoint source monitoring and evaluation efforts, particularly when soil erosion, water runoff, and sedimentation are estimated with models. For example, the USLE includes both slope length and slope steepness factors (Wischmeier and Smith, 1978). AGNPS input includes a slope shape factor, field slope length, channel slope, and channel side slope (Young et al., 1985).

2.4.6 Soil Characteristics Data

Other data such as soil chemistry and soil physical characteristics might be required for some monitoring and evaluation efforts. Recent approaches to assessing the potential for ground water contamination from nonpoint sources have emphasized the need for data such as hydrologic soil group, soil organic carbon content, depth to water, net recharge, aquifer media, and vadose zone characteristics (Aller et al., 1985; Dean et al., 1984).

2.5 Program Design

Numerous program designs can be used to evaluate the monitoring objectives identified earlier in this chapter. To select the program design, the researcher should develop clear, quantitative monitoring objectives; understand the watershed or waterbody to be monitored; and know something about the locations of and pollutant transport from point and nonpoint sources. In developing the information expectations report described earlier in this chapter, the technical staff will typically decide whether parameter estimation or hypothesis testing is the primary evaluation tool. This choice has an impact on the program design. As an example, balanced designs (e.g., two sets of data with the same number of observations in each set) are generally more desirable for hypothesis testing, whereas parameter estimation might require unbalanced sample allocations to account for spatial and temporal variabilities (Gaugush, 1986). Hypothesis testing is likely to be used in a program evaluation (e.g., water quality before and after pollution controls are implemented), whereas parameter estimation can be applied in assessments when determining pollutant loads from various sources. Hypothesis testing will typically require more intensive databases than those needed for objectives that entail general water quality assessments. As a result, the sampling methodologies required to meet different objectives for the same waterbody may differ considerably.

Most monitoring programs are based on either a probabilistic or a targeted design, or some combination of the two. Probabilistic designs include random selection of station locations and/or sampling events to provide an unbiased assessment of the waterbody. In targeted designs, monitoring sites are selected based on known existing problems or knowledge of upcoming events in the watershed such as installation of a BMP. The most common types of targeted designs employed for the evaluation of nonpoint source pollution sources and BMP systems include monitoring single watersheds, nested watersheds (e.g., above-and-below implementation), two watersheds, paired watersheds, multiple watersheds, and trend stations. Statistical procedures to analyze the data from these study designs are presented in Chapter 4.

Simply identifying the site location and sampling frequency is not sufficient to describe the where and when of sampling programs. Additional considerations include the depth of sampling, the origins of the aliquot(s) taken in each sample bottle, the time frame over which measurements are made, and others. For example, if a stream is well mixed, a single grab sample from the center of the stream might be sufficient, whereas it

Example Objective: *Determine the annual loading of phosphorus from a watershed with no point sources.*

Sampling Methodology: *Assuming no snowmelt inputs and that the majority of phosphorus is delivered under high-flow conditions, the investigator should perform flow-proportional sampling during events. This, of course, assumes that a stage-discharge relationship has been established. Vertical and horizontal concentration and flow profiles should be assessed to determine the need for transect and/or depth-integrated sampling.*

might be more appropriate to take an integrated sample from a wider stream. In deeper estuaries, it is a common practice to collect samples near the top and bottom of the waterbody as well as just above and just below the pycnocline. Frequency of sampling should be based on several factors (Sherwani and Moreau, 1975):

- Response time of the system
- Expected variability of the parameter
- Half-life and response time of constituents
- Seasonal fluctuation and random effects
- Representativeness under different conditions of flow
- Short-term pollution events
- Variability and types of the inputs
- Magnitude of response

Examples of sample type classifications include instantaneous and continuous; discrete and composite; surface, soil profile, and bottom; time-integrated, depth-integrated, and flow-integrated; and biological, physical, and chemical. Several existing guidance manuals (Brakensiek et al., 1979; Koterba et al., 1995; Lapham et al., 1995; Platts et al., 1983; Scalf et al., 1981; Shelley, 1979; Shelton, 1994; Shelton and Capel, 1994; USDA-NRCS, 1996; USEPA, 1978b, 1981, 1987a; USGS, 1977) and other reference materials (Wetzel and Likens, 1979) describe these various sample types and the equipment used to collect them.

Selecting an appropriate sampling design for nonpoint source monitoring and evaluation efforts can be a complicated and frustrating experience for the program manager. In addition to balancing multiple (and sometimes competing) objectives, program managers must contend with large variabilities in measured parameters. These variabilities are caused by several factors, including distance to the pollutant source; nonuniform distribution of the pollutant due to physical, biological, or chemical influences; buildup or degradation over time; temporal and spatial variation in background levels; diversity in the biological community; and other nonuniformities such as those in topology, climatic conditions, and waterbody geometry. These factors, in turn, make collecting accurate and unbiased environmental samples more difficult. Biased samples are those which result in consistently higher or lower values than what exists in the waterbody. For example, suspended solids samples taken only during base flow conditions will most likely result in low estimates of annual solids loadings. Accuracy is a measure of how close the sample value is to the true population value. It is necessary to design sampling efforts that meet accuracy requirements while not placing unreasonable burdens on personnel or budgets. Data that are biased or do not meet the project's accuracy requirements are of little use to program managers. An exception might be volunteer data, which often do not meet accuracy requirements but are highly useful in gaining public support for projects.

Other types of sampling uncertainty include random sampling errors and gross errors. Random sampling errors arise from the variability of population units (Gilbert, 1987) and explain why the sample means from two surveys are

never equal. Gross mistakes can occur at any point in the process beginning with sample collection and ending with the reporting of study results. Adherence to accepted sampling and laboratory protocols combined with thorough quality control and data screening procedures and experience, dedication, and care will minimize the chances for gross errors.

2.5.1 Probabilistic Designs

In a probabilistic sampling program, the entity about which inferences are made (e.g., watershed) is the population or target population and consists of population units. The sample population is the set of population units that are directly available for measurement. As an example, in a watershed impacted by nonpoint sources, the target population could be defined as storm-event dissolved phosphorus concentrations at the inlets to all impoundments, and phosphorus concentrations in 1-liter grab samples could be population units. Note that both spatial and temporal limits of the water quality variable should be established in defining the target population (Gaugush, 1986). This focuses the sampling program better, in this case eliminating the need to monitor at upstream and in-lake sites, and during baseflow conditions. As a further refinement, the technical staff may define the population units as the dissolved phosphorus concentrations in half-hour composite samples taken during all storms. By sampling and statistically evaluating selected population units, inferences can be made about the entire waterbody.

Simple random sampling

In simple random sampling, each unit of the target population has an equal chance of being selected (Figure 2-3). This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population (Gilbert, 1987). Random sampling can be applied in a variety of ways, including site selection along the length of a river or areally throughout a lake. Samples may also be taken at a single station using random time intervals. The number of random samples required to achieve a desired margin of error when estimating the mean is (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2,n-1}s/d)^2}{1 + (t_{1-\alpha/2,n-1}s/d)^2/N} \qquad (2\text{-}1)$$

where

n = number of samples,
t = Student's t value,
s = sample standard deviation,

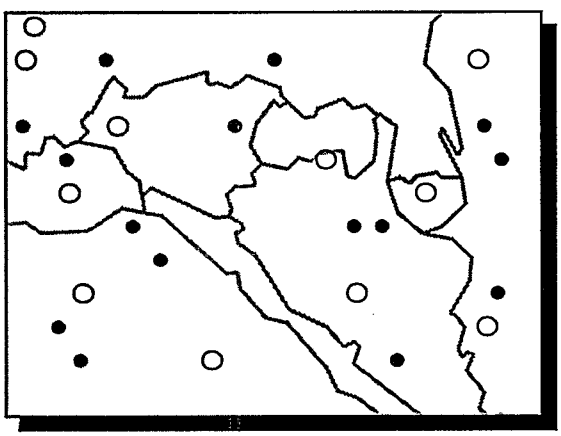


Figure 2-3. Simple random sampling for silviculture. Dots represent harvest sites. All harvest sites of interest are represented on the map, and the sites to be sampled (open dots—O) were selected randomly from all harvest sites on the map. The shaded lines on the map could represent county, watershed, hydrologic, or some other boundary, but they are ignored for the purposes of simple random sampling.

d = absolute margin of error,
N = number of population units, and
α = confidence interval.

If N is large, the above equation can be simplified to

$$n = (t_{1-\alpha/2,n-1}s/d)^2 \qquad (2\text{-}2)$$

Since the Student's t value is a function of n, both of the above equations are applied iteratively. If the population standard deviation is known, rather than estimated, Equation 2-2 can be further simplified to

$$n = (Z_{1-\alpha/2}\sigma/d)^2 \qquad (2\text{-}3)$$

where Z is the standard normal deviate and σ is the population standard deviation. In most cases, N is large enough to apply Equation 2-2 or 2-3. Values of Z and t can be found in Appendix D.

Suppose, for example, that the monitoring objective is to estimate the mean dissolved orthophosphate concentration (mg/L as P) during August in a waterbody segment such that there is a 95 percent chance that the mean concentration is within ± 0.025 mg/L of the estimated mean. Assuming a population standard deviation of 0.05 mg/L, the number of samples can be estimated using Equation 2-3 as

$$\left(1.96 \times \frac{0.05}{0.025}\right)^2 = 15.4 \approx 16 \ \textit{samples}$$

In most cases the standard deviation is not known and Equation 2-2 would be applied. Intuitively, more samples are required due to the uncertainty associated with the standard deviation. To apply Equation 2-2, it is reasonable to initially assume that n is equal to some value greater than 16, say 18, which will correspond to a t statistic of 2.110. Substituting the above values into Equation 2-2 where the standard deviation now refers to the sample standard deviation yields

$$\left(2.110 \times \frac{0.05}{0.025}\right)^2 = 17.8 \approx 18 \ \textit{samples}$$

Since the computed 18 samples correspond to the initial assumption, no iterations are necessary. In practice, this type of analysis would be performed for several variables and a judgment between sampling size, allowable error, and cost would be made.

Applying any of these equations is difficult when no historical data set exists to quantify the standard deviation. To estimate the population standard deviation, Cochran (1977) recommends four sources:

- Existing information on the same population or a similar population.
- Informed judgment, or an educated guess.
- A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of the second step. Use data from both steps to estimate the final precision of the characteristic(s) sampled.
- A "pilot study" on a "convenient" or "meaningful" subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because the pilot sample often is not representative of the entire population to be sampled.

Gilbert (1987) and Cochran (1977) address additional aspects of simple random sampling. Included in these texts are estimation of the mean and total for sampling with and without

Example Objective: *Determine the monthly mean total suspended solids concentration (to within ± 15 mg/L at the 95 percent confidence level) for a tributary from an agricultural watershed.*

Sampling Methodology: *Since the concentration may vary with stream depth, width, and flow, the investigator should select a site that is well mixed so that a single grab sample can be taken. If a well-mixed site cannot be found, an integrated sample would be required. Samples would be collected during high and low flow conditions to obtain a representative mean. Random or stratified random samples would then be collected as grab or composite samples depending on the averaging time selected.*

replacement, equations for determining the number of samples required for both independent and correlated data, and the impact of measurement errors. In most cases, environmental sampling is done without replacement (e.g., aliquots of stream water are not placed back into the stream), N is relatively large, the samples are assumed to be independent, and measurement error is ignored, thus making many of these specialized cases less critical. However, the reader should be aware that these issues might become paramount depending on the monitoring objectives and sampling design.

Stratified Random Sampling

In stratified random sampling, the target population is divided into groups called strata for the purpose of obtaining a better estimate of the mean or total for the entire population (Figure 2-4). Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables (e.g., season, flow condition) to group observations into more units that reduce the variability of observations within each unit. As an example, stratified random sampling can be used to evaluate chemical concentrations in waterbodies when evaluating nonpoint source loadings. One approach would be to stratify stream flow into base and various storm flow periods to account for the energy relationship between precipitation and pollutant generation. Random sampling would then be performed in each stratum.

Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a linear trend, followed in

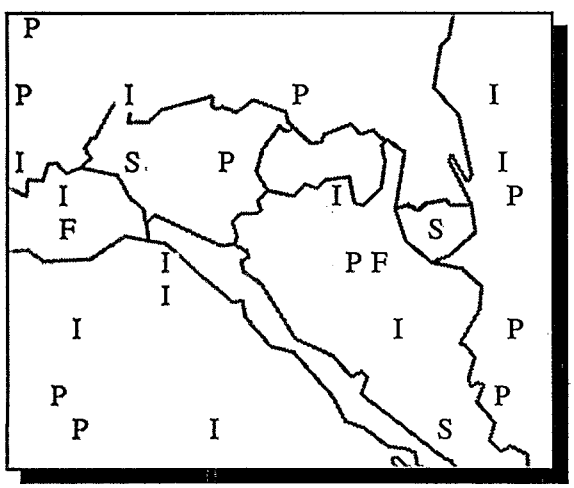


Figure 2-4. Stratified random sampling for silviculture. Letters represent harvest sites, subdivided by type of ownership (P = private nonindustrial, I = industrial, F = federal, S = state). All harvest sites of interest are represented on the map. From all of the sites in one ownership category, sites were randomly selected for sampling (highlighted sites). The process was repeated for each ownership category. The shaded lines on the map could represent county, soil type, or some other boundary, and could have been used as a means for separating the harvest sites into categories for the sampling process.

order by systematic sampling (discussed later) and simple random sampling. He also states that stratification normally results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample.

In a stratified random sampling program when N, the number of population units, is large, the optimum number of samples can be estimated with (Cochran, 1977)

$$n = Z^2_{1-\alpha/2} \sum_{h=1}^{L} \frac{W_h s_h^2}{d^2} \quad (2\text{-}4)$$

where

n = number of samples across all strata,
Z = standard normal variate,
L = number of strata,
W_h = stratum weight,
s_h = sample standard deviation for stratum h,
d = absolute margin of error for weighted mean, and
α = confidence interval.

The stratum weight, W_h, is the relative size of each stratum. Once the total number of samples is determined, the samples may be allocated to each stratum by (Gilbert, 1987)

$$n_h = \frac{n\ W_h s_h}{\sum_{h=1}^{L} W_h s_h} \quad (2\text{-}5)$$

Alternatively, the samples may be proportionally allocated, with each stratum given a percentage of the total samples in accordance with the stratum size. The above equation allocates more samples to a stratum that is larger or has a higher variability. Cochran (1977) provides an approach for optimizing the sampling when the sampling cost per population unit, c_h, is different among the strata:

$$n_h = \frac{n\ W_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}} \quad (2\text{-}6)$$

In general, a larger number of samples would be taken in a stratum that is more variable, larger, or less costly to sample than other strata.

The mean for stratum h, $\bar{x}_h$, is the simple mean of all samples within the stratum. The weighted mean, $\bar{x}_{st}$, is given by

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \quad (2\text{-}7)$$

Systematic Sampling

Systematic sampling is used extensively in water quality monitoring programs, usually because it is relatively easy to do from a management perspective. In systematic sampling the first sample is taken from a random starting point (or at a random starting time) and each subsequent sample is taken at a set distance (or time interval) from the first sample (Figure 2-5). For example, if budgetary constraints limit the number of samples to 10 and the objective is to characterize a 10-mile river using systematic sampling, the first observation would be taken randomly in the first river mile. Subsequent samples would be taken at 1-mile increments up the river. In comparison, a stratified random sampling approach would divide the river into 10 1-mile segments (strata) and one random sample would be taken in each segment.

Gilbert (1987) recommends systematic sampling when estimating long-term trends, defining

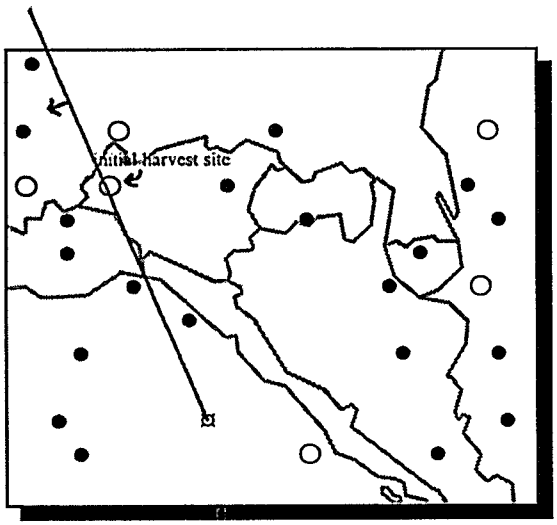


Figure 2-5. Systematic sampling for silviculture. Dots (● and O) represent harvest sites of interest. A single point on the map (¤) and one of the harvest sites were randomly selected. A line was stretched outward from the point to (and beyond) the selected harvest site. The line was then rotated about the map and every fifth dot that it touched was selected for sampling (open dots—O). The direction of rotation was determined prior to selection of the point of the line's origin and the beginning harvest site. The shaded lines on the map could represent county boundaries, soil type, watershed, or some other boundary, but were not used for the sampling process.

seasonal or other cycles, or forecasting pollution concentrations. In general, systematic sampling is superior to stratified random sampling with one or two samples per stratum for estimating the mean (Cochran, 1977). Gilbert (1987) reports that systematic sampling is equivalent to simple random sampling in estimating the mean if the target population has no trends, strata, or correlations among the population units. Estimates of variance from systematic samples may differ from those determined from random samples. Cochran (1977) notes that "on the average the two variances are equal." However, Cochran also states that for any single population for which the number of sampling units is small, the variance from systematic sampling is erratic and may be smaller or larger than the variance from simple random sampling.

Gilbert (1987) cautions that any periodic variation in the target population should be known before establishing a systematic sampling program.

Sampling intervals equal to or multiples of the target population's cycle of variation may result in biased estimates of the population mean. Systematic sampling can be designed to capitalize on a periodic structure if that structure can be characterized sufficiently (Cochran, 1977). A simple or stratified random sample is recommended, however, in cases where the periodic structure is not well known or where the randomly selected starting point is likely to have an impact on the results (Cochran, 1977). Quantitative procedures for estimating the population mean and variance from systematic sampling data are presented by Gilbert (1987).

Gilbert (1987) notes that assumptions about the population are required in estimating population variance from a single systematic sample of a given size. However, there are systematic sampling approaches that do support unbiased estimation of population variance, including multiple systematic sampling, systematic stratified sampling, and two-stage sampling (Gilbert, 1987). In multiple systematic sampling more than one systematic sample is taken from the target population. Systematic stratified sampling involves the collection of two or more systematic samples within each stratum.

Cluster Sampling

Cluster sampling is applied in cases where it is more practical to measure randomly selected groups of individual units than to measure randomly selected individual units (Gilbert, 1987). In cluster sampling, the total population is divided into a number of relatively small

subdivisions, or clusters, and then some of these subdivisions are randomly selected for sampling (Figure 2-6). For one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is then performed within each cluster (Gaugush, 1986). An example of one-stage cluster sampling is the collection of all macroinvertebrates on randomly selected rocks within a specified sampling area. The stream bottom might contain hundreds of rocks with thousands of organisms attached to them, thus making it difficult to sample the organisms as individual units. However, it is often possible to randomly select rocks and then inspect every organism on each selected rock.

Gaugush (1986) states that the "analysis of cluster samples requires the estimation of variance at two levels, the between-cluster variability and the within-cluster variability. The total variability is a recombination of these two levels." Freund (1973) notes that estimates based on cluster sampling are generally not as good as those based on simple random samples, but they are more cost-effective. As a result, Gaugush believes that the difficulty associated with analyzing cluster samples is compensated for by the reduced sampling requirements and cost. Cochran (1977) discusses one-stage cluster sampling for clusters of either equal or unequal sizes and provides equations for determining the optimal population unit size using the relative sizes of possible population units, the variance among the population unit totals, and the relative cost of measuring one population unit. He notes that many factors come into play when determining optimal population size, including cost versus unit size.

Two-stage Sampling

Two-stage sampling involves dividing the target population into primary units, randomly selecting a subset of these primary units, and then taking random samples (second-stage units) within each of the selected primary units. This is a common practice when a large sample is taken and then a smaller aliquot is actually measured from the original sample. The process of subsampling introduces additional uncertainty and becomes significant if the pollutant is in particulate form and very small subsamples are used (Gilbert, 1987).

Two-stage sampling might also include systematic sampling within a randomly selected subset of the population primary units. For example, if the target population is the average annual pollutant concentration in a stream, the primary units could be daily average concentrations (n = 365). A

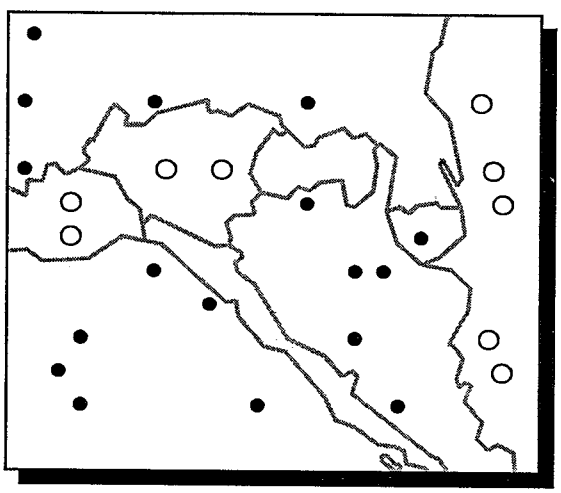


Figure 2-6. Cluster sampling for silviculture. All harvest sites in the area of interest are represented on the map (closed {●} and open {O} dots). The shaded lines on the map represent county boundaries. Some of the counties were randomly selected, and all harvest sites within those counties (open dots - O) were selected for sampling. Some other type of boundary, such as soil type or watershed, could have been used to separate the harvest sites for the sampling process.

subset of these daily concentrations (e.g., n = 24) could be selected at random for further systematic sampling of hourly concentrations. For example if four systematic, hourly samples could be taken on each of 24 different days, with the hour for the first sample determined randomly, followed by three more hourly samples taken every sixth hour, 96 hourly composite samples would be available for the calculation of the population mean and variance.

Cochran (1977) describes two-stage sampling in great detail and presents methods for determining the mean and variance in two-stage sampling with units of equal size. In Cochran's discussion, he notes that if all population units are sampled, the formula for estimating the variance is the same as that used to estimate the variance for proportional stratified random sampling. This means that two-stage sampling is a type of incomplete stratification, with the primary units treated as strata.

For further information regarding two-stage (and three-stage) sampling, the reader is referred to Gilbert (1987) and Cochran (1977). The authors provide equations for estimating the number of samples (primary units) and subsamples for two conditions: (1) primary units of equal size and (2) primary units of unequal size. Equations for estimating the mean and total values in composite samples of equal- and unequal-sized units are also provided. The authors also provide equations for calculating the number of composites and composite subsamples needed.

Double Sampling

Double sampling is often used when two techniques exist for measuring a pollutant. Initially, both methods are used. Then, after a correlation has been established, only the cheaper or simpler technique is used. Gilbert (1987) provides an approach for calculating the sample size when the cost and variability associated with both methods has been determined during the initial sampling. This same procedure can also be used when it is less expensive to measure a surrogate variable (Gilbert, 1987). This technique can be used for stratification, ratio estimates, and regression estimates (Cochran, 1977).

Regression analyses are used to predict values for one variable (i.e., the dependent variable) using one or more independent variables based on a mathematical relationship. As an example, total suspended solids concentration is typically a covariate of total phosphorus concentration in watersheds impacted by agricultural runoff. Measurement of total suspended solids may help increase the precision of total phosphorus estimates. Gaugush (1986) discusses sampling to support regression analyses using spatial or temporal gradients as the independent variable, the latter being for trends over time. Some key points in his discussion related to using a spatial independent variable are as follows:

- Whenever the type of relationship (e.g., linear, log-linear) is known, relatively few sampling points are needed along the gradient. More samples may then be used as replicates.

- Whenever the relationship is not known, more sampling points are needed along the gradient. More replicates are also needed to test the proposed model.

- It is usually acceptable to place sampling points equal distances from each other along the gradient as long as the sampling does not fall in step with some natural phenomenon that would bias the data collected.

Some key points in the discussion regarding time sampling are as follows:

- Time can be used either as a covariate or as a grouping variable. Grouping by time might be desirable when changes in the variable of interest either are small over time or occur only during short periods with long periods of little or no change.

- Considerations in using time as a covariate are similar to those for spatial gradients, but (1) time is usually only a surrogate for other variables that truly affect the variable of interest, and (2) the relationship with time is likely to be complex.

- If time is to be used as a covariate, relatively frequent sampling will be needed, with some replication within sampling periods. Random sampling within the periods is also recommended.

The sampling designs most common to environmental monitoring are summarized in Table 2-2.

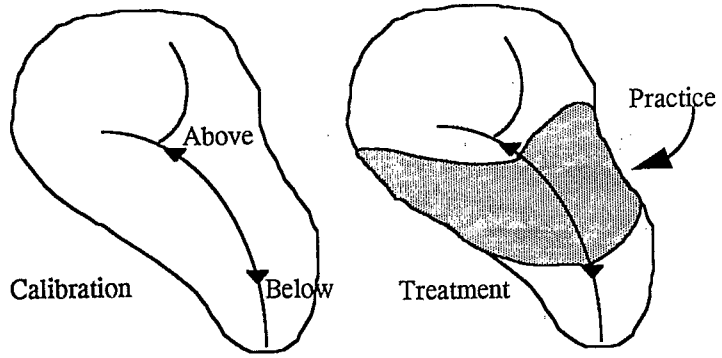
2.5.2 Targeted Site Location Study Designs

Paired and nested paired watershed approaches are the two most appropriate approaches when trying to evaluate the impact or benefit of a BMP or system of BMPs at the watershed scale (Spooner et al., 1985). A nested paired watershed design (Figure 2-7A) is sometimes referred to as an "above-and-below" design where one monitoring station is located above the treatment area and one station is located below the treatment area. The paired watershed design (Figure 2-7B) is based on identifying two watersheds where one watershed is the control and the second is the treatment. In both study designs, data are collected before treatment (calibration) and after treatment is implemented so that differences between watersheds (or nested watersheds) can be evaluated. The key advantage of these two approaches is that the variation due to year-to-year climatic differences and differences between watersheds are statistically

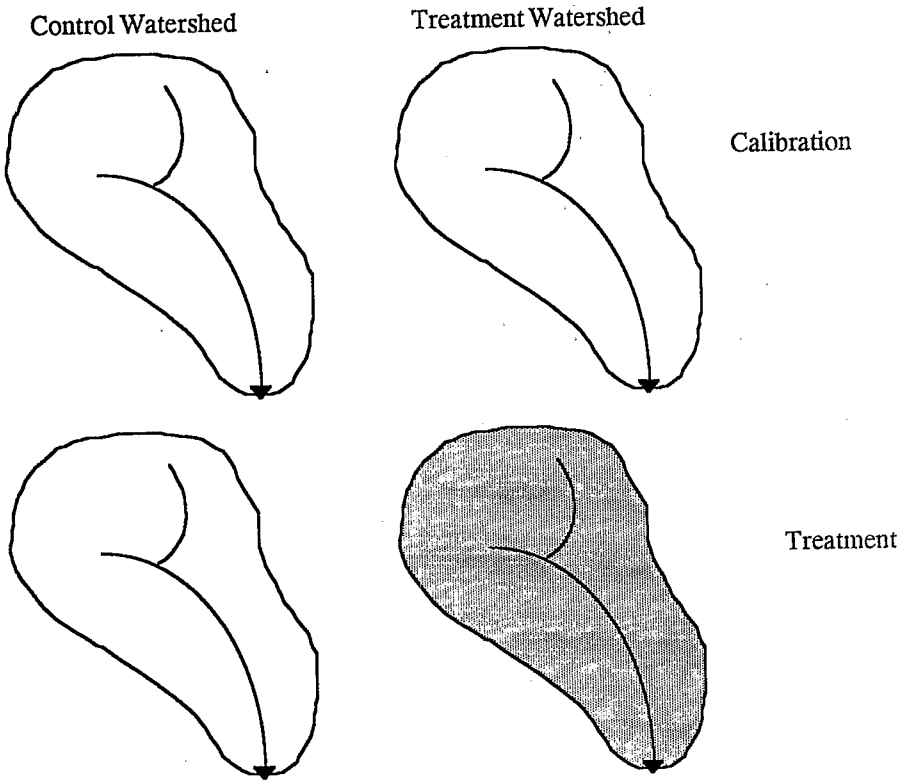
Table 2-2. Applications of six sampling designs to estimate means and totals.

Sampling Design	Conditions for Application
Simple Random Sampling	Population does not contain major trends, cycles, or patterns of contamination.
Stratified Random Sampling	Useful when a heterogeneous population can be broken down into parts that are internally homogeneous.
Two-stage Sampling	Needed when measurements are made on subsamples or aliquots of the field sample.
Cluster Sampling	Useful when population units cluster together and every unit in each randomly selected cluster can be measured.
Systematic Sampling	Usually the method of choice when estimating trends or patterns of contamination over space. Also useful for estimating the mean when trends and patterns in concentrations are not present.
Double Sampling	Useful when there is a strong linear relationship between the variable of interest and a less expensive or more easily measured variable.

Source: After Gilbert, 1987.



A) Nested Paired Watershed Study Design



B) Paired Watershed Study Design

Figure 2-7. Nested paired and paired watershed study designs.

controlled, provided that a sufficient calibration period has been used. Clausen (1991) states that the cost of conducting a paired watershed experiment in Vermont ranged from $30,000 to $50,000 per year for 3 or 4 years. This cost included continuous discharge and water sampling, as well as the analysis of approximately six water quality characteristics.

In St. Albans Bay, Vermont, in another RCWP, two small watersheds received proper manure management during a 2-year calibration period, followed by a period in which one watershed received winter-spread manure (Clausen, 1985). This is an interesting example of the paired watershed approach since BMPs were removed from, instead of applied to, a watershed after the calibration period. Data from this type of nested paired or paired watershed design can be evaluated by an analysis of covariance as described by USEPA (1993c). Unfortunately, both study designs are limited because the experiment is not repeated to account for spatial variability, and transferability of BMP effectiveness to other regional watersheds is not appropriate (MacDonald et al., 1991).

Nested watershed designs can also be used to document the severity of a nonpoint source pollution problem. In an example from the Rock Creek, Idaho, RCWP, paired data were collected using an upstream-downstream approach. These data were used in regressions of water quality against time.

The downstream concentrations (below the nonpoint pollution source) were adjusted for upstream concentrations (above the nonpoint pollution source), transformed, and then regressed against time as a continuous variable (Spooner et al., 1986). Results of this approach indicated that decreasing pollutant concentrations from nonpoint pollution sources were due to implementation of BMPs.

Single-watershed designs, which collect data before and after BMP implementation, and two-watershed designs, which collect data after BMP implementation in one watershed, should generally be avoided for evaluating BMP effectiveness. The single-watershed design does not account for year-to-year climatic variability. The two-watershed design does not account for differences between watersheds since no calibration data are collected.

An alternative approach, when collecting data during a calibration period is not viable, is to use a multiple-watershed design, in which numerous watersheds are monitored. In this design, multiple watersheds in a region are selected, including some that have a particular BMP implemented and others that do not have the BMP implemented. Alternatively, numerous paired upstream and downstream stations (i.e., nested watersheds) are selected. In the case of paired upstream and downstream stations, the designation of controls or treatments is not random, and it is necessary to add additional station pairs where no treatment or BMP is implemented (MacDonald et al., 1991). By monitoring numerous watersheds, the true variability between watersheds is considered and the results from this study design can be transferred to other watersheds in the region. Fifteen paired stations were established in the Snohomish River basin (Washington State) to determine the effect of commercial agriculture on water quality along with other objectives over a 3-year period (Luchetti et al., 1987). The pairs varied considerably in terms of stream size and agricultural activity. Combining the monitoring data with land use and BMP implementation data, the project documented the impact of commercial agriculture on water quality.

Use of trend stations, or long-term ambient monitoring, is based on establishing monitoring stations that are routinely monitored. This type of study design is generally most appropriate for

watersheds where a variety of BMPs are being implemented over a period of time or gradual water quality changes are expected. The difficulty in using trend stations is developing a causal link between water quality and the various land use activities. To use trend stations, variables associated with land treatment, hydrology, and meteorology should be accounted for to increase the likelihood of successful documentation of water quality-BMP relationships. The long-term commitment required from management to monitor these stations is one of the key disadvantages of this approach. The U.S. Geological Survey has systematically sampled the national stream quality accounting network (NASQAN) once a month for more than 20 years to monitor the water quantity and quality (Smith et al., 1987).

One key to establishing the study design, which is often overlooked, is site selection. Site location and establishment are discussed in several existing monitoring guides and texts (Brakensiek et al., 1979; Ponce, 1980a; USEPA, 1978b, 1981; USGS, 1977; Wetzel and Likens, 1979). Few differences exist between nonpoint source site location strategies and the approaches discussed in these documents. Within any given budget, site location is a function of water resource type, monitoring objectives, and data analysis plans. When evaluating the effectiveness of nonpoint source control measures, it might be necessary to locate monitoring sites above known point sources to remove them as confounding influences in the study. Additional considerations in site selection are site accessibility and landowner cooperation in data collection efforts (e.g., farm management records). It is strongly recommended that nonpoint source monitoring stations be located near or at USGS gaging stations, when possible, due to the extreme importance of obtaining accurate flow records for estimating pollutant loads. In the absence of a USGS gaging station, monitoring stations should be located at sites that offer adequate flow monitoring capabilities.

Some station requirements may be such that, with careful station siting, one particular station can meet multiple monitoring objectives. Caution should be exercised, however, to avoid compromising the worth of a station for the sake of false economy.

For evaluating the overall background or performing a problem assessment, a panel of federal and state monitoring professionals (USEPA, 1975) determining several points for establishing site locations for physical and chemical water column sampling, which should be considered as appropriate. The process of site selection for biological monitoring is described in Chapter 3.

- Sites should be located at representative sites in mainstem rivers, estuaries, coastal areas, lakes, and impoundments. These sites can be used to characterize the overall quality of the area's surface waters and will provide water quality baselines against which progress can be measured.

- Sites should be located in water quality-limited and major water use areas. Sites in water quality-limited areas can be used to evaluate the overall pollution control strategy and BMP system effectiveness. Sites in major water use areas, such as public water supply intakes, commercial fishing areas, and recreational areas, serve a dual purpose—public health protection and overall water quality characterization.

- Sites should be located upstream and downstream from representative land use areas (e.g., mining, silviculture) and morphologic zones. These sites can be used to compare the relative effects of pollution sources and morphologic zones on water quality and to document baseline water quality.

- Sites should be located at the mouths of major or significant tributaries to mainstem streams, lakes, impoundments, estuaries, or coastal areas. Data from these sites, when taken in concert with permit monitoring data and intensive survey data, can be used to determine the major sources of pollutants to the area's major waterbodies. By comparison with other tributary data, the relative magnitude of the pollution sources can be evaluated and problem areas can be identified.

- Sites should be located to measure the input and output of nutrients and other pertinent substances into and from waterbodies (i.e., lakes, impoundments, estuaries, or coastal areas) that exhibit eutrophic characteristics, as well as at critical locations within the waterbody. The information from these stations, when taken in combination with the pollution source data, can be used to establish cause-and-effect relationships, identify problem areas, and indicate appropriate corrective measures.

Sediment sampling sites should be located in sink areas as determined by intensive surveys, reconnaissance surveys, and historical data. A major concern of sediment monitoring is to assess the accumulation of toxic substances and sediment-bound nutrients. The location for a sediment sampling site should be chosen by considering the sediment mechanics and the hydrological characteristics of the waterbody (USEPA, 1975).

2.6 *Example Program Design*

The RCWP includes several examples of nonpoint source monitoring and evaluation strategies. Two project strategies are described here. Several additional examples are provided in Appendix C.

The Idaho RCWP's major focus was to control sediment from irrigation return flows. Using a targeted study design, seven ambient monitoring stations (Figure 2-8) were used (Clark, 1986):

S-1: Near mouth - integrated all pollution sources flowing into Rock Creek and measured the pollutant load that going into the Snake River (river mile (RM) 0.75). Water quality, benthic macroinvertebrates, and fisheries data were collected.

S-2: At Poleline Road - a benthic invertebrate and fisheries monitoring site as well as water quality (RM 3.75).

S-3: Above Highway 93 - below the confluence of the high-priority agricultural drains and city of Twin Falls urban runoff (RM 7.3). Water quality, benthic macroinvertebrates, and fisheries data were collected.

S-4: At Twelvemile - above the influence of Twin Falls urban area and the high-priority drains (RM 13.5). Water quality, benthic macroinvertebrates, and fisheries data were collected.

S-5: At 3500 East Road - a benthic invertebrate and fisheries monitoring site only (RM 21.1).

S-6: Near Rock Creek townsite - measured the quality of the natural surface water above the irrigation tract (RM 30.3). Water quality, benthic macroinvertebrates, and fisheries data were collected.

C-1: Twin Falls Main Canal - source of water for the irrigation tract. Only water quality data were collected.

Intensive monitoring stations were placed on irrigation drains to track changes in sediment load and associated pollutants close to their source and

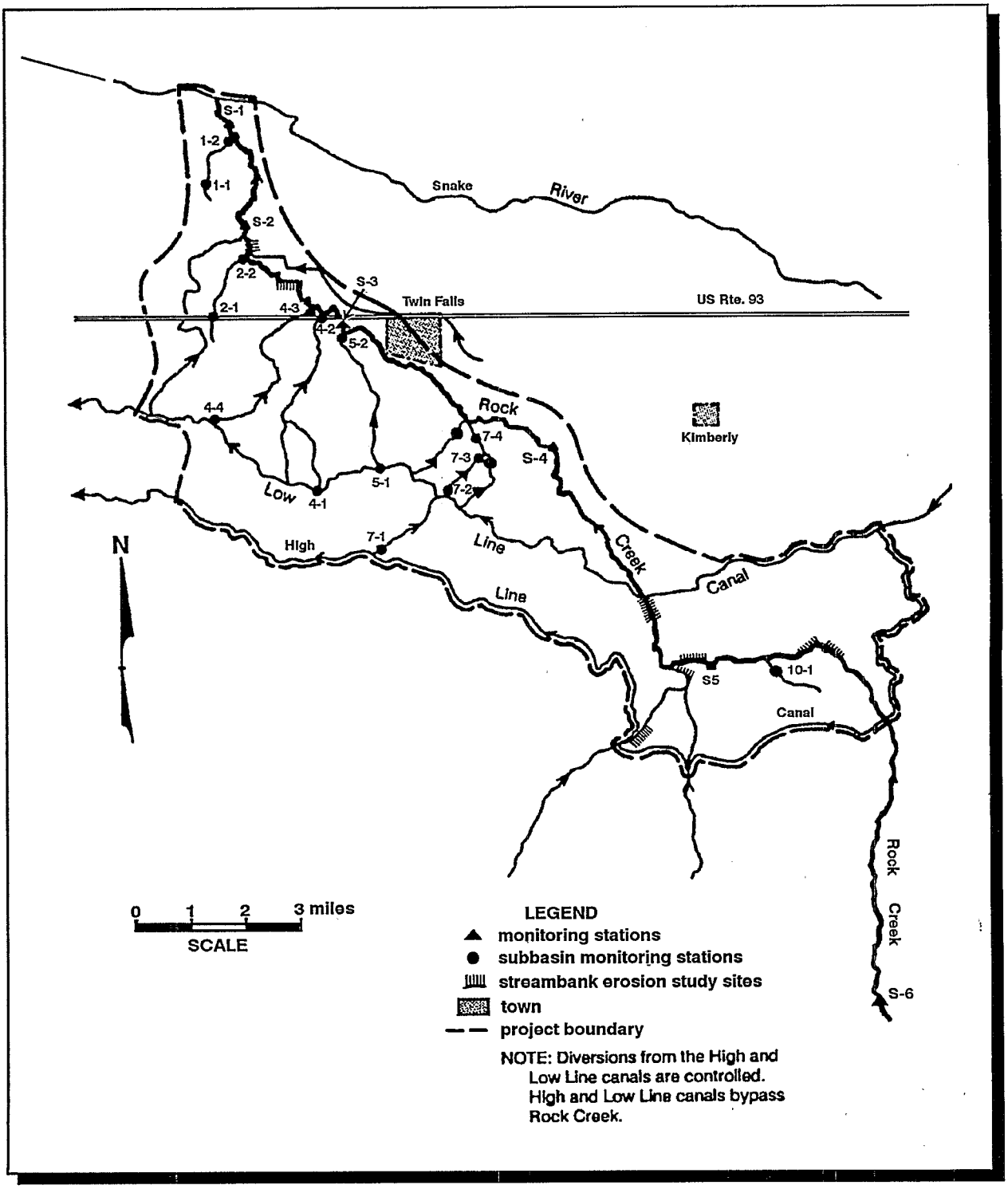


Figure 2-8. Map of the Rock Creek Rural Clean Water Program study area, Twin Falls County, Idaho. (Source: Clark, 1986)

associated BMPs. In this way, changes in water quality due to the RCWP could be detected. Nineteen stations were located in six subbasins (Figure 2-8). Stations measured the source of water to the subbasins (7-1, 5-1, 4-1, 4-3, 2-1, and 1-1), the input of the subbasins to Rock Creek (7-7, 7-4, 5-2, 4-2, 4-3, 2-2, and 1-2), and key intermediate sites (7-2, 7-3, and 7-6). Additional stations were added in other subbasins as they were needed (2-3, 2-4, and 10-1).

The St. Albans Bay, Vermont, RCWP project used a four-level monitoring and evaluation program to meet three objectives (Vermont RCWP Coordinating Committee, 1986):

- Document changes in the water quality of specific tributaries within the watershed resulting from implementation of manure management practices.

- Measure changes in suspended sediment and nutrients entering St. Albans Bay resulting from implementation of water quality management programs within the watershed.

- Evaluate trends in the water quality of St. Albans Bay and the surface waters within the St. Albans Bay watershed during the period of the St. Albans Bay RCWP Watershed Project. Monitoring sites for all four levels of monitoring and evaluation are shown in Figure 2-9. The Level 1 bay sampling was designed to determine long-term water quality trends in St. Albans Bay over the life of the project (Vermont RCWP Coordinating Committee, 1984). The Level 2 tributary sampling was designed to determine the long-term water quality trends for the major tributaries including the Bay and the St. Albans City wastewater treatment plant (Vermont RCWP Coordinating Committee, 1984). The Level 3 monitoring was directed toward evaluating the effect of best manure management practices on the quality of surface runoff from individual fields; Level 4 was designed to supplement the Level 2 monitoring by sampling additional tributaries to St. Albans Bay and to isolate subunits within the Level 2 subwatersheds (Vermont RCWP Coordinating Committee, 1984).

2.7 Roles and Responsibilities

Designing and implementing a monitoring program is an interdisciplinary and interagency activity. In many cases, technical staff will need to integrate "new" monitoring with what is already being done in order to demonstrate to program managers that duplicate work is not proposed. The most effective way to achieve this goal is to bring all the involved agencies and other stakeholders in the monitoring effort together. One or a few agencies acting as project coordinator(s) should seek to obtain an agreement from each involved party with respect to their role(s) and responsibilities in the performance of the project. These agreements can be formalized as commitments and specified in the quality assurance project plan, which is discussed at greater length in Chapter 5.

Such coordinated cooperation permits each involved party to offer the results of its ongoing activities to the monitoring effort and lessens the burden on the proposed budget. For example, the U.S. Geological Survey might already have a gaging station in place and the Natural Resources Conservation Service might already have a tracking system for BMPs in place. Other agencies, including the U.S. Fish and Wildlife Service and EPA, might have other ongoing monitoring programs. When multiple agencies are involved in the monitoring program, each can benefit from the efforts of the others.

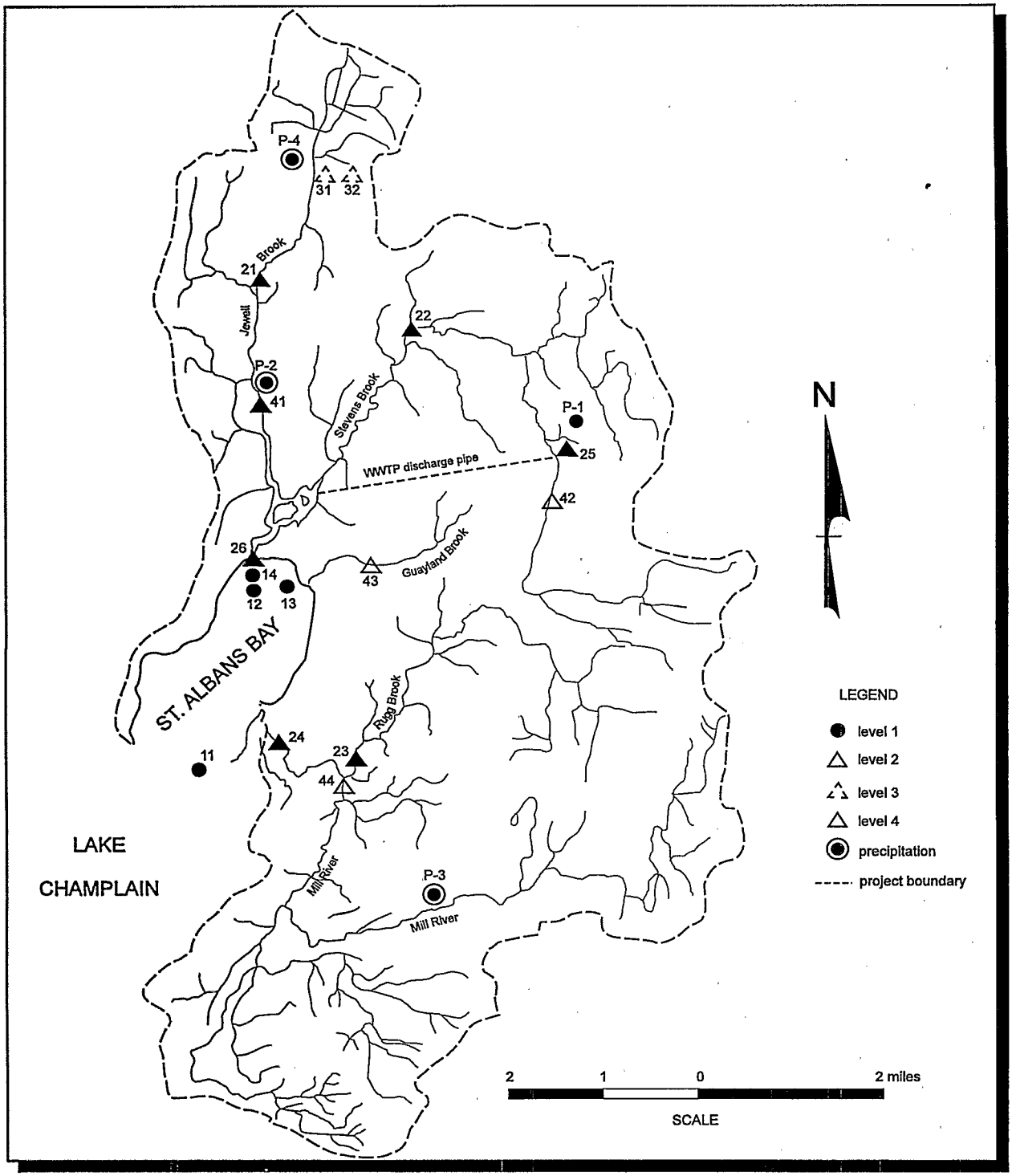


Figure 2-9. St. Albans Bay watershed, Franklin County, Vermont, sampling locations. (Source: Vermont RCWP Coordinating Committee, 1986)

2.8 Quality Assurance Project Planning

An integral part of the design phase of any nonpoint source pollution monitoring project is the development of a quality assurance project plan (QAPP). The QAPP is a critical document for the data collection effort inasmuch as it integrates the technical and quality aspects of the planning, implementation, and assessment phases of the project. The QAPP documents how quality assurance (QA) and quality control (QC) elements will be implemented throughout the life of a project. It contains statements about the expectations and requirements of those for whom the data are being collected (i.e., the decision makers) and provides details on project-specific data collection and data management procedures that are designed to ensure that these requirements are met. Development and implementation of a QA/QC program, including preparation of a QAPP, can require up to 10 to 20 percent of project resources (Cross-Smiecinski and Stetzenback, 1994), but this cost is recaptured in lower overall costs due to the project's being well planned and executed. A thorough discussion of QA/QC is provided in Chapter 5.

2.9. Chemical and Physical Monitoring

Chemical and physical monitoring and the mechanics of sampling are important topics and need to be considered as carefully as other monitoring topics discussed in this guide, such as data analysis and biological monitoring. However, these aspects of monitoring are covered in detail in other documents (e.g., USDA-NRCS, 1996; USGS, 1977) and it would be redundant to duplicate the information here. Therefore, these types of monitoring are only briefly mentioned here.

Important topics related to chemical and physical monitoring and sampling procedures that managers of nonpoint source pollution monitoring programs should consider include the following:

- **Type of sample.** Water quality varies temporally and spatially, and samples must be taken that will accurately reflect overall water quality and overall water quality impacts of nonpoint source pollutants. There are four basic types of samples to consider—grab, composite, integrated, and continuous (USDA-NRCS, 1996):

Typically, a grab sample is a sample taken at one place a single time. Care should be taken to make sure that a grab sample is representative. If there is spatial variability (e.g., across a stream, at different depths in a lake) or temporal variability (e.g., during a storm event) it might be more appropriate to take a composite or time-integrated sample rather than a grab sample.

Composite samples consist of a series of grab samples, usually collected in the same location but at different times with the results averaged. Composite samples are usually either time-weighted or flow-weighted. Time-weighting means that a fixed volume is collected at a predetermined time interval. Flow-weighting means that a sample is taken after a specified quantity of water has passed the monitoring station. Both types of composite sampling are amenable to automatic sampling equipment. Composite samples are appropriate for most monitoring objectives.

Integrated samples account for variations in water quality with depth or distance from a stream bank at a monitoring station. Subsamples are taken at various depths or distances from the stream bank, and integrated into a single sample.

Continuous sampling requires electronic measuring devices and is therefore limited to variables that are amenable to this type of sampling, such as dissolved oxygen,

conductivity, pH, and salinity. It is generally not suitable for measurements of metals, organics, or pesticides. Continuous sampling is typically used for research and fate and transport studies.

Some factors that influence the type of sample to collect include the objectives of the study, waterbody type, and variables to be sampled.

- **Type of sample collection.** Samples can be collected manually or with automated equipment. Sampling location, sample site accessibility, and staffing are factors to consider when determining which approach to use.

- **Type of sample collection equipment.** Sampling equipment can be either mechanically operated or powered, and the use of one or the other approach again depends on project-specific considerations and constraints. Commonly used sampling equipment includes flow recorders, staff gauges, and precipitation gauges.

- **Station type.** Various monitoring stations might be necessary to measure the variables of interest. Discharge stations might be installed to measure runoff from a sampling plot in a field or at the edge of a field, or to measure stream discharge. Other monitoring stations might be necessary to collect water samples, record precipitation, analyze soil water, assess biological factors, or monitor sediment.

- **Sampling equipment operation and maintenance.** It is important to ensure that all sampling equipment is in good operational condition prior to sampling and during sampling to ensure that reliable data are being collected. The use of automated sampling equipment does not mean that project staff are relieved of the responsibility to regularly check equipment operation. Staff should be thoroughly trained to use and maintain sampling equipment properly.

- **Record keeping.** Proper record keeping is important to make the process of data analysis less burdensome and to aid in tracking any anomalies in data to possible influences, such as equipment malfunctions or variations in sample collection timing. Detailed records are also valuable when writing reports and preparing presentations.

2.10 RECOMMENDED REFERENCES

Important monitoring references that should be consulted include the following:

American Public Health Administration. 1995. *Standard methods for the examination of water and wastewater.* 19th ed. American Public Health Association, Washington, DC.

Discussion of how to collect samples and the required volume of sample material for numerous water quality parameters.

Bauer, S.B., and T.A. Burton. 1993. *Monitoring protocols to evaluate water quality effects of grazing management of western rangeland streams.* EPA 910/R-93-017. Submitted to U.S. Environmental Protection Agency, Region 10, Water Division, Surface Water Branch, by Idaho Water Resources Research Institute, University of Idaho, Moscow, ID. October.

Temperature, nutrients, bacteria, stream channel morphology, stream bank stability, sediment, streamside vegetation. For each, parameters to measure, sample collection procedures, sample analysis.

Clark, W.H. 1990. *Coordinated nonpoint source water quality monitoring program for Idaho.*

Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, ID. January.

Appendix with suggested monitoring parameters and protocols, including suggested protocols for various types of BMP implementation and pollutant sources and transport mechanisms.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska.* EPA/910/9-91-001. U.S. Environmental Protection Agency, Region 10, Seattle, WA.

Forestry focus; parameter selection; discussion of many parameters, including a definition, relation to designated uses, how the parameter responds to management activities, parameter-specific measurement notes, applicable standards, present uses of the parameter, and parameter assessment. Parameter recommendations for various land treatments.

USDA. 1979. *Field manual for research in agricultural hydrology.* Agricultural Handbook 224. U.S. Department of Agriculture, Washington, DC.

USDA-NRCS. 1996. *Water quality monitoring.* U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC.

Numerous recommendations for good references on a variety of sampling topics. Also includes tables with recommendations of variables to measure based on the above considerations. Topics covered include variable selection, sample types (grab, composite, integrated, continuous), station type (discharge, concentration, precipitation, soil water, biotic, sediment), sample collection (volume), sample preservation.

USGS. 1977. *National handbook of recommended methods for water-data acquisition.* U.S. Geological Service, Office of Water Data Coordination, Reston, Virginia.

3 BIOLOGICAL MONITORING OF AQUATIC COMMUNITIES

3.1 INTRODUCTION

Biological monitoring, as described here, consists of assessing the condition of the physical habitat and specific biological assemblages—typically benthic macroinvertebrates and fish— that inhabit the aquatic environment. In the true definition of the term, biological monitoring includes toxicity testing, fish tissue analyses, and single population surveys conducted over time. However, for purposes of this document, the definition of biological monitoring is limited to the concept of community-level assessments.

> **Community:** *Any group of organisms of different species that co-occur in the same habitat or area.*

This chapter discusses the rationale behind using biological monitoring as part of a nonpoint source (NPS) monitoring program, gives basic guidance on conducting biological assessments, provides biological monitoring program design considerations, and discusses ways in which biological assessment data can be used to detect trends in NPS impacts. Numerous texts and papers have been written on biological monitoring methods (see references), but methods and means of interpreting the results from them are still under development for many habitats and regions of the country. The U.S. Environmental Protection Agency (EPA) has produced two guidance documents that are the foundation for the majority of state biological monitoring programs:

- *Biological criteria: Technical guidance for streams and small rivers*. EPA 822-B-96-001. (Gibson et al., 1996)
- *Rapid bioassessment protocols for use in streams and wadable rivers*. EPA/444/4-89-001. (Plafkin et al., 1989)

Many state agencies, such as the Ohio Environmental Protection Agency (OEPA), Illinois Environmental Protection Agency (IEPA), Delaware Department of Natural Resource and Environmental Control (DNREC), Florida Department of Environmental Protection (DEP), Connecticut DEP, and New York Department of Environmental Conservation (DEC), are incorporating biological monitoring into ongoing and new monitoring programs. In addition, numerous ongoing biological monitoring programs have been implemented by federal, local, and tribal entities, including the U.S. Geological Survey (USGS), National Water Quality Assessment Program (NAWQA); U.S. Department of the Interior, Bureau of Land Management; U.S. Department of Agriculture (USDA), Forest Service; King County, Washington; Prince George's County, Maryland; and the Yakama Tribe. The methods for biological monitoring are thereby being improved, and this in turn is making biological assessment a more widely accepted and applicable tool for monitoring programs.

The common use of EPA methods for sampling and analysis is documented in a recent work that reviews state programs (*Summary of state biological assessment programs for streams and wadable rivers*, Davis et al., 1996; EPA230-R-96-007). It shows that 45 states are using rapid bioassessment protocol (RBP)-type sampling and analysis methods or are in the process of establishing such programs.

Biological survey approaches differ depending on the waterbody, i.e., stream, river, lake, estuary, or

wetland. EPA has developed and or is currently developing bioassessment survey methods appropriate for use in these different waterbody types.

3.1.1 Rationale and Strengths of Biological Assessment

The central purpose of assessing biological condition is to determine how well a waterbody supports aquatic life and what kind of aquatic life it supports. Biological communities reflect the cumulative effects of different pollutant stressors—excess nutrients, toxic chemicals, increased temperature, excessive sediment loading, and others—and thus provide an overall measure of the aggregate impact of the stressors. Although biological communities respond to changes in water quality more slowly than water quality actually changes, they respond to stresses of various degrees over time. Because of this, monitoring changes in biological communities can be particularly useful for determining the impacts of infrequent or low-level stresses, such as highly variable NPS pollutant inputs, which are not always detected with episodic water chemistry measurements. Improvements in waterbody condition after the implementation of best management practices (BMPs) can sometimes be difficult to detect, and biological assessment can be useful for measuring such improvements. Several biosurvey techniques that can be used for detecting aquatic life impairments and assessing their relative severity are discussed in this chapter.

A small number of factors are often key in determining the structure of a community and its response to stress; e.g., the type of substrate or the riparian vegetation that provides organic material to the stream, regulates temperature, and provides bank stability (National Research Council, 1986). Landscape features such as soil type, vegetation, surrounding land use, and climate also have a well-documented influence on water chemistry and hydrologic characteristics. Finally, water quality, as influenced by landscape features and anthropogenic sources of pollutants, has a direct effect on aquatic biological communities.

The quality of the physical habitat is an important factor in determining the structure of benthic macroinvertebrate, fish, and periphyton assemblages. The physical features of a habitat include substrate type, amount of debris in the waterbody, amount of sunlight entering the habitat, water flow regime (in streams and rivers), and type and extent of aquatic and riparian vegetation. Even though there might not be sharp boundaries between habitat features in a stream, such as riffles and pools, the biota inhabiting them are often taxonomically and biologically distinct (Hawkins et al., 1993). Habitat quality is assessed during biological assessment and is a measure of the extent to which the habitat provides a suitable environment for healthy biological communities.

Biological assessment: *An evaluation of the biological condition of a waterbody using biological surveys and other direct measurements of biota in surface waters.*

Biological monitoring: *Multiple, routine biological assessments over time using consistent sampling and analysis methods for the detection of changes in biological condition.*

Natural biological communities are usually diverse, comprising species at various trophic levels (e.g., primary producers, secondary producers, carnivores) and levels of sensitivity to environmental changes. Adverse impacts from NPS pollution or other stressors such as habitat

alteration can reduce the number of species in a community, change the relative abundances of species within a community, or alter the trophic structure of a community. Biological surveys of select species or types of organisms that are particularly sensitive to stressors, such as fish, periphyton, or benthic macroinvertebrates, take advantage of this sensitivity as a means to evaluate the collective influence of the stressors on the biota (Cummins, 1994).

Some NPS pollutant inputs such as wet-weather runoff of chemical contaminants or sediment are highly variable in time, and biological monitoring can be a useful approach for monitoring this type of NPS impact. Biological monitoring can be used to assess the overall impact of multiple stressors, although it might not provide information about the relative magnitude of each stressor.

Knowledge of the natural physical habitat and biological communities in an area is important for interpreting biological assessment data. Biological and habitat data collected from numerous sites that are in good or near-natural condition can be used to determine the type of biological community that should be found in a particular aquatic habitat. In areas where natural conditions do not exist (due to past disturbances), historical data or the best professional judgment of knowledgeable experts can sometimes be used to select reference sites and define the reference condition. This natural condition has been referred to as reflecting biological integrity, defined by Karr and Dudley (1981) as "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region." Highly detailed biological assessments are comparisons of biological conditions at a test site to the expected natural community and are thus a measure of the degree to which a site supports (or does not support) its "ideal" or potential biological community (Gibson et al., 1996). Other types of biological assessment involve comparisons of impacted sites to control sites, the latter being sites that are similar to monitored sites but are not affected by the stresses that affect the monitored sites (Skalski and McKenzie, 1982). Paired watersheds or upstream-downstream approaches are examples. A knowledge of the natural condition is still valuable for accurate data interpretation when control sites are used (Cowie et al., 1991).

3.1.2 Limitations of Biological Assessment

Although biological assessment is useful for detecting impairments to aquatic life and assessing the severity of the impairments, it is not necessarily a measure of specific stressors. Thus, it usually does not provide information about the cause of impairment, i.e., specific pollutants or their sources. Certain biological indicators do provide information about the types of stress affecting a biological community. However, if stress to a stream community is chemical, chemical monitoring in addition to biological monitoring is required to determine the actual pollutants responsible for biological or water quality impairments and their sources. Chemical monitoring and toxicity tests are also necessary to design appropriate pollution control programs.

> ***Biological integrity***: *"The capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region" (Karr and Dudley, 1981).*

A detailed biological monitoring program requires the development of the analysis protocols for biological assessments (including reference conditions and metrics, which are discussed later). Establishing the analysis approach can require a considerable level of effort and therefore can be somewhat expensive. Also, detailed biological assessments generally require precise training and broad experience with taxonomic identification of the samples. Experience in the region where sampling is to occur is helpful, but not required. The biological assessment methods and the means of interpreting the results from assessments need to be tailored to many habitats and regions of the country. Establishing a biological assessment and monitoring program can require a significant investment of time, staff, and money. However, the majority of these are one-time, up-front investments dedicated to the establishment of reference conditions, standard operating procedures, and a programmatic quality assurance and quality control plan.

Finally, there can often be a lag between the time at which a toxic contaminant or some other stressor is introduced into a waterbody and a detectable biological response. Consequently, biological monitoring is not appropriate for determining system response due to short-term stresses, such as storms. Similarly, there is often a lag time in the improvement of biological communities following habitat restoration or pollution problem abatement. The extent of this lag time is difficult, if not impossible, to predict. Other factors also determine the rate at which a biological community recovers, e.g., the availability of nearby populations of species for recolonization following pollution mitigation and the extent or magnitude of ecological damage done during the period of perturbation (Richards and Minshall, 1992). In extreme cases, a biological community might not recover following pollution abatement or habitat restoration. Both the possibility of the failure of a biological community to recover from perturbation and unpredictable lag times before improvement is noticeable have obvious implications for the applicability of biological monitoring to some NPS pollution monitoring objectives. Table 3-1 summarizes the strengths and limitations of the biomonitoring approach.

3.2 Habitat Assessment

Habitat assessment is an important component of biological assessment and monitoring, both in describing the biological potential of a system and in addressing the Clean Water Act emphasis on "physical integrity." As mentioned in the introduction, the quality of the physical habitat is important in determining the structure of benthic macroinvertebrate and fish communities. Habitat quality refers to the extent to which habitat structure provides a suitable environment for healthy biological communities to exist. Habitat quality encompasses the three factors habitat structure, flow regime, and energy source. Habitat structure refers to the physical characteristics of stream environments. It comprises channel morphology (width, depth, sinuosity); floodplain shape and size; channel gradient; instream cover (boulders, woody debris); substrate types and diversity; riparian vegetation and canopy cover; and bank stability. Flow regime is defined by the velocity and volume of water moving through a stream. Energy enters streams as the input of nutrients in runoff or ground water, as debris (e.g., leaves) falling into streams, or from photosynthesis by aquatic plants and algae.

These three factors—habitat structure, flow regime, and energy source—are interrelated and make stream environments naturally heterogeneous. Habitat structural features that

Table 3-1. General strengths and limitations of biological monitoring and assessment approaches.

Strengths	Limitations
Properly developed methods, metrics, and reference conditions provide a tool that provides a means to assess the ecological condition of a waterbody	Development of regional methods, metrics, and reference conditions takes considerable effort and an organized and well-thought-out design
Simpler bioassessments can be relatively inexpensive and easily performed with minimal training	Rigorous bioassessment can be expensive and requires a higher level of training and expertise to implement
Bioassessment indicates the cumulative impacts of multiple stressors on biological communities, not only water quality	Basic biological assessment information does not provide information on specific cause-effect relationships
Biological assessment data can be interpreted based on regional reference conditions where reference sites for the immediate area being monitored are not available	There may be a lag time between pollution abatement or BMP installation and community recovery, so monitoring over time is required for trend detection
Bioassessments involving 2 or more organism groups at different trophic levels provide a reasonable assessment of ecosystem health	The optimal season for biological sampling season varies regionally, and sampling during multiple seasons may be required in some areas
Biological community condition reflects both short- and long-term effects	Biological assessment does not always distinguish between the effects of different stressors in a system impacted by more than one stressor

determine the assemblages of macroinvertebrates can differ greatly within small areas—or microhabitats—or in short stretches of a stream. For instance, woody debris in a stream affects the flow in the immediate area, provides a source of energy, and offers protection to aquatic organisms. Curvature (sinuosity) in a stream affects currents and thereby deposition of sediment on the inner and outer banks. Rocks and boulders create turbulence, which affects dissolved oxygen levels; deep, wide portions are areas of lowered velocity where material can settle out of the water and increased decomposition occurs.

Microhabitat: *In streams, any small-scale physical feature contributing to the texture of the habitat such as the type and structure of substrate particles; submerged, emergent, or floating aquatic vegetation; algal growths; snags and woody debris; or leaf litter.*

The aspects of habitat structure mentioned above (channel morphology, floodplain size and shape, etc.) should be inspected during habitat assessment. The aspects are separated into primary, secondary, and tertiary groupings corresponding to their influence on small-, medium-, and large-scale aquatic habitat features (Plafkin et al., 1989). The status or condition of each aspect of habitat is characterized as falling somewhere on a continuum from optimal to poor. An optimal condition would be one that is in a natural state. A less than optimal condition, but one that satisfies most expectations, is suboptimal. Slightly worse is a marginal condition, where degradation is for the most part moderate but is severe in some instances. Severe degradation is characterized as a poor condition. Habitat assessment field data sheets (see Plafkin et al., 1989) provide narrative descriptions of the condition categories for each parameter. Habitat can be assessed visually, and a number of biological assessment methods incorporate assessments of the surrounding habitat (Ball, 1982; Ohio EPA, 1987; Plafkin et al., 1989; Platts et al., 1983).

The relationship between habitat quality and biological condition is generally one of three types (Barbour and Stribling, 1991):

- The biological community varies directly with habitat quality—water quality is not the principal factor affecting the biota.
- The biological community is degraded relative to the potential of its actual habitat—water quality degradation is implicated as a cause of the biotic condition.
- The biological community is elevated above what actual habitat conditions should support—organic enrichment in the water or alteration of energy source is suspected as a cause.

A clear distinction between impacts due to watershed (i.e., large-scale habitat), stream habitat, and water quality degradation is often not possible, so it is difficult to determine with certainty the extent to which biological condition will improve with specific improvements in either habitat or water quality.

3.3 Overview of Biological Assessment Approaches

3.3.1 Screening-Level or Reconnaissance Bioassessment

The simplest bioassessment approach that can be used to obtain useful information about the status of an aquatic community and the condition of a site is a screening-level, or reconnaissance, bioassessment (Plafkin et al., 1989; USEPA, 1994a). This type of survey can be done inexpensively and with few resources. If conducted by a trained and experienced biologist with a knowledge of aquatic ecology, taxonomy, and field sampling techniques, the results of screening-level bioassessment will have the greatest validity. This bioassessment method is most often conducted using benthic macroinvertebrates and is described in detail in Plafkin et al. (1989) and by the U.S. Environmental Protection Agency (USEPA, 1994a).

The first element of the screening-level approach, as in many biological assessment approaches, is an assessment of physical habitat. The instream habitat should be inspected for the amount of embeddedness, type of bottom substrate, depth, flow velocity, presence of scoured areas or areas of sediment deposition, relative abundance of different habitat types (pools, riffles, runs),

presence of woody debris, and aquatic vegetation. When conducting the assessment in a stream, record whether the stream channel has been altered. If the assessment is in a lake, reservoir, or pond, determine whether artificial bottoms or shorelines (beach sand, cement) have been installed. The riparian habitat must also be inspected for the amount of riparian cover, evidence of bank erosion, areas where livestock enter to water, and proximity of altered land uses (e.g., residential, agricultural, silvicultural, or urban). Determine the width of any natural vegetation buffer areas. The surrounding land use should be noted as a percentage of each type (e.g., 40 percent agriculture, 40 percent wooded, 20 percent residential).

The biological sampling portion of the streamside bioassessment is relatively simple. No laboratory work is involved, and it can be conducted by a person with a basic knowledge of aquatic biology. Macroinvertebrates should be collected from different instream habitats, with data from each habitat kept separate. Calculations of relative abundance and number of orders/families represented are then made. Calculations of basic community structure can also be made if specimen identifications are sufficiently detailed to allow determination of the functional feeding group the organisms occupy. Sample calculations of relative abundance and community structure are presented in Figure 3-1. Different functional feeding groups dominate in different habitats (filterers and scrapers dominate in riffle/run habitats whereas shredders dominate areas with large amounts of woody debris), so these calculations require that distinct habitats be sampled. Samples of invertebrates from woody debris of all types should be taken, including sticks, twigs, leaves, needles, and so forth. Freshly fallen debris will generally support a less representative macrobenthos than debris that is at least 50 percent decomposed.

A reference collection of biological organisms, usually available at a museum or university, should be used to positively identify any specimens whose identification is in doubt. A reference collection is a collection of preserved specimens of organisms from an area that is the same as or similar to that where monitoring is done.

> ***Reference collection:*** *A biological collection of positively identified specimens with one or more individuals representing each taxon likely to be sampled in the study area.*

Judgment of biological condition is made using the presence or absence of indicator taxa, the dominance of nuisance or sensitive taxa in the sampled habitats, or the evenness of taxonomic distribution and comparison with what is expected at unimpaired locations. A trained biologist will be able to determine whether the biota at a site are moderately or severely impaired using this approach, but subsequent sampling is often necessary to confirm any findings. The most useful application of this approach is for problem identification or screening and for setting pollution abatement priorities. Florida DEP has developed a biological screening tool, the BioRecon, that is used for this purpose in its nonpoint source pollution control program.

3.3.2 Paired-Site Approach

The paired-site approach for biological monitoring involves the use of control and treatment sites for the detection of changes in biological condition. This approach is useful for the detection of changes due to changes in water quality, habitat quality, or land use features. A

Example of transforming benthic macroinvertebrate data into biological metrics

The data presented here are from a single sampling event at one site in New England. They are a 200-organism subsample of organisms collected with the 20-jab method (USEPA, 1997) for low-gradient streams. The 11 metrics calculated from the data (note that 7 metrics fall under the category "Percent of Total Sample") are described below.

Taxa Richness - the number of distinct taxa in the sample.

Hilsenhoff Biotic Index - measures the abundance of tolerant and intolerant individuals in a sample by the following formula: $HBI = \sum x_i t_i / n$, where x_i is the number of individuals in the ith species, t_i is the tolerance value of the ith species, and n is the total number of species in the sample.

EPT Index - the number of taxa in the insect groups Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies).

Percent Dominance - the number of individuals in the numerically most dominant taxon as a proportion of the total sample [(Number individuals in dominant taxon / Total individuals in sample) x 100].

Percent of Total Sample - (1) *Pisidium*, (2) Simuliidae (= *Prosimulium* + *Simulium*), (3) Isopoda (= *Caecidotea*), (4) Diptera, (5) EPT (= mayflies, stoneflies, caddisflies), (6) filters FFG, and (7) collectors FFG.

Figure 3-1. Sample calculations of biological metrics.

key element of the approach, as the name implies, is the simultaneous monitoring of sites that are not affected by the changes for which the monitoring is being conducted (control sites) and separate sites that are affected by a "treatment" (treatment sites), which might be BMP implementation or another form of NPS pollution control. To provide reliable and valuable data, the control and treatment sites must be as similar as possible. "Similarity" in this context means that the biological populations to be monitored at both the control and treatment sites must respond similarly to changes in environmental parameters (Richards and Minshall, 1992; Skalski and McKenzie, 1982). Paired sites can be similar watersheds within a region or separate sites within a watershed that are located upstream and downstream from a nonpoint source of pollution.

Habitat assessment is as important in the paired-site approach as it is in the other biological assessment approaches. Because of the influence of surrounding landscape features on aquatic biota, control and treatment sites that are influenced by the same habitat features should be chosen. This implies that the hydrologic characteristics (flow, waterbody type, channel width, etc.) of the waterbodies in which the

Example of transforming benthic macroinvertebrate data into biological metrics

RAW DATA				CALCULATED METRICS	
Taxa	**Number**	**FFG**	**TV**	**Metrics**	**Values**
Oligochaeta	2	col	8	Taxa Richness	30
Valvata	4	scr	6	Hilsenhoff Biotic Index	5.7
Fossaria	3	col	6	EPT Index	5
Gyraulus parvus	15	scr	8	Percent Dominance	17
Pisidium	32	fil	8	Percent Pisidium	14
Hydracarina	1	pre	6	Percent Simuliidae	19
Caecidotea	6	col	6	Percent Isopoda	3
Hyallela azteca	35	col	8	Percent Diptera	38
Arthroplea	2	fil	3	Percent EPT	18
Ameletus	34	col	0	Percent Filterers (fil)	33
Amphinemura	1	shr	3	Percent Collectors (col)	45
Anabolia	4	shr	5		
Neophylax	2	scr	3		
Anomalagrion/Ischnura	1	pre	9		
Hygrotus	1	pre	5		
Hydrobius	1	pre	5		
Sialis	3	pre	4		
Nymphuliella	1	shr	5		
Diptera	1	fil	8		
Culicoides	1	pre	2		
Simulium	40	fil	6		
Prosimulium	4	fil	2		
Chrysogaster	1	col	10		
Molophilus	1	shr	4		
Pseudolimnophila	1	pre	2		
Bittacomorpha	1	col	8		
Ptychoptera	2	col	8		
Orthocladiinae	20	col	5		
Tanypodinae	15	pre	7		
Chironominae	2	col	6		
TOTAL NUMBER	**237**				

LEGEND

FFG	functional feeding group
TV	tolerance value
scr	scrapers
pre	predators
shr	shredders
fil	shredders
col	collectors

Figure 3-1. (continued)

control and treatment sites are located, surrounding land use, slope and soil type, and riparian vegetation should be as nearly identical as possible. For the same reasons, similar habitats should be sampled at the control and treatment sites. Types of substrates, locations in the streams (e.g., center or edge of stream), and nature of surrounding aquatic vegetation and debris should be nearly identical. All habitat features should be thoroughly investigated before the final selection of the control and treatment sites to be monitored to ensure their similarity.

Determination that the biota at control and treatment sites respond similarly to environmental factors is extremely important and usually requires separate sampling before any treatment is introduced at the treatment sites. It is important that the biota at control and treatment sites vary similarly both spatially and temporally, and a critical assumption of the paired-site approach is that the control and treatment populations *do* respond similarly to environmental parameters (Skalski and McKenzie, 1982). It is this similarity of response that enables one to detect changes due to the treatment. If control and treatment populations that respond differently to environmental factors are chosen, then the effect of the treatment cannot be determined.

Identification of specimens to the genus or species level should be sufficient to determine significant changes in biological communities at pairs of sites. Some laboratory work may be necessary or desirable to be certain that accurate identifications have been made.

Pretreatment sampling establishes the pattern of changes at the control and treatment sites. Skalski and McKenzie (1982) recommend that the proportional abundance of populations of macroinvertebrates at control and treatment sites be the parameter used to determine any change attributable to the treatment. Further discussion of monitoring program design and data analysis for the paired-site approach can be found in Skalski and McKenzie (1982) and Richards and Minshall (1992).

3.3.3 Composited Reference Site Bioassessment

Composited reference site bioassessment is an approach wherein biological communities at monitored sites are compared to "reference" biological communities, or reference conditions, which represent biological communities in unimpaired or minimally impaired waterbodies in the region of interest. Reference conditions are discussed in greater detail below. The approach is useful for ranking sites according to the degree by which they differ from the reference condition, which is equated with the degree of impairment at the monitored site. The composited reference site approach integrates characteristics among broad geographical areas and watersheds and thus is a more comprehensive assessment and monitoring approach than the paired-site approach. Regional reference conditions are useful for providing ecological realism in impairment criteria because they incorporate the geographic distritution, or biogeography, of organisms. The composited reference site approach also requires the greatest amount of time, specialized expertise, and field and laboratory effort to perform, largely because to conduct a composited reference site biological assessment and monitoring program, it is

***Biogeography**: The geographic distribution of plants and animals that results from a combination of their evolutionary history, mobility, and ability to adapt to changing conditions.*

necessary to initially establish a reference database for the region in which monitoring will be conducted. However, it is recognized as being the most accurate approach.

Biological sampling for the composited reference site approach is conducted at each of the reference sites on a periodic basis, which can vary from region to region. Once a composite of reference sites has been established, monitoring can be conducted on a randomly-selected subset of the reference sites, thereby reducing the intensity of monitoring. All monitoring of reference sites and assessment sites (unknown condition) is done within a specific index period, which reduces temporal (seasonal) variation. This approach can include more than one index period (as in the Florida nonpoint source program), but it is usually based on a single index period established to optimize the evaluation of biological communities (as is done by Ohio EPA and Delaware DNREC).

The habitat assessment phase of composited reference site bioassessment is not different from that of the screening-level bioassessment approach. Refer to section 3.3.1 for a description of what is involved.

The biota collection phase for composited reference site bioassessment is similar to that of a screening-level bioassessment, but involves the collection of additional samples to detect subtle differences in NPS pollution impacts. Specimen identification is generally done in the laboratory to the genus or species level. This level of detail allows for a more accurate analysis of community structure and biological condition. Data analysis using genus- and species-level identifications can provide information on the generic cause of impairment (nutrient enrichment, toxic pollutants, or habitat degradation). To gain this level of insight, however, it is necessary to be able to distinguish the effects of NPS pollution impairment from natural variability of the populations being sampled. Reference conditions must be established for this purpose.

Additionally, area- or region-specific metrics must be established before the composited reference site bioassessment approach can be used effectively. During the process of establishing reference conditions for an area, metrics specific to the area are selected and calibrated. Figure 3-2 describes the development of metrics and associated reference conditions in a step-by-step manner.

> ***Metric:*** *An enumerated or calculated term that represents some aspect of biological assemblage structure, function, or other measurable feature and that changes in a predictable way in response to environmental (including human) influences.*

3.4 Reference Sites and Conditions

In biological assessment, macroinvertebrates and fish are commonly used as indicators of the condition of biological communities. Comparisons are made between macroinvertebrates and fish found at undisturbed sites and those found at monitored sites to determine how closely they resemble one another. The undisturbed, or reference sites, are aquatic habitats that are assumed to fully support natural biological communities. The greater the difference is between reference and monitored sites, the more disturbed the monitored sites are considered to be. The disturbance responsible for the difference might be a habitat change, pollution, or some other stress.

A reference condition is a composite characterization of the natural biological condition in an ecologically homogeneous region created

The Process for Metric Selection, Validation, and Development of Reference Conditions

Step 1. ***Reference Site Selection.*** Select candidate reference sites from maps and other available information and confirm through reconnaissance. Sites are confirmed through the existence of non-degraded physical habitat *and* the absence of known contaminant sources.

Assuming that reference sites are available (see explanation in Section 3.4) if reference sites are not available), candidate sites are selected to represent the "natural" condition within a region or area. These sites should be representative of:

- Extensive natural riparian vegetation
- Natural channel structures typical of region
- Natural hydrograph (typical flow patterns and discharges)
- Absence (or minimal presence) of sources of perturbation

These sites can be identified from existing GIS or land use maps, historical data, or local "expert" knowledge, and confirmed through site reconnaissance.

Step 2. ***Site Classification.*** Determine site classes based on mapped information or regional water quality characteristics such as, e.g., ecoregion or subecoregion, gradient, alkalinity, and hardness.

The purpose of site classification is to partition the variability within each biological metric to enhance the ability to discriminate impairment from nonimpairment, or to improve the interpretation of change in monitoring. Physicochemical aspects, e.g., ecoregions, alkalinity, pH, elevation, drainage area, etc., are analyzed to derive site classes. Then, the biological metrics are used to confirm site classes and to partition variability.

For example, the number of taxa in a set of reference sites might range between 10 and 40 species. However, in ecoregion A, the number of taxa is between 10 and 25 to represent a natural community. In ecoregion B, the number of species ranges between 20 and 40. The classification of sites by ecoregion, in this case, allows for a better understanding of natural variability than a universal compositing of all reference sites.

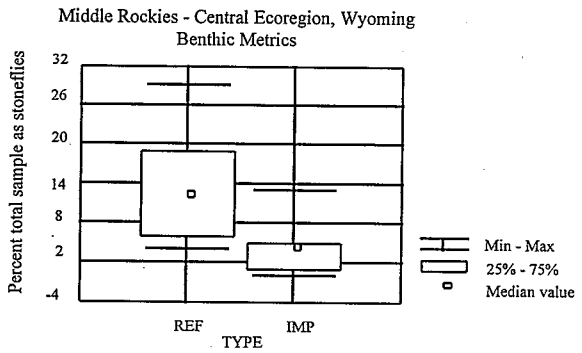
Step 3. ***Candidate Metric Selection.*** List all metrics that are relevant to the biological communities being used for assessment of a site or waterbody.

Metrics allow the investigator to use meaningful indicator attributes in assessing the status of communities in response to perturbation, or to monitor trends in the health of the communities. All metrics that have relevancy to the assemblage under study and will respond to the targeted stressors are potential metrics for consideration. For example, the number of taxa as a measure of diversity can be identified for various groups of organisms that are relatively sensitive to environmental change (i.e., mayflies, darters, diatoms, etc.); the relative dominance of a single taxon is informative of a pollutant situation; an imbalance in trophic structure is suggestive of an adverse effect on food source.

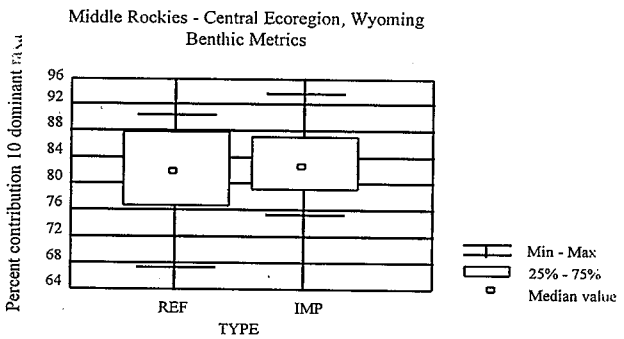
Figure 3-2. The process for metric selection and validation and development of reference conditions.

Step 4. ***Determination of Core Metrics.*** Calculate metrics based on biosurvey data (see Figure 3-1). Compare value range of each candidate metric from reference sites to those from impaired sites. Metrics become part of the core analysis if the data show them able to discriminate between reference and impaired sites.

Core metrics are those remaining following initial candidate metric screening that will discriminate between good and poor quality ecological conditions, or will provide a basis for monitoring changes over time. Metrics that use the relative sensitivity of the monitored assemblage to specific pollutants or stressors, where these relationships are well characterized, can be useful as a diagnostic tool. The discriminatory ability of metrics can be evaluated by comparing the distribution of each metric at a set of assumed reference sites with the distribution of the same metric values from a set of known impaired sites within each site class. This is done to calibrate the metrics. If there is minimal or no overlap between the percentile distributions, the metric can be considered to be a strong discriminator between reference and impaired conditions. The following two figures graphically demonstrate the difference between strong and weak metrics:



(a) Strong metric: Percent Total Sample as Stoneflies



(b) Weak metric: Percent Contribution of 10 Dominant Taxa

Figure 3-2. (continued)

ecoregions in the conterminous United States. The size of each ecoregion is a function of its within-region homogeneity relative to between-region variation. The ecoregion concept is useful for water quality management because waterbodies within ecoregions are relatively homogeneous and can therefore be managed similarly.

Omernik (1987, 1995) found that hydrologic units such as river basins cannot be used to accurately delineate ecoregions, but that within an ecoregion there might be separate watersheds or subwatersheds. Therefore, surveys and monitoring conducted in several watersheds are strengthened by the ecoregion framework.

Characteristics other than ecoregion are also helpful in classifying sites. For example, the Wyoming Department of Environmental Quality (DEQ) found that elevation distinguishes stream classes within the Middle Rockies Ecoregion. EPA's *Biological Criteria: Technical Guidance for Streams and Small Rivers* (Gibson et al., 1996) describes the process for classifying sites and selecting reference sites.

In a landscape that is heavily altered by agricultural activity, silviculture, industrial-commercial development, or urbanization, undisturbed streams or reaches might not exist and reference conditions might need to be determined based on best professional judgment of that which is likely attainable, historical records, or another means of estimation. The most appropriate approach to establishing reference conditions is to conduct a preliminary resource assessment to determine the feasibility of using reference sites (Figure 3-3). If acceptable, minimally impaired reference sites cannot be found for a region, some form of simulation modeling might be the best alternative. Biological attributes can be modeled from neighboring regional site classes, expert consensus, and/or a composite of "best" ecological information. Such models might be the only viable means of examining significantly altered systems. The expectations derived from these models should be regarded as hypothetical until more reliable information is obtained.

3.5 Rapid Bioassessment Protocols

EPA has recommended a set of rapid bioassessment protocols, or RBPs, that use benthic macroinvertebrate and fish communities to assess biological condition in streams and wadable rivers. The five protocols differ in the level of effort, taxonomic level, and expertise required to perform them, and in the applicability of the data obtained (Table 3-2). More intensive bioassessments (RBPs III and V) give the most useful information for trend analysis and establishment of a baseline for problem diagnosis. RBPs I and II are less intensive bioassessment approaches and are useful for setting priorities for more intensive study. RBP IV, not described here, is a screening technique used to survey persons knowledgeable about the fish in an area. For further information about RBP IV, consult Plafkin et al. (1989).

Selection of appropriate organisms and protocols for biological assessment depends on the objectives of the monitoring study (Figure 3-4). RBP benthic protocols have been applied in freshwater streams and wadable rivers, and their applicability is presently limited to these waterbodies. Fish RBP protocols have been used in freshwater streams and larger rivers and are applicable to both. RBP-type methods for fish and invertebrates have been adapted for use by many states and federal agencies and are in use across the country (Southerland and Stribling, 1995).

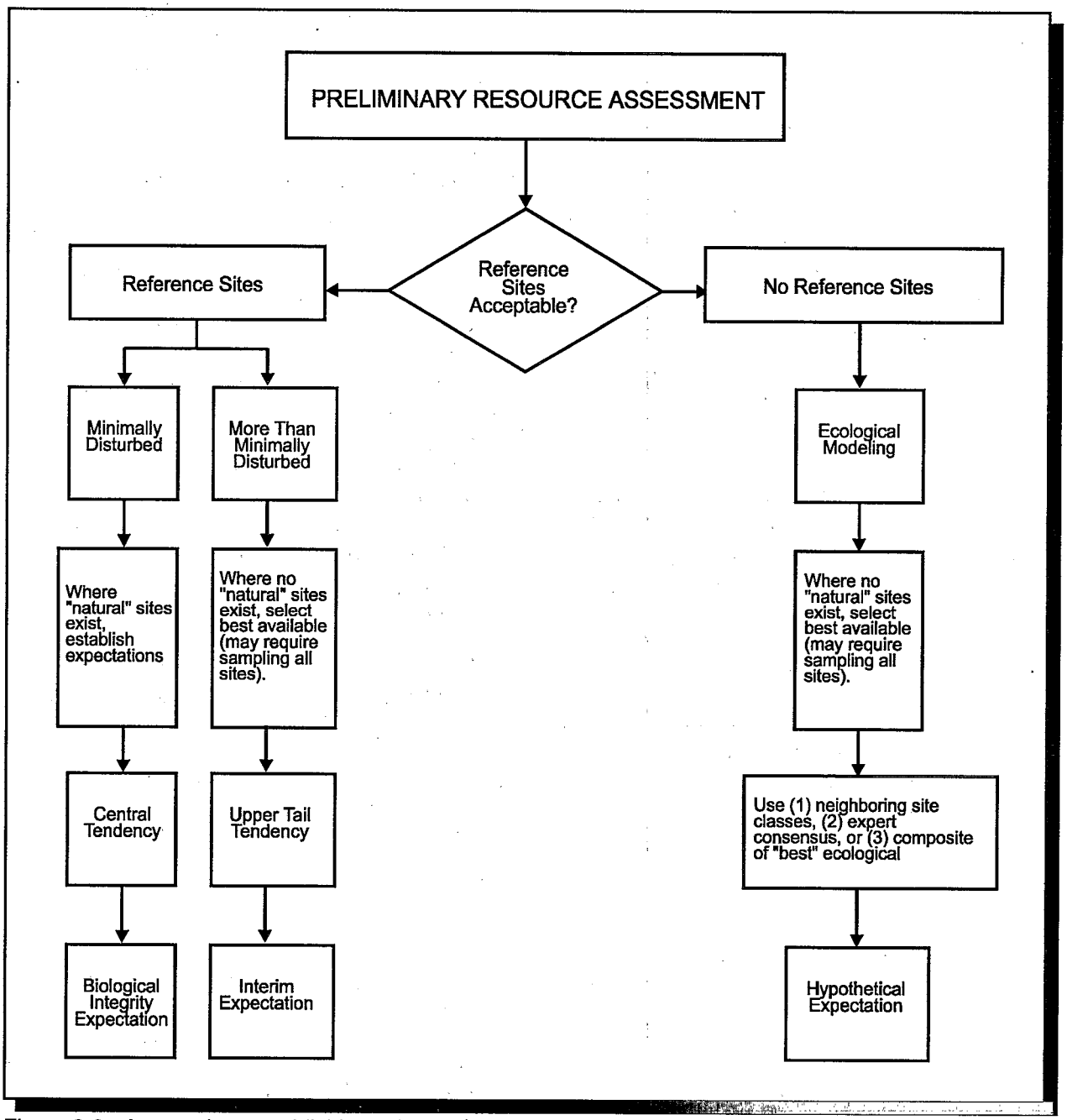


Figure 3-3. Approach to establishing reference conditions (Gibson et al., 1994).

Table 3-2. Five tiers of the rapid bioassessment protocols (Plafkin et al., 1989).

Level or Tier	Organism Group	Relative Level of Effort	Level of Taxonomy/Where Performed	Level of Expertise Required
I	benthic invertebrates	low; 1-2 hr per site (no standardized sampling)	order, family/field	one highly trained biologist
II	benthic invertebrates	intermediate; 1.5-2.5 hr per site (all taxonomy performed in field)	family/field	one highly trained biologist and one technician
III	benthic invertebrates	most rigorous; 3-5 hr per site (2-3 hr of total are for lab taxonomy)	genus or species/ laboratory	one highly trained biologist and one technician
IV	fish	low; 1-3 hr per site (no fieldwork involved)	not applicable	one highly trained biologist
V	fish	most rigorous; 2-7 hr per site (1-2 hr per site are for data analysis)	species/ field	one highly trained biologist and 1-2 technicians

3.6 *THE MULTIMETRIC APPROACH FOR BIOLOGICAL ASSESSMENT*

Accurate assessment of biological condition requires a method that integrates biotic responses through an examination of patterns and processes from the organism to ecosystem level (Karr et al., 1986). The rapid bioassessment protocols (Plafkin et al., 1989) discussed above make use of an array of measures that individually provide information on diverse biological attributes and, when integrated, provide an overall indication of biological condition.

The raw biological data collected during a survey consist entirely of taxonomic identifications and numbers of individuals within each taxon. The level of identification—whether to family, genus, or species—depends on the method being used. For instance, RBP II involves identification to the family level, whereas RBP III involves identification to the lowest practical level, generally genus or species. These data are used to calculate or enumerate a variety of values, or metrics. Each reflects a different characteristic of community structure and has a different range of sensitivity to pollution stress (Plafkin et al., 1989). Appropriately developed metrics can be used to draw conclusions about different aspects of the biological condition at a site, and measurements of multiple metrics in a biological assessment will yield a more accurate representation of the overall biological condition at a site. Gray (1989) stated that the three best-documented biological responses to environmental stressors are a reduction in species richness, a change in species composition to dominance by opportunistic species, and a reduction in the mean

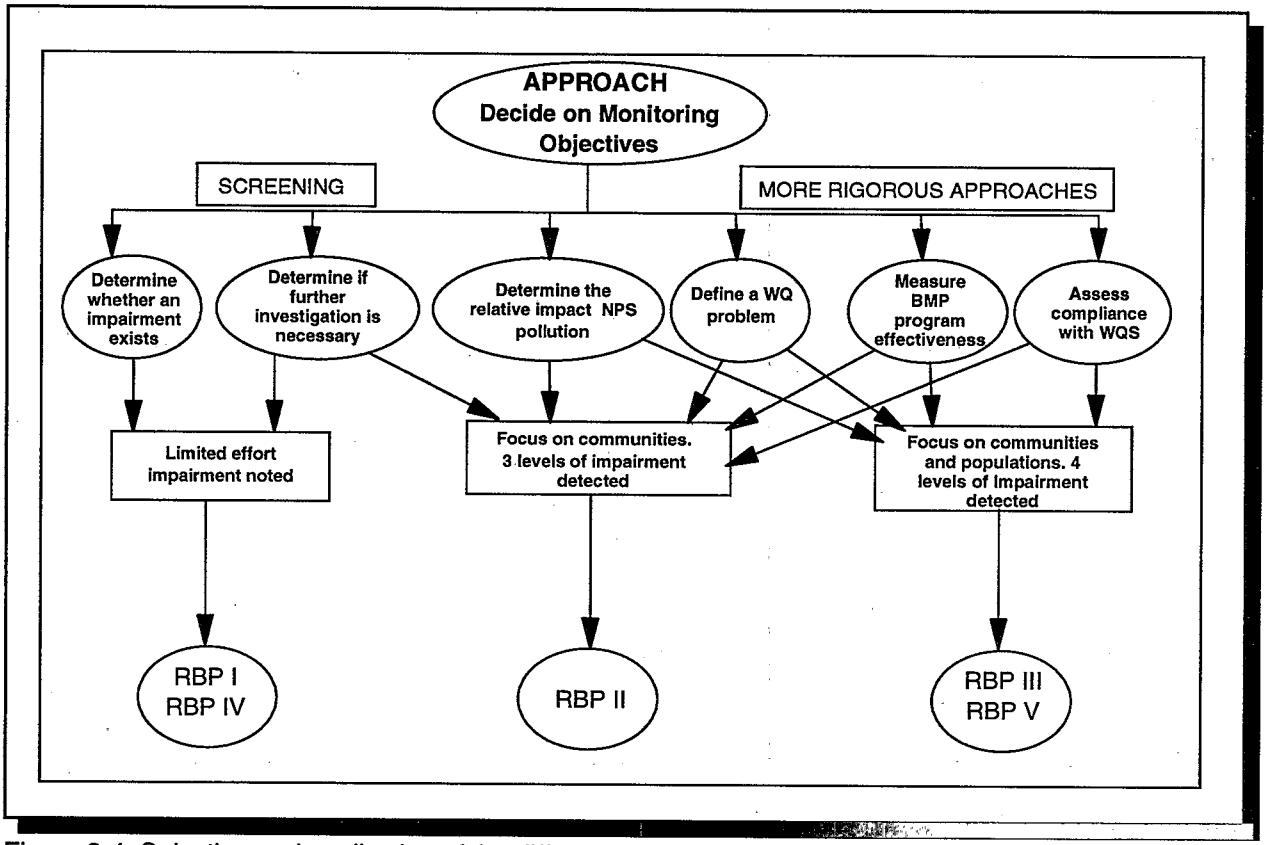


Figure 3-4. Selection and application of the different tiers of RBP depend on monitoring objectives.

body size of organisms. Though the last type of biological response (change in mean body size) may be well documented, it is rarely used in the more common bioassessment protocols because the level of effort for an accurate interpretation can be prohibitive.

Figure 3-5 illustrates a conceptual structure for the attributes calculated or measured for a biological assemblage during a biological assessment. Generally, the biological assemblage at a site can be characterized by metrics organized into four classes—community structure, taxonomic composition, individual condition, and biological processes. These are described below.

Community structure is characterized by measurements of the variety of taxa and the distribution of individuals among taxa. *Taxa richness* is the number of distinct taxa in a sample and reflects the diversity of the sample. The *relative abundance* of each taxon is a comparison of the number of individuals in one taxon to the total number of individuals in the sample. *Dominance* is calculated as the percent composition of the dominant taxon within the total sample. It indicates balance within the community.

Taxonomic composition refers to the types of taxa in the sample. *Sensitivity* to pollution and other environmental disturbance is the number of

pollution-tolerant and intolerant species in the sample. The presence of exotic and nuisance species is also noted because they can play important ecological roles and indicate stressed conditions.

Individual condition is more easily measured in fish than in benthos and periphyton; it refers to the presence or absence and frequency of diseases and anomalies. Contaminant levels in the tissues of individuals can also be measured. The frequency of head capsule deformities in midges (Chironomidae) has been used by some researchers.

Biological processes occurring at the sample site are indicated by measurements of species that perform specific functions within the community. For instance, the functional feeding groups (e.g., detritivores, filter feeders) indicate the primary source of energy for the biological system.

Numerous biological metrics have been tested in various regions of the country (Figure 3-6), primarily for fish and benthos. Summaries of those used have recently been presented in tabular form (Gibson et al., 1996; Barbour et al., 1995) and are reproduced in Tables 3-13 and 3-14 at the end of this chapter. Examples of metrics that have been tested and have had scoring criteria established are those for the montane region of Wyoming (Table 3-3) and the plains streams of Florida (Table 3-4). Figure 3-1 explains five common metrics and presents sample data and calculated values for each of the metrics. Readers should calculate the metrics themselves to be certain that use of the data for metric calculation is understood.

3.7 Sampling Considerations

The large influence that small environmental factors, such as amount of sunlight or presence of woody debris in a waterbody, can have on aquatic communities means that even though there might not be easily distinguished boundaries between habitats, such as those between riffles and pools, the biota inhabiting them are often taxonomically and biologically distinct (Hawkins et al., 1993). The distribution of benthic fauna in lakes and streams is also heterogeneous because of variable requirements among species for feeding, growth, and reproduction, which are satisfied for different species by different substrata, water chemistry, and inputs of woody debris (Wetzel, 1983). This leads to a patchy, nonrandom distribution of animals.

Because of the influence that habitat has on biological communities, sampling similar habitats at all sampling stations is important for data comparability and for data interpretation (Plafkin et al., 1989). Collection of habitat quality data each time biological data are collected helps to establish the correlation between the two for a particular ecoregion. Obviously, the more correlative data that are collected, the more useful they will be in interpreting sampling data, that is, in separating water quality and habitat quality effects as they relate to biological condition. The ability to separate the two influences is important for determining the expected or potential improvement in biological condition from water quality improvement programs such as point or nonpoint source pollution control (Barbour and Stribling, 1991).

When sampling multiple habitats, it is important to establish consistency in sampling procedures. Sampling protocols should be standardized, and the same level of effort should be applied at each sampling station. Because differences in gear efficiency and techniques may affect results, standardized sampling is needed if direct comparisons are to be made between stations or between data from a single station at different times.

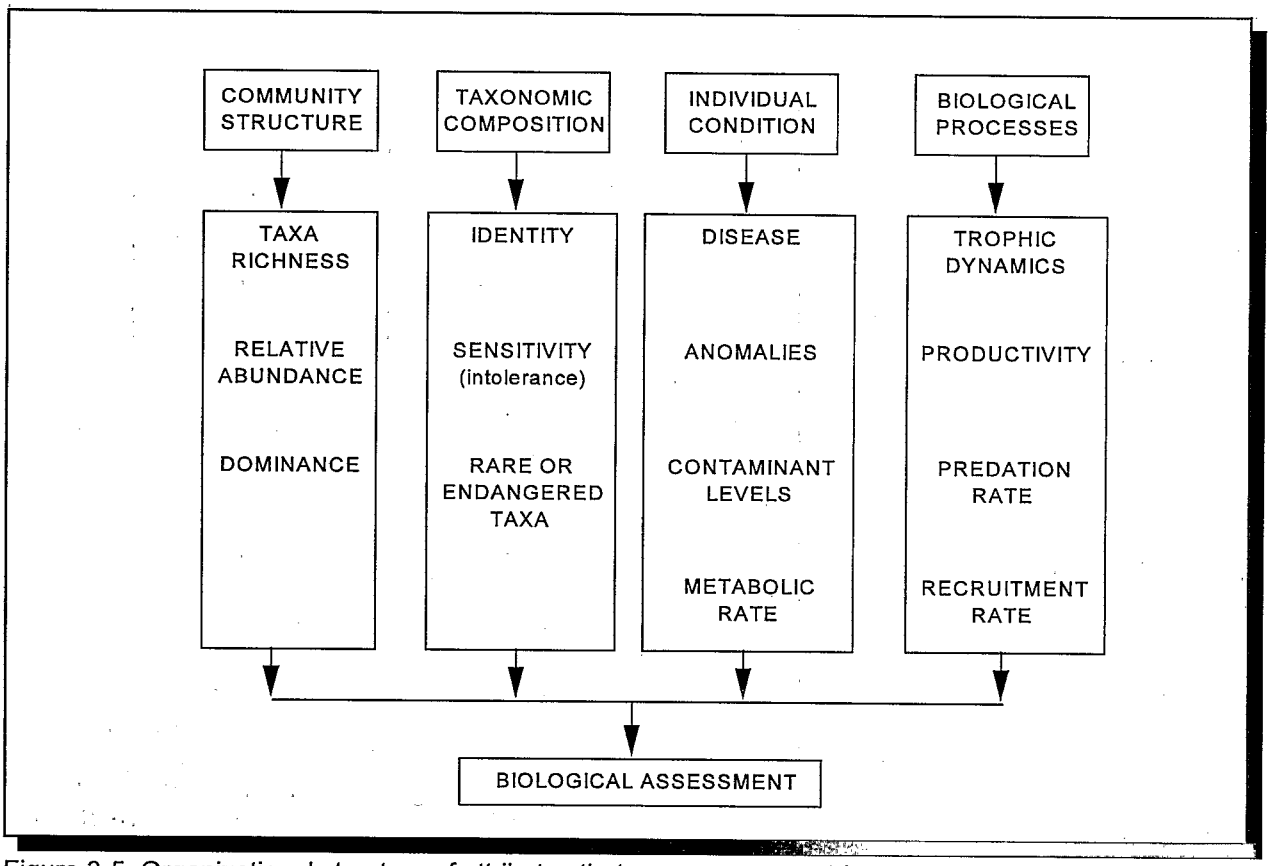


Figure 3-5. Organizational structure of attributes that can serve as metrics.

3.7.1 *Benthic Macroinvertebrate Sampling*

Stream environments contain a variety of macro- and microhabitat types including pools, riffles, and runs of various substrate types; snags; and macrophyte beds. Relatively distinct assemblages of benthic macroinvertebrates inhabit various habitats (Hawkins et al., 1993), and it is unlikely that most sampling programs would have the time and resources to sample all habitat types. Decisions on the habitats selected for sampling should be made with consideration of the regional characteristics of the streams. For instance, high-gradient mountain streams are best sampled from the cobble substrate of riffles for macroinvertebrates, whereas low-gradient coastal streams lack riffles and are appropriately sampled from snags and shorezone vegetation. These two different stream types might be sampled with different methods, during different times of the year, or with different biological index periods. The seasonal variability of the biota and stream environment are key factors that determine the proper index period. Established sampling protocols that are part of existing monitoring programs should be considered for NPS bioassessment. However, current and/or historical sampling approaches should be evaluated to determine whether they will provide the required data to address the program objectives.

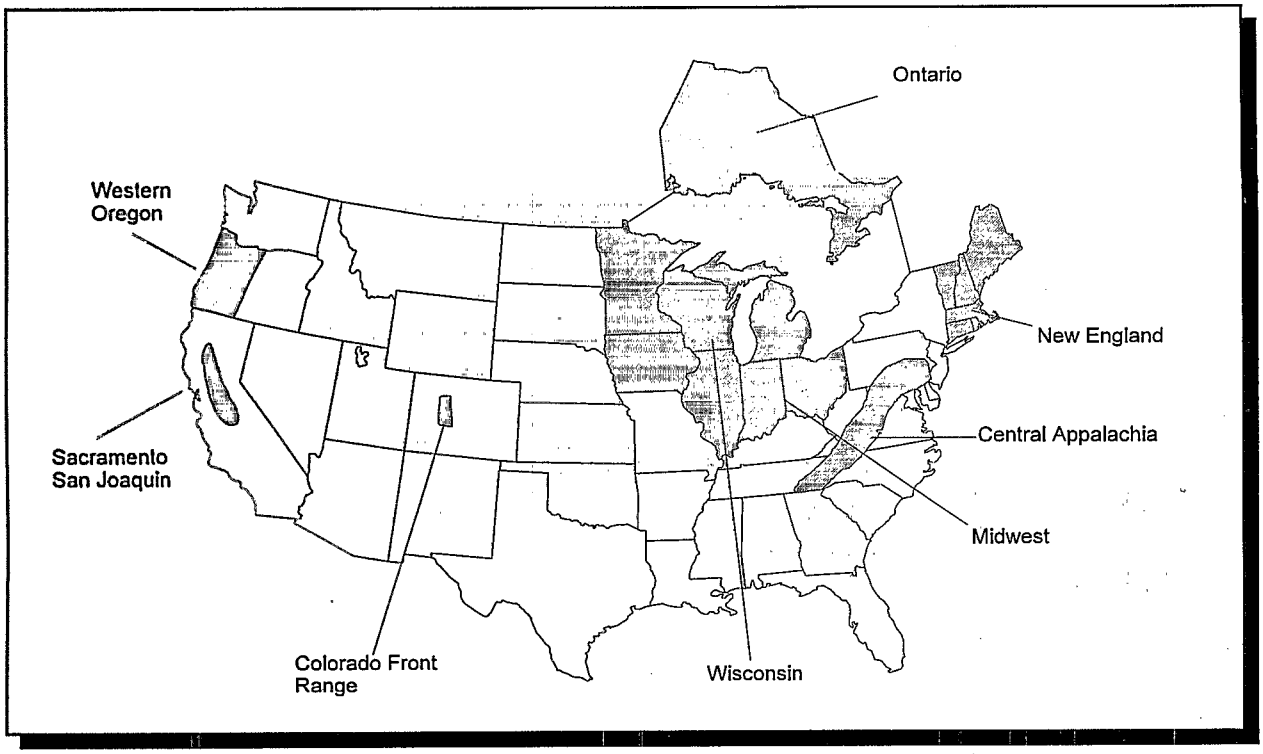


Figure 3-6. Areas in which various fish IBI metrics (see Table 3-4) have been used.

Sampling a single habitat, such as riffles, limits the variability inherent in sampling natural habitats. This can produce a more repeatable characterization of the biological condition of a stream because sampling bias is reduced, whereas sampling of multiple habitats must be carefully standardized to reduce investigator bias and to control for sampling efficiency problems.

If the biological assessment strategy is to sample a single habitat, the most representative stable habitat conducive to macroinvertebrate colonization should be chosen. The most suitable habitat choice will vary regionally. The key is to select one habitat that supports a similar assemblage for benthos within a range of stream sizes, is the most representative of the stream type or class under investigation, and is likely to reflect anthropogenic disturbances within the watershed. Suitable habitat alternatives to riffles for sampling benthos include snags, downed trees, submerged aquatic vegetation beds, emergent shoreline vegetation, and the most prevalent substrate.

The RBPs recommended by EPA specify that a subsample of 100 organisms be used for a biological assessment; however, several states use 200 to 300 organisms or more. Agencies should evaluate the level of subsampling required to meet their objectives. The level of taxonomic identification should be specified in the study design and is determined by the study objectives. Identification to the species level gives the most accurate information on pollution tolerances and sensitivities, though some metrics or analytical techniques might require identification only to the order, family, or genus. Verification of

Table 3-3. Scoring criteria for the core metrics as determined by the 25th percentile of the metric values from the Middle Rockies-Central Ecoregion, Wyoming.

Metric	Stream Class	Elevation > 6,500 ft			Elevation < 6,500 ft		
	Score	5	3	1	5	3	1
EPT Taxa		>14	14-8	<8	>18	18-10	<9
% Plecoptera		>7%	7%-4%	<4%	--	--	--
% Ephemeroptera		--	--	--	>22%	22%-11%	<11%
% Chironomidae		<36%	36%-68%	>68%	>6%	6%-4%	<4%
Predator Taxa		>5	5-3	<3	<12%	12%-39%	>39%
% Scrapers		>5%	5%-3%	<3%	>7	7-4	<4
MHBI		<4.3	4.3-4.8	>4.8	>8%	8%-5%	<5%
BCI		>74	74-42	<42	<3.7	3.7-4.7	>4.7
CTQD		<99	99-112	>112	>79	79-46	<46
Shannon H		>3.5	3.5-1.8	<1.8	--	--	--
% Multivoltine		<48%	48%-57%	>57%	--	--	--
% Univoltine		>40%	40%-20%	<20%	--	--	--
% Collector-Filterers		--	--	--	<2.6%	2.6-23.2	>23.2%
Range of Aggregate Score		11-55			9-45		

taxonomic identifications is critical and can be accomplished by (1) comparing specimens with a reference specimen collection or (2) sending specimens to taxonomic experts familiar with the group in question.

Benthic macroinvertebrates can be collected actively or passively. Two of the more commonly used active methods use a square-meter kicknet or a long-handle D-frame. The former is typically used at sites that are considered to be in higher-gradient (riffle-prevalent) streams (Plafkin et al., 1989); the latter is used primarily in coastal plains streams and is standardized as the 20-jab method (USEPA, 1997). In both of these methods, organisms are dislodged from their substrate by the sampler and captured in a net. Passive collection approaches include the Hester-Dendy multiplate sampler and rock baskets. These are considered artificial substrates. They are placed in the stream or stream bottom and left for a standardized amount of time. Upon retrieval, the invertebrates are removed from the sampling units in the laboratory. For further information on sampling methods, see Klemm et al. (1990).

Table 3-4. Scoring criteria for the metrics as determined by the 25th percentile of the metric values for the two aggregated subecoregions for Florida streams.

Metric	Stream Class	Panhandle			Peninsula		
	Score	5	3	1	5	3	1
# of Total Taxa		≥31	16-30	0-15	≥27	14-26	0-13
EPT Index		≥7	4-6	0-3	--	≥4	0-3
# Crustacea + Mollusca Taxa		--	--	--	--	≥4	0-3
% Dominant Taxon		--	≤20	>20	--	≤37	>37
% Diptera		--	≤38	>38	--	≤32	>32
% Crustacea + Mollusca		--	--	--	--	≥16	0-15
Florida Index		≥18	0-8	0-8	≥7	4-6	0-3
% Filterers		≥12	0-6	0-6	≥8	4-7	0-3
% Shredders		--	≥0-9	0-9	--	≥13	0-12
Range of Aggregate Score		7-29			9-33		

3.7.2 *Fish Sampling*

Fish surveys should yield a representative sample of the species present at all habitats within a sampling reach that is representative of the stream. Sampling reaches should ensure that generally comparable habitats will exist at each station. If comparable physical habitat is not sampled at all stations, it will be difficult to separate degraded habitat from degraded water quality as the factor limiting the fish community (Klemm et al., 1992).

At least two of each of the major habitat types (i.e., riffles, runs, and pools) should be incorporated into the sampling as long as they are typical of the stream being sampled. Most species will be successfully sampled in areas where there is adequate cover, such as macrophytes, boulders, snags, or brush.

Sampling near modified sites, such as channelized stretches or impoundments, should be avoided unless it is conducted to assess the impact of those habitat alterations on the fish community. Sampling at mouths of tributaries entering larger waterbodies should also be avoided because these areas will have habitat characteristics more typical of the larger waterbody (Karr et al., 1986). Sampling station lengths range from 100 to 200 meters for small streams and 500 to 1000 meters for rivers. Some agencies identify their sampling reach by measuring a length of stream that is 20 to 40 times the stream width.

Fish are generally identified to the species or subspecies level. For biological assessments of the entire assemblage, the gear and methods used

should ensure that a representative sample is collected.

Fish can be collected actively or passively. Active collection methods involve the use of seines, trawls, electrofishing equipment, or hook and line. Passive collection can be conducted either by entanglement using gill nets, trammel nets, or tow nets, or by entrapment with hoop nets or traps. For a discussion on the advantages and limitations of the different gear types, see Klemm et al. (1992). The Index of Biotic Integrity (IBI) emphasizes active gear, and electrofishing is the most widely used active collection method. Ohio EPA (1987) discusses appropriate electrofishing techniques for bioassessment. Other sources for sampling method discussions are Allen et al. (1992), Dauble and Gray (1980), Dewey et al. (1989), Hayes (1983), Hubert (1983), Meador et al. (1993), and USFWS (1991).

Length and Weight Measurements

Length and weight measurements can provide estimations of growth, standing crop, and production of fish. The three most commonly used length measurements are standard length, fork length, and total length. Total length is the measurement most often used.

Age may be determined using the length-frequency method, which assumes that fish increase in size with age. However, this method is not considered reliable for aging fish beyond their second or third growing season. Length can also be converted to age by using a growth equation (Gulland, 1983).

Annulus formation is a commonly used method for aging fish. Annuli (bands formed on hard bony structures) form when fish go through differential growth patterns due to the seasonal temperature changes of the water. Scales are generally used for age determination, and each species of fish has a specific location on the body for scale removal that yields the clearest view for identifying the annuli. More information on the annulus formation method and most appropriate scale locations by species can be found in Jerald (1983) and Weatherley (1972).

Fish External Anomalies

The physical appearance of fish usually indicates their general state of well-being and therefore gives a broad indication of the quality of their environment. Fish captured in a biological assessment should be examined to determine overall condition such as health (whether they appear emaciated or plump), occurrence of external anomalies, disease, parasites, fungus, reddening, lesions, eroded fins, tumors, and gill condition. Specimens may be retained for further laboratory analysis of internal organs and stomach contents if desired.

Periphyton

Of the three biological assemblages discussed in this chapter, periphyton is perhaps the least used, though the information potential can be dramatic as well as cost-effective. Laboratory analysis of species composition is labor-intensive. Because species within a genus can display varying tolerances to a disturbance, diatoms must be identified to species. Rosen (1995) estimates an average of 2 hours per sample to identify 500 organisms to species, with processing time decreasing as taxonomic expertise is gained.

Periphyton are a community of organisms that adhere to and form a surface coating on stones, plants, and other submerged objects. These can take the form of soft algae, algal or filamentous mats, or diatoms. As the primary producers in

the stream ecosystem, their importance to the food web cannot be overstated. The advantages for using the periphyton assemblage in a bioassessment program are many:

- They have rapid reproduction rates and short life cycles and thus respond quickly to perturbation, which makes them valuable indicators of short-term impacts.
- Because they are primary producers and ubiquitous in all waters, they are directly affected by water quality.
- Periphyton sampling is rapid and requires few personnel, and results are easily quantified.
- A list of the taxa present and their proportionate abundance can be analyzed using several metrics or indices to determine biotic condition and diagnose specific stressors.
- The periphyton community contains a naturally high number of taxa that can usually be identified to species.
- Tolerance of or sensitivity to changes in environmental conditions are known for many species or assemblages of diatoms.
- Periphyton are sensitive to many abiotic factors that might not be detectable in the insect and fish assemblages.

The state of Kentucky has developed a Diatom Bioassessment Index (DBI), currently used in water quality assessments (Kentucky Department of Environmental Protection, 1993). Metrics use to construct the DBI include diatom species richness, species diversity, percent community similarity to reference sites, a pollution tolerance index, and percent sensitive species. Scores for each metric range from 1 to 5. The scores are then translated into descriptive site bioassessments, which are used to determine aquatic life use.

Network design refers to the array, or network, of sampling sites selected for a monitoring program. It usually takes one of two forms:

- **Probabilistic design:** Network that includes sampling sites selected randomly to provide an unbiased assessment of the condition of the waterbody at a scale above the individual site or stream; can address questions at multiple scales.
- **Targeted design:** Network that includes sampling sites selected based on known existing problems, knowledge of upcoming events in the watershed, or a surrounding area that will adversely affect the waterbody such as development or deforestation; or installation of BMPs or habitat restoration that is intended to improve waterbody quality; provides assessments of individual sites or reaches.

An **integrated design** combines these two approaches and incorporates multiple sampling scales and monitoring objectives.

For diatoms, Montana (Bahls, 1993) uses a diversity index, a similarity index, and a siltation index. Three other metrics—dominant phylum, indicator taxa, and number of genera—are used for soft-bodied algae to support the diatom assessment.

3.8 Biomonitoring Program Design

The design of a biomonitoring program (similar to other types of monitoring programs) will depend ultimately on the goals and objectives of the program. Several of the objectives identified

in Chapter 2 of this document can be directly addressed using a properly designed biomonitoring program. These objectives may differ in spatial and temporal scales, therefore requiring different monitoring designs as reflected in differences in the site selection process, number of sites sampled, and time and frequency of sampling. The sampling design used in nonpoint source biological monitoring consists of one of three types of network designs depending on the objectives of the programs—probabilistic, targeted, or integrated design. Objectives that are site-specific, such as determining whether biological impairment exists at a given site, are addressed using a targeted monitoring design (Table 3-5). Objectives that address questions of large-scale status and trends, for example, require a probabilistic design. For many nonpoint source objectives (see Chapter 2), an integrated network monitoring design is most appropriate.

Monitoring performed at different spatial scales can provide different types of information on the quality and status of water resources. Conquest et al. (1994) discuss a hierarchical landscape classification system, originally developed by Cupp (1989) for drainage basins in Washington state, that provides an organizing framework for integrating data from diverse sources and at different resolution levels. The framework focuses on river and stream resources at its higher resolutions, but could be modified for other waterbody types such as lakes and wetlands. In its simplest form, the nested hierarchy consists of five levels:

- Ecoregions
- Watersheds/subwatersheds
- Valley segments
- Habitat complexes (e.g., stream reach)
- Habitat units (e.g., riffle)

Assessments of waterbodies on a large scale such as an ecoregion, subregion, state, or county provide information on the overall condition of waterbodies in the respective unit. Appropriately designed probabilistic sampling can provide results such as the percentage of waterbodies in a geographic area that are impaired (status), or, if the sampling is repeated at regular intervals, an assessment of the trends in the percentage of impaired waters. Probabilistic site selection is most appropriate for an unbiased estimate of the status and temporal behavior of waterbodies on a large geographic scale.

Assessments of waterbodies and subsequent monitoring often occur at the watershed scale, within which both targeted and probabilistic sites could be selected. A probabilistic design would yield information on the watershed scale as well as on the site- or stream-specific scale; these locations might or might not be impaired. The targeted design would ensure that known problem sites or sites of special interest are evaluated and their response over time is assessed.

Assessment on a small geographic scale may involve a whole stream, river, or bay or a segment (reach) of the waterbody. A targeted sampling design applies to monitoring waterbodies within a watershed that are exposed to known stressors. Known disturbances, such as point sources, specific NPS inputs, or urban stormwater runoff, can all be targeted for small-scale assessments. It is at this scale that the effectiveness of specific pollution controls, BMP installation/implementation, natural resource management activities, or physical habitat restoration can be monitored.

Since a target population is recognized to consist of groups that each have internal homogeneity (relative to other groups), it can be stratified to minimize within-group variance and maximize

Table 3-5. Comparison of probabilistic and targeted monitoring designs.

	Advantages	Disadvantages
Probabilistic Design	Provides unbiased estimates of status for a valid assessment on a scale larger than that of the sample location. Can provide large-scale assessment of status and trends of resource or geographic area that can be used to evaluate effectiveness of environmental management decisions for watersheds, counties, or states over time. Stratified random sampling can improve sampling efficiency, provide separate data on each stratum, and enhance statistical test sensitivity by separating variance among strata from variance within strata.	Small-scale problems will not necessarily be identified unless the waterbody or site happens to be chosen in the random selection process. Cannot track restoration progress at an individual site or site-specific management goals. Stratified random sampling requires prior knowledge for delineating the strata (MacDonald et al., 1991).
Targeted Design	Systematic sampling along a stream or river can be an efficient means of detecting pollution sources (Gilbert, 1987). Identifies small-scale status and trends of individual sites, which can be used to assess potential improvements due to restoration projects and other management activities. Contributes to understanding of responses of biological resources to environmental impact.	A targeted design will not yield information on the condition of a large-scale area such as the watershed, county, state, or region. It cannot specifically monitor changes from management activities on a scale larger than site-specific. Resource limitations usually make it impossible to monitor effects of all pollutant sources using a targeted design. Systematic sampling can result in biased results if there is a systematic variation in the sampled population.

among-group variance (Gilbert, 1987). Table 3-6 summarizes a waterbody stratification hierarchy for streams and rivers, lakes, reservoirs, estuaries, wetlands, and ground water. With the exception of estuaries, the highest-level strata would be ecoregions and subecoregions. Biogeographic provinces (e.g., Virginian and Louisianian Provinces used in the Environmental Monitoring and Assessment Program (EMAP), as described by Weisberg et al. (1993)) are more appropriate as the highest stratification level for estuaries because of the relatively large size of their watersheds and the fact that they are influenced directly by marine processes.

Depending on the waterbody, subsequent stratification levels may vary in number and may be quite different across waterbodies at a given level in the hierarchy. For example, a state or regional monitoring program designed to assess the status of biological communities in streams might need to be stratified to the level of segments, whereas monitoring to assess the efficacy of specific stream restoration measures might need to be stratified to the macro- or microhabitat level. If data collected by a particular design are so variable that meaningful conclusions cannot be drawn, poststratification of the data set might be required. If stratification to the level of microhabitat is needed, the sampling and analysis methods used at higher levels might be inappropriate or inadequate.

3.8.1 Process of Randomized Sampling Site Selection

Probabilistic sampling designs require the random selection of sampling sites within the basic design (e.g., simple random, stratified random, multistage; Chapter 4). Three major steps are involved in selecting sampling sites using a probabilistic design:

(1) Identify the level of site classification. Monitoring program objectives will dictate the geographic extent to which monitoring is to be done, for instance statewide, county-wide, within an ecoregion, or within a watershed. This level of site classification should be identified initially.

(2) Stratify the site classes. A sampling design appropriate to the monitoring objectives must be selected. For probabilistic designs, simple random sampling is not usually the optimal method. It can produce clusters of sampling sites that might not be representative of the larger scale area of interest (e.g., Hurlburt, 1984). Therefore, some sort of stratification is preferred for ensuring a dispersed distribution of site locations. Stratification can begin at an ecoregion site classification level and proceed to more specific levels of resolution as necessary to meet project objectives (Table 3-6). If there are clear clusters of differing land use among watersheds, the watersheds may be further stratified to ensure inclusion of an even distribution of land use types (e.g., subwatersheds having different levels of urban development). Waterbodies can be further stratified by section or segments. For instance, streams can be stratified by stream order (first, second, third, etc.), size of drainage area, or specific sections of a bay or lake.

(3) Select sampling sites. Sampling units or sites within each stratum (or other delineation as dictated by the general design) are randomly selected. This approach provides a basis for making

Table 3-6. Waterbody stratification hierarchy.

Population of Interest	Streams and Rivers[a]	Lakes [b, c]	Reservoirs [b, d]	Estuaries [e]	Wetlands [f]	Ground Water [g]
Operational Sampling Unit (SU)	Channel segment (i.e., a length of river channel into which no tributaries flow)	Self-contained basin	Self-contained basin (hydrologically isolated from other basins)	Self-contained basin	Transect upland or deep water boundaries	Transect
Strata (or higher stages) comprising Sus	Ecoregion	Ecoregion	Ecoregion	Biogeographic province	Wetland system type (marine, estuarine, riverine, lacustrine, palustrine)	Ecoregion
	Watershed	Size/ surface area of lake (km^2)	Size/surface area of reservoir (km^2) [watershed area/basin surface area]	Watershed	Watershed recharge, discharge or both	Watershed
	Stream/river channel (ordinal/ areal)	Lake hydrology (retention time, thermal stratification)	Hydrology (water level fluctuation/ drawdown; retention time stratification)	Watershed area (km^2)	Class (based on vegetative type; substrate and flooding regime; hydroperiod)	Aquifer type (unconfined (water table) aquifers, confined (artesian well) aquifers)
	Segment	Characteristic water quality (natural conditions)	Characteristic water quality (natural conditions)	Zones (tidal basin, depth, salinity)	Flooding regime water chemistry soil type	Aquifer material type (sedimentary rock; marl; unconsolidated rock, etc.)
	Characteristic water quality (natural conditions)					Hydraulic conductivity (homogeneous, heterogeneous, isotropic, anisotropic)

Table 3-6. (continued)

Population of Interest	Streams and Rivers[a]	Lakes [b, c]	Reservoirs [b, d]	Estuaries [e]	Wetlands [f]	Ground Water [g]
Habitats within Sus	Macrohabitat (pool/riffle; shorezone vegetation; submerged aquatic macrophytes)		Longitudinal zone (riverine; transitional; lacustrine; tailwaters (can be more riverine but always associated with dams))			Depth zone
		Depth zone (eulittoral → profundal)	Depth	Substrate/habitat	Subsystem (subtidal, intertidal, tidal, lower perennial, upper perennial, intermittent, littoral, limnetic)	Temperature
	Microhabitat	Substrate/ microhabitat	Substrate/ microhabitat		Substrate/ microhabitat	

[a] Frissell et al., 1986; [b] Gerritsen et al., 1996 [Draft]; [c] Wetzel, 1975; [d] Thornton et al., 1990; [e] Day et al., 1989; [f] Cowardin et al., 1979; [g] Heath, 1989.

general statements about the condition of the entire stratum, including sites not sampled.

3.8.2 Targeted Site Selection

In a targeted site selection design, sites are selected based on the location of known or suspected perturbations (stressors), planned point source controls, or BMPs in the geographic area of concern. NPDES permits, urban stormwater sites, timber harvest areas, rangeland, row crop farming, construction sites, and Superfund sites are all examples of known stressor sites. Upstream/downstream sampling stations, before-and-after site alterations, or recovery zones (sampling at established distances from sources) are types of sampling locations for known stressor sites. Ecologically sensitive sites that may or may not be affected by stressors might also be chosen as targeted sites.

3.8.3 Integrated Network Design

Integrated network design consists of integrating multiple monitoring subprograms to effectively meet various monitoring objectives and improve the applicability of data. An emerging biological monitoring program in Prince George's County, Maryland, is used here as a useful example demonstrating the design components reviewed above (see box on next page).

Prince George's County is interested in assessing the status and trends of biological stream resources in the county with known confidence. Assessment levels include county-wide, watershed-wide, and stream-specific. Biological assessment, based on sampling benthic macroinvertebrates, is used as the indicator of stream condition. Judgment of impairment or nonimpairment results from comparison to reference conditions (see Section 3.4). Two major components of this program are volunteer monitoring of selected sites and non-volunteer monitoring of probability-based and targeted sites. The volunteer sites serve for public education and some initial screening of stream conditions. Those sites monitored by the non-volunteer program are intended to provide unbiased estimates of biological status of streams throughout the county, trends in their condition, problem identification, evaluation of management activities (e.g., restoration, BMP installation, chemical controls, and altered land use practices), and data for eventual establishment of cause-and-effect associations.

Three types of sites are monitored in this program to address the county's multiple goals: targeted sites, reference sites, and probability sites. Each site type addresses specific questions on stream status in Prince George's County. Targeted sites are sampled semiannually during two program index periods—spring and fall. Annual sampling allows estimation of intra- and interannual variation in the measured variables and indices, and it allows estimation of trends after several years of monitoring. After the first 2 or 3 years of monitoring, the data will be reevaluated to determine whether sampling at longer intervals (for example, every other year) is sufficient to detect trends. Data from targeted sites can document the decline or recovery of streams subject to specific stresses and will allow assessment of restoration and mitigation efforts.

A set of 15 to 20 reference sites (having over 50 percent forest cover) are monitored semiannually as are the targeted sites. These data allow estimation of annual variation and trends in the biological characteristics of the reference sites. These estimates are critical for determining the biological status of test sites.

Forty-one watersheds have been identified to address status and trends for the county. Site

Case Study

Development of an Integrated Biological Monitoring and Assessment Program

in Prince George's County, Maryland

Prince George's County, Maryland, immediately east of Washington, DC, is developing a biological monitoring program to assess the status and trends in ecological condition and the physical habitat quality in county streams. Program goals include:

- Document and monitor the biological status and trends of county streams.
- Integrate data from biological, chemical, and physical monitoring programs to make a comprehensive assessment of the county's stream resources.
- Use biological monitoring data to identify and characterize impairment to the ecological system.
- Further public education in environmental problems through a component of the biological monitoring program tailored for layperson involvement.
- Evaluate the effectiveness of environmental management and mitigation activities.

The program design includes both probabilistic and targeted elements that will allow specific questions to be addressed at three spatial scales—county-wide, watershed-wide, and stream-specific. Approximately 100 sampling locations have been selected for the initial year of monitoring (1995). In two separate sampling periods (early spring and fall), approximately 50 probability sites, 20 known problem sites, 20 reference sites, and another 15 sites will be used for either confirmation of volunteer monitoring results or quality control samples.

The sampling units are stream segments between confluences. Segments are sampled at accessible points. Sampling is two-stage: for the first stage, watersheds are selected randomly so that one-fifth of the watersheds in the county are sampled in any one year and all watersheds will be sampled in the fifth year of the program. The second stage is random selection of stream segments within watersheds, stratified by stream order, so that sampling effort is allocated optimally among three stream orders.

Used by permission of Watershed Protection Branch, Department of Environmental Resources, Prince George's County, Maryland.

selection is in two stages. In the first stage, a set of watersheds are selected randomly. In the second stage, stream segments within the selected watersheds are chosen at random for sampling. In each year of the monitoring program, a set of 6 to 10 watersheds are selected (depending on size) and approximately 45 stream segments of those watersheds are sampled. After 5 years, all watersheds in the county will have been sampled, and in the sixth year the program will return to the originally sampled watersheds. An estimate of status can be made for each watershed every 5 years. The status of streams county-wide can be estimated from the first year on. The first year's estimate will be based on a small sample of watersheds and segments and will have greater uncertainty than the estimates developed in later years.

Prior knowledge of land use in the county was applied to stratify watersheds and sites. Urbanization is known to affect stream hydrology, water quality, habitat, and biota, and the northern watersheds of the county are urban and suburban, being close to Washington, DC. Therefore, Prince George's County was divided (stratified) into northern and southern watersheds so that in any given year, an equal number of watersheds would be selected in the more urban north and in the more rural southern parts.

3.9 Monitoring Trends in Biological Condition

Two separate factors affect our ability to distinguish trends over time in biological monitoring. The first factor, common to any trend analysis, are those sources of variability inherent in the measurements obtained from the monitoring sites. A significant challenge is distinguishing random changes in biological monitoring data over time from actual trends. Section 4.4 outlines statistical considerations regarding this issue, and Figures 4-4 to 4-7 illustrate several common types of data patterns that might be obtained over time in a water quality monitoring program. One of the solutions to this challenge is to have either a long time series of data or a sufficient number of monitoring sites (probabilistic, targeted, or integrated design) so that the data can be statistically evaluated.

A second factor affecting the ability to detect trends in a biomonitoring program is the observed relationship between reference site data and monitoring site data over time. This factor is somewhat unique to biomonitoring programs because a biological assessment at any given time is dependent on data from two sources—the population of reference sites and the monitoring sites. Thus, an accurate interpretation of trends in a biological monitoring nonpoint source program is accomplished, in part, by using a comparative analysis of reference site and monitoring site data. Figure 3-7 illustrates some possible trends that could be observed in a biological monitoring program. These types of data can be analyzed in a variety of ways as described in Section 4.4 and in Green (1993) and Smith et al. (1988).

Accurate characterization of trends in biological data over time depends on the degree to which sources of variability in the data are defined and characterized. In general, there are two major sources of variability—natural and experimental. Natural sources of variation include seasonal effects such as species life cycles, natural disturbances such as floods and fires, and microhabitat differences among sites or over time. Experimental sources of variability include collection method and gear variation. The previous discussion of biological assessment methods in this chapter addresses many of the common natural and experimental sources of variability. All field collection methods should be documented and standardized in approved SOPs. The use of a well-defined and appropriate field collection method, for example, reduces experimental sources of variability. Incorporation of a population of ecoregional reference sites into a biological monitoring design serves as a control for natural sources of variability over time, in addition to providing a sound data assessment framework. An apparently downward trend for certain sites in watersheds over time, for example, may be open to interpretation if the same trend is observed in reference sites (Figure 3-7). Repeated sampling during the same index period over time also provides some control for natural sources of variability.

Although natural sources of variability are not often amenable to characterization, experimental

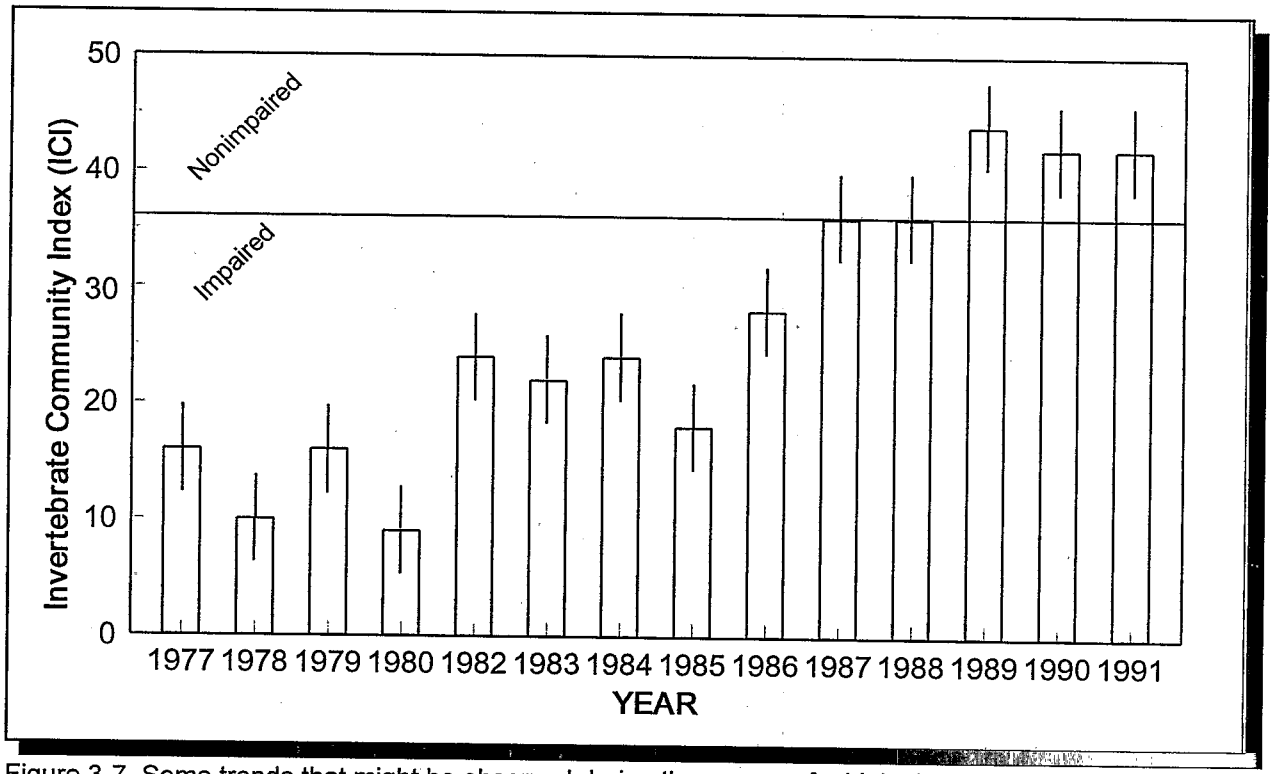


Figure 3-7. Some trends that might be observed during the course of a biological monitoring program.

or method-sources of variation can and should be defined. Factors such as precision and sensitivity of the assessment scores are achieved by collecting and analyzing multiple samples from the same site using consistent procedures and by sampling and analyzing samples from multiple sites that are similar, particularly ecoregional reference sites. One of the results of this method characterization process is the ability to define the statistical power of a sampling and bioassessment method. Statistical power is the degree to which a Type 2 error has a given probability of occurrence. Put simply, this is the probability of assessing a site as unimpaired when in fact it is impaired; or concluding that there is no trend or change in water quality over time when in fact there has been a change.

Defining the statistical power of a given bioassessment method allows one to rigorously determine the number of sites necessary in a monitoring program to detect a given level of impairment, or a certain trend over time, with a known degree of confidence. A power analysis achieves this objective by constructing an empirical relationship between the number of sites or measurements observed and the resultant difference in assessment score detected between the reference and test sites. A bioassessment method that has substantial intra-site variability will have reduced statistical power resulting in a greater number of sampling sites to distinguish a given level of impairment. Similarly, a very heterogeneous population of ecoregional reference sites, with widely ranging bioassessment scores, will also yield reduced statistical power.

An example of power analysis is shown for the Delaware coastal plain streams in which the benthic macroinvertebrate community was the biomonitoring indicator used to detect status and trends of water quality (Maxted and Dickey 1993; Figure 3-8). The power analysis graph shows a steep decline in the number of monitoring sites necessary as percent difference in assessment score (between reference and monitoring sites) decreases. In other words, if one is interested in detecting a relatively small difference in assessment score between the reference and monitoring sites or if, alternatively, one wants to detect a relatively small change in assessment score over time with a high degree of confidence, a greater number of monitoring sites within a

Determining the Appropriate Sampling Effort: An Example Using DNREC Data from the Mid-Atlantic Coastal Plain Ecoregion

Most environmental monitoring programs are designed to avoid detecting a problem or noncompliance that does not really exist. These false positives are called Type I errors. It is traditional to accept a 5 percent probability of false positives as occurring (see Section 4.1 for further discussion). The probability of false negatives (Type II error, failing to detect a problem that does exist) is evaluated through power analysis. A power analysis provides an estimate of the number of measurements (sample size) required to detect a change for a given significance level (usually, $\propto = 0.05$), with the power typically set at 80 percent. Power analysis requires prior knowledge of, or rational assumptions on, the statistical properties of the data, in particular, the nature of the variability associated with them. Two types of variance are used in power analysis: variance of total bioassessment scores among all reference sites and variance among total bioassessment scores within a site.

Maxted and Dickey (1993), as part of their nonpoint source biological monitoring program, produced replicated bioassessment data from 23 sites. Total bioassessment scores were developed for each using an RBP - type approach for normalization of six metrics: (1) taxa richness, (2) EPT index, (3) percent EPT abundance, (4) percent Chironomidae abundance, (5) percent dominant family, and (6) family biotic index.

The *sampling error* takes into account the natural variability among multiple sites; *measurement error* is the variability observed among multiple samples at the same site, that is, it tells us how well the site is being characterized. The power analysis estimates the number of samples required for a given percent difference in total bioassessment score. Using the sampling error, it estimated that in order to detect a 20 percent difference in the index value (between a population of reference sites and a population of test sites) a minimum of eight reference and eight test sites (subjected to NPS pollution) should be sampled (solid line in accompanying figure). Power analysis using the measurement error indicates that three repeated measurements at each site are necessary to detect a 20 percent difference in total bioassessment score (broken line in the accompanying figure).

Figure 3-8. Sample power analysis of a bioassessment method.

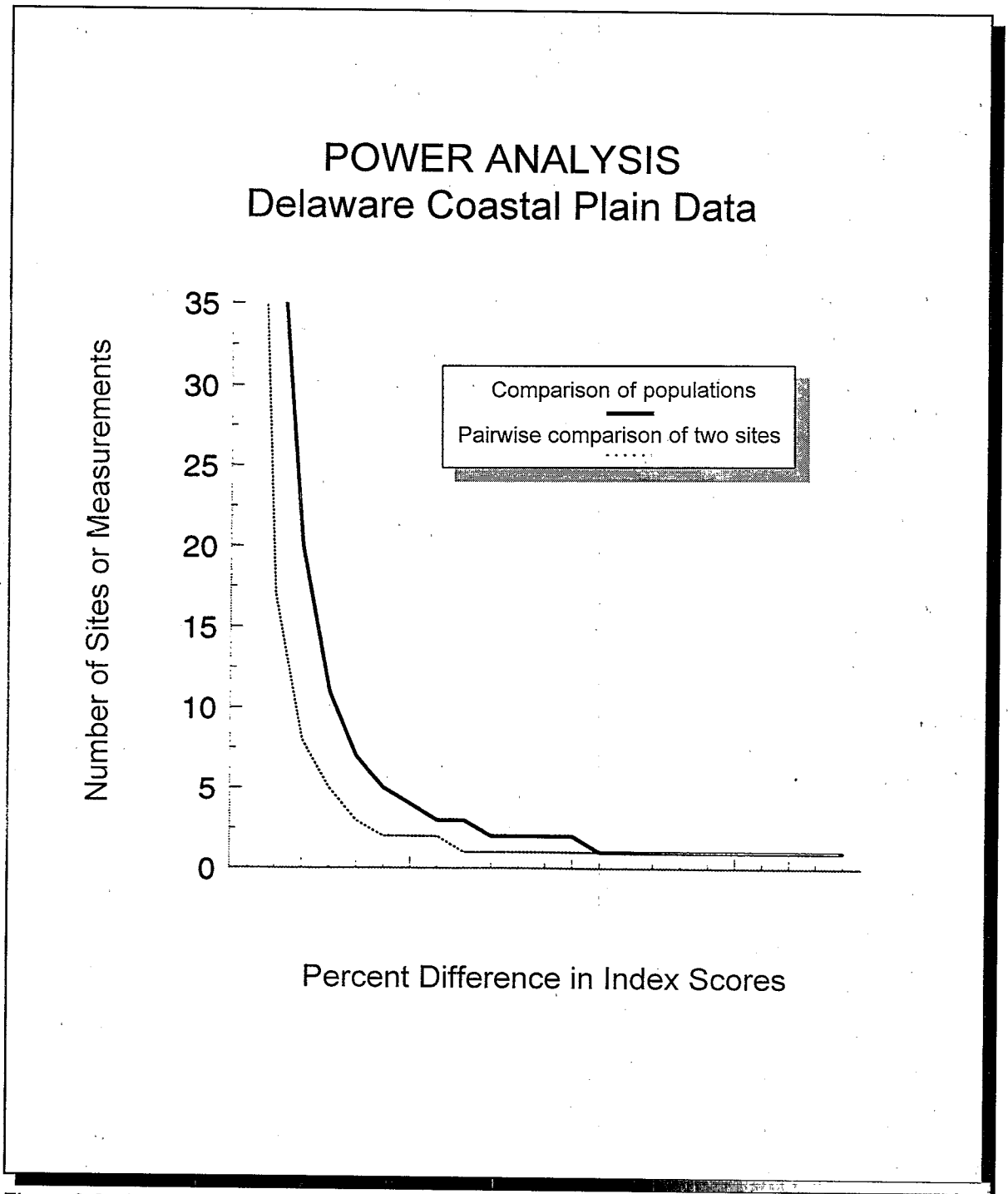


Figure 3-8. (continued)

stratum (i.e., stream order, urban versus agricultural) will need to be sampled.

The Prince George's County case study reviewed earlier illustrates how power analysis was used to determine the number of monitoring sites needed to obtain meaningful biomonitoring data. A design consideration was that impairment, as defined by a 50 percent reduction of the reference condition, be detected with 80 percent probability of detection, and 95 percent confidence that observed differences are significant. Power analysis (Fiugre 3-8) revealed that a sample of two sites on a stream class could be used to detect a difference of 50 percent using bioassessment procedures in coastal plain streams. Assessment of a single watershed can be done by sampling at least two sites of each stream order present in the watershed. Most of the 41 watersheds in this county are third-order coastal plain streams. Therefore, an average of 6 sites per watershed, or approximately 246 sites, should be sampled for an assessment of all watersheds. This represents approximately 25 percent of the total population of stream segments in the county.

The above considerations led to the target sampling design of 25 percent of stream segments. Because county-wide assessment is one of the goals of this program, selection probability was kept at 25 percent. The sampling rule used in this program is at least 25 percent of stream segments in each stream order are selected in each watershed. For example, one segment is selected if there are one to four segments of a given order in the watershed, two segments are selected if there are five to eight, and so on. The probability that a watershed will be included in sampling thus varies slightly among watersheds, and this probability is used as a weighting factor in county-wide assessment and estimation. Given an annual sampling effort of approximately 50 probability sites per year, a 5-year rotation of sampling will allow assessment of all watersheds over the 5-year period.

Repeated sampling of monitoring and reference sites over time, along with adequate characterization of bioassessment method precision, can yield significant management information as demonstrated by the Ohio EPA bioassessment program. Since 1977, Ohio EPA (1992) has used assessments of the benthic macroinvertebrate assemblage to document changes in the biological condition of a waterbody. Figure 3-7 illustrates annual results of biological assessments using the Invertebrate Community Index (ICI) (Ohio EPA, 1992) over a 14-year period in the Cuyahoga River. A 95 percent confidence interval of ± 4 ICI units was determined based on analysis of intra- and inter-site variance in ecoregional reference scores. The graph demonstrates several features, including a general improvement in water quality over time. Furthermore, certain apparent changes in score between 1977 and 1980 are in fact illusory since the confidence intervals overlap. Thus, actual trends can be statistically differentiated from random changes in the data over time.

3.10 Overview of Some State Programs

Biological monitoring programs of five states are highlighted to illustrate technical components of biological monitoring programs, which are summarized for nonpoint source evaluations in Table 3-7. The Delaware Department of Natural Resources and Environmental Conservation (DNREC) led the Mid-Atlantic Coastal Streams (MACS) workgroup in adapting the RBPs to low-gradient streams. The MACS workgroup consists of technical staff from biomonitoring programs of Delaware, Maryland, Virginia, and North Carolina (USEPA, 1997). They were able to determine that the most appropriate method for obtaining a representative

Table 3-7. Summary of the primary technical issues related to biological monitoring for nonpoint source evaluations.

Biomonitoring for Nonpoint-source Evaluations	
Sampling Area/ Size	Sampling is conducted over a stream reach, such that a composite of different habitats or different parts of a habitat is created. Natural biological heterogeneity is accounted for by sampling a relatively large area. Fish sampling may extend from 100 yards to 20 or 30 times the stream width. Benthic sampling may extend over several riffles or a composite of habitats taken from a 100-yard length of stream.
Replication	Replication at a site is important to evaluate within-site variability. However, several samples taken at a single site would be pseudoreplication if used to evaluate effects of impacts, such as nonpoint source pollution. Sites from different streams and watersheds are considered replicates to assess status or condition of regions or watershed basins. Monitoring can be accomplished with single large samples that sufficiently represent the stream reach under study.
Sampling Gear	Gear can be as rigorous and quantitative as a program deems necessary. However, gear must efficiently sample the targeted assemblage and specified habitat and be maintained in good operational condition. Electrofishing is the preferred gear for fish assemblages. Most state agencies have selected a kick seine or D-frame, and artificial substrates are used for periphyton.
Biological Index Period	The investigator must find a compromise between selecting a sampling period that is representative of the biological community and selecting one that reflects the worst-case conditions of pollutant stress. Seasonality is an important consideration because the taxonomic and functional feeding group compositions change naturally throughout the year in response to emergence and reproductive cycles. The optimal biological index period will vary throughout the United States. Some states, such as Florida, have more than one index period for sampling.

sample of benthic macroinvertebrates from low-gradient, sandy-bottomed, coastal plain streams was a multihabitat approach. The approach uses a standard D-frame net and samples the most dominant habitat types from a 100-meter reach in proportion to their frequency. The Delaware DNREC samples approximately 300 sites on an annual basis using this approach (Table 3-8). The Florida Department of Environmental Protection has adopted this method for its nonpoint source monitoring program (Table 3-9) and has established reference conditions. The State of Montana, Department of Health and Environmental Sciences, uses a different approach for its higher-gradient streams, which are generally cobble-bottomed (Table 3-10). It has adapted a traveling kick method for macroinvertebrates and is developing reference conditions. The health of

Table 3-8. Selected biomonitoring program components, Delaware Department of Natural Resources and Environmental Conservation (DNREC).

	Macroinvertebrates
Habitat Selection	Stable habitats that are 5% of the 100-meter assessment area. Habitats include snags, submerged macrophytes, banks/root systems, and riffles. Sampled in proportion to representation.
Sample Gear/ Preservation	Standard aquatic D-frame dip net (0.3-meter width, 700- to 900-micron mesh), sieve bucket (600-micron mesh), 70% ethanol, storage containers (1- to 2-liter).
Sampling Method	20-jab method; a single jab consists of thrusting the net into a productive habitat for a distance of approximately 1 meter. 20 jabs composited across habitats (6.2 m^2). Samples are cleaned by running stream water through the net, then transferring to a sieve bucket for further cleaning. Transferred to a storage container and preserved in 70% ethanol.
Subsampling and Enumeration	Preserved samples returned to the lab for processing. Subsampling to 100-organism level.
Taxonomic Level	Currently at family level. Investigating efficacy of doing genus-level identifications.
QA procedures	Same as RBPs (Plafkin et al., 1989); habitat assessments are evaluated by all investigators while reviewing the slides and field notes. Field investigators must have proper training.
Habitat Assessment	Field data sheet: 7 parameters numerically scored (1-20) similar to RBPs; documents other physical and water quality data.
Comments	Standard operating procedures in draft form and prepared by Mid-Atlantic Coastal Streams Workgroup.

the fish assemblage is the primary biological monitoring indicator for the North Dakota Department of Environmental Health (Table 3-11), which is planning to further develop its monitoring program to include benthic macroinvertebrates. The State of Vermont, Department of Environmental Conservation, does three types of sampling (Table 3-12)—fish, benthic macroinvertebrates from riffles in high-gradient streams, and benthic macroinvertebrates from multihabitats in low-gradient streams.

Table 3-9. Selected biomonitoring program components, Florida Department of Environmental Protection (Florida DEP).

	Macroinvertebrates
Habitat Selection	Presence or absence of major productive habitats at each sampling location is established during preliminary reconnaissance. Habitats include: riffles, snags, aquatic vegetation, leaf packs, undercut banks/root systems, leaf mat, rocky outcrops, muck/silt, sand. Major habitats sampled equally; Group of minor habitats treated as a single major habitat.
Sample Gear/ Preservative	Standard D frame dip net (0.3 meter width 600 micron mesh), wide mouth jars, formalin.
Sampling Method	20 jab dip net sample, composite sample across habitats. Individual jabs are approximately 0.5 m making a total composite of 3 m^2.
Subsampling and Enumeration	Entire sample in gridded pan (5 X 3 cm grids), randomly select grids (1/72 of sample), remove contents, sort into taxonomic groups, continue until a minimum of 100 organism are counted; a grid's entire contents must be sorted.
Taxonomic Level	Lowest taxonomic level (genus or species).
QA Procedures	Replicate sampling for 10% of samples collected on an annual basis have not been implemented but are planned. Resorting of 10% of samples. Field crew undergo periodic training.
Habitat Assessment	Field data sheet, 7 visually-estimated habitat parameters, weighted equally. Physical/Chemical characterization field data sheet
Comments	Stream classification factors for establishing reference conditions based on Ecoregion and subecoregion. Site selection factors: Availability of least impaired and reference sites, specific monitoring issues, accessibility and safety, compatibility of habitat. Standard operating procedures and report describing state-wide nonpoint source program prepared by Florida DEP and support contractors.

Table 3-10. Selected biomonitoring program components, Montana Department of Health and Environmental Sciences (Montana DHES).

Macroinvertebrates

Region of State	Entire
Site Selection	Riffles
Sample Gear	12-inch D-frame net
Sampling Method	Travelling kick across riffles, 1-2 diagonal collections (time and distance recorded)
Subsampling and Enumeration	300-organism subsample in laboratory
Taxonomic Level	Genus/species
QA Procedures	Replication on selected sites and projects
Habitat Assessment	Follows RBP habitat assessment approach
Comments	Lab processing done by outside contractors

Table 3-11. Selected biomonitoring program components, North Dakota Department of Environmental Health (North Dakota DEH).

	Macroinvertebrates	*Fish*
Region of State	Not developed as of current date	Red River
Site Selection		25 sites to date; randomized selection, but dictated by access logistics
Sample Gear		Electrofishing; shore-based longline
Sampling Method		Minimum of 100 meters; 20 times river width on wide rivers
Subsampling and Enumeration		Not applicable
Taxonomic Level		Lowest positive taxon
QA Procedures		10 percent repeat sampling
Habitat Assessment		Based on RBPs
Comments	Planned to be started in 1997	In development

Table 3-12. Selected biomonitoring program components, Vermont Department of Environmental Conservation (Vermont DEC).

	Macroinvertebrates		*Fish*
Region of State	Entire state; high gradient	Entire state; low gradient	Entire state
Site Selection	Riffles	Woody debris Macrophytes Boulders	Representative of reach; mix of riffle, run, pool
Sample Gear	Rectangular kick net 18 in x 8 in 500-μm mesh	Rectangular kick net 18 in x 8 in 500-μm mesh	Electro-shocking
Sampling Method	2-min timed composite; 30 sec in each of 2 points in slow area; 2 points in fast area	2-min sample; all substrate materials are hand scrubbed	1-3 upstream passes into blocknet or vertical drop instream
Sub-sampling and Enumeration	Preserve in field; subsample gridded tray 1/4 sample and min 300 animals	Preserve in field subsample gridded tray (technique allows density estimate for site)	Total count
Taxonomic Level	Specified in protocols; family-genus-species, depending on order	Protocols list taxonomic level and key to be used	Species
QA Procedures	All samples archived; replicate samples always collected; two people check pick; two people trained in each taxonomic order	All samples archived; replicate samples always collected; two people check pick; two people trained in each taxonomic order	Same person conducts collection and taxonomy

Table 3-12. (continued)

	Macroinvertebrates		*Fish*
Habitat Assessment	Includes substrate composition, embeddedness, periphyton cover type, canopy cover, immediate riparian zone info ave. stream velocity, depth	Includes substrate composition, embeddedness, periphyton cover type, canopy cover, immediate riparian zone information, average stream velocity, depth	Qualitative or quantitative transect method; depth, velocity, substrate
Comments	Also record temperature, pH, alkalinity, condition and characterization of site, elevation, DO; also will look at stream low flow characteristics and surficial and bedrock geology and land use cover to help differentiate sites as "ecotype"; set biocriteria for individual metrics		Has modified the Index of Biotic Integrity to fit Vermont's wadable streams

Table 3-13. Fish Index of Biotic Integrity metrics used in various regions of North America.

Alternative IBI metrics	Metric Category	Midwest	New England	Ontario	Central Appalachia	Colorado Front Range	Western Oregon	Sacramento-San Joaquin	Wisconsin
1. Total number of species	A	X	X		X	X		X	
# Native fish species	A	X		X			X		X
# Salmonid age classes[a]	A					X	X		
2. Number of darter species	B	X			X	X			
# Sculpin species	B						X		
# Benthic insectivore species	B								
# Darter and sculpin species	B	X	X						
# Salmonid yearlings (individuals)[a]	B		X				X	X	
% Round-bodied suckers	B	X							
# Sculpins (individuals)	B							X	
3. Number of sunfish species	B	X				X			X
# Cyprinid species	B						X		
# Water column species	B		X						
# Sunfish and trout species	B			X					
# Salmonid species	B							X	
# Headwater species	B	X							
4. Number of sucker species	B	X	X				X		X
# Adult trout species[a]	B						X	X	
# Minnow species	B	X				X			
# Sucker and catfish species	B			X					
5. Number of intolerant species	B	X	X			X	X		X
# Sensitive species	B	X							
Presence of brook trout	B			X					

Table 3-13. (continued)

Alternative IBI metrics		Metric Category	Midwest	New England	Ontario	Central Appalachia	Colorado Front Range	Western Oregon	Sacramento-San Joaquin	Wisconsin
6.	% Green sunfish	B	X							
	% Common carp	B						X		
	% White sucker	B		X			X			
	% Tolerant species	B	X							X
	% Creek chub	B				X				
	% Dace species	B			X					
7.	% Omnivores	D	X	X	X	X	X			X
	% Yearling salmonids[a]	D					X	X		
8.	% Insectivorous cyprinids	D	X							
	% Insectivores	D		X				X		X
	% Specialized insectivores	D				X	X			
	# Juvenile trout	D							X	
	% Insectivorous species	D	X							
9.	% Top Carnivores	D	X	X	X					X
	% Catchable salmonids	D						X		
	% Catchable trout	D							X	
	% Pioneering species	D	X							
	Density catch-able wild trout	D							X	

Table 3-13. (continued)

Alternative IBI metrics	Metric Category	Midwest	New England	Ontario	Central Appalachia	Colorado Front Range	Western Oregon	Sacramento-San Joaquin	Wisconsin
10. Number of individuals	A	X		X	X	X	X	X	X
Density of individuals	A		X						
11. % Hybrids	C	X	X						
% Introduced species	C					X	X		
% Simple lithophills	C	X							X
# Simple lithophills species	C	X							
% Native species	C							X	
% Native wild individuals	C							X	
12. % Diseased individuals	C	X	X	X	X	X	X		X

Note: X = metric used in region. Many of these variations are applicable elsewhere.

Metric Categories: A = Community structure, B = Taxonomic composition, C = Individual condition, D = Biological processes

[a]Metric suggested by Moyle or Hughes as a provisional replacement metric in small western salmonid streams.

Taken from Karr et al. (1986), Hughes and Gammon (1987), Ohio EPA (1987), Miller et al. (1988), Steedman (1988), and Lyons (1992).

Table 3-14. Examples of metric suites used for analysis of macroinvertebrate assemblages.

	Alternative Benthic Metrics	Metric Category	ICI (a)	RBP (b)	RBP (c)	RBP (d) ID	RBP (d) OR	RBP (d) WA	B-IBI (e)
1.	Total No. Taxa	A	X	X	X	X	X	X	X
	% Change in Total Taxa Richness	A				X	X	X	
2.	No. EPT Taxa	B	X	X		X	X	X	
	No. Mayfly Taxa	B	X						X
	No. Caddisfly Taxa	B	X						X
	No. Stonefly Taxa	B							X
	Missing Taxa (EPT)	B			X				
3.	No. Diptera Taxa	B	X						
	No. of Chironomidae	B				X		X	
4.	No. Intolerant Snail and Mussel Species	B							X
5.	Ratio EPT/Chironomidae Abundance	B				X	X	X	
	Indicator Assemlage Index	B			X	X	X		
	% EPT Taxa	B				X			
	% Mayfly Composition	B	X						
	% Caddisfly Composition	B	X						
6.	% Tribe Tanytarsini	B	X						
7.	% Other Diptera and Noninsect Composition	B	X						
8.	% Tolerant Organisms	B	X						
	% Corbicula Composition	B							X
	% Oligochaete Composition	B							X
	Ratio Hydropsychidae/ Tricoptera	B		X			X		
9.	% Ind. Dominant Taxon	A		X		X	X	X	
	% Ind. Two Dominant Taxa	A							X
	Five Dominant Taxa in Common	A		X	X		X		
	Common Taxa Index	A			X				

Table 3-14. (continued)

	Alternative Benthic Metrics	Metric Category	ICI (a)	RBP (b)	RBP (c)	RBP (d) ID	RBP (d) OR	RBP (d) WA	B-IBI (e)
10.	Indicator Groups	B				X		X	
11.	% Ind. Omnivores and Scavengers	D							X
12.	% Ind. Collector Gatherers and Filterers	D							X
	% Ind. Filterers	D				X		X	
13.	% Ind. Grazers and Scrapers	D				X			X
	Ratio Scrapers/Filterer Collectors	D				X	X	X	
	Ratio Scrapers/(Scrapers + Filterer Collectors)	D		X					
14.	% Ind. Strict Predators	D							X
15.	Ratio Shredders/Total Ind. (% shredders)	D		X		X		X	
16.	% Similarity Functional Feeding Groups (QSI)	D		X	X				
17.	Total Abundance	A				X			
18.	Pinkham-Pearson Community Similarity Index	A		X					
	Community Loss Index	A					X	X	
	Jaccard Similarity Index	A				X			
19.	Quantitative Similarity Index (Taxa)	A		X	X				
20.	Hilsenhoff Biotic Index	B		X		X	X	X	
	Chandler Biotic Score	B				X			
21.	Shannon-Weiner Diversity Index	A					X		
	Equitability	A				X			
	Index of Community Integrity	A				X			

Metric Categories: A = Community structure, B = Taxonomic composition, C = Individual condition, D = Biological processes
a: Invertebrate Community Index, Ohio EPA (1987).
b: Rapid Bioassessment Protocols, Barbour et al. (1992) revised from Plafkin et al. (1989).
c: Rapid Bioassessment Protocols, Shackelford (1988).
d: Rapid Bioassessment Protocols, Hayslip (1993); ID = Idaho, OR = Oregon, WA = Washington.
(Note: these metrics in ID, OR, and WA are currently under evaluation)
e: Benthic Index of Biotic Integrity, Kerans et al. (1992).

4. DATA ANALYSIS

Data analysis begins in the monitoring program design phase. Those responsible for monitoring should identify the goals and objectives for monitoring and the methods to be used for analyzing the collected data. Monitoring objectives should be specific statements of measurable results to be achieved within a stated time period (Ponce, 1980b). Chapter 2 provides an overview of commonly encountered monitoring objectives. Once goals and objectives have been clearly established, data analysis approaches can be explored.

Typical data analysis procedures usually begin with screening and graphical methods, followed by evaluating statistical assumptions, computing summary statistics, and comparing groups of data. The analyst should take care in addressing the issues identified in the information expectations report (Section 2.2). By selecting and applying suitable methods, the data analyst responsible for evaluating the data can prevent the "data rich–information poor syndrome" (Ward 1996; Ward et al., 1986).

This chapter provides detailed information on the statistical analysis of environmental monitoring data. The first section of the chapter is intended for both the manager and data analyst. Its goal is to acquaint the reader with key concepts and issues related to data analysis. This section also provides recommendations for selecting statistical procedures for routine analyses and can be used to guide the reader in selecting additional portions of the chapter for more in-depth reading.

4.1 INTRODUCTION

4.1.1 Estimation and Hypothesis Testing

Instead of presenting every observation collected, the data analyst usually summarizes major characteristics with a few descriptive statistics. Descriptive statistics include any characteristic designed to summarize an important feature of a data set or sample (Freund, 1973). The reader should note that a sample in this context refers to a group of observations selected from the target population. In the case of water quality monitoring, descriptive statistics of samples are used almost invariably to formulate conclusions or statistical inferences regarding populations (MacDonald et al., 1991; Mendenhall, 1971; Remington and Schork, 1970; Sokal and Rohlf, 1981). A *point estimate* is a single number representing the descriptive statistic that is computed from the sample or group of observations (Freund, 1973). For example, the mean total suspended solids concentration during baseflow is 35 mg/L. Point estimates such as the mean (as in this example), median, mode, or geometric mean from a sample describe the central tendency or location of the sample. The standard deviation and interquartile range could likewise be used as point estimates of spread or variability.

The use of point estimates is warranted in some cases, but in nonpoint source analyses point estimates of central tendency should be coupled with an *interval estimate* because of the large spatial and temporal variability of nonpoint source pollution (Freund, 1973). For example, the sample mean and standard deviation could be used to report that the mean total suspended solids concentration during baseflow is 35 $\pm$ 10 mg/L using a 95 percent confidence interval. Stated in other words, there is a 95 percent chance that the actual mean baseflow concentration is between 25 and 45 mg/L. There is a 5 percent chance that the mean baseflow concentration is outside this range. The confidence interval is a function of the variability of the data, the number of observations, and the probability (e.g., 95 percent) selected by the data analyst. This sort of estimation can be useful in developing baseline information, developing or verifying models, or

determining the load of a single nonpoint source runoff event.

Evaluating the effectiveness of controls and changing environmental conditions is one of the key monitoring program objectives described in Chapter 2. In addition to summarizing key statistics that describe the central tendency and spread of water quality variables and biological metrics, statistical analysis usually involves hypothesis testing. Two common types of hypothesis testing done in environmental monitoring are step changes and monotonic trends. Step changes are typically evaluated when comparing at least two different sample populations such as an impacted site and a reference site or when comparing one sample population to an action level. Step changes can also be evaluated when comparing samples collected during different time periods. Monotonic trends (e.g., consistently increasing or decreasing concentrations) are typically evaluated when the analyst is investigating long-term gradual changes over time.

The null hypothesis (H_o) is the root of hypothesis testing. Traditionally, null hypotheses are statements of no change, no effect, or no difference. For example, the flow-averaged mean total suspended solids concentration after BMP implementation is equal to the flow-averaged mean total suspended solids concentration before BMP implementation. The alternative hypothesis (H_a) is counter to the null hypothesis, traditionally being statements of change, effect, or difference. Upon rejecting H_o, H_a would be accepted. Regardless of the statistical test selected for analyzing the data, the analyst must select the significance level of the test. That is, the analyst must determine what error level is acceptable. There are two types of errors in hypothesis testing:

Type I: The null hypothesis (H_o) is rejected when H_o is really true.

Type II: The null hypothesis (H_o) is accepted when H_o is really false.

Table 4-1 depicts these errors, with the magnitude of Type I errors represented by α and the magnitude of Type II errors represented by β. The probability of making a Type I error is equal to the significance level (α) of the test and is selected by the data analyst. In most cases, managers or analysts define $1-\alpha$ to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99 percent), although there have been environmental applications where $1-\alpha$ has been set to 0.80. Selecting a 95 percent confidence level implies that the analyst will incorrectly reject the H_o (i.e., a false positive) 5 percent of the time.

Type II error depends on the significance level, sample size, and variability, and which alternative hypothesis is true. The power of a test ($1-\beta$) is defined as the probability of correctly rejecting H_o when H_o is false. In general, for a fixed sample size, α and β vary inversely. For a fixed value of α, β can be reduced by increasing the sample size

Table 4-1. Errors in hypothesis testing.

Decision	State of affairs in the population	
	H_o is True	H_o is False
Accept H_o	$1-\alpha$ (Confidence level)	β (Type II error)
Reject H_o	α (Significance level) (Type I error)	$1-\beta$ (Power)

(Remington and Schork, 1970). Figure 4-1 illustrates this relationship. Suppose this interest is in testing whether there is a significant difference between the means from two independent random samples. As the difference in the two sample means increases (as indicated on the x-axis), the probability of rejecting H_o, the power, increases. If the real difference between the two sample means is zero, the probability of rejecting H_o is equal to the significance level, α. Figure 4-1A shows the general relationship between α and β if α is changed. Figure 4-1B shows the relationship between α and β if the sample size is increased.

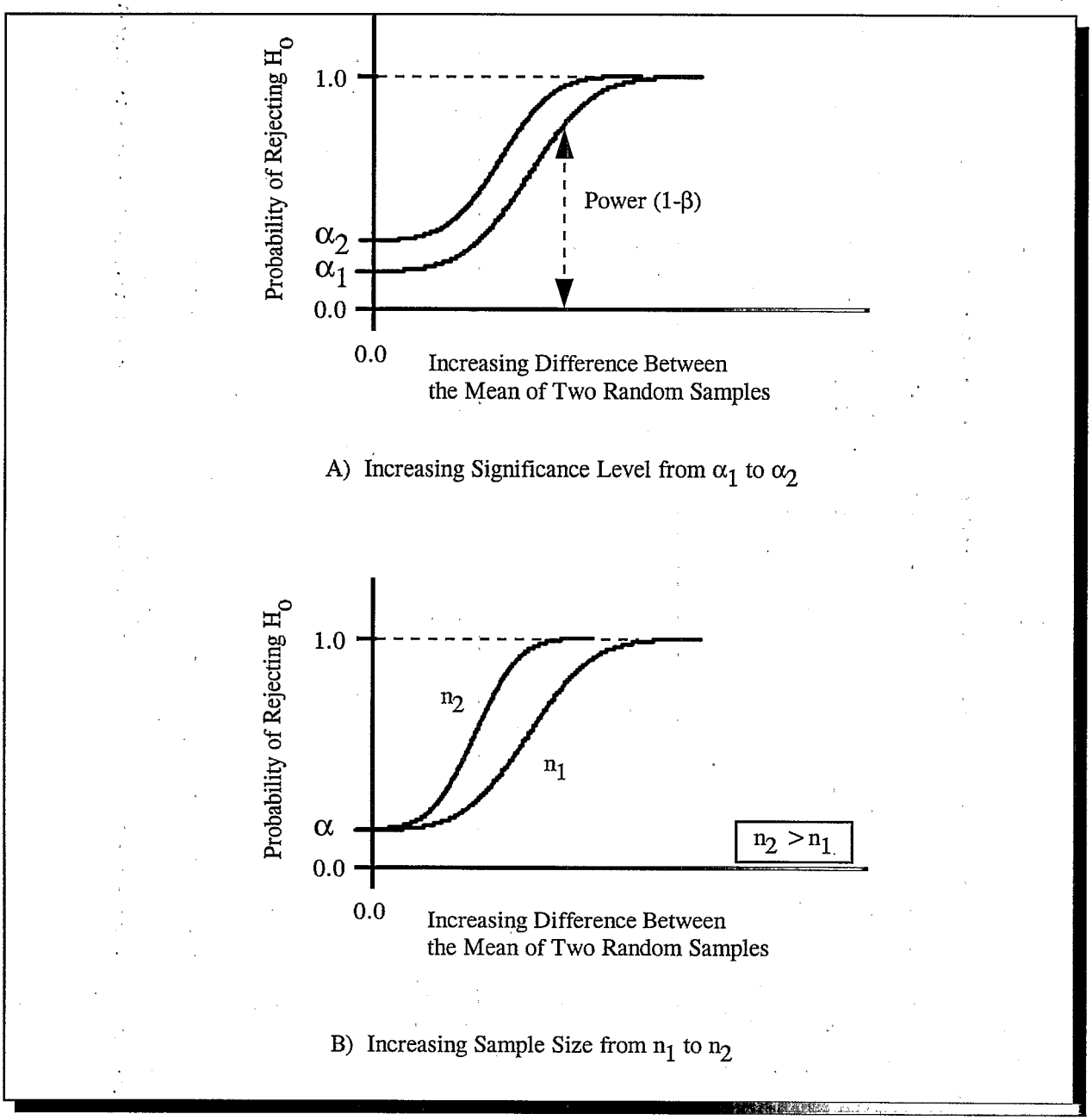


Figure 4-1. Comparison of α and β.

4.1.2 Characteristics of Environmental Data

The selected statistical method must match the type of environmental data collected and the decisions to be made. Although summarizing the mean annual dissolved oxygen concentration along an impaired stream might provide an indication of habitat quality, evaluating the minimum dissolved oxygen during summer months over the same time period might have a greater impact on subsequent management decisions since that is when critical conditions often occur. Environmental managers and data analysts must collectively determine which statistical methods will result in the most useful information for decision makers.

The selection of appropriate statistical methods must be based on the attributes of the data (Harcum, 1990). Two main types of attributes important to environmental monitoring are data record limitations and statistical characteristics. Common data record limitations include missing values, changing sampling frequencies over time, different numbers of samples during different sampling periods, measurement uncertainty, censored data (e.g., "less-thans"), small sample sizes, and outliers. Data limitations are, for the most part, human-induced attributes that often result in less reliable observations and less information for a given data set. The presence of data limitations also increases the complexity in applying standard statistical methods (and using commercially available software).

Common statistical characteristics include location (central tendency), variability (scale or spread), distribution shape, seasonality, and serial correlation. Table 4-2 presents a variety of methods for characterizing data that are helpful in providing a general understanding of water quality data and selecting appropriate statistical methods. Cross-references for each method are provided in the last column in Table 4-2.

4.1.3 Recommendations for Selecting Statistical Methods

The statistical methods discussed in this manual include parametric and nonparametric procedures. Parametric procedures assume that the data being analyzed have a specific distribution (usually normal), and they are appropriate when the underlying distribution is known (or is assumed with confidence). For data with an unknown distribution, nonparametric methods should be used since these methods do not require that the data have a defined distribution.

Nonparametric methods can directly handle special data commonly found in the nonpoint source area, such as censored data or outliers. Censored data are those observations without an exact numerical value, such as a value of less than 10 μg/L (<10 μg/L) or not-detected (ND). Censored data often appear in laboratory reports when the concentration being analyzed is lower than the detection limit or higher than the allowable range for a particular type of laboratory equipment or procedure (Dakins et al., 1996; Gilliom and Helsel, 1986). Censored data can cause problems in parametric methods because these methods often require that all data have numerical values. In this case, nonparametric methods can be used because they often deal with the ranking of the data, not the data themselves. For example, for data "below the detection limit," any value that is less than the smallest value of all the data being analyzed can be assigned. This assignment does not affect the ranking of the data even though the exact value of the "below the detection limit" is unknown. Nonparametric procedures are also less affected by outliers (Spooner, 1994a).

On the other hand, nonparametric procedures are not as powerful as their parametric counterparts when the assumptions of the parametric procedure are met. Thus, when the underlying distributions of the data being analyzed are known or can be

Table 4-2. Methods for characterizing data.

Data Characteristic	Method	Method Type	Section
Central tendency	Sample mean	P	4.2.1
	Sample median	N	4.2.1
	Sample geometric mean	P	4.2.1
	Boxplot	G	4.3
Spread	Sample standard deviation	P	4.2.2
	Interquartile range	N	4.2.2
	Sample geometric standard deviation	P	4.2.2
	Range, maximum-minimum	P,N	4.2.2
	Interquartile range	N	4.2.2
	Boxplot	G	4.3
Distribution shape	Histogram	G	4.3
	Percentiles	N	4.2.2
	Sample skewness	P	4.4.1
	Sample kurtosis	P	4.4.1
	Shapiro-Wilk test	N	4.4.1
Seasonal variation	Time series plots	G	4.3
	Seasonal boxplot	G	4.3
	ANOVA	P	4.6
	Kruskal-Wallis test	N	4.6.1
Serial correlation	Sample autocorrelation	P	4.9.2
	Spearman's rho	N	4.9.2
Key to Method Type: P = Parametric, N = Nonparametric, G = Graphical			

Adapted from Ward et al., 1990.

transformed to the form in which standard theory can be applied, parametric methods might be preferred. As a matter of fact, to improve the analytical power, nonparametric methods are often modified to include more assumptions and requirements. This makes the nonparametric methods more powerful, and the difference between nonparametric and parametric methods becomes smaller (Hipel, 1988). For example, the hypotheses associated with the Mann-Whitney test (for comparing two independent random samples) vary depending on which assumptions are valid.

The remainder of this section provides recommendations for selecting statistical methods that can be applied on a routine basis for evaluating the average, changing, and extreme conditions of environmental variables (Table 4-3, adapted from Ward et al., 1990). In some instances, more appropriate methods might be available depending on the specific information needs. For routine analyses, both parametric and nonparametric methods are recommended. Nonparametric procedures are recommended together with parametric procedures since nonparametric procedures tend to be resilient to characteristics commonly found in nonpoint

Table 4-3. Methods for routine data analysis.

Information Need	Method (P = parametric, N = nonparametric)		
	Graphical	**Estimation**	**Hypothesis Testing**
Average conditions	Boxplots (4.3) Time series plots (4.3)	**(P)** Sample mean (or geometric mean) and sample standard deviation (or geometric standard deviation) with confidence limits (4.2.1, 4.2.2) **(N)** Sample median and interquartile range with confidence limits (4.2.1, 4.2.2) **(N)** (Seasonal) Hodges-Lehman estimator (4.5.3)[a]	**Two random samples** **(P)** Student's *t* test (4.5.2) **(N)** Mann-Whitney test (4.5.2) **Matched samples** **(P)** Paired *t* test (4.5.1) **(N)** Wilcoxon signed rank test (4.5.1) **Three or more random samples** **(P)** ANOVA (4.6.1, 4.6.2) **(N)** Kruskal-Wallis test (4.6.1)
Changing conditions	Annual boxplots (4.3) Time series plots (4.3)	**(P)** Linear regression (4.7) **(N)** Sen (Seasonal Kendall) slope estimator (4.9.1)[a]	**(P)** *t* test for significance of slope (4.7) **(N)** Mann-Kendall (Seasonal Kendall) test (4.9.1)[a]
Extreme conditions	Time series plots with excursion limit	**(P,N)** Proportion (frequency) of Excursions (4.11.3)	**(P)** Test for equality of proportions (4.11.3) **(N)** Confidence limits on proportions (4.11.3) **(N)** Tolerance intervals (4.11.3)

Adapted from Ward et al., 1990.

[a] If seasonality is not present, the nonseasonal form of the test may be used.

source monitoring data (Berryman et al., 1988; Gilliom and Helsel, 1986; Harcum et al., 1992; Harris et al., 1987; Helsel and Hirsch, 1995; Hirsch et al., 1982; van Belle and Hughes, 1984; Lettenmaier, 1988). However, the data analyst must be aware that violating assumptions associated with parametric or nonparametric tests can lead to incorrect conclusions about the collected data.

Average conditions

What is the quality of water? What were the phosphorus loadings from the last storm? To answer these types of questions the data analyst is typically faced with describing the average conditions. Measures of central tendency and spread are the most common measures of average conditions. As suggested earlier, using the mean, geometric mean, or median is recommended for summarizing the central tendency and the standard deviation, geometric standard deviation, and interquartile range are recommended measures of spread or dispersion. Each parameter (mean, median, etc.) is a useful point estimate; however, no information on the parameter's accuracy is given. Therefore, it is also recommended that point estimates of central tendency be reported with confidence limits.

The selection of the mean (and standard deviation) versus the median (and interquartile range) should be based on the objective and type of data. The mean and standard deviation are sensitive to a few large observations. This is particularly true for the small sample sizes and skewed data that are common in nonpoint source monitoring. If the goal is to estimate pollutant loadings, an average concentration would be appropriate (Helsel and Hirsch, 1995). In general, parametric and nonparametric parameters are acceptable when the data are symmetrically distributed. Notwithstanding the pollutant loading example above, data that are not symmetrically distributed (skewed) should typically be summarized with the median and interquartile range. The geometric mean and standard deviation are most appropriate when the data typically range over a couple orders of magnitude. The presentation of geometric means is also called for in some regulations such as those for coliform bacteria. In many cases, simple graphical displays such as time series or box-and-whiskers plots will convey more information than tables of numerical results.

Changing conditions

One of the most frequently asked questions related to the evaluation of monitoring data is whether conditions have improved or degraded. The data collected for evaluating changes will typically come as (1) two or more sets of random samples or (2) a time series at a single station. In the first case, the analyst will test for a shift or step change. This would be typical for data collected from a nested paired and paired watershed design. Or when performing a biological assessment, for example, the goal might be to determine whether there is a significant difference (i.e., a step change) in the biological metric between the reference and test (targeted) sites.

The Mann-Whitney test is recommended for comparing two random samples when the distribution of the data is unknown or sufficiently nonnormal. The Student's *t* test can be used when the data are normally distributed. It has been demonstrated that the Student's *t* test can be successfully applied when the data are not normally distributed and might be more powerful under selected circumstances (Montgomery and Loftis, 1987), but that approach is not recommended here. The Kruskal-Wallis test (an extension of the Mann-Whitney test) is recommended for when there are three or more random samples. For example, numerous biological surveys are initiated by collecting data during the spring, summer, and fall. The

hypothesis might be to determine whether there is a significant difference in key biological indices between the different seasons (index periods). An analysis of variance could be used if the data were normally distributed. Applying the Mann-Whitney or Student's *t* test to each pair of random samples is not appropriate.

A special case of random sampling is when the random samples from one population (e.g., the upstream location) are paired with random samples from the second population (e.g., the downstream location). This situation is referred to as paired or matched sampling. The Wilcoxon signed rank test is recommended for paired samples. The paired *t* test can be used if the data are normally distributed.

In the second case we commonly test for monotonic or gradual changes at a single station. In this case, observations are typically taken on a regular basis (e.g., weekly, monthly, quarterly). The seasonal Kendall test is recommended for hypothesis testing. Linear regression might also be used but is generally discouraged. If the data do not have seasonal cycles, the Mann-Kendall test could be used.

Determining only the existence of a change is sometimes not sufficient for decision makers. It is also necessary to estimate the magnitude of the change. The seasonal Hodges-Lehman estimator is recommended for estimating the magnitude when comparing two random samples. The seasonal Kendall slope estimator is recommended when estimating the magnitude of monotonic trends. The difference in means and the Hodges-Lehman estimator are recommended for changes between two independent random samples, and the Sen slope estimator is recommended for estimating the magnitude of changes when seasonality is not present.

Extreme values

The most effective means for summarizing extreme values is to compute the proportion (or frequency) of observations exceeding some threshold value. This can be accomplished by plotting a time series with the threshold value or dividing the number of excursions by the total number of observations. A common analysis would be to compare the proportion of excursions from one year or station to the proportion of excursions from another year or station. A test for equality of proportions can be performed, or the confidence limits on proportions can be compared.

The evaluation of extreme values related to nonpoint source monitoring and other rain-induced impacts (e.g., combined sewer overflows (CSOs)) may require greater care. For example, when evaluating the number of overflows in a year or comparing storms, it is important to make sure that the data are comparable (similar rainfall, antecedent conditions, etc.). This may result in selecting portions of data sets for analysis.

4.1.4 Data Stratification

Lumping measurements over a period of time has limited use in water quality evaluations unless the period of time is defined in more specific terms and is directly related to the source of the identified problem. This is particularly true when comparing the effectiveness of management measures. If the implemented management measure is designed to reduce pollutant loadings during storm events, lumping baseflow and storm event data together for analysis makes little sense and might mask the effectiveness of the management measure.

In urban areas the time periods should be set to correspond to the pollutant of concern and urban activities. Depending on the monitoring objectives, it might be necessary to consider

periods of activity and nonactivity. If phosphorus is the pollutant of concern, periods that correspond to lawn maintenance activities and spring flush should be considered. If sediment is the problem, periods that correspond to the construction season should be considered. For irrigated agriculture, two periods should be established to correspond to irrigation and nonirrigation time.

In nonirrigated agricultural settings the periods selected should conform to the normal agricultural management pattern of the watershed. These periods should be based on amount of surface covered, precipitation patterns, and the timing of land and/or water management activities. By defining time periods, the analyst can evaluate a hypothesis regarding whether significant differences in nitrogen and phosphorus losses occur during different agricultural seasons. Alberts et al. (1978) used this concept to examine seasonal losses of nitrogen and phosphorus in Missouri during three periods:

- Fertilizer, seedbed, and establishment period (March-June).
- Reproduction and maturation period (July-October).
- Residue period (November-February).

Once temporal stratification has been completed, and if sufficient data are available, the water quality variable being examined could be categorized by initiation/transport mechanisms. In a sediment-related problem, for example, three categories were devised (Davenport, 1984b) to relate the principal detachment process of sediment particles:

(1) **Baseflow** (no rainfall or overland runoff to the stream). This category consists of non-precipitation-induced flow and is considered as the normal day-to-day flow (Viessman et al., 1977). Sediment concentrations are dependent on available material in the channel network and the carrying capacity of the flow.

(2) **Rainfall and snowmelt runoff.** This category consists of runoff events where the sediment concentrations are dependent on flowing water detachment or reentrainment of previously detached soil particles, together with sufficient overland flow to transport them to the stream network.

(3) **Event.** This category consists of rainfall-runoff events where the sediment concentrations are dependent on the detachment of soil particles due to the impact of raindrops and flowing water detachment or reentrainment of previously detached soil particles, together with overland flow to transport them to the stream network.

Data categorized by detachment category can then be examined in terms of resource management systems implemented to control the various types of detachment. It should be noted that data stratification results in smaller data sets. These new data sets must be checked for normality before performing any statistical analyses on them. It is also important to note that due to the smaller data set size the differences between data sets must be more pronounced to be significant.

4.1.5 Recommended Reading List and Available Software

Recommended reading list

Over the last 20 years, considerable effort by researchers and practitioners has gone into the development of improved statistical methods for analyzing environmental data. Nonetheless, there is probably no single reference that fully covers all of the issues that the data analyst must consider when selecting methods for analyzing environmental data. The following list provides a summary of selected references that provide more

details about a wider variety of issues. These references are strongly recommended for those who need a more in-depth discussion than that provided in this chapter.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. *Graphical Methods for Data Analysis*. Duxbury Press, Boston, 395 pp.

Conover, W.J. 1980. *Practical Nonparametric Statistics,* 2nd ed. Wiley, New York, 493 pp.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Company, New York, 320 pp.

Helsel, D.R. and R.M. Hirsch. 1995. *Statistical Methods in Water Resources*. Elsevier, Amsterdam, 529 pp.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical Methods,* 7th ed. The Iowa State University Press, Ames, Iowa, 507 pp.

Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. *Design of Water Quality Monitoring Systems*. Van Nostrand Reinhold Company, New York, 231 pp.

Available software

Many statistical methods have been computerized in easy-to-use software that is available for use on personal computers. Inclusion or exclusion in this section does not imply an endorsement or lack thereof by the U.S. Environmental Protection Agency. Commercial off-the-shelf software that covers a wide range of statistical and graphical support includes SAS, Statistica, Statgraphics, Systat, Data Desk (Macintosh only), BMDP, and JMP. Numerous spreadsheets, database management packages, and graphics programs can also be used to perform many of the needed analyses. In addition, the following programs, written specifically for environmental analyses, are available:

SCOUT: A Data Analysis Program, EPA, NTIS Order Number PB93-505303.

WQHYDRO (WATER QUALITY/HYDROLOGY GRAPHICS/ANALYSIS SYSTEM), Eric R. Aroner, Environmental Engineer, P.O. Box 18149, Portland, OR 97218.

WQSTAT, Jim C. Loftis, Department of Chemical and Bioresource Engineering, Colorado State University, Fort Collins, CO 80524.

4.2 Summary (Descriptive) Statistics

4.2.1 Point Estimation

Central tendency

The central tendency of a data set is the most important and widely used statistic (Gaugush, 1986; Ponce, 1980a). The mean, median, and mode are three common measures of central tendency. The arithmetic mean ($\bar{x}$) is the sum of the individual observations (x_i) divided by the number of observations (n):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (4\text{-}1)$$

The median ($P_{.50}$) is the middle value when all observations are ordered by magnitude ($x_1 \le x_2 \ldots \le x_n$). When there is an even number of observations, the median is the arithmetic mean of the two middle observations:

$$P_{.50} = \begin{cases} x_{(n+1)/2} & when\ n\ is\ odd \\ 0.5(x_{(n/2)} + x_{(n/2)+1}) & when\ n\ is\ even \end{cases} \qquad (4\text{-}2)$$

The mode is the most frequently occurring value in the set of observations. Comparison of these measures of central tendency reveals that the mean is sensitive to extreme values, whereas the median is not (Helsel and Hirsch, 1995; Remington and Schork, 1970). When the data are symmetrically distributed, the mean and median are comparable. In the case of nonpoint source pollution where storm events generate very large pollutant loadings, it is clear that the event mean and median may be very different. It is important that the data analyst consider the ramifications of relying on just one of these statistics when reporting results.

Other measures of central tendency include the midrange, geometric mean (GM_x), harmonic mean (HM_x), and weighted mean (Remington and Schork, 1970). The midrange is the arithmetic mean of the smallest and largest values and is influenced by extreme values. The geometric mean can be computed by

$$GM_x = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right) \qquad (4\text{-}3)$$

where $\ln(x)$ and $\exp(x)$ represent the natural log and exponential of the quantity x. It is the mean of the logarithms, transformed back to its original units. If the log-transformed data (i.e., $y_i = \ln x_i$) are symmetric, GM_x is a an unbiased estimate of the median (Helsel and Hirsch, 1995; Gaugush, 1986). It is common to report the GM_x for coliform data. It has also become common practice to estimate the HM_x flow for performing chronic risk assessments. It is computed as the reciprocal of the mean of the reciprocals using the following formula:

$$HM_x = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \qquad (4\text{-}4)$$

The weighted mean is a mean for which all observations do not have equal importance. For example, a common application of weighted means is the use of flow-weighted means for water quality variables measured during a storm event or when comparing water quality between two stream systems with different volumes of water flowing through them. The weight can be based on the portion of the population that the observation represents, either spatially or temporally (Gilbert, 1987). This may occur when the monitoring program has used a stratified sampling strategy and the strata have different sample sizes. In general, a weighted mean is computed where each observation is accorded its own weight (w_i):

$$Weighted\ mean = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad (4\text{-}5)$$

Summarizing storm event data

Three approaches for summarizing storm event data, which are applications of the weighted mean described above, are the flow-weighted mean concentration (FWMC), the time-weighted mean concentration (TWMC), and the event mean concentration (EMC). The FWMC and TWMC are calculated as (USEPA, 1990)

$$FWMC = \frac{\sum_{i=1}^{n} C_i T_i Q_i}{\sum_{i=1}^{n} T_i Q_i} \qquad (4\text{-}6)$$

$$TWMC = \frac{\sum_{i=1}^{n} C_i T_i}{\sum_{i=1}^{n} T_i} \qquad (4\text{-}7)$$

where

C_i = concentration of the i^{th} sample;
T_i = time period for which the i^{th} sample is used to characterize the concentration; and
Q_i = instantaneous discharge at the time of the i^{th} sample.

The numerator of Equation 4-6 is equal to the total loading. The EMC can be estimated with the following equation and is similar to the TWMC except for end effects:

$$EMC = \frac{\sum_{i=1}^{n} 0.5\,(T_{i+1}-T_i)(C_iQ_i+C_{i+1}Q_{i+1})}{\sum_{i=1}^{n} 0.5\,(T_{i+1}-T_i)(Q_i+Q_{i+1})} \qquad (4\text{-}8)$$

Figure 4-2 presents a summary of hte rainfall, runoff, and total nitrogen data collected from a storm event in Florida. Runoff (1,780 ft^3) from this 0.2-inch storm lasted for approximately 2.4 hours. The total runoff volume and precipitation depth can be computed by integrating the representative curves in Figure 4-2 or directly from the data. The nitrogen concentrations are typical of a "first flush" in which the concentrations are higher during the early part of the runoff. Tables 4-4 and 4-5 present the raw nitrogen values from Figure 4-2 together with the example calculations for computing the FWMC and EMC, respectively.

The first column in Table 4-4 is the time since the beginning of the storm. The fourth column is the time interval, T_i, represented by each sample. For example, the first entry, T_1, of 540 seconds is computed as (0.24 hours - 0.09 hours) times 3600 seconds/hour. The value of 0.24 is halfway between 0.20 and 0.28 hours. Selecting the halfway point between 0.20 and 0.28 hours centers the water quality observation in the time period being evaluated. The second entry, T_2, of 306 seconds is computed as (0.325 hours - 0.24 hours) times 3,600 seconds/hour. The value of

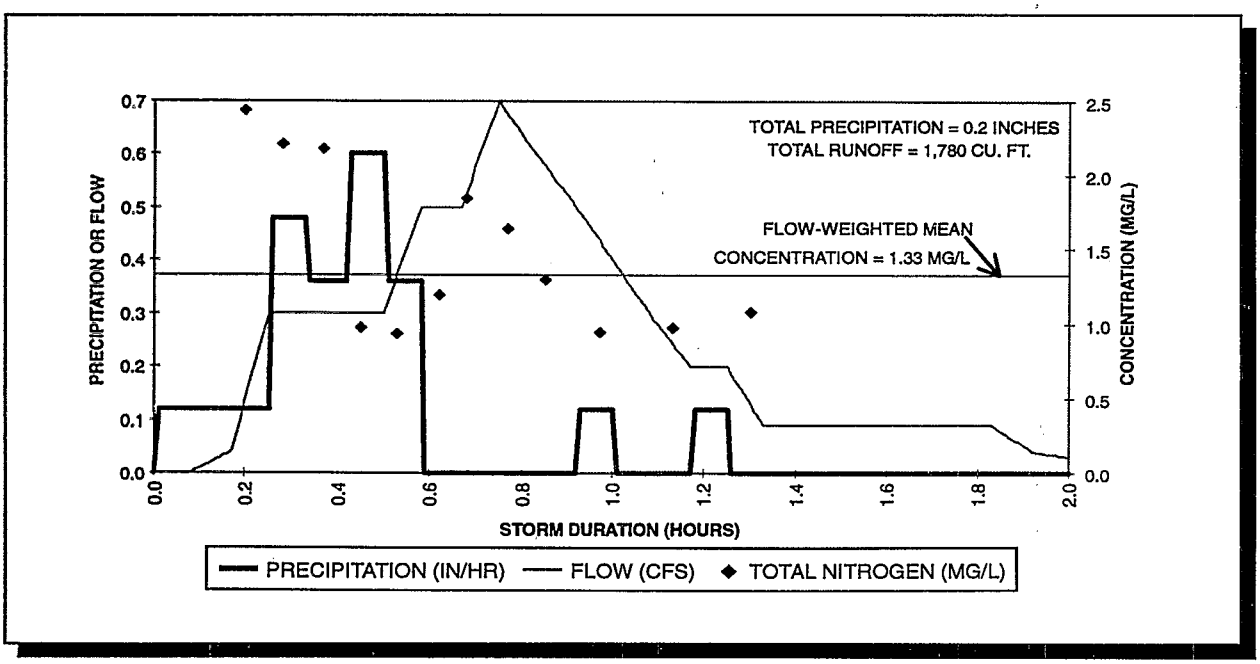


Figure 4-2. Precipitation, runoff, total nitrogen, and total phosphorus from a single storm event in Florida.

Table 4-4. Total nitrogen (TN) runoff concentrations for a single storm event in Florida.

	Q_i	C_i	T_i	T_iQ_i	$C_iT_iQ_i$
Time (hr)	Flow (cfs)	TN (mg/L)	Int. (sec)	(ft^3)	(mg-ft^3/L)
0.09	0.00	-	-	-	-
0.20	0.14	2.44	540	75.60	184.46
0.28	0.30	2.21	306	91.80	202.88
0.37	0.30	2.18	306	91.80	200.12
0.45	0.30	0.97	288	86.40	83.81
0.53	0.38	0.93	306	116.28	108.14
0.62	0.50	1.19	270	135.00	160.65
0.68	0.53	1.85	270	143.10	264.74
0.77	0.68	1.64	306	208.08	341.25
0.85	0.58	1.30	360	208.80	271.44
0.97	0.44	0.94	504	221.76	208.45
1.13	0.24	0.97	594	142.56	138.28
1.30	0.13	1.08	4302	559.26	604.00
2.41	0.00				
			Sum	2,080.44	2,768.23

FWMC = 2,768.23 / 2,080.44 = 1.33 mg/L

0.325 is halfway between 0.28 and 0.37 hours. The value of 0.24 is halfway between 0.20 and 0.28 hours. The fifth column is equal to flow (column 2) multiplied by the time interval (column 4). For example, the entry of 75.60 ft^3 is equal to 0.14 cfs times 540 seconds. The sum of the fifth column is equal to the denominator of Equation 4-6. The sixth column is equal to the volume (column 5) multiplied by the nitrogen concentration (column 3). For example, the entry of 184.46 mg-ft^3/L is equal to 75.60 ft^3 times 2.44 mg/L. The sum of this column is equal to the total nitrogen loading for the storm (and the numerator in Equation 4-6). Using conversions, the total nitrogen loading for this storm is 78.4 grams. As shown in Table 4-4, the FWMC is equal to 1.33 mg/L. Because different analysts use different conventions for analyzing storms, it is important that the analyst exercise care when comparing the storm summaries computed by different analysts.

Table 4-5 demonstrates the use of Equation 4-8 with the same storm event presented in Figure 4-2 and Table 4-4. The first three columns of Table 4-5 are the same as Table 4-4. The next four columns correspond to intermediate calculations needed for Equation 4-8. For example, the values of 0.11, 0.000, 0.342, and 0.14 in the first data row are computed from 0.20-0.09, 0.00 x 0.000, 0.14 x 2.44, and 0.00 + 0.14, respectively. The last two columns correspond to intermediate calculations for the numerator and denominator of Equation 4-8, respectively. Finally, the EMC can be calculated as 0.6722/0.4981 or 1.35 mg/L, as shown in Table 4-5.

Table 4-5. Total nitrogen (TN) runoff concentrations for a single storm event in Florida and example calculations for the EMC.

Time (hr)	Flow (cfs)	TN (mg/L)	T_{i+1} $-T_i$	$C_i \cdot Q_i$	$C_{i+1} \cdot Q_{i+1}$	$Q_i + Q_{i+1}$	Num.	Den.
0.09	0.00	0.00	0.11	0.000	0.342	0.14	0.0188	0.0077
0.20	0.14	2.44	0.08	0.342	0.663	0.44	0.0402	0.0176
0.28	0.30	2.21	0.09	0.663	0.654	0.60	0.0593	0.0270
0.37	0.30	2.18	0.08	0.654	0.291	0.60	0.0378	0.0240
0.45	0.30	0.97	0.08	0.291	0.353	0.68	0.0258	0.0272
0.53	0.38	0.93	0.09	0.353	0.595	0.88	0.0427	0.0396
0.62	0.50	1.19	0.06	0.595	0.981	1.03	0.0473	0.0309
0.68	0.53	1.85	0.09	0.981	1.115	1.21	0.0943	0.0545
0.77	0.68	1.64	0.08	1.115	0.754	1.26	0.0748	0.0504
0.85	0.58	1.30	0.12	0.754	0.414	1.02	0.0701	0.0612
0.97	0.44	0.94	0.16	0.414	0.233	0.68	0.0517	0.0544
1.13	0.24	0.97	0.17	0.233	0.140	0.37	0.0317	0.0315
1.30	0.13	1.08	1.11	0.140	0.000	0.13	0.0779	0.0722
2.41	0.00							
						Sum	0.6722	0.4981

The event mean concentration (EMC) = 0.6722 / 0.4981 = 1.35 mg/L

Loading rates

Converting data into a loading rate is a very common practice in nonpoint source evaluations. Computing loading rates results in factoring out activities that are related to the data collection or generation process. The most common conversions are related to time period (kg/yr), unit area (kg/ha), or a combination of unit area and time period (kg/ha/month). The other major type of conversion is related to parameter generation or transport factors such as rainfall and runoff; examples are kilograms per centimeter of precipitation or kilograms per cubic liter of streamflow.

Examples of raw data and normalized data are provided in Tables 4-6 and 4-7, respectively. The watershed is 20 ha and has three consecutive years of pre- and post-implementation sediment loading, precipitation, and runoff data. Review of Table 4-7 indicates that there has been a 20 percent reduction in sediment generated per centimeter of rainfall and a 22 percent reduction in annual loading. This indicates that sediment loading, adjusted for runoff and total precipitation, has decreased. A more detailed frequency analysis would be required to test for statistical significance. It might also be useful to consider other issues such as rainfall intensity.

Summarizing data with censored observations

Observations reported as less-than or nondetect are often troublesome for many statistical procedures. Quite simply, it is difficult to compute the mean (or any number of other statistics) when one or more of the values is

Table 4-6. Raw data by time period.

1971-1973		
Total sediment loading	48	kg
Total precipitation	120	cm
Total runoff	15	L^3
1974: Implementation of terraces and conservation tillage		
1975-1977		
Total sediment loading	45	kg
Total precipitation	180	cm
Total runoff	18	L^3

Table 4-7. Loadings rate data.

1971-1973		
Average annual loading	12	kg/year
Average annual loading	0.10	kg/cm/year
Average annual loading	1.07	kg/L^3/year
1974: Implementation of terraces and conservation tillage		
1975-1977		
Average annual loading	15	kg/year
Average annual loading	0.08	kg/cm/year
Average annual loading	0.83	kg/L^3/year

reported as less than the detection limit. Some authors have recommended not censoring the data (Dakins et al., 1996; Porter et al., 1988), but this concept has not been adopted too often in practice. One approach is to substitute one-half the detection limit for the censored observations. This practice is discouraged by Helsel and Hirsch (1995), Although it is widely used due to quick implementation in spreadsheet software.

Gilbert (1987) describes the trimmed mean and the Winsorized mean for use when there are censored data in the data set. The trimmed mean is a useful estimator of the mean when the data are symmetrically distributed and it is necessary to guard against erroneous data or when censored observations are present (Gilbert, 1987). The trimmed mean is equal to the arithmetic mean after equal proportions of the smallest and largest observations have been dropped from the analysis. Research has suggested that for symmetric distributions, no more than 50 percent of all data should be dropped (Hoaglin et al., 1983). If the data are not symmetric, no more than 30 percent of all data should be dropped (Mosteller and Rourke, 1973). In all cases, the percentage of observations trimmed should be reported.

The Winsorized mean can be computed by estimating the mean after substituting an equal proportion of the smallest observations with the next largest observation and the largest observations with the next smallest observation. Two final approaches for estimating summary statistics with censored data include maximum likelihood estimation (Cohen, 1959) and probability plotting procedures (Travis and Land, 1990). Helsel and Hirsch (1995) describe these methods and their shortcomings, particularly with small sample sizes. Helsel and Cohn (1988) provide

approaches estimating summary statistics when there are multiple censoring levels in the same data set.

Dispersion

Measures of dispersion or measures of variation describe the extent to which the data are spread out from the central tendency (Freund, 1973). The measures of dispersion described in this manual are the range, variance, standard deviation, and interquartile range. The variance (and standard deviation) are acceptable measures of dispersion when the data are normally distributed or can be transformed into normally distributed data. Even more so than the mean, the variance can be influenced by a few outliers. The interquartile range is a stable estimate of dispersion.

The *range* of a set of observations is simply the difference between the largest and smallest values and should be considered only as a rough estimate of dispersion due to its dependence on extreme values (Gaugush, 1986; Ponce, 1980a; Remington and Schork, 1970).

The *variance* (s^2) is given by the following:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \qquad (4\text{-}9)$$

The *standard deviation* (s) is the square root of the variance. For observations that come from a normal distribution, about 68 percent of the observations are within $\pm$ one standard deviation of the mean (Figure 4-3A). Figure 4-3B demonstrates the effect of changing the mean and variance for a normal distribution.

In cases where it is necessary to compare standard deviations for samples with different means, a measure of relative variation is needed. The variation in a population can also be measured using the *coefficient of variation* (*CV*) and is defined as:

$$CV = s/\bar{x} \qquad (4\text{-}10)$$

Since *CV* is unitless, it does not matter what units (e.g., mg/L, μg/L) are used, making qualitative comparisons of different studies easier. In Figure 4-3B, the *CV*s for the two normal distributions are nearly the same (0.25 and 0.236). The *CV* can also be used to compare the dispersions of two or more data sets that are measured in different units. It is recommended that analysts use the above equation for computing *CV* although some analysts commonly multiply the above result by 100.

The *interquartile range* is a robust alternative (i.e., it changes little in the presence of outliers) to the standard deviation (Gaugush, 1986; Helsel and Hirsch, 1995). It is the difference between the observation at the upper quartile, Q_3 ($P_{.75}$), and the observation at the lower quartile, Q_1 ($P_{.25}$). The upper quartile is the observation value for which 75 percent of the observation values are lower, and the lower quartile is the value for which 25 percent of the observation values are lower.

To compute a quartile, the data must be ordered from smallest to largest observation. Then compute $p(n+1)$ where p corresponds to the quartile (as a fraction), either 0.25 or 0.75, and n is the number of observations. Consider the following example of 10 observations that have been ordered from low to high:

<0.10, 0.11, 0.16, 0.51, 0.59, 0.68, 0.79, 0.85, 0.98, 3.00

For n equal to 10, the lower and upper quartile are equal to the 2.75th (0.25 x 11) and 8.25th

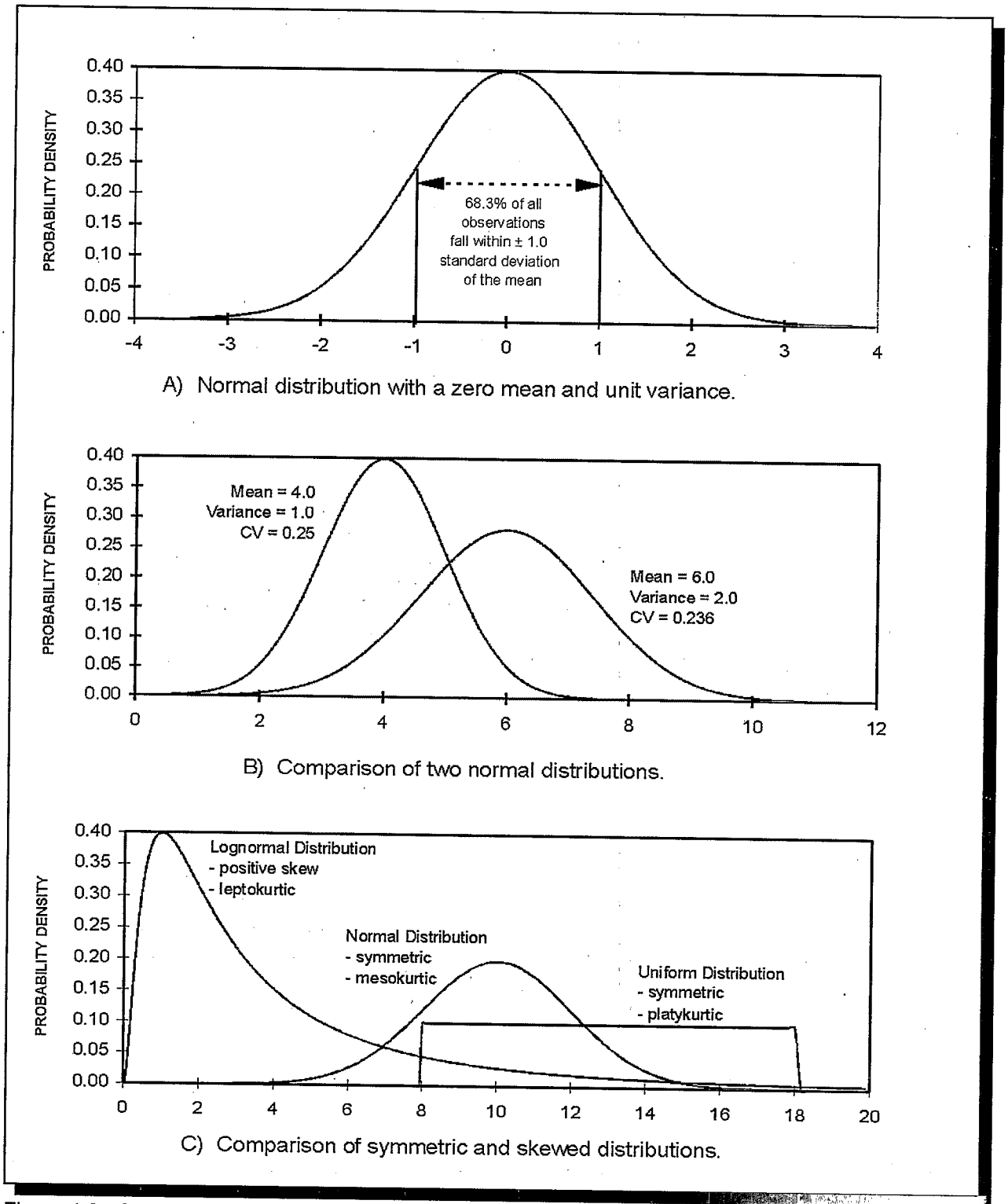


Figure 4-3. Comparison of several theoretical distributions.

(0.75 x 11) ordered observation. Using the data from above, Q_1 is equal to 0.11 + 0.75 x (0.16-0.11) or 0.1475 and Q_3 is equal to 0.85 + 0.25 x (0.98-0.85) or 0.8825. Similar to the *CV*, the coeffi-cient of quartile variation (*V*) can be used to compare different data sets:

$$V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (4\text{-}11)$$

Skewness and Kurtosis

Skewness (γ) is a measure of distribution symmetry and is given by the following formula:

$$\gamma = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{s^3} \qquad (4\text{-}12)$$

Figure 4-3C is a comparison of a lognormal distribution (positively skewed) and two symmetric distributions. The kurtosis (*k*) of a distribution describes its peakedness relative to the length and size of its tails (Remington and Schork, 1970). It has been argued, however, that kurtosis measures tail heaviness, not the peakedness of a distribution (SAS Institute, Inc., 1985a). The normal distribution is considered to have intermediate kurtosis (mesokurtic). Flat distributions with short tails have low kurtosis (platykurtic), whereas distributions with sharp peaks and long tails have high kurtosis (leptokurtic). These types of distributions are also shown in Figure 4-3C. Kurtosis can be estimated with the following equation:

$$k = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad (4\text{-}13)$$

4.2.2 *Interval Estimation*

In practice, the real mean and standard deviation of the target population are never known. We take random samples from the target population, compute the mean from the random samples, and infer the target population mean. Since we cannot sample all of the waterbody, some error will always be associated with the estimate. To report the reliability of estimated statistics, it is recommended that the confidence interval also be computed. This section describes procedures for estimating the confidence interval for the mean, standard deviation, median, and quartiles.

Mean

For large sample sizes or samples that are normally distributed, a symmetric confidence interval for the mean is appropriate. This is because the distribution of the sample mean will approach a normal distribution even if the data from which the mean is estimated are not normally distributed. The Student's *t* statistic ($t_{\alpha/2,n-1}$) is used to compute a symmetric confidence interval for the population mean, μ:

$$\bar{x} - t_{\alpha/2,n-1}\sqrt{s^2/n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\sqrt{s^2/n} \qquad (4\text{-}14)$$

Values for the *t* statistic can be found in Table D2. This equation is appropriate if the samples are normally distributed or the sample size is greater than 30 (Freund, 1973), although Helsel and Hirsch (1995) suggest that highly skewed data might require more than 100 observations.

Problem:

Fifty-four samples were collected to determine the fraction of water collected (i.e., the split) by a water and sediment sampler for plot and field studies (Dressing et al., 1987). The data were tested and found to be normally distributed with a

mean split of 0.0265 and a standard deviation of 0.0040. Determine the 95 and 99 percent confidence intervals for the population mean, μ.

Solution:

For the 95 and 99 percent confidence intervals, $\alpha/2$ is equal to 0.025 and 0.005, respectively. There are 53 degrees of freedom. The t value is then estimated by interpolation between the values for 50 and 60 degrees of freedom (Table D2) using the columns $\alpha = 0.025$ and $\alpha = 0.005$, respectively. We obtained t values of 2.0061 and 2.6726.

The 95 percent confidence interval about the mean can then be estimated as

$$\bar{x} - t_{.025,53}\sqrt{s^2/n} \le \mu \le \bar{x} + t_{.025,53}\sqrt{s^2/n}$$

$$\mu \ge 0.0265 - 2.0061\sqrt{0.004^2/54}$$
$$\mu \le 0.0265 + 2.0061\sqrt{0.004^2/54}$$

$$0.0254 \le \mu \le 0.0276$$

There is a 95 percent chance that the population mean, μ, will fall between 0.0254 and 0.0276.

The 99 percent confidence interval about the mean can then be estimated as

$$\bar{x} - t_{.005,53}\sqrt{s^2/n} \le \mu \le \bar{x} + t_{.005,53}\sqrt{s^2/n}$$

$$\mu \ge 0.0265 - 2.6726\sqrt{0.004^2/54}$$
$$\mu \le 0.0265 + 2.6726\sqrt{0.004^2/54}$$

$$0.0250 \le \mu \le 0.0280$$

There is a 99 percent chance that the population mean, μ, will fall between 0.0250 and 0.0280. Note that to have a higher confidence (99 versus 95 percent), a bigger interval is required.

Standard deviation

The confidence interval for the standard deviation of a normal distribution for small sample size can be estimated as (Freund, 1973)

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right]^2 \le \sigma \le \left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]^2 \qquad (4\text{-}15)$$

where χ^2 is the chi-square distribution. Values of χ^2 can be found in Table D3. Note that since the χ^2 is not symmetric, the above inequality requires a different chi-square value for each end of the confidence interval, i.e., values for $\alpha/2$ and $(1-\alpha/2)$. For large samples the following formula may be used (Freund, 1973):

$$\frac{s}{1+\frac{Z_{\alpha/2}}{\sqrt{2n}}} \le \sigma \le \frac{s}{1-\frac{Z_{\alpha/2}}{\sqrt{2n}}} \qquad (4\text{-}16)$$

Note that the confidence interval for the variance can be obtained by squaring the confidence interval for the standard deviation (Remington and Schork, 1970).

Median and Quartiles

Although several approaches exist to estimate confidence intervals for any percentile, many rely on assuming a normal or lognormal distribution. The approach presented here (Conover, 1980) for more than 20 observations does not rely on these assumptions. Conover (1980) also provides a procedure for smaller sample sizes. To calculate the confidence interval corresponding to the median, lower quartile, or upper quartile, the following procedure is used.

1. Order the data from smallest to largest observation such that

$$x_1 \le \ldots \le x_r \le \ldots \le x_p \le \ldots \le x_s \le \ldots \le x_n$$

where x_p corresponds to the median, lower quartile, or upper quartile.

2. Compute the values of r^* and s^* as

$$r^* = np - Z_{\alpha/2}(np(1-p))^{0.5}$$

$$s^* = np + Z_{\alpha/2}(np(1-p))^{0.5}$$

where $Z_{\alpha/2}$ is selected from Table D1.

3. Round r^* and s^* up to the next highest integers r and s. The 1-α lower and upper confidence limits for x_p are x_r and x_s, respectively.

Problem:

Compute the 90 percent confidence interval for the median using the 25 observations presented below.

0.08, 0.09, 0.10, 0.23, 0.29, 0.32, 0.38, 0.48, 0.49, 0.61, 0.62, 0.62, 0.68, 0.70, 0.72, 0.75, 0.76, 0.77, 0.80, 0.83, 0.84, 0.87, 0.96, 0.98, 1.00

Solution:

Note that the data have already been ordered and the median is equal to 0.68.

r^* and s^* can then be computed as follows:

$$r^* = np - Z_{\alpha/2}(np(1-p))^{0.5}$$
$$= 25 \times 0.5 - 1.645(25 \times 0.5 \times 0.5)^{0.5} = 8.4$$

$$s^* = np + Z_{\alpha/2}(np(1-p))^{0.5}$$
$$= 25 \times 0.5 + 1.645(25 \times 0.5 \times 0.5)^{0.5} = 16.6$$

r and s are therefore 9 and 17, respectively. From the above listing, x_9 and x_{17} can be estimated as 0.49 and 0.76 mg/L, respectively.

4.3 Graphical Data Display

Graphical data display is an important aspect of data analysis. Gaugush (1986) recommends beginning an analysis with a graphical display of data. This is an excellent approach, though in this document graphical displays are discussed after Section 4.2, Summary Statistics, so that basic terminology is provided first.

Based on an inspection of the data, the analyst should be able to make a qualitative assessment of seasonality, variance homogeneity, distributions, data gaps, unusual sampling patterns, the presence of censored data, and a general characterization of the available data. All of these features might have an influence on the type of statistical analyses to be performed. By using graphical methods to examine the data, the data analyst can more appropriately select statistical methods. The reader is cautioned, however, that visual inspection of the results cannot be used to group data into the categories before and after BMP implementation. This decision must be made based on the analyst's knowledge of the system.

Figures 4-4 to 4-7 illustrate various graphical displays of dissolved oxygen (DO) data for a monitoring station in the Delaware River at Reedy Island, Delaware. Each figure reveals different features of the data. The DO time series plot (Figure 4-4) demonstrates a seasonal nature to the data. In this case, the time series includes data from a 10-year time span. Similar plots can also be made over shorter time periods such as intensive data collection efforts during a storm event. In the case of a storm event, the investigator may plot precipitation and runoff volume together with pollutant concentrations (see Figure 4-2). It is also apparent from Figure 4-4

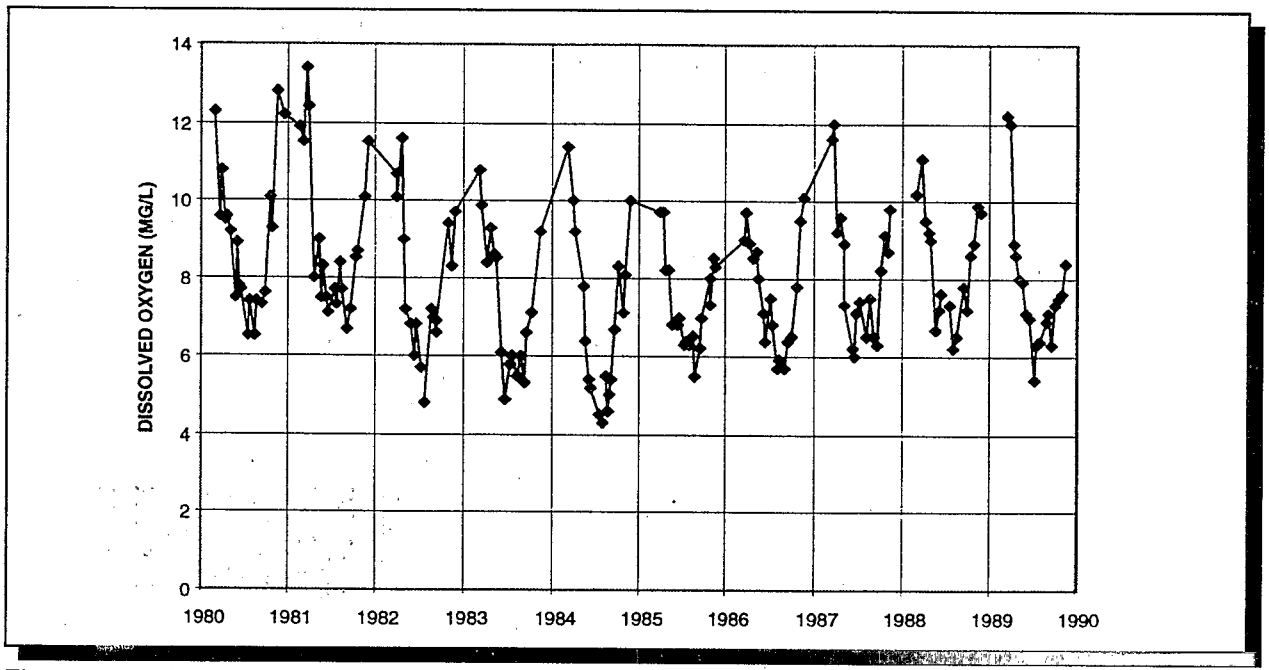


Figure 4-4. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a time series plot.

that data are collected more frequently in the summer months. Inspection of the raw data show that DO was typically sampled twice a month during the summer, once a month during the spring and autumn months, and less often during the winter months. It is also clear that since the summer of 1984, the DO has not dropped below 5.0 mg/L.

Figures 4-5 and 4-6 are a DO histogram and stem-and-leaf plot, respectively. In Figure 4-5, the height of the bar indicates the number of observations falling within a certain DO range. For example there are 15 observations between 7.5 and 8.0 mg/L. The stem-and-leaf plot (Figure 4-6) displays the raw data instead of a bar. The values on the left side of the vertical axis indicates the DO concentration in a whole number (e.g., 11| represents 11 mg/L). The values on the right side of the vertical axis indicate the DO concentration to the tenths of a mg/L. Thus 11|14566 indicates that there is one value of 11.1 mg/L, one value of 11.4 mg/L, one value of 11.5 mg/L, and two values of 11.6 mg/L. These figures demonstrate that most of the observations fall between 6.0 and 10.0 mg/L. Typically, the analyst would select the histogram for less technical audiences and the stem-and-leaf plot for technical audiences.

Figure 4-7 is a boxplot. For each month along the horizontal axis, the box indicates the middle 50 percent of the data (which corresponds to the interquartile range). The lower and upper ends of the box represent the 25th and 75th percentiles ($P_{.25}$ and $P_{.75}$), respectively. The horizontal line inside the box represents the median. The whiskers extending from the box represent the range of the remaining observations. In this case, the whiskers extend to the minimum and maximum observations for a given month. Some software packages use different rules for creating the whiskers (Chambers et al., 1983), and the analyst should be aware of such differences when mixing

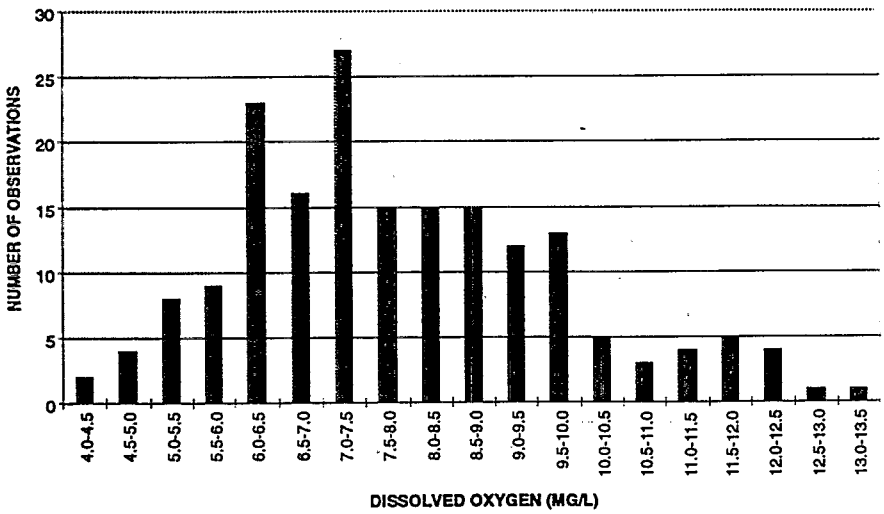


Figure 4-5. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a histogram.

```
 4 | 35689
 5 | 02344455577789
 6 | 000012223333344444455555556677788888899
 7 | 000011111112222233333344455555566677788889
 8 | 00001222333344455566677799999
 9 | 000012222233455566677777899
10 | 0011112788
11 | 14566
12 | 00148
13 | 1

4|3 = 4.3 MG/L
```

Figure 4-6. Stem-and-leaf plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

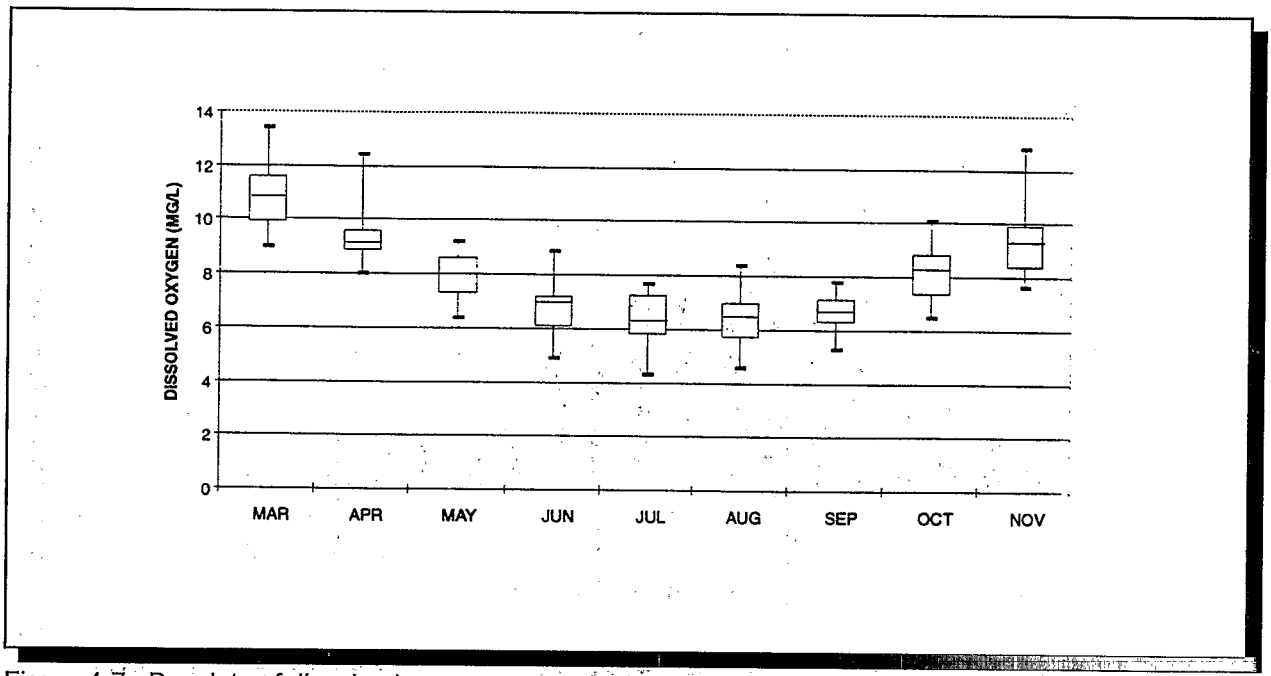


Figure 4-7. Boxplots of dissolved oxygen concentrations by month from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

and matching analyses from different software packages. Some software packages plot observations that exceed $P_{.75}$ (or are less than $P_{.25}$) by more than 1.5 times the interquartile range as individual points, which is perhaps a more desirable approach than others. Depending on how far the observations exceed this range, different symbols may be displayed.

The expected seasonal nature of DO is strongly depicted in Figure 4-7, confirming the suspicions developed from visual inspection of Figure 4-4. This figure also allows the analyst to evaluate how much variability there is in the data. It may be interesting to note, for example, that in November the lower and upper 25 percent of the data (represented by the whiskers) are drastically different lengths while the whiskers (and the box) for August appear symmetric. In this case, DO was plotted as a function of month. Similar plots as a function of year could also have been made with these data. Alternatively, the analyst may compare data by station. Figure 4-8 is a boxplot of sulfate concentrations. Stations 16 and 17 are roughly 20 miles downstream from Stations 14 and 15. Based on visual inspection, it appears that the sulfate concentration increases at the downstream stations; however, a statistical test is required. In this case, the stream receives significant irrigation return flows between the upstream and downstream stations, which might be the cause of the increased sulfate concentrations.

In other cases, it might be helpful to plot water quality data as a function of other explanatory variables such as flow. Figure 4-9 is a log-log plot of total suspended solids measured at a storm sewer in Denver, Colorado, as a function of instantaneous flow. Depending on the nature of the source loading, the correlation between pollutant concentrations and flow could be positive (as in Figure 4-9) or negative, or no correlation might exist. Typically, a negative

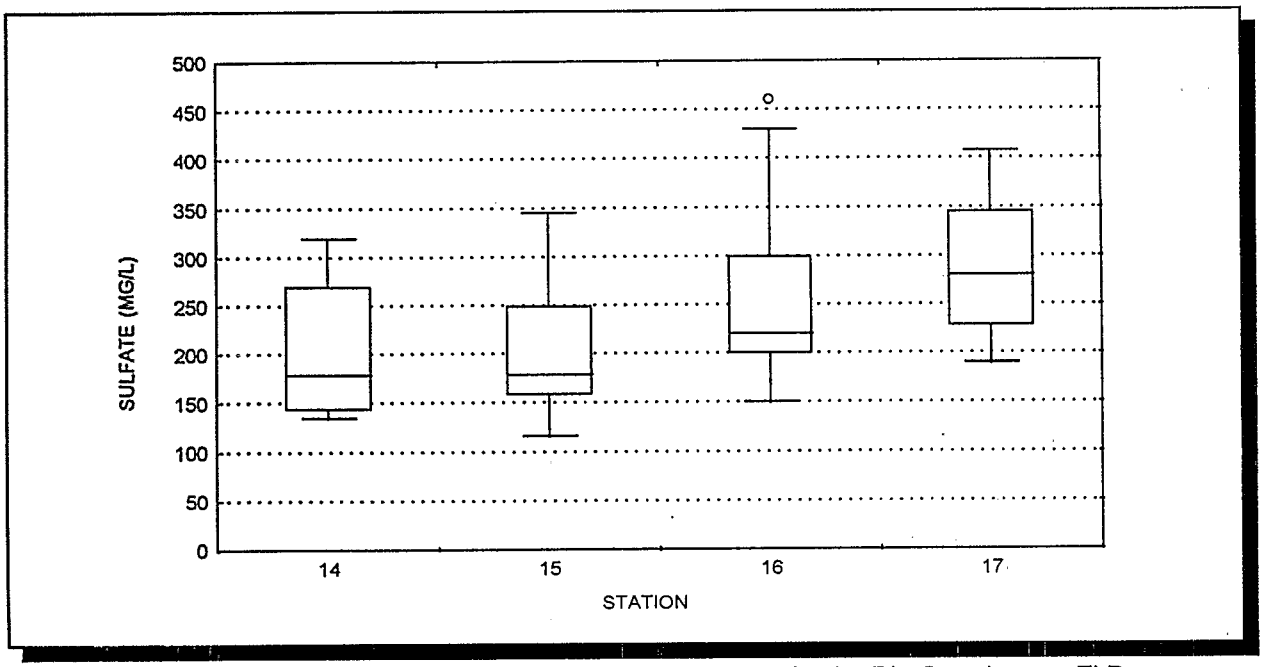


Figure 4-8. Boxplot of sulfate concentrations from 1993 and 1994 for the Rio Grande near El Paso, Texas.

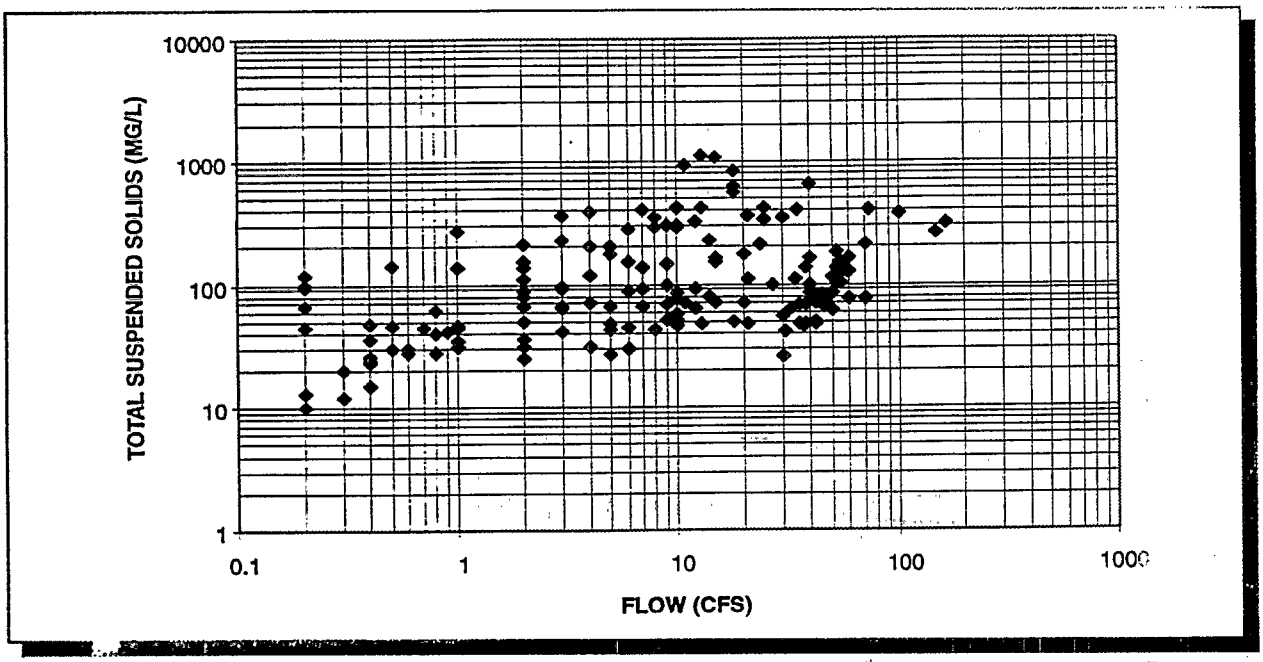


Figure 4-9. Bivariate scatter plot of total suspended solids and flow at 36th Street storm sewer in Denver, Colorado.

correlation (decreasing concentrations with increasing flows) is indicative of constant pollutant sources (e.g., traditional point sources) while a positive correlation (increasing concentrations with increasing flows) is indicative of nonpoint source loadings. It is critically important that the analyst know what is going on in the field before jumping to any conclusion about the meaning of concentration and flow correlations.

Figure 4-10 is a scatter plot of orthophosphate for several stations along the Delaware River. In addition to the seasonal cycles during each year, some unusually high values that exceed 0.2 mg/L as phosphorus on September 23, 1991, can be observed. In this case, one potential cause might be unit conversions. The data were stored as milligrams per liter of phosphorus; however, another common set of units for orthophosphate is milligrams per liter of phosphate. If one were to multiply the data collected on September 23, 1991, by one-third (approximate conversion from phosphate to phosphorus), the data would fall in line with the rest of the observations. Ideally, the analyst would go back to the original data to determine what type of error occurred and perform corrective action before proceeding with the statistical analysis. These types of errors also occur while converting data from parts per million to parts per billion, converting from wet-weight to dry-weight basis, normalizing for organic carbon, and so forth. It might also be helpful to plot this orthophosphate data as a function of suspended solids for corroborative evidence. Data visualization is a good method for picking out gross errors; however, it cannot be relied on for more subtle errors. *The likelihood of correcting data errors decreases significantly with time.*

4.4 Evaluation of Test Assumptions

One of the basic criteria for selecting between parametric tests is whether the data being analyzed have a specific distribution (usually

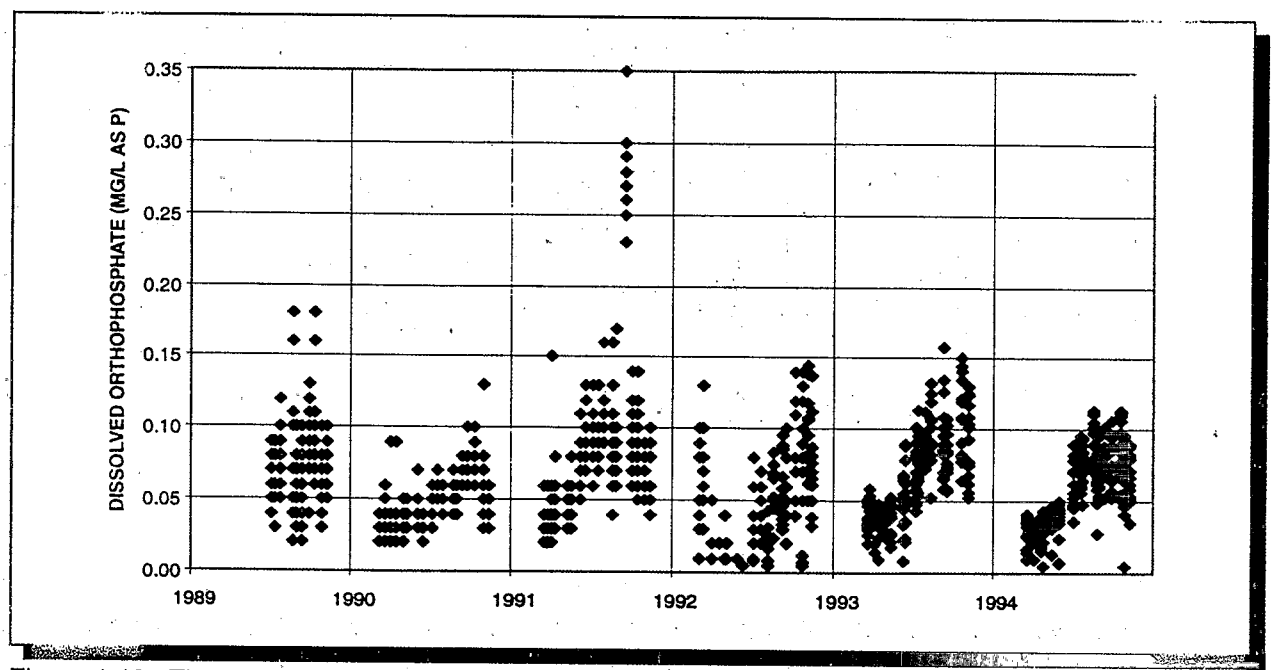


Figure 4-10. Time series plot of dissolved orthophosphate from 1989 through 1994 for portions of the Delaware River.

normal). For data with unknown distributions, nonparametric methods should be used since these methods do not require that the data have a defined distribution. In addition, numerous tests require that the observations be independent (that is, randomly collected) and that the variances of the populations being compared be equal or of known ratio (Ponce, 1980a).

This section describes tests that can be used to determine whether a data set satisfies some of the assumptions and requirements of statistical tests. Analysts are referred to statistics texts such as Snedecor and Cochran (1980) for further information regarding test assumptions.

4.4.1 Tests of Normality

There are a variety of methods for evaluating normality that range from graphical methods to statistical tests. If the sample data set does not pass the normality tests, there are several options including data transformation. Data transformation can (Gaugush, 1986):

- Straighten (linearize) a nonlinear relationship between two variables.
- Reduce skew (achieve symmetry) in a data set for a single value.
- Stabilize variance (create constant variance) for a particular variance across two or more data sets.

Log transformations are the most common in water quality and hydrologic variables (Gaugush, 1986; Ponce, 1980a; Spooner et al., 1986; USEPA, 1983a) because these data typically have a positive skew. The reader is encouraged to study the examples of log transformations presented by Ponce (1980a) and USEPA (1983a). Additional information regarding other transformations such as Box-Cox transformations is provided by Snedecor and Cochran (1980).

The transformed data should also be tested for normality before proceeding with further statistical analyses (Spooner et al., 1986).

Graphical Methods

Examining boxplots can be useful in developing a qualitative opinion regarding normality. Another graphical approach is to prepare probability plots. The cumulative frequency can be plotted on normal probability graph paper. If the graphics software does not provide for probability plots, the following method can be used. First, sort the data from low to high. For each observation, compute a plotting position using

$$p_i = \frac{i - 0.375}{n + 0.25} \qquad (4\text{-}17)$$

Helsel and Hirsch (1995) identify several other formulas that could be used for plotting position, but note that this approach is the most appropriate for comparing data to normal distributions in probability plots. The plotting positions are then converted to normal quantiles (Z_p) using Table D1.

Consider, for example, the sulfate data from Station 16 (see Figure 4-8). Table 4-8 presents the 42 observations ordered from low to high. For i equal to 1, p_1 is equal to (1-0.375)/42.25 or 0.0148. Using Table D1, it is necessary to look up p equal to 1.0-0.0148 or 0.9852. The corresponding Z_p for p equal to 0.9852 is 2.176.

Therefore, the corrresponding Z_p for p equal to 0.0148 is -2.176. The same procedure is followed for the remaining observations. Sulfate concentrations are then plotted as a function of the normal quantile as shown in Figure 4-11A. The straight line in Figure 4-11A corresponds to the theoretical shape of the normal distribution with a mean and standard deviation equal to those computed from the raw sulfate data. If the data were normally distributed, the data would tend to

Table 4-8. Calculation of plotting position for the sulfate data from Station 16 in Figure 4-8.

Ordered Obs. Num Quantile	Sulfate (mg/L)	Plotting Position	Normal Quantile	Ordered Obs. Num	Sulfate (mg/L)	Plotting Position	Normal
(i)	p_i		Z_p	(i)		p_i	Z_p
1	150	0.0148	-2.176	22	220	0.5118	0.030
2	150	0.0385	-1.769	23	220	0.5355	0.089
3	160	0.0621	-1.537	24	240	0.5592	0.149
4	170	0.0858	-1.367	25	240	0.5828	0.209
5	170	0.1095	-1.229	26	240	0.6065	0.270
6	180	0.1331	-1.112	27	250	0.6302	0.332
7	190	0.1568	-1.008	28	260	0.6538	0.396
8	200	0.1805	-0.914	29	270	0.6775	0.461
9	200	0.2041	-0.827	30	290	0.7012	0.528
10	200	0.2278	-0.746	31	300	0.7249	0.597
11	200	0.2515	-0.670	32	300	0.7485	0.670
12	200	0.2751	-0.597	33	310	0.7722	0.746
13	200	0.2988	-0.528	34	310	0.7959	0.827
14	210	0.3225	-0.461	35	320	0.8195	0.914
15	210	0.3462	-0.396	36	330	0.8432	1.008
16	210	0.3698	-0.332	37	360	0.8669	1.112
17	210	0.3935	-0.270	38	380	0.8905	1.229
18	210	0.4172	-0.209	39	400	0.9142	1.367
19	210	0.4408	-0.149	40	420	0.9379	1.537
20	220	0.4645	-0.089	41	430	0.9615	1.769
21	220	0.4882	-0.030	42	460	0.9852	2.176

fall along the straight line. Clearly, the data do not fit a normal distribution, but are more typical of a positively skewed data set. As an alternative, the data can be log-transformed and the same analysis performed. In this case, the log-transformed data are less skewed (Figure 4-11B). The conclusion from this analysis that the data are not normal. Visually, it is difficult to determine whether the data are lognormally distributed.

Skewness

The approach used in testing for normality using skewness (Equation 4-12) is that a nonnormal distribution may be skewed, whereas a normal distribution is not skewed. If there are more than 150 observations and the data are normally distributed, the confidence limits on skewness from a normal distribution are given by (Salas et al., 1980)

$$-Z_{1-\alpha/2}\sqrt{\frac{6}{n}} \le \gamma \le Z_{1-\alpha/2}\sqrt{\frac{6}{n}} \qquad (4\text{-}18)$$

where Z is from Table D1. If the estimated skewness exceeds this range, the data are not normally distributed. Typically, the sample size is much smaller than 150 and the estimated skewness should be compared to the values in Table 4-9. If the absolute value of the estimated

Both Remington and Schork (1970) and the SAS Institute (1985a) caution that the test for skewness is only a partial indicator of normality. With small samples (less than 25), the test is particularly unreliable. That is, because of the small sample size, very large departures from normality are required before statistical tests will reject the null hypothesis of normality. Cochran (1977) proposed a general rule for determining how large n must be (i.e., n in the equation below) to allow safe use of the normal approximation in computing confidence limits for the mean. This rule is used most effectively for distributions with positive skewness, which are most common for environmental data.

$$n' > 25\gamma_1 \qquad (4\text{-}19)$$

where γ_1

$$\gamma_1 = \frac{1}{ns^3}\sum_{i=1}^{n}(x_i-\bar{x})^3 \qquad (4\text{-}20)$$

Applying these equations to the data summarized in Table 4-10 yields a γ_1 of 0.99, and therefore more than 25 ($=25 \times 0.99^2$) samples are needed. The example data set contains 42 samples. Therefore, there are sufficient data to allow safe use of the normal approximation in computing confidence limits for the mean.

Kurtosis

The test for kurtosis is similar to the test for skewness since it measures only one attribute of normality and requires large samples for meaningful results. Remington and Schork (1970) recommend the following equation to evaluate kurtosis:

$$k_1 = \frac{1}{s\sqrt{n(n-1)}}\sum_{i=1}^{n}|x_i-\bar{x}| \qquad (4\text{-}21)$$

For any normally distributed population, k_1 would be 0.7979. Table 4-11 presents lower and upper limits for k_1.

If the calculated value of k_1 falls outside the values in Table 4-11 for the selected level of confidence, there is evidence of non-normal kurtosis. Using the same example data, k_1 can be computed as 0.80 and 0.82 for the raw and log-

Table 4-11. Values of kurtosis test for normality for small sample sizes.

	α = 0.02		α = 0.10	
n	Lower	Upper	Lower	Upper
11	0.6675	0.9359	0.7153	0.9073
21	0.6950	0.9001	0.7304	0.8768
31	0.7110	0.8827	0.7404	0.8625
41	0.7216	0.8722	0.7470	0.8540
51	0.7291	0.8648	0.7518	0.8481
61	0.7347	0.8592	0.7554	0.8434
71	0.7393	0.8549	0.7583	0.8403

After Remington and Schork, 1970.

transformed data, respectively. From this analysis, it is concluded that the raw and the log-transformed data have a kurtosis that is consistent with a normal distribution since k_1 is between the range of 0.7470 to 0.8540 for α equal to 0.10.

Shapiro-Wilk *W* test

The Shapiro-Wilk *W* test can be used to test the distribution of a data set for sample sizes of less than 2,000 (SAS Institute, Inc., 1990). This test uses the *W* statistic, which is "the ratio of the best estimator of the variance to the usual corrected sum of squares estimator of the variance" (SAS Institute, Inc., 1990). The null hypothesis for this test is that the data set is a random sample from a normal distribution. Values of *W* are greater than zero and less than or equal to one. The null hypothesis is rejected with small values. For sample sizes greater than 2,000, the Kolmogorov *D* statistic may be used (SAS Institute, Inc., 1990).

Anderson and McLean (1974) recommend the Shapiro-Wilk *W* test for normality and note that it is superior to the Kolmogorov-Smirnov and chi-squared tests in detecting non-normality over sample sizes ranging from 10 to 50. The following procedure for using the test is adapted from Anderson and McLean (1974) and Gilbert (1987):

1. Order the n observations as $x_1 \le x_2 \le \ldots \le x_n$.

2. Compute $d = (n\text{-}1)s^2$.

3. Compute k. If n is even, $k = n/2$. If n is odd, $k = (n\text{-}1)/2$.

4. Compute

$$W = \frac{1}{d}\left[\sum_{i=1}^{k} a_i (x_{n-i+1} - x_i)\right]^2 \qquad (4\text{-}22)$$

where the values of a_i appear in Table D4. The value x_{n-i+1} is equal to x_n when i is equal to 1 and x_{n-k+1} when i is equal to k.

5. Reject H_o (of normality) at the α significance level if *W* is less than the quantile given in Table D5.

Table 4-12 presents the sulfate data from Station 16 in Figure 4-8 in a format ready for analysis. The results for step 2 can be computed from the statistics in Table 4-10. Since there are 42 observations, k is equal to 21. The first column in Table 4-12 indicates the value of i for each row in the table. The second column corresponds to the values of a_i from Table D4. (Note that the values in Table D4 are for a_{n-i+1} and are exactly the same as a_i.) The third and fourth column, x_i and x_{n-i+1}, represent the raw sulfate data. The third column represents the first half of the observations, and the fourth column represents the last half of the data in reverse order (e.g., 460 is the largest sample observation). The fifth and sixth columns correspond to the log-transformed data from columns 3 and 4. For example, log(150) is equal to 5.01. The last two columns provide intermediate calculations associated with Equation 4-22 (i.e., $a_i(x_{n-i+1} - x_i)$) for the raw and log-transformed data, respectively.

Summing the last two columns results in completing the summation specified in Equation 4-22. The *W* statistic may now be computed using Equation 4-22 to yield 0.88 and 0.89 for the raw and log-transformed data, respectively. From Table D5, the quantile for 42 observations (95 percent confidence level) is 0.942. As a result, it can be concluded that the raw data and the log-transformed data are normally distributed.

4.4.2 Tests of Equal Variance

When performing hypothesis tests of two samples using parametric procedures, it is typically necessary to make sure that the two data sets have

Table 4-12. Example analysis of the Shapiro-Wilk *W* test using the sulfate data from Station 16 in Figure 4-8.

				log	log	Intermediate Calculations	
i	a_i	x_i	x_{n-i+1}	x_i	x_{n-i+1}		
1	0.3917	150	460	5.01	6.13	121.43	0.44
2	0.2701	150	430	5.01	6.06	75.63	0.28
3	0.2345	160	420	5.08	6.04	60.97	0.23
4	0.2085	170	400	5.14	5.99	47.96	0.18
5	0.1874	170	380	5.14	5.94	39.35	0.15
6	0.1694	180	360	5.19	5.89	30.49	0.12
7	0.1535	190	330	5.25	5.80	21.49	0.08
8	0.1392	200	320	5.30	5.77	16.70	0.07
9	0.1259	200	310	5.30	5.74	13.85	0.06
10	0.1136	200	310	5.30	5.74	12.50	0.05
11	0.1020	200	300	5.30	5.70	10.20	0.04
12	0.0909	200	300	5.30	5.70	9.09	0.04
13	0.0804	200	290	5.30	5.67	7.24	0.03
14	0.0701	210	270	5.35	5.60	4.21	0.02
15	0.0602	210	260	5.35	5.56	3.01	0.01
16	0.0506	210	250	5.35	5.52	2.02	0.01
17	0.0411	210	240	5.35	5.48	1.23	0.01
18	0.0318	210	240	5.35	5.48	0.95	0.00
19	0.0227	210	240	5.35	5.48	0.68	0.00
20	0.0136	220	220	5.39	5.39	0.00	0.00
21	0.0045	220	220	5.39	5.39	0.00	0.00
					SUM	479.00	1.81

the same variance. Testing for equal variances between two populations can be done by evaluating the ratio of the two sample variances (F_1) with the following equation:

$$F_1 = \frac{s_a^2}{s_b^2} \qquad (4\text{-}23)$$

where

$$s_a^2 \geq s_b^2$$

The null hypothesis in this test is that the variance ratio is equal to 1, and the alternative hypothesis is that the ratio is not equal to 1. The ratio is compared to a critical value from the F distribution (Table D6) that is based on the sample sizes (n_a and n_b) and the selected level of significance (α). Since the numerator is selected to be the variance with the larger value, it is necessary to look at only one critical value even though a two-sided test is being used.

For the sulfate data from Stations 16 and 17 in Figure 4-8, F_1 can be computed as 6,342.9/5,536.3 or 1.15 with 41 (42-1) and 10 (11-1) degrees of freedom. Using Table D6, the critical F value (for a two-sided 95 percent confidence level test where $\alpha/2$ is equal to 0.025) is approximately 3.25. Therefore, the null

hypothesis is accepted and it is concluded that the variances of the sulfate data from Stations 16 and 17 are the same.

4.4.3 Tests of Randomness

Another type of hypothesis testing involves time series at a single station. The DO data plotted in Figure 4-4 are one example. An approach to evaluate randomness is to compute the total number of runs (u) above and below the median (Freund, 1973). A run is a string of values all above or all below the median. A string of one value is acceptable. In this test, the median is determined, all values are placed in chronological order, and each value is assigned an "a" if it is above the median and a "b" if it is below the median. For example, the following is a set of data in chronological order:

5, 5, 6, 9, 13, 12, 2, 3, 2, 8, 14, 13, 11, 20, 4, 6, 9, 1, 7, 11, 12.

The median for this set of values (n=21) is 8. The series of values in terms of "a" and "b" is

b, b, b, a, a, a, b, b, b, omit, a, a, a, a, b, b, a, b, b, a, a

The number of runs (u) in the example data set is 8. Note that in this test all values equal to the median are omitted. Also, the number of values above (n_1) and below (n_2) must each be 10 or more to allow use of the following statistics. For n_1 and n_2 less than 10, special tables are required (Freund, 1973). The test statistic (derived from the normal distribution) is:

$$Z = \frac{u - \mu_u}{\sigma_u} \qquad (4\text{-}24)$$

where

$$\mu_u = \frac{2 n_1 n_2}{n_1 + n_2} + 1 \qquad (4\text{-}25)$$

and

$$\sigma_u = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} \qquad (4\text{-}26)$$

Applying these equations to the above example data,

$$\mu_u = \frac{2(10)(10)}{10 + 10} + 1 = 11$$

$$\sigma_u = \sqrt{\frac{2(10)(10)[2(10)(10) - (10) - (10)]}{(10 + 10)^2 (10 + 10 - 1)}} = 2.1764$$

$$Z = \frac{8 - 11}{2.1764} = -1.38$$

With α equal to 0.05 in a two-tailed test, the Z values (for $\alpha/2$) are 1.96 and -1.96 (Table D1). Since -1.38 falls within this range, the null hypothesis that the sample is random is accepted.

4.5 Evaluation of One or Two Independent Random Samples

The data collected for evaluating changes will typically come as (1) two or more sets of random samples or (2) a time series at a single station. In the first case, the analyst will test for a shift or step change (e.g., a significant difference between conditions before and after treatment). This might be typical for data collected from two stations along a stream segment. Or, when performing a biological assessment, for example, the goal might be to determine whether there is a

significant difference (i.e., a step change) between biological metrics for data collected at randomly selected reference and test (targeted) sites. It is also possible to compare a single random sample to a particular value. This might be the case when comparing data to a standard or reference condition. This section describes common approaches for comparing one or two independent random samples. Comparing more than two independent random samples or time series is discussed later.

Depending on the objective, it is appropriate to select a one- or two-sided test. For example, if the analyst knows that TSS would only decrease as a result of BMP implementation or is interested only if the TSS decreases, a one-sided test can be formulated. Alternatively, if the analyst does not know whether TSS will go up or down, a two-sided test is necessary. If the analyst simply wants to compare two random samples to decide if they are significantly different, a two-sided test can be used. Appropriate uses of a one-sided test include testing for decreased sediment or nutrient loads after implementing a flood control dam or best management practice, or comparing a suspected contaminated site to an upstream or control site. Typical null hypotheses (H_o) and alternative hypotheses (H_a) for one- and two-sided tests are provided below:

One-sided test

H_o: TSS (postimplementation) $\geq$ TSS (pre-implementation)

H_a: TSS (postimplementation) $<$ TSS (pre-implementation)

Two-sided test

H_o: TSS (postimplementation) $=$ TSS (pre-implementation)

H_a: TSS (postimplementation) $\neq$ TSS (pre-implementation)

Selecting a one-sided test instead of a two-sided test results in an increased power for the same significance level (Winer, 1971). That is, if the conditions are appropriate, a corresponding one-sided test is more desirable than a two-sided test given the same level of significance (α) and sample size. The manager and analyst should take great care in selecting one- or two-sided tests.

4.5.1 Tests for One Sample or Paired Data

Suppose the analyst is interested in evaluating compliance with a water quality standard or reference condition, e.g., a target determined from a load allocation or a percent substrate embeddedness less than the amount that hinders fisheries. In these situations the analyst might collect a random sample and compare it to a reference value. The Student's *t* and the Wilcoxon Signed Ranks tests are the two most appropriate tests when evaluating one *independent* random sample. The sign test can also be used, but it is generally limited to random samples that cannot be transformed into a symmetric distribution.

In addition, the analyst might be interested in determining whether a water quality variable increased between two sites located along a stream. In this situation the analyst might collect two random samples with matched or paired

Tests for One Sample or Paired Data

Test[a]	Additional Assumptions
Student's *t* (paired *t*)	Normal distribution
Wilcoxon Signed Ranks	Symmetric distribution
Sign	None

[a] The standard forms of these tests require independent random samples.

observations. Paired observations are a series of data collected as pairs at a given time or location. For example, if BOD_5 is sampled at two stream locations at a regular time interval, the result is a pair of BOD_5 observations for each time period. The same statistical tests used for one independent sample can be used to compare paired observations. The tests are adjusted by computing and analyzing the difference between the paired observations. The associated *t* test is referred to as the paired *t* test.

Student's *t* test

The participants in the Highland Silver Lake RCWP project (Jamieson, 1986) formulated a null hypothesis that a BMP would not reduce the post-implementation mean TSS concentrations to less than 25 mg/L. (Presumably, the participants hoped that the mean TSS concentration would be less than 25 mg/L so that H_0 could be rejected.) A formalized statement of the null and alternative hypotheses using a one-sided test would be:

H_o: $\mu \geq 25$ mg/L
H_a: $\mu < 25$ mg/L

In this case it is assumed that the mean TSS concentration is a good measure of central tendency and is the best measure for evaluation. It is also assumed that any change in TSS mean concentration is due to the BMP alone. H_a is stated such that a one-sided test can be applied because there is concern specifically about whether the postimplementation mean TSS concentration is lower than 25 mg/L since this might have been the target in a load allocation.

The Student's *t* test statistic (*t*) with *n*-1 degrees of freedom (df) can be used if the data are independent and normally distributed:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad (4\text{-}27)$$

Modification for Paired *t* Test

The sample mean, $\bar{x}$, and standard deviation, s, in Equation 4-27 refer to the mean and standard deviation of the differenced data (i.e., $d_i = x_i - y_i$).

The differenced data must be normally distributed.

The number of degrees of freedom is equal to the number of paired observations minus one.

where μ would be equal to the hypothesized value, 25 mg/L in this case. Assuming a one-sided test is used, the critical value for *t* would be obtained from Table D2 with *n*-1 degrees of freedom and a significance level of α. If a two-sided test were used (H_o: $\mu = 25$ mg/L; H_a: $\mu \neq 25$ mg/L), a value corresponding to a significance level of $\alpha/2$ would be obtained from Table D2.

The TSS data from the Highland Silver Lake RCWP project (Table 4-13) are from May 21, 1981, through October 31, 1984. The period after April 1, 1983, is the postimplementation period. Before testing H_o with a statistical test, the data must be inspected and the assumptions of randomness and normality must be tested. These tests are performed on the preimplementation and postimplementation data sets although only the postimplementation data in the current example are used. Using the SAS Univariate procedure (SAS Institute, Inc., 1985a), summary statistics and graphical presentations can be generated for the two data sets (Figures 4-12 and 4-13).

The values for skewness (0.82) and kurtosis (- 0.42) indicate positive skew and low kurtosis in the pre-BMP sample distribution. The Shapiro-Wilk *W* statistic (0.893) and associated probability

PRE-BMP IMPLEMENTATION DATA SET

Univariate Procedure

Variable=PRETSS

Moments

N	16	Sum Wgts	16
Mean	29.375	Sum	470
Std Dev	16.14879	Variance	260.7833
Skewness	0.817104	Kurtosis	-0.41702
USS	17718	CSS	3911.75
CV	54.9746	Std Mean	4.037197
T:Mean=0	7.276088	Pr>|T|	0.0001
Num ^= 0	16	Num > 0	16
M(Sign)	8	Pr>=|M|	0.0001
Sgn Rank	68	Pr>=|S|	0.0001
W:Normal	0.893033	Pr<W	0.0630

Quantiles(Def=4)

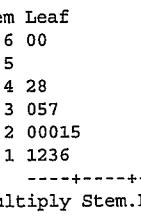
100% Max	60	99%	60
75% Q3	40.75	95%	60
50% Med	23	90%	60
25% Q1	17	10%	11.7
0% Min	11	5%	11
		1%	11
Range	49		
Q3-Q1	23.75		
Mode	20		

Extremes

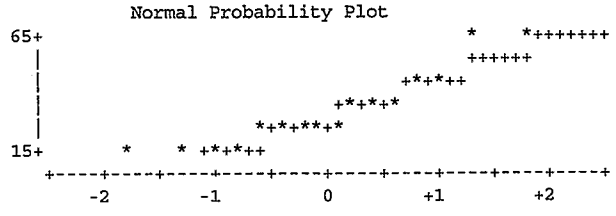
Lowest	Obs	Highest	Obs
11(	1)	37(	6)
12(	4)	42(	16)
13(	3)	48(	13)
16(	12)	60(	8)
20(	11)	60(	9)

```
     Stem Leaf                     #
        6 00                       2
        5                          
        4 28                       2
        3 057                      3
        2 00015                    5
        1 1236                     4
          ----+----+----+----+
     Multiply Stem.Leaf by 10**+1
```

Boxplot



Normal Probability Plot



Frequency Table

Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum
11	1	6.3	6.3	20	3	18.8	43.8	35	1	6.3	68.8	60	2	12.5	100.0
12	1	6.3	12.5	21	1	6.3	50.0	37	1	6.3	75.0				
13	1	6.3	18.8	25	1	6.3	56.3	42	1	6.3	81.3				
16	1	6.3	25.0	30	1	6.3	62.5	48	1	6.3	87.5				

Figure 4-12. Preimplementation data set.

POST-BMP IMPLEMENTATION DATA SET

Univariate Procedure

Variable=POSTTSS

Moments

N	15	Sum Wgts	15
Mean	19.86667	Sum	298
Std Dev	12.17061	Variance	148.1238
Skewness	0.697811	Kurtosis	-0.99379
USS	7994	CSS	2073.733
CV	61.26147	Std Mean	3.142439
T:Mean=0	6.322054	Pr>|T|	0.0001
Num ~= 0	15	Num > 0	15
M(Sign)	7.5	Pr>=|M|	0.0001
Sgn Rank	60	Pr>=|S|	0.0001
W:Normal	0.877487	Pr<W	0.0438

Quantiles(Def=4)

100% Max	42	99%	42
75% Q3	32	95%	42
50% Med	14	90%	40.8
25% Q1	10	10%	6.6
0% Min	6	5%	6
		1%	6
Range	36		
Q3-Q1	22		
Mode	10		

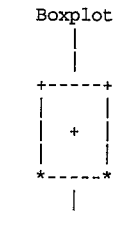
Extremes

Lowest	Obs	Highest	Obs
6(	2)	30(	3)
7(	4)	32(	5)
10(	13)	32(	7)
10(	8)	40(	11)
11(	15)	42(	9)

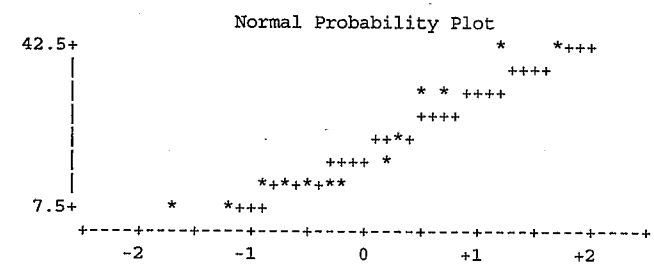
Stem	Leaf	#
4	02	2
3		
3	022	3
2		
2	2	1
1	6	1
1	001244	6
0	67	2

Multiply Stem.Leaf by 10**+1

Boxplot



Normal Probability Plot



Frequency Table

Value	Count	Percents Cell	Percents Cum
6	1	6.7	6.7
7	1	6.7	13.3
10	2	13.3	26.7
11	1	6.7	33.3
12	1	6.7	40.0
14	2	13.3	53.3
16	1	6.7	60.0
22	1	6.7	66.7
30	1	6.7	73.3
32	2	13.3	86.7
40	1	6.7	93.3
42	1	6.7	100.0

Figure 4-13. Postimplementation data set.

Table 4-13. Highland Silver Lake TSS data for site 1.

Preimplementation		Postimplementation	
Date	TSS (mg/L)	Date	TSS (mg/L)
5/21/81	11	4/12/83	14
6/18/81	21	5/10/83	6
7/30/81	13	6/7/83	30
9/3/81	12	7/27/83	7
10/6/81	30	10/5/83	32
11/5/81	37	11/18/83	14
12/8/81	20	12/29/83	32
4/8/82	60	1/25/84	10
4/27/82	60	2/20/84	42
5/25/82	20	4/11/84	22
6/22/82	20	5/15/84	40
7/20/82	16	7/17/84	12
9/20/82	48	8/21/84	10
10/26/82	35	9/26/84	16
11/23/82	25	10/31/84	11
12/2/82	42		

Overall:	n = 31 mean = 24.77	s = 14.93 median = 20
PreImplementation:	n = 16 mean = 29.38	s = 16.15 median = 23
PostImplementation:	n = 15 mean = 19.87	s = 12.17 median = 14

(0.063) show that the null hypothesis (that the sample is normally distributed) can be rejected with 93.7 percent confidence. In other words, there is only a 6.3 percent chance that a lower W value could be obtained if the sample were indeed taken from a normal distribution. Hence, the assumption of a normal distribution is rejected and the alternative hypothesis that the distribution is non-normal is accepted.

In the post-BMP sample distribution, the values for skewness (0.70) and kurtosis (-0.99) again indicate positive skew and low kurtosis. The Shapiro-Wilk W statistic (0.88) and associated probability (0.044) show that the null hypothesis (that the sample is from a normal distribution) can be rejected with 95.6 percent confidence. Also rejected is the assumption of a normal distribution for the post-BMP data set.

Taking the logarithm (base 10) of each data point for the pre-BMP and post-BMP data sets, the SAS Univariate procedure is run to see if the assumption of normality would be appropriate for the log-transformed data set. The output plots and statistics are shown in Figures 4-14 and 4-15. Note that the skewness (0.10) is much less pronounced, but the kurtosis (-1.09) is more negative for the transformed pre-BMP data set. The higher W statistic (0.951) and associated probability (0.493) indicate that the null hypothesis that the transformed data are normally distributed should be accepted.

For the log-transformed post-BMP data, the skewness (0.072) is also reduced and the kurtosis

PRE-BMP IMPLEMENTATION DATA SET (LOG TRANSFORMED)

Univariate Procedure

Variable=PRELTSS

Moments

N	16	Sum Wgts	16
Mean	1.407004	Sum	22.51207
Std Dev	0.238929	Variance	0.057087
Skewness	0.102176	Kurtosis	-1.08679
USS	32.53089	CSS	0.856304
CV	16.98138	Std Mean	0.059732
T:Mean=0	23.55521	Pr>|T|	0.0001
Num ~= 0	16	Num > 0	16
M(Sign)	8	Pr>=|M|	0.0001
Sgn Rank	68	Pr>=|S|	0.0001
W:Normal	0.951026	Pr<W	0.4929

Quantiles(Def=4)

100% Max	1.778151	99%	1.778151
75% Q3	1.609487	95%	1.778151
50% Med	1.36008	90%	1.778151
25% Q1	1.228347	10%	1.067845
0% Min	1.041393	5%	1.041393
		1%	1.041393
Range	0.736759		
Q3-Q1	0.38114		
Mode	1.30103		

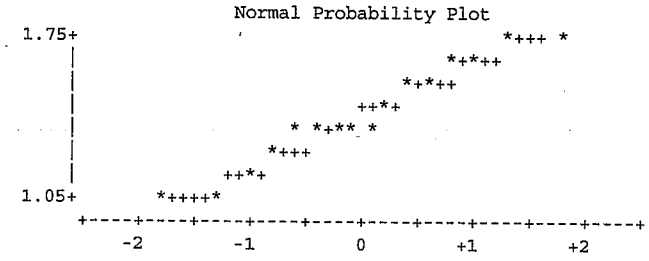
Extremes

Lowest	Obs	Highest	Obs
1.041393(	1)	1.568202(	6)
1.079181(	4)	1.623249(	16)
1.113943(	3)	1.681241(	13)
1.20412(	12)	1.778151(	8)
1.30103(	11)	1.778151(	9)

Stem Leaf # Boxplot

```
17 88      2
16 28      2
15 47      2
14 08      2
13 0002    4
12 0       1
11 1       1
10 48      2
   ----+----+----+----+
```

Multiply Stem.Leaf by 10**-1



Frequency Table

Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum
1.041393	1	6.3	6.3	1.30103	3	18.8	43.8	1.544068	1	6.3	68.8	1.778151	2	12.5	100.0
1.079181	1	6.3	12.5	1.322219	1	6.3	50.0	1.568202	1	6.3	75.0				
1.113943	1	6.3	18.8	1.39794	1	6.3	56.3	1.623249	1	6.3	81.3				
1.20412	1	6.3	25.0	1.477121	1	6.3	62.5	1.681241	1	6.3	87.5				

Figure 4-14. Log-transformed preimplementation data set.

POST-BMP IMPLEMENTATION DATA SET (LOG TRANSFORMED)

Univariate Procedure

Variable=POSTLTSS

Moments

N	15	Sum Wgts	15
Mean	1.21969	Sum	18.29535
Std Dev	0.273571	Variance	0.074841
Skewness	0.072331	Kurtosis	-1.22633
USS	23.36244	CSS	1.047777
CV	22.42956	Std Mean	0.070636
T:Mean=0	17.26731	Pr>|T|	0.0001
Num ^= 0	15	Num > 0	15
M(Sign)	7.5	Pr>=|M|	0.0001
Sgn Rank	60	Pr>=|S|	0.0001
W:Normal	0.939469	Pr<W	0.3661

Quantiles(Def=4)

100% Max	1.623249	99%	1.623249
75% Q3	1.50515	95%	1.623249
50% Med	1.146128	90%	1.610536
25% Q1	1	10%	0.818319
0% Min	0.778151	5%	0.778151
		1%	0.778151
Range	0.845098		
Q3-Q1	0.50515		
Mode	1		

Extremes

Lowest	Obs	Highest	Obs
0.778151(	2)	1.477121(	3)
0.845098(	4)	1.50515(	5)
1(	13)	1.50515(	7)
1(	8)	1.60206(	11)
1.041393(	15)	1.623249(	9)

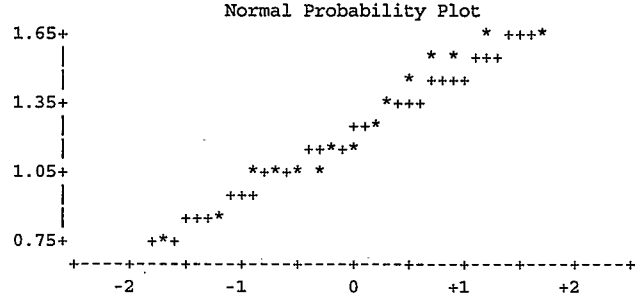
Stem	Leaf	#
16	02	2
15	11	2
14	8	1
13	4	1
12	0	1
11	55	2
10	0048	4
9		
8	5	1
7	8	1

----+----+----+----+

Multiply Stem.Leaf by 10**-1

Boxplot

Normal Probability Plot



Frequency Table

Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum	Value	Count	Percents Cell	Percents Cum
0.778151	1	6.7	6.7	1.041393	1	6.7	33.3	1.20412	1	6.7	60.0	1.50515	2	13.3	86.7
0.845098	1	6.7	13.3	1.079181	1	6.7	40.0	1.342423	1	6.7	66.7	1.60206	1	6.7	93.3
1	2	13.3	26.7	1.146128	2	13.3	53.3	1.477121	1	6.7	73.3	1.623249	1	6.7	100.0

Figure 4-15. Log-transformed postimplementation data set.

(-1.23) is more negative than for the raw data set. The *W* statistic (0.939) and associated probability (0.367) indicate that the null hypothesis that the transformed data are normally distributed should be accepted. In fact, there is a 63.3 percent probability that a lower *W* statistic could be obtained if the sample is from a normal distribution.

To test the randomness of the data sets, the test described in Section 4.4.3 can be used. Since the test requires only the number of runs and the number of values above and below the median, it does not matter whether the raw data or transformed data are used. Using the raw data in Table 4-13, the number of runs for the pre-implementation data set is 6 while the number for the postimplementation data set is 9. The resulting *z* statistics (from Equation 4-24) for the preimplementation and postimplementation data sets are 1.5526 and 0.8971, respectively. These values are compared to a critical value of 1.96 (using $\alpha/2 = 0.025$) from Table D1 and the null hypothesis is accepted. Both samples are random.

Once the data sets are randomly sampled and normally distributed (after log-transformations), the one-sample hypothesis test using the log-transformed post-BMP data set can be performed. As shown in Figure 4-15, the mean of the log-transformed post-BMP data set is 1.21969 and the standard deviation is 0.273571. The log of the hypothesized value (25 mg/L) is 1.3979. Note that it is recommended that these values be rounded to the correct number of significant digits when reporting the results. The *t* statistic (Equation 4-27) is used to determine whether the post-BMP mean TSS concentration is less than 25 mg/L.

$$t = \frac{1.21929 - 1.3979}{0.273571 / \sqrt{15}} = -2.53$$

The schematic representation of this test is shown in Figure 4-16A, where the critical *t* value (-1.761) for the one-sided test (df = 14, α = 0.05) is taken from Table D2. The computed *t* statistic falls to the left of the critical value, so the null hypothesis is rejected. In turn, the alternative hypothesis that the post-BMP mean TSS concentration is less than 25 mg/L is accepted.

Alternatively, had the participants in the Highland Silver Lake RCWP project selected a two-sided test where H_o and H_a are given as

H_o: $\mu = 25$ mg/L
H_a: $\mu \neq 25$ mg/L

a two-sided *t* test would be appropriate. The critical *t* value for the two-sided test from Table D2 (df = 14, $\alpha/2 = 0.025$) would be $\pm$ 2.145. In this case, the computed *t* statistic (-2.52) still falls outside this range and it is concluded that the post-BMP mean TSS concentration is less than 25 mg/L. Notice how the rejection region (shaded portion) in Figure 4-16B differs from Figure 4-16A. The total shaded area in the two curves is the same (i.e., 5 percent); however, it is in one piece in Figure 4-16A and is split into two parts in Figure 4-16B.

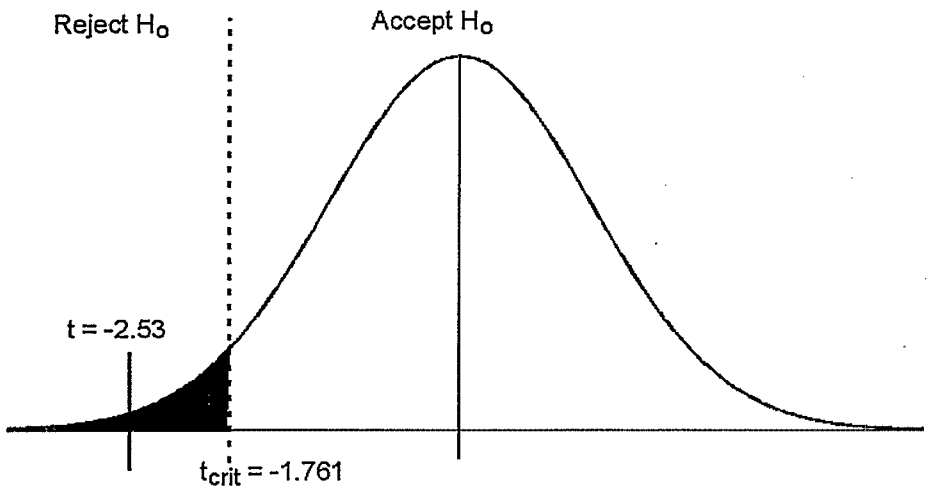
The power of this test can be evaluated using the noncentral *t* distribution with respect to various alternative hypotheses. The noncentral *T* statistic with *n*-1 degrees of freedom is given by

$$T_\Delta = \frac{\bar{x} - \mu_1 + \Delta}{s/\sqrt{n}} = \frac{\bar{x} - u_0}{s/\sqrt{n}} \qquad (4\text{-}28)$$

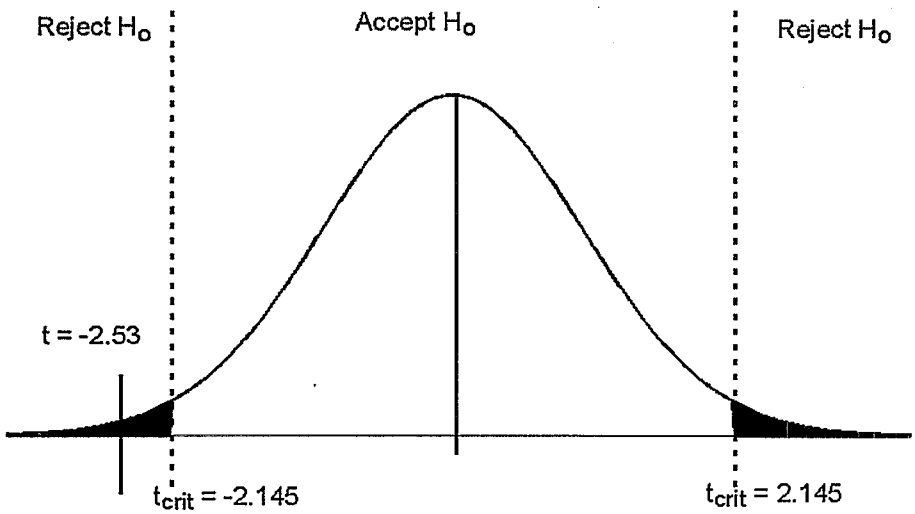
where $\Delta = \mu_1 - \mu_0$, the difference between the real and hypothesized mean. The noncentrality parameter (δ) is given by

$$\delta = \frac{\sqrt{n}\Delta}{s} \qquad (4\text{-}29)$$

Values of δ are given in Table D7 for a one-sided noncentral *t* distribution. Continuing with the



A) One-sided *t*-test with 14 degrees of freedom and α equal to 0.5.



B) Two-sided *t*-test with 14 degrees of freedom and α equal to 0.5.

Figure 4-16. One- and two-sided *t* test for post-BMP mean TSS concentration.

current example, it is possible to develop a power curve that indicates the trade-offs between Type I and II errors. (Background discussion on power curves is provided in Section 4.1.1.) From Table D7 (df = 14, α = 0.05), one value of δ is obtained for each level of β (Table 4-14). In Table 4-14, power is computed as 1-β and Δ is obtained by rearranging Equation 4-29 and using *s* equal to 0.273571 and *n* equal to 15. Note that Δ, referred to as the minimum detectable difference, is in log-transformed units.

Power can be plotted as a function of the minimum detectable difference (see Figure 4-17). The dotted line indicates an approximate extrapolation back to α when the minimum detectable difference is equal to zero. Using the log-transformed postimplementation data, Δ is equal to 0.178 (= 1.3979 - 1.21969). Interpolating from Table 4-14 or Figure 4-17 yields that there is a 77 percent probability (i.e., power = 0.77) that a significant difference would be detected (i.e., reject H_o) if the difference between the estimated mean and true mean using log-transformed data were 0.178. For Δ less than 0.027, there is only a 10 percent chance of detecting a significant difference, whereas for Δ greater than 0.3 there is almost a 100 percent chance of detecting a significant difference.

Table 4-14. Evaluation of power using the post-implementation TSS data.

Power (1-β)	β	δ	Δ
0.10	0.90	0.38	0.027
0.20	0.80	0.84	0.059
0.30	0.70	1.18	0.083
0.40	0.60	1.46	0.103
0.50	0.50	1.73	0.122
0.60	0.40	2.00	0.141
0.70	0.30	2.28	0.161
0.80	0.20	2.62	0.185
0.90	0.10	3.08	0.218
0.95	0.05	3.46	0.244
0.99	0.01	4.18	0.295

Wilcoxon Signed Ranks test

Alternatively, if the log (or some other) transformation did not result in normally

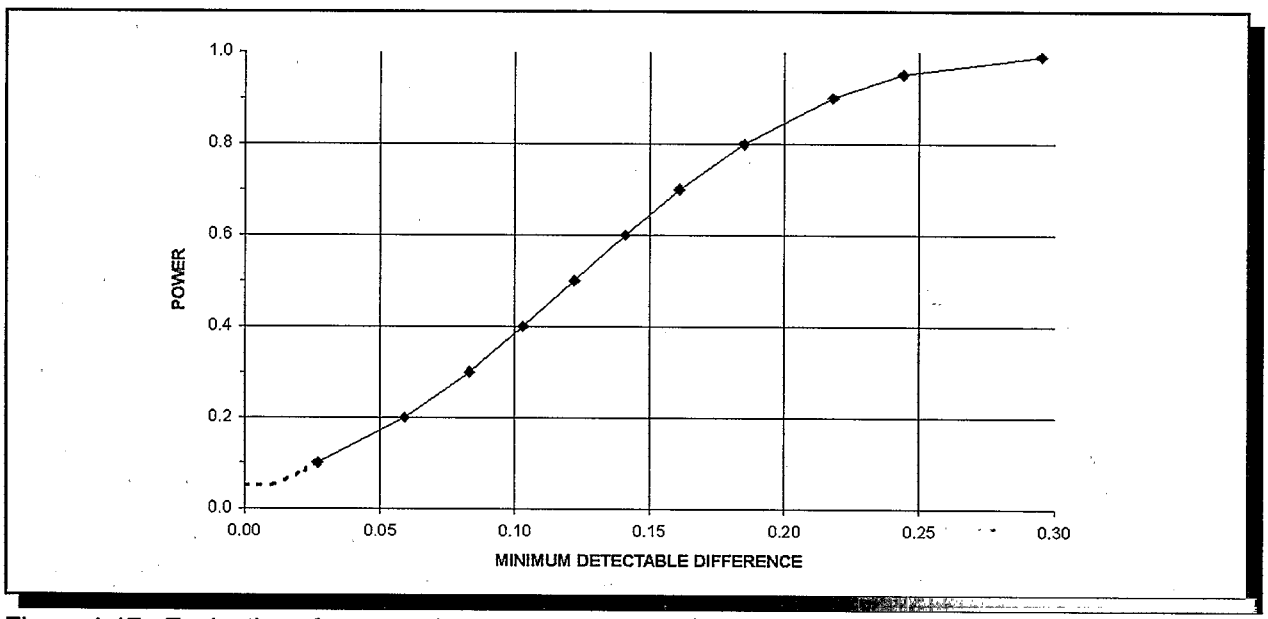


Figure 4-17. Evaluation of power using the log-transformed postimplementation TSS data.

distributed data, the analyst could consider the Wilcoxon Signed Ranks test. Although less restrictive than the *t* test, this test requires that the data are independent and come from a symmetric distribution. As the name implies, a symmetric distribution is one in which the distribution of data above the midpoint is a mirror image of the data distribution below the midpoint. (The normal distribution is a special case of a symmetric distribution.) When the data distribution is symmetric, the mean and median coincide and therefore inferences about the median are also valid for the mean (Conover, 1980). For this presentation, the median concentration is evaluated rather than the mean using the following hypotheses:

H_o: $P_{.50} \geq 25$ or H_o: $25\text{-}P_{.50} \leq 0$
H_a: $P_{.50} < 25$ H_a: $25\text{-}P_{.50} > 0$

The test statistic, T, is normally distributed and is given by Conover (1980) as

$$T = \frac{\sum_{i=1}^{n} Rank\,|d_i|}{\left[\sum_{i=1}^{n} Rank\,|d_i|^2\right]^{0.5}} \qquad (4\text{-}30)$$

where d_i is equal to the difference between the hypothesized value (25 mg/L) and the actual data and the rank is assigned a negative value if d_i is negative. El-Shaarawi and Damsleth (1988) provide a modified version of the Wilcoxon Signed Ranks test for use with serially correlated data.

Modification for Paired Data

d_i corresponds to the difference between paired observations (e.g., $d_i = x_i - y_i$).

From the previous example, it is already known that the raw postimplementation data are lognormal and thus not symmetric. Therefore, the log-transformed data are analyzed since it has already been determined that the log-transformed observations are symmetric as well as independent. Table 4-15 shows the calculations used to evaluate the log-transformed post-implementation data set. For convenience the data are sorted from smallest to largest observation. The difference, d_i, is computed as log(25) - log(TSS_i). For example, the first entry is equal to log(25) - log(6) or 0.620. Since the log-transformed data were symmetric, d_i will also be symmetric. The fourth column is the absolute value of the difference, $|d_i|$. The last two columns are the rank and rank-squared of $|d_i|$ where the rank is assigned a negative value if d_i is negative. T is equal to $76/(1238.5)^{0.5}$ or 2.16. Since 2.16 is greater than 1.645 (which is obtained from Table D1 using $\alpha = 0.05$), the null hypothesis is rejected and it is concluded that the median concentration is less than 25 mg/L. Had the raw data that are not symmetric been incorrectly used, T would have been equal to 1.54 and the null hypothesis would have been incorrectly accepted.

Sign test

Suppose that the postimplementation data could not be transformed into a symmetric distribution. By using the sign test, the symmetric distribution assumption can be relaxed (i.e., it is not required). In this case, the appropriate hypotheses for a one-sided test are

H_o: $P(+) \geq P(-)$
H_a: $P(+) < P(-)$

where P(+) is defined as the probability of an observation's being greater than the hypothesized value (in this case 25 mg/L). As stated, H_o implies that 50 percent or more of the population is greater than or equal to the hypothesized value.

Table 4-15. Nonparametric evaluation of postimplementation data using the Wilcoxon Signed Ranks test.

TSS (mg/L)	Log(TSS)	d_i=Log(25) -Log(TSS$_i$)	$\lvert d_i \rvert$	rank $\lvert d_i \rvert$ [a]	rank $\lvert d_i \rvert^2$
6	0.778	0.620	0.620	15	225
7	0.845	0.553	0.553	14	196
10	1.000	0.398	0.398	12.5	156.25
10	1.000	0.398	0.398	12.5	156.25
11	1.041	0.357	0.357	11	121
12	1.079	0.319	0.319	10	100
14	1.146	0.252	0.252	8.5	72.25
14	1.146	0.252	0.252	8.5	72.25
16	1.204	0.194	0.194	5	25
22	1.342	0.056	0.056	1	1
30	1.477	-0.079	0.079	-2	4
32	1.505	-0.107	0.107	-3.5	12.25
32	1.505	-0.107	0.107	-3.5	12.25
40	1.602	-0.204	0.204	-6	36
42	1.623	-0.225	0.225	-7	49
			SUM =	76	1238.5

[a] Assign the negative of the rank if d_i is negative.

Modification for Paired Data

The comparison is made between the paired observations rather than with a hypothesized value.

By comparing each observation from the random sample to the hypothesized value, the data set is converted into a series of "+," "-," and ties. The test statistic, T, is equal to the number of "+." The more "+" that result from the comparisons, the more H_o is supported.

Using the raw postimplementation data, T is equal to 5 and n is equal to 15. There are no ties. In this one-sided test, small values of T indicate that "-" are more probable. For sample sizes less than 20, use Table D8 with p equal to 0.5 and n equal to the number of "+" and "-" (ties are excluded). Find the table entry, y, that approximately equals α, rejecting H_o if $T \leq y$. If n is greater than 20, y can be computed as

$$y = \frac{1}{2}(n + Z_\alpha \sqrt{n}) \qquad (4\text{-}31)$$

were Z_α is obtained from Table D1. For example, if α is equal to 0.05 in a one-sided test, $Z_{0.05}$ is equal to -1.645. Using the example data, a y equal to 4 (α=0.0592) is obtained from Table D7. T is greater than 4, so H_o is accepted.

Had the hypotheses been stated in the other direction (i.e., H_o: P(+) ≤ P(-); H_a: P(+) > P(-)), H_o would be rejected if $T \geq n - y$. Had this been a two-sided test, the rejection region would

be for $T \leq y$ or $T \geq n - y$ where y is obtained from Table D8 or Equation 4-31 using $\alpha/2$.

Table 4-16 presents paired observations for BOD_5 collected at two locations from the same stream. In this case, the hypothesis that there is no difference in BOD_5 concentrations between the two locations with a $\alpha = 0.10$ is being tested:

Hypotheses	Description
H_o: P(+) = P(-)	BOD_5 concentrations at the two locations are the same.
H_a: P(+) ≠ P(-)	BOD_5 concentrations at location 1 tends to be larger or smaller than the BOD_5 concentration at location 2.

In this case, a two-sided test is appropriate where P(+) indicates the probability that an observation from location 1 is greater than an observation from location 2. The fourth column indicates whether the BOD_5 concentration at location 1 is larger (+), smaller (-), or equal to (tie) the BOD_5 concentration at location 2. In this analysis there are 8 "+" and a total of 13 observation pairs without ties. From Table D8 with $\alpha/2 = 0.05$ and $n = 13$, $y = 3$ ($\alpha = 0.0461$) is obtained. H_o is accepted since $3 \leq 8 \leq (13-3)$.

Comparison of example results

In this case, the Student's t test and Wilcoxon Signed Ranks test give the same conclusion. It is proposed that the results from the t test are more appropriate for this example since all of the assumptions of the parametric test were met. Had the assumptions not been met, the results from the Wilcoxon Signed Ranks test would have been more appropriate. That is, if all assumptions are met, parametric procedures are more powerful than their nonparametric alternative. The sign test, while not incorrect, was not a good choice for the example data because the distributional assumptions were met and more powerful tests could be applied. Applying the Wilcoxon Signed Ranks test to data that are not symmetric results in a level of significance (α) that is somewhat lower than what is specified, whereas applying

Table 4-16. Sign test for comparing paired BOD_5 concentrations.

Day	Conc. at Location 1 (mg/L)	Conc. at Location 2 (mg/L)	Sign of Difference
1	29	19	+
2	22	20	+
3	10	5	+
4	26	24	+
5	12	15	-
6	32	24	+
7	23	25	-
8	11	23	-
9	32	32	tie
10	27	30	-
11	28	20	+
12	23	16	+
13	18	33	-
14	35	25	+
15	20	20	tie

the *t* test to data that are not normally distributed results in an α that is much larger than specified (Helsel and Hirsch, 1995).

4.5.2 Two-sample Tests

In many instances, paired observations are not a practical or appropriate sampling methodology. Instead, two random samples are collected. The pre- and postimplementation data in Table 4-13 from the Highland Silver Lake RCWP are one example. The Student's *t* test for two samples and the Mann-Whitney test are the most appropriate tests for these types of data.

Two-sample *t* test

Suppose that a comparison of the pre- and post-implementation TSS data sets is desired to see if the BMPs have had an effect on TSS levels in Highland Silver Lake. Remembering the assumptions made earlier about using the mean TSS concentration as a good measure of central tendency and assuming that any change in TSS mean concentration is due to the BMP alone, the pre- and postimplementation data sets can be used in a one-sided hypothesis:

H_o: TSS (Post) $\geq$ TSS (Pre) or
H_o: TSS (Post) - TSS (Pre) ≥ 0

H_a: TSS (Post) $<$ TSS (Pre) or
H_a: TSS (Post) - TSS (Pre) < 0

Note that in this case the H_o that the post-implementation TSS is greater than or equal to the preimplementation TSS concentration is tested with an H_a that postimplementation TSS is lower. The results from this analysis will be interpreted as simply indicating whether the BMPs worked. This could also have been set up as a two-sided test where H_o and H_a would be

H_o: TSS (Post) = TSS (Pre) or
H_o: TSS (Post) - TSS (Pre) = 0

H_a: TSS (Post) $\neq$ TSS (Pre) or
H_a: TSS (Post) - TSS (Pre) $\neq 0$

With confidence that the BMP would have only an effect of reducing TSS concentrations, H_o is tested using a one-sided *t* test. Both the pre-implementation and postimplementation data sets are random samples and normal when log-transformed. However, the two-sample *t* test also requires that the variances of the two populations be equal (Gaugush, 1986). Since a major effect of many nonpoint source control practices is to reduce the occurrence of large loading events, it is very likely that these practices will have an effect on the variance of nonpoint source loads. Thus, an *F* test is performed to evaluate variance homogeneity before proceeding with the *t* test even though the *t* test is robust with respect to moderate departures from homogeneous variance (Winer, 1971).

Since the log-transformed data (Figures 4-14 and 4-15) are being used, the variance of the transformed data must also be used in the *F* test. The resulting *F* statistic is computed from Equation 4-23:

Tests for Two Independent Random Samples

Test[a]	Key Assumptions
Two-Sample *t*	• Both data sets must be normally distributed • Data sets should equal variances[b]
Mann-Whitney	• None

[a] The standard form of these tests requires independent random samples.
[b] The variance homogeneity assumption can be relaxed (see Table 4-17).

$F_1 = 0.075/0.057 = 1.32$

The variances are substituted into Equation 4-23 so that the F statistic is greater than unity to account for the organization of Table D6. The critical F value from Table D6 ($f_n = 14, f_d = 15$, $\alpha/2 = 0.025$) is 2.89. The value 1.32 is compared to 2.89, and the null hypothesis of equal variance is accepted.

Satisfied that the data meet all of the assumptions required of the two-sample hypothesis test, H_o (TSS (Post) ≥ TSS (Pre)) is now tested. The two-sample t statistic with n_1+n_2-2 degrees of freedom is (Remington and Schork, 1970)

$$t = \frac{(\overline{x_1} - \overline{x_2}) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (4\text{-}32)$$

where s_p is the pooled standard deviation, which is defined by

$$s_p = \left[\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}\right]^{0.5} \qquad (4\text{-}33)$$

The difference quantity (Δ_o) can be any value, but in this case it is set to zero. Δ_o can be set to a non-zero value to test whether the difference between the two data sets is greater than a selected value. Using the transformed data for preimplementation ($n_1 = 16$, $s_1^2 = 0.057087$, $\overline{x}_1 = 1.407$) and postimplementation conditions ($n_2 = 15$, $s_2^2 = 0.074812$, $\overline{x}_2 = 1.21969$), s_p is calculated as

$$s_p = \left[\frac{0.057087(16-1) + 0.074812(15-1)}{16 + 15 - 2}\right]^{0.5} = 0.2562$$

and the t statistic is calculated as

$$t = \frac{(1.407 - 1.21969) - 0}{0.2562 \sqrt{\frac{1}{16} + \frac{1}{15}}} = 2.034$$

Comparing this t statistic in a one-tailed test to the t value from Table D2 ($\alpha = 0.05$, df $= n_1 + n_2 - 2 = 29$), it is found that the 2.034 exceeds the table value of 1.6991. Therefore, the null hypothesis is rejected and it is concluded that the postimplementation mean log-transformed TSS concentration is lower than the preimplementation level (i.e., the BMPs worked given earlier assumptions). Note that if a two-tailed test had been used, the null hypothesis would have been accepted since the corresponding t value from Table D2 is 2.0452. Remington and Schork (1970) give test statistics for other cases in which the difference between means is being tested. These cases and corresponding equations are given in Table 4-17. In particular, note Case #3, which allows for unequal variances.

The power of this test can be estimated using the noncentrality parameter (Larsen and Marx, 1981):

$$\delta = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{2}\,\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (4\text{-}34)$$

where σ is approximated with the pooled standard deviation. Using the data in this example,

$$\delta = \frac{1.407 - 1.21969}{\sqrt{2}\ 0.2562 \sqrt{\frac{1}{16} + \frac{1}{15}}} = 1.438$$

From Table D7 (df = 29, $\alpha = 0.05$), a β approximately equal to 0.60 is obtained, so the power is equal to 0.40. Had the difference in

Table 4-17. Summary of parametric tests used to evaluate difference between means (Remington and Schork, 1970).

Null Hypothesis	Test Statistic	Assumptions
Case 1: Difference between means when variances are known (test statistic is standard normal distribution)		
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{(\sigma_1^2/n_1 + \sigma_2^2/n_2)^{0.5}}$	Independent, random samples of size n_1 and n_2 from two normally distributed populations.
Case 2: Difference between means when variances are unknown but equal (test statistic is Student's t distribution with n_1+n_2-2 degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{s_p(1/n_1 + 1/n_2)^{0.5}}$	Independent, random samples of size n_1 and n_2 from two normally distributed populations with equal variances
Case 3: Difference between means when variances are known and unequal (test statistic is approximately Student's t; see below for degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{(s_1^2/n_1 + s_2^2/n_2)^{0.5}}$ $df^* = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 + 1} + \dfrac{(s_2^2/n_2)^2}{n_2 + 1}}$ $df = df^* - 2$	Independent, random samples of size n_1 and n_2 from two normally distributed populations with unknown and presumably unequal variances
Case 4: Pairing—the mean difference (test statistic is Student's t distribution with n-1 degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_d = \Delta_o$ $\mu_d = \mu_2 - \mu_1$	$t = \dfrac{\overline{d} - \Delta_o}{s_d/n^{0.5}}$	Random sample of size n paired differences from a normally distributed populations of differences

means (of the log-transformed data) been larger (e.g., 0.30), δ would be 2.31 and the power would be equal to 73 percent.

Mann-Whitney (Wilcoxon's rank sum) test

The Mann-Whitney test can also be used to compare two independent random samples. This test is very flexible since there are no assumptions about the distribution of either sample or whether the distributions have to be the same (Helsel and Hirsch, 1995). Wilcoxon (1945) first introduced this test for equal-sized samples. Mann and Whitney (1947) modified the original Wilcoxon's test to apply it to different sample sizes. This test tests whether one data set tends to have larger observations than the other. Example two- and one-sided hypotheses are as follows:

Two-sided

H_o: Prob [TSS (Post) > TSS (Pre)] = 0.5

Description: The probability that the post-implementation TSS is larger than the pre-implementation TSS is equal to 50 percent.

H_a: Prob [TSS (Post) > TSS (Pre)] $\neq$ 0.5

Description: The postimplementation TSS is larger or smaller than the preimplementation TSS.

One-sided

H_o: Prob [TSS (Post) > TSS (Pre)] $\geq$ 0.5

Description: The probability that the post-implementation TSS is larger than the pre-implementation TSS is equal to or greater than 50 percent.

H_a: Prob [TSS (Post) > TSS (Pre)] < 0.5

Description: The postimplementation TSS is smaller than the preimplementation TSS.

If the distributions of the two samples are similar except for location (i.e., similar spread and skew), H_a can be refined to imply that the median concentration from one sample is "greater than," "less than," or "not equal to" the median concentration from the second sample. To achieve this greater detail in H_a, transformations such as logs can be used.

Table 4-18 shows the intermediate calculations using the same TSS data presented earlier. First, all observations from the pre- and post-implementation are sorted together and ranks assigned. Note that ties are assigned the average rank. The test statistic is equal to the sum of the ranks for the group with the smaller number of observations—in this case, the postimplementation data set.

Tables of Mann-Whitney test statistics (e.g., Conover, 1980) may be consulted to determine whether to reject H_o for small sample sizes. If n_1 and n_2 are greater than or equal to 10 observations, the test statistic can be computed from the following equation (Conover, 1980):

$$T_1 = \frac{T - n_1\frac{N+1}{2}}{\sqrt{\frac{n_1 n_2}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \frac{n_1 n_2 (N+1)^2}{4(N-1)}}} \quad (4\text{-}35)$$

where

n_1 = number of observations in sample with fewer observations (e.g., post-implementation);
n_2 = number of observations in sample with more observations (e.g., pre-implementation);
N = $n_1 + n_2$;

Table 4-18. Nonparametric evaluation of post-implementation data using the Mann-Whitney test.

Rank	Pre-Implement. TSS (mg/L)	Post-Implement. TSS (mg/L)	Rank	Pre-Implement. TSS (mg/L)	Post-Implement. TSS (mg/L)
1	-	6	17	21	-
2	-	7	18	-	22
3.5	-	10	19	25	-
3.5	-	10	20.5	30	-
5.5	11	-	20.5	-	30
5.5	-	11	22.5	-	32
7.5	12	-	22.5	-	32
7.5	-	12	24	35	-
9	13	-	25	37	-
10.5	-	14	26	-	40
10.5	-	14	27.5	42	-
12.5	16	-	27.5	-	42
12.5	-	16	29	48	-
15	20	-	30.5	60	-
15	20	-	30.5	60	-
15	20	-			
Sum of ranks for post-implementation, T = 193.5 Sum of all ranks squared, ΣR_i^2 = 10,409.5					

T = sum of ranks for sample with fewer observations; and

R_i = rank for the ith ordered observation used in both samples.

This equation is appropriate for situations when there are many ties. Applying this equation yields

$$T_1 = \frac{193.5 - 15\frac{31+1}{2}}{\sqrt{\frac{15 \cdot 16}{31(31-1)} 10409.5 - \frac{15 \cdot 16(31+1)^2}{4(31-1)}}} = -1.84$$

T_1 is normally distributed, and Table D1 can be used to determine the appropriate quantile. Since the test was one-sided and α is equal to 0.05, the appropriate quantile from Table D1 is -1.645. T_1 is less than -1.645, and therefore the null hypothesis is rejected. The post-implementation TSS concentrations are significantly less than the pre-implementation TSS concentrations. Had a two-sided test been used, the appropriate quantile from Table D1 would have been -1.96 and the H_o would have been accepted. In this case, the two-sample t test and the Mann-Whitney test result in the same conclusion.

4.5.3 Magnitude of Differences

So far, Section 4.5 has described statistical tests for comparing one and two random samples for significant differences. A question remains: How big is the difference? For data that are normally distributed, the difference can be computed as the difference between the two sample means. The confidence interval (CI) for

the differences can be computed under the equal variance scenario as (Winer, 1971):

$$CI = (\overline{x_1} - \overline{x_2}) \pm t_{(\alpha/2, n_1+n_2-2)}\, s_p \sqrt{1/n_1 + 1/n_2} \quad (4\text{-}36)$$

If the standard deviations were not similar, the CI would be

$$CI = (\overline{x_1} - \overline{x_2}) \pm t_{(\alpha/2, df)} \sqrt{s_1^2/n_1 + s_2^2/n_2} \quad (4\text{-}37)$$

where df is from Table 4-17 (Case 3).

Helsel and Hirsch (1995) recommend that a Hodges-Lehmann estimator ($\hat{\Delta}$) be used if the data have been transformed for testing or if the data are not normally distributed. The Hodges-Lehmann estimator (Hodges and Lehmann, 1963) can be used as a nonparametric estimator of the difference between the two samples. To compute the Hodges-Lehman estimate, the analyst computes the difference between all n_1 and n_2 observations. Using the TSS data used earlier, there are 16•15 or 240 differences to compute. The Hodges-Lehmann estimator is the median of these differences or 8 mg/L. This estimator is preferred to the difference between the medians of the random samples (Helsel and Hirsch, 1995). For sample sizes larger than 10, the upper and lower confidence intervals for $\hat{\Delta}$ can be estimated:

$$R_l = \frac{n_1 n_2 - z_{\alpha/2}\sqrt{\dfrac{n_1 n_2 (n_1+n_2+1)}{3}}}{2} \quad (4\text{-}38)$$

$$R_u = \frac{n_1 n_2 + z_{\alpha/2}\sqrt{\dfrac{n_1 n_2 (n_1+n_2+1)}{3}}}{2} + 1 = n_1 n_2 - R_l + 1 \quad (4\text{-}39)$$

where R_l and R_u correspond to the l^{th} and u^{th} ranked difference. The 95 percent confidence interval for the difference between the pre- and post-implementation data would be computed as

$$R_l = \frac{16 \cdot 15 - 1.96\sqrt{\dfrac{16 \cdot 15(16+15+1)}{3}}}{2} = 70.4 \approx 70$$

$$R_u = 16 \cdot 15 - 70.4 + 1 = 170.6 \approx 171$$

Therefore, the confidence interval on the median difference is equal to the 70^{th} and 171^{st} ranked difference or $-1 \le \hat{\Delta} \le 19$.

4.6 Comparison of More Than Two Independent Random Samples

The analysis of variance (ANOVA) and Kruskal-Wallis are extensions of the two-sample *t* and Mann-Whitney tests, respectively, and can be used for analyzing more than two independent random samples. Unlike the *t* test described earlier, the ANOVA can have more than one factor or explanatory variable. In the Highland Silver Lake RCWP project example used in Section 4.5, one factor described whether the data were collected before or after implementation of a BMP. In the example that will be analyzed in this section, trout population, there are two factors. One factor is based on the stream from which the trout were collected; the other factor is based on the region from which the trout were collected. The Kruskal-Wallis test accommodates only one factor, whereas the Friedman test can be used for two factors. In addition to applying one of the above tests to determine whether one of the samples is significantly different from the others, it is also necessary to do postevaluations to determine which of the samples is different. This section recommends Tukey's method to analyze the raw or rank-transformed data only if one of the previous tests (ANOVA, rank-transformed

ANOVA, Kruskal-Wallis, or Friedman) indicates a significant difference between groups. The reader is cautioned that when performing an ANOVA using standard software, the ANOVA test used must match the data.

4.6.1 One-Factor Comparisons

ANOVA

The ANOVA for one factor is a procedure for comparing the mean value from each group with the overall mean. H_o is typically stated that there are no differences between the group means, whereas H_a states that at least one group's mean is significantly different from the overall mean or

H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$.
H_a: At least one group mean is different.

The basic assumptions made in using an ANOVA are as follows (Remington and Schork, 1970):

- Each sample is a random sample from the corresponding population, and observations from different populations are independent.
- The measurement variable is normally distributed in each of the k groups.
- The groups have the same variance (homoscedasticity).

The variation (or total noise) in the data can be split into the treatment sum of squares (SST) and the errors sum of squares (SSE) (see Equation 4-40) (Helsel and Hirsch, 1995) where

k = number of groups,
n_j = number of observations in the j^{th} group,
x_{ij} = i^{th} observation in the j^{th} group,
$\bar{x}$ = overall mean, and
$\bar{x}_j$ = mean of the j^{th} group.

This notation is also used in Table 4-19, which indicates each observation, group sample size, group sample mean, and group true mean. Note that sample sizes for the different groups need not be the same. The reader should compare the notation in Table 4-19 to that used in Equation 4-40.

The observations (x_{ij}) within each group are assumed normally distributed about the mean, μ_j and variance, σ^2. The variance is the same for all classes, but the mean can vary among classes. The overall mean is denoted as μ, and the corresponding linear model is expressed as (Snedecor and Cochran, 1980)

$$x_{ij} = \mu + \alpha_j + \epsilon_{ij} \quad [i = 1,\ldots n_j;\ j = 1,\ldots k;\ \epsilon_{ij} = N(0,\sigma^2)] \tag{4-41}$$

This fixed effects model shows that each observed value is the sum of an overall mean (μ), a treatment or class deviation (α_j), and a random element (ϵ_{ij}) from a normally distributed population with a zero mean and a standard deviation equal to σ. The model is referred to as "fixed" because the α_j, while unknown, are constant for a group. The random element represents variations due to such factors as unit-to-unit variation in treatment effect, measurement errors, or individual characteristics of the unit (Snedecor and Cochran, 1980). To detect a significant difference, the variation within the group (i.e., ϵ_{ij}) must be sufficiently smaller than the variation between groups.

$$\textit{Total sum of squares} = \textit{SST} + \textit{SSE}$$
$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 = \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{x})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2 \tag{4-40}$$

Table 4-19. ANOVA notation.

	Factor Level			
	1	2	...	k
	x_{11}	x_{12}	...	x_{1k}
	x_{21}	x_{22}		x_{2k}
	.	.		.
	.	.	...	.
	.	.		.
	$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_k k}$
Sample size:	n_1	n_2	...	n_k
Sample mean:	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$
True mean:	μ_1	μ_2	...	μ_k

The ANOVA test statistic, F, is based on a ratio of the treatment mean squares (MST) and error mean squares (MSE):

$$F = \frac{MST}{MSE} \tag{4-42}$$

where

$$MST = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2}{k-1} = \frac{SST}{k-1} \tag{4-43}$$

$$MSE = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{N-k} = \frac{SSE}{N-k} \tag{4-44}$$

and N is the total number of observations.

An F value of 1 represents the condition where H_o is true, and large F values indicate differences among the μ_j. Snedecor and Cochran (1980) note that the F test is more affected by nonnormality and heterogeneity of variances when sample sizes, n_j, are not equal.

Table 4-20 presents a common format for the results from a one-factor ANOVA analysis generated by typical software. The first column identifies which portion of the linear model is being displayed and corresponds to the top portion of Equation 4-40. The second column presents the sum of squares for each source of variation, the third column presents the degrees of freedom, and the fourth column presents the treatment and error mean squares (Equations 4-43 and 4-44). F is calculated using Equation 4-42. The p value corresponds to the significance level associated with the computed F. The "F crit" corresponds to the critical value from Table D6 using k-1 and N-k degrees of freedom and a selected α. Note that some software packages do not present "F crit." If the p value is less than the selected α, H_o is rejected because at least one of the groups has a different mean.

As an example one-factor ANOVA, consider the situation where the trout populations of three streams are measured by the multiple-step Zippin

Table 4-20. Common one-way ANOVA output format.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Between Groups (Treatment)	SST	k-1	MST = SST/(k-1)	MST/MSE	p	F value for selected α
Within Groups (Error)	SSE	N-k	MSE = SSE/(N-k)			
Total	SST + SSE	N-1				

approach for electrofishing at five randomly selected sites in the Coastal Plain region (Platts et al., 1983). The data from this monitoring effort are shown in Table 4-21.

Using the one-factor ANOVA procedure from a standard spreadsheet, trout population as a function of stream was modeled to test the null hypothesis that stream has no effect on trout population (i.e., the treatment effect is zero). The results of this test are shown in Table 4-22. Note that the *F* value of 6.332 is equal to MST (92.867) divided by MSE (14.667). The *p* value is 0.013. The critical value from Table D6 with 2 and 12 degrees of freedom and $\alpha = 0.05$ is 3.885. H_o is rejected since at least one of the stream's trout populations has a different mean. Since H_o is rejected, it is appropriate to continue with postevaluations to determine which group has a different mean. Had H_o not been rejected, postevaluations would be meaningless and inappropriate.

One approach (Least Significant Difference) to determining which of the means is different is to compare each pair of means. To do a

Table 4-21. Trout population from streams in the coastal plain region.

	Stream		
Site	Black Creek	Blue Creek	Red Creek
	Trout Population (Pounds/Acre/Year - Year Class 2)		
1	60	49	50
2	65	60	56
3	64	54	51
4	63	58	60
5	58	57	52
$\bar{x}_i$	62.0	55.6	53.8

Table 4-22. One-way ANOVA of stream trout data from the coastal plain region using stream as the treatment.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Between Groups (Treatment)	185.733	2	92.867	6.332	. 0.013	3.885
Within Groups (Error)	176.000	12	14.667			
Total	361.733	14				

pairwise comparison, the standard error of the difference between two means is calculated as (Snedecor and Cochran, 1980)

$$s_{\overline{D}} = \sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \qquad (4\text{-}45)$$

with *N-k* degrees of freedom. Using the data from Table 4-22, $s_{\overline{D}}$ is equal to 2.422 with 12 degrees of freedom. For $\alpha/2$ equal to 0.025 (and df = 12), the value of *t* from Table D2 is 2.1788. Therefore, if any pair of means exceeds a difference of 2.422 x 2.1788 or 5.3, the difference is significant. The mean trout populations for Black, Blue, and Red Creeks are 62, 55.6, and 53.8 pounds/acre/year, respectively. The trout population in Black Creek is significantly higher than the trout population in Blue Creek or Red Creek. Note that a pairwise comparison was made between the three groups (i.e., three pairwise comparisons) with $\alpha = 0.05$; therefore, the overall error rate is $1-(1-0.05)^3$, or about 14 percent. Other approaches for multiple comparisons are discussed in Section 4.6.4.

Kruskal-Wallis test

The Kruskal-Wallis test is an extension of the Mann-Whitney test described earlier. This test can be used when there are several independent samples that do not have the same distribution. In this case, H_o and H_a are as follows:

H_o: All *k* groups have identical distributions.
H_a: At least one of the groups tends to yield larger observations than at least one other group.

If the distributions of all groups are similar except for location (i.e., similar spread and skew), H_a can be refined to imply that the median concentration from one group is different from the median concentration from at least one other group. To achieve this greater detail in H_a, transformations such as logs can be used.

Again consider the notation used in Table 4-19 where there are *k* groups and each group has n_j observations. *N* is the total number of observations. To compute the Kruskal-Wallis statistic, the following steps (Conover, 1980) can be used:

- Rank all of the data from lowest to highest, assigning the average of ranks to ties. The rank of observation x_{ij} is denoted as $R(x_{ij})$.
- Compute R_j for all k random samples using

$$R_j = \sum_{i=1}^{n_j} R(x_{ij}) \quad for\ j = 1,2,...,k \qquad (4\text{-}46)$$

- Compute the test statistic, T:

$$T = \frac{1}{S^2}\left[\sum_{j=1}^{k}\frac{R_j^2}{n_j} - \frac{N(N+1)^2}{4}\right] \qquad (4\text{-}47)$$

where

$$S^2 = \frac{1}{N-1}\left[\sum_{j=1}^{k}\sum_{i=1}^{n_j} R(x_{ij})^2 - N\frac{(N+1)^2}{4}\right] \qquad (4\text{-}48)$$

For $k = 3$, all n_j are 5 or less, and there are no ties, special tables should be used to determine the rejection region for T (see Conover, 1980). If these criteria do not apply, Table D3 with $p = 1\text{-}\alpha$ and $k\text{-}1$ degrees of freedom should be used. If the computed T statistic from Equation 4-47 is greater than the value obtained from the table, H_o is rejected.

Table 4-23 presents the rank of the trout population data used in the previous example; R_j for each group has already been computed. Applying Equation 4-48 with the individual ranks from Table 4-23 and $N = 15$, S^2 is equal to 19.82. Substituting S^2, $N = 15$, $n_j = 5$ (for all j) into Equation 4-47 along with the R_j summarized in Table 4-23, T is equal to 7.21. From Table D3 with $\alpha = 0.05$ and 2 degrees of freedom, the critical value is 5.991. H_o is rejected. Had there been no ties, the exact critical value would be 5.66 (Conover, 1980).

Since H_o has been rejected, it is acceptable to do a multiple comparisons evaluation. One approach is to compare the ranks from each pairwise group. The groups i and j are different if the following inequality is satisfied (Conover, 1980):

$$\left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right| > t_{1-(\alpha/2)}\left(S^2\frac{N-1-T}{N-k}\right)^{0.5}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{0.5} \qquad (4\text{-}49)$$

In this example, all n_j are equal to 5 and the above equation can be reduced to

Table 4-23. Rank of trout population from streams in the coastal plain region.

Site	Stream		
	Black Creek	Blue Creek	Red Creek
	Rank		
1	11	1	2
2	15	11	6
3	14	5	3
4	13	8.5	11
5	8.5	7	4
R_j	61.5	32.5	26

$$\left|\frac{R_i}{5}-\frac{R_j}{5}\right| > 2.1788\left(19.82\frac{15-1-7.21}{15-3}\right)^{0.5}\left(\frac{1}{5}+\frac{1}{5}\right)^{0.5}$$

or

$$|R_i - R_j| > 23.08$$

where t is obtained from Table D2 with 15-3 degrees of freedom. By comparing the above result with R_j in Table 4-23, it can be concluded that the trout population in Black Creek is significantly greater than the trout population in Blue Creek or Red Creek.

4.6.2 Two-Factor Comparisons

ANOVA

In a two-way ANOVA the variation due to two factors is quantified. One factor cannot be a subset of the other factor. Subsetted factors are referred to as nested factors, a subject that is not considered here. The reader is referred to Gaugush (1986) and Snedecor and Cochran (1980) for more thorough discussions regarding factorial experiments and hierarchical arrangements for fixed effects models. In this section, Equation 4-41 is extended to include a second factor (Helsel and Hirsch, 1995; Snedecor and Cochran, 1980),

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad (4\text{-}50)$$

where $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, n$. The number of levels in factors A and B are represented by a and b, respectively. There are axb treatment groups. The number of replicates is equal to n and is constant across all treatment levels i and j. That is, there are the same number of observations for each unique combination of factors A and B. In this case, each observed value is the sum of an overall mean (μ), the influence of the i^{th} category of factor A (α_i), the influence of the j^{th} category of factor B (β_j), the interaction effect between factors A and B ($(\alpha\beta)_{ij}$), and a residual error (ϵ_{ijk}). If $(\alpha\beta)_{ij}$ is equal to zero, there is no interaction. No interaction means that a change in factor B has the same impact on x_{ijk} regardless of factor A (and vice versa).

H_o is that all treatment groups have the same mean, whereas H_a indicates that at least one treatment group mean has a different mean. The two assumptions made using this model (Equation 4-50) is that (1) the effects are additive, and (2) the residuals are independent, random variables normally distributed with a zero mean and constant variance across all treatment groups (Snedecor and Cochran, 1980).

Helsel and Hirsch (1995) caution the practitioner that when evaluating data with unequal numbers of observations some smaller statistical packages incorrectly apply the balanced equations (equal number of observations) presented here to unbalanced data sets (unequal number of observations) without notice. Packages such as SAS and Minitab provide options for analyzing unbalanced data sets. Two-way ANOVA can be performed for two cases, one in which there is no interaction between the two variables and one in which there is an interaction between the two variables. The sum of squares for factor A (SSA), factor B (SSB), and the interaction between A and B (SSI) for a balanced data set including interaction can be computed using Equations 51 through 55 (Helsel and Hirsch, 1995).

Table 4-24 is an ANOVA table that incorporates the above equations into the second column, presents the degrees of freedom in column three, and provides the equations for the mean squared error terms and F statistics in the fourth and fifth columns.

$$SSA = \sum_{i=1}^{a} \frac{\left(\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{bn} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \quad (4\text{-}51)$$

$$SSB = \sum_{j=1}^{b} \frac{\left(\sum_{i=1}^{a}\sum_{k=1}^{n} x_{ijk}\right)^2}{an} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \quad (4\text{-}52)$$

$$SSI = Total\ SS - SSA - SSB - SSE \quad (4\text{-}53)$$

where

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (x_{ijk})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(\sum_{k=1}^{n} x_{ijk}\right)^2}{n} \quad (4\text{-}54)$$

and

$$Total\ SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (x_{ijk})^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \quad (4\text{-}55)$$

Table 4-24. Common two-way ANOVA output format.

Source of Variation	SS	df	MS	F	p-value	F criteria
Factor A	SSA	a-1	MSA = SSA/ (a-1)	MSA/MSE	p	F value for selected α
Factor B	SSB	b-1	MSB = SSB/ (b-1)	MSB/MSE	p	F value for selected α
Interaction (Factor AxFactor B)	SSI	(a-1)x (b-1)	MSI = SSI/ [(a-1)x(b-1)]	MSI/MSE	p	F value for selected α
Error	SSE	ab(n-1)	MSE = SSE/ [ab(n-1)]			
Total	Total SS	abn-1				

To demonstrate this procedure, the two-way ANOVA procedure is applied to the data in Table 4-25. This data set includes the trout population data from three streams and three regions (i.e., $a = b = 3$). This test could reflect, for example, the hunch that regional effects on trout population differ across streams (e.g., perhaps the streams are impacted differently by point and nonpoint sources). In this experimental design, factor A is the region and factor B is the stream. Using standard statistical software, Table 4-26 presents the results of the two-way ANOVA calculations. The p values for the region, stream, and region x stream factors are 1.0×10^{-10}, 0.001, and 0.1458, respectively. Using $\alpha = 0.05$, H_o is rejected; there is a significant difference between treatment group means due to region and stream. The interaction of region and stream is not significant at the 95 percent confidence level. Based on this analysis, it is acceptable to perform a multiple comparisons analysis for regions and streams.

This ANOVA discussion is simple in many respects. For example, a balanced data set and a fixed effects model were analyzed. In situations

Table 4-25. Stream trout population.

		Site				
		1	2	3	4	5
Stream	Region	Trout Population (Pounds/Acre/Year - Year Class 2)				
Black Creek	Mountain	75	70	65	72	68
	Piedmont	68	72	70	70	67
	Coastal Plain	60	65	64	63	58
Blue Creek	Mountain	70	76	69	67	74
	Piedmont	64	66	60	69	62
	Coastal Plain	49	60	54	58	57
Red Creek	Mountain	68	70	63	65	70
	Piedmont	62	66	58	69	67
	Coastal Plain	50	56	51	60	52

Table 4-26. Two-way ANOVA of trout population data using an interaction term.

Source of Variation	SS	df	MS	F	p-value	F crit (α=0.05)
Region	1213.73	2	606.87	46.60	1.0E-10	3.26
Stream	219.73	2	109.87	8.44	0.0010	3.26
Region x Stream	94.93	4	23.73	1.82	0.1458	2.63
Error	468.80	36	13.02			
Total	1997.20	44				

where multiple variables are examined, a balanced data set is not likely to be feasible or economical. A key limitation of the fixed effects model is that inferences cannot be made beyond the groups being tested. In the trout population example, only statements about the three streams and three regions analyzed can be made. Nothing about a fourth stream or region can be inferred. If the three streams had been randomly selected from across the state with the intent of determining whether there was a spatial difference in trout population, the stream factor would have been a random factor rather than a fixed factor, and the calculation of the *F* statistics would be different. If both factors were random, the *F* statistics would use the mean squares for interaction (*MSI*) rather than the *MSE* as the denominator. If there were a mixture of fixed and random factors, the *F* statistic for the fixed factor would be computed with the *MSI* and the random factor would be computed with the *MSE* in the denominator (Helsel and Hirsch, 1995).

Ranked transformed ANOVA

To perform the ANOVA described in Section 4.6.2, the data in each treatment group must be normally distributed with a constant variance. If the data do not meet this requirement, it is possible to use transformations of the data such as logarithms to convert the data to a normal distribution with constant variance. The use of logarithms implies that the influences of each factor are multiplicative in the original units (Helsel and Hirsch, 1995; Snedecor and Cochran, 1980). Alternatively, the data can be rank-transformed (i.e., a rank from 1 to *N* can be assigned to the data) and a two-way ANOVA can be performed on the ranks. Rejection of H_o using an ANOVA on the rank-transformed data indicates that the medians differ between treatment groups. Helsel and Hirsch (1995) state that "rank transformation results in tests which are more robust to non-normality, and resistant to outliers and non-constant variance, than is ANOVA without transformations."

4.6.3 Matched Data

Collecting paired data to mask or block out unwanted noise due to meteorological or geographical differences is a common practice when comparing "before" and "after" data. Comparing just two groups was described in Section 4.5.1. Comparing matched data with more than two groups is described here. In this case, the objective is to compare one factor (referred to as the treatment) while blocking out the other factor (referred to as the block).

The linear model for this analysis is (Helsel and Hirsch, 1995)

$$x_{ij} = \mu + \alpha_j + \beta_i + \epsilon_{ij} \qquad (4\text{-}56)$$

where $j = 1, \ldots, k$ and $i = 1, \ldots, n$. In this case, each observed value is the sum of an overall mean (μ), the influence of the j^{th} group effect (α_j), the influence of the i^{th} block effect (β_i), and a residual error (ϵ_{ij}). In addition to the two-way ANOVA without replication and the Friedman test described here, Helsel and Hirsch (1995) also describe the median polish and the median aligned-ranks ANOVA.

Two-way ANOVA without replication

In the ANOVA model, ϵ_{ij} is assumed to be normally distributed. The sums of squares for the two-way ANOVA without replication are computed using Equations 57 through 60 (Helsel and Hirsch, 1995). Table 4-27 presents a common format for a two-way ANOVA without replication. Removing the block effect from the calculation of the *SSE* results in a higher *F* statistic, thus improving the detection of significant differences between groups. H_o is rejected if the computed *F* is greater than the

$$SST = \frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{n} x_{ij}\right)^2}{n} - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \quad (4\text{-}57)$$

$$SSB = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} x_{ij}\right)^2}{k} - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \quad (4\text{-}58)$$

$$SSE = Total\ SS - SST - SSB \quad (4\text{-}59)$$

where

$$Total\ SS = \sum_{j=1}^{k}\sum_{i=1}^{n} (x_{ij})^2 - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \quad (4\text{-}60)$$

Table 4-27. Common two-way ANOVA without replication output format.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Treatment	SST	k-1	MST = SST/ (k-1)	MST/MSE	p	F value for selected α
Block	SSB	n-1	MSB = SSB/ (n-1)			
Error	SSE	(k-1) x (n-1)	MSE = SSE/ [(k-1)x(n-1)]			
Total	Total SS	kn-1				

critical F value from Table D6 with (k-1) and (k-1)(n-1) degrees of freedom.

Friedman test

The Friedman test is the most common nonparametric test for randomized complete block designs. It is an extension of the sign test (Helsel and Hirsch, 1995). H_o is that the median value of the k groups are identical, whereas H_a states that at least one median is different. To compute the test statistic, the following steps are used:

- Rank the data in each block from 1 to k.
- Compute the average rank for each group ($\bar{R}_j$).
- Compute χf using the following formula, which accounts for ties:

$$\chi f = \frac{12n}{k(k+1) - \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} (t_{ij}(j^3 - j))} \sum_{j=1}^{k} \left[\bar{R}_j - \frac{k+1}{2} \right]^2 \quad (4\text{-}61)$$

where t_{ij} equals the number of ties of the extent j in block i. For $k+n \leq 9$, exact tables should be used (see Helsel and Hirsch, 1995). Otherwise, H_o is rejected if χf greater than or equal to the critical F value from Table D6 with (k-1) and (n-1)(k-1) degrees of freedom and $p = 1-\alpha$.

4.6.4 Multiple Comparisons

All of the hypothesis tests featured to this point allow the analyst to determine whether at least one treatment results in a mean or median that is significantly different from that which results from the other treatments. It does not indicate which treatment is different or whether there are multiple differences. Multiple comparisons should be done only if the analyses performed under Section 4.6.1, 4.6.2, or 4.6.3 indicate a significant difference.

Two key features distinguish multiple comparisons: (1) whether α is based on a pairwise or overall comparison and (2) whether the test is a multiple-stage test (MST) or a simultaneous inference method (SIM). An important distinction should be made about whether a pairwise or overall α is used. The α level indicates the probability of making an incorrect comparison. Helsel and Hirsch (1995) cite an example of a one-factor analysis with six groups (in which there are 15 pairwise comparisons). If $\alpha = 0.05$, the potential for making at least one error is equal to $1-(1-.05)^{15}$ or 0.54, a 54 percent chance of making one error. MSTs are valid for groups with constant sample size, whereas SIMs are valid for equal and unequal sample sizes.

For these reasons, Helsel and Hirsch (1995) recommend using Tukey's method, which uses an overall α and is a SIM. Other tests include the Bonferroni t tests, Duncan's multiple range test, Gabriel's multiple-comparison procedure, the Ryan-Einot-Gabriel-Welsch (REGW) multiple F test, the REGW multiple range test, Scheffe's multiple-comparison procedure, and the Waller-Duncan k-ratio test. The reader should consult statistics texts (e.g., Snedecor and Cochran, 1980) to learn more about these procedures, with preference given to Tukey's method for equal or unequal sample sizes and the REGW tests when the sample sizes are equal. If a nonparametric analysis was performed, the most appropriate approach is to rank-transform the data and apply a test based on the above discussion.

Tukey's method indicates that the mean between two groups can be considered different if (Helsel and Hirsch, 1995)

$$\left| \bar{x}_i - \bar{x}_j \right| > q_{(1-\alpha),k,N-k} \sqrt{MSE \frac{n_i + n_j}{2 n_i n_j}} \quad (4\text{-}62)$$

where

q = studentized range statistic from Neter, Wasserman, and Kutner (1985);
α = overall significance level;
k = number of treatment group means compared;
$N\text{-}k$ = MSE's degrees of freedom; and
n_i, n_j = sample size of group i and j respectively.

Typically, the results of a multiple comparison are commonly displayed using letters to distinguish groups. For example,

$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
A	AB	B	C

indicates that $\bar{x}_1 \approx \bar{x}_2$; $\bar{x}_1 < \bar{x}_3$; $\bar{x}_2 \approx \bar{x}_3$; and $\bar{x}_4$ is greater than $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$. The letter groupings could also be placed in boxplots.

4.7 REGRESSION TECHNIQUES

4.7.1 Overview

Regression can be used to model or predict the behavior of one or more variables. The general regression model, where ϵ is an error term, is given as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \epsilon \qquad (4\text{-}63)$$

In this equation, the behavior of a single dependent variable (y) is modeled with one or more independent variables (x_1, ..., x_n). The x's may be linear or nonlinear (e.g., x_i can represent x^2, x^3, x^{-1}, etc.). β_0, ..., β_n are numerical constants that are computed using equations described later. Nonlinear models are commonly applied to physical systems, but they are somewhat more difficult to analyze because iterative techniques are involved when the model cannot be transformed to a linear model. The use of two or more independent variables (x) in a linear function to describe the behavior of y is referred to as multiple linear regression. In either case, regression techniques attempt to explain as much of the variation in the dependent variable as possible.

In nonpoint source analyses, linear regression is often used to determine the extent to which the value of a water quality variable (y) is influenced by land use or hydrologic factors (x) such as crop type, soil type, percentage of land treatment, rainfall, or stream flow, or by another water quality variable. Practical applications of these regression results include the ability to predict the water quality impacts due to changes in the independent variables.

In developing a regression model, the analyst will want to select from a set of variables, normally selecting those independent variables that are most strongly correlated with the dependent variable. To begin, therefore, the analyst might want to compute correlation coefficients between numerous monitored variables at the exploratory phase of the analysis (sometimes referred to as correlation matrices). In fact, determining which variables are most strongly correlated might be the entire goal of the analysis. The correlation matrix can then be used to guide the analyst, to some extent, in selecting appropriate independent variables. In adding additional variables to a model, the analyst must be aware of correlations among different independent variables (multicollinearity) that can mask the relationship of one x to the y variable due to the correlation of this independent variable with another in the model.

In contrast to the univariate models discussed above, where only one dependent response variable (y) is involved, multivariate models can have several dependent variables. Multivariate analyses (which include MANOVA and principal component analysis, among others) are designed to take into account the correlation structure of

the x's and y's to reduce the overall variance. For example, a nonpoint source application might be to examine the effect of different BMP implementation programs on several water quality parameters.

Analysts are encouraged to read the detailed discussion of regression in statistics texts such as Snedecor and Cochran (1980), Cochran (1977), and Srivastava and Khatri (1979) for a more complete discussion of this important statistical procedure.

4.7.2 Simple Linear Regression

The simplest form of regression is to consider just one dependent variable and one independent variable using

$$y = \beta_0 + \beta_1 x + \epsilon \quad (4\text{-}64)$$

where y is the dependent variable, x is the independent variable, and β_0 and β_1 are numerical constants representing the y-intercept and slope, respectively. Helsel and Hirsch (1995) summarize the key assumptions regarding application of linear regression (Table 4-28). The uses of a regression analysis should not be extended beyond those supported by the assumptions that are met. Note that the normality assumption (assumption 5) can be relaxed when testing hypotheses and estimating confidence intervals if the sample size is relatively large.

The first step in applying linear regression is to examine the data to see if linear regression makes sense—that is, to use a bivariate scatter plot to see if the points approximate a straight line. If they fall in a straight line, linear regression makes sense; if they do not, data transformation might be needed, or perhaps a nonlinear relationship should be used.

To illustrate the use of linear regression, the data in Table 4-29, which are a subset of calibration data for a plot-size runoff sampler (Dressing et al., 1987), can be used. In this data set the sampling percentage (split) was measured for a range of flow rates. The scatter plot in Figure 4-18 shows that linear regression can be applied to the data.

Presuming that the data are representative (assumption 2 in Table 4-28), the next step is to develop the regression line using the method of least squares, which minimizes the sum of the squares of the vertical deviations from the points to the line (Freund, 1973). To determine the values of β_0 and β_1 in Equation 4-64, the following equations can be used (Helsel and Hirsch, 1995):

$$\beta_1 = \frac{S_{xy}}{SS_x} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \quad (4\text{-}65)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (4\text{-}66)$$

For the data in Table 4-29, the above equations were used to compute a slope of -0.0119 and an intercept of 3.1317. Thus, the linear model for predicting split versus flow rate is

Split = 3.1317 - 0.0119 • Flow rate

Assumption evaluation

The top section of Table 4-30 provides the same information along with additional characteristics about the β_0 and β_1 that were computed using standard spreadsheet software. Before looking at these additional characteristics, the analyst must make sure that β_0 and β_1 make sense. In this case, perhaps the best approach is to plot the regression line with the raw data as shown in Figure 4-18. The bottom portion of Table 4-30 contains the predicted split (data for the

Table 4-28. Assumptions necessary for the purposes of linear regression..

Assumption	Purpose			
	Predict y given x	Predict y and a variance for the prediction	Obtain best linear unbiased estimator of y	Test hypotheses, estimate confidence or prediction intervals
(1) Model form is correct: y is linearly related to x	✓	✓	✓	✓
(2) Data used to fit the model are representative of data of interest	✓	✓	✓	✓
(3) Variance of the residuals is constant and does not depend on x or anything else		✓	✓	✓
(4) The residuals are independent			✓	✓
(5) The residuals are normally distributed				✓
✓ Indicates that assumption is required.				

Source: Helsel and Hirsch, 1995.

regression line in Figure 4-18) for each flow rate as well as the residual, e_i, defined as $y_i - \hat{y}_i$.

Residuals plotted as a function of predicted values of y and time, and normal probability plots of residuals, are the most effective approaches to evaluate the last three assumptions listed in Table 4-28, respectively. As shown in Case A of Figure 4-19, the plot of residuals versus predicted values of y or time should appear to be a uniform band of points around 0 (Ponce, 1980a). The analyst should look for two types of patterns when evaluating assumption 3 from Table 4-28 (e.g., constant variance). The first is a pattern of increasing or decreasing variance with predicted values of y, as depicted in Case B of Figure 4-19. The second is a pattern (e.g., a trend, a curved line) of the residual with predicted values of y. Both characteristics are usually assessed based on a review of the residual plots and professional judgment alone. The analyst may also need to examine variables other than predicted values of y to fully evaluate assumption 3.

Independence of residuals (assumption 4 from Table 4-28) can be evaluated by examining residuals plotted as a function of time. The analyst should look for the same patterns as before. As an alternative for evaluating independence, the analyst can also plot the ith residual, e_i, as a function of the (i-1)th residual, e_{i-1}. One word of caution is in order when

Table 4-29. Runoff sampler calibration data.

X Flow Rate (gpm)	Y Split (%)	X Flow Rate (gpm)	Y Split (%)
52.1	2.65	17.6	2.84
19.2	3.12	37.6	2.60
4.8	3.05	41.4	2.54
4.9	2.86	40.1	2.58
35.2	2.72	47.4	2.49
44.4	2.70	35.7	2.60
13.2	3.04	13.9	3.19
25.8	2.83		
n = 15	Σx = 433.30	Σy = 41.81	
	$\bar{x}$ = 28.89	$\bar{y}$ = 2.79	
	Σx^2 = 15,940.33	Σy^2 = 117.25	
	Σxy = 1,166.93		
S_{xy} = - 40.817533			
SS_x = 3423.73733			
SS_y = 0.70929333			

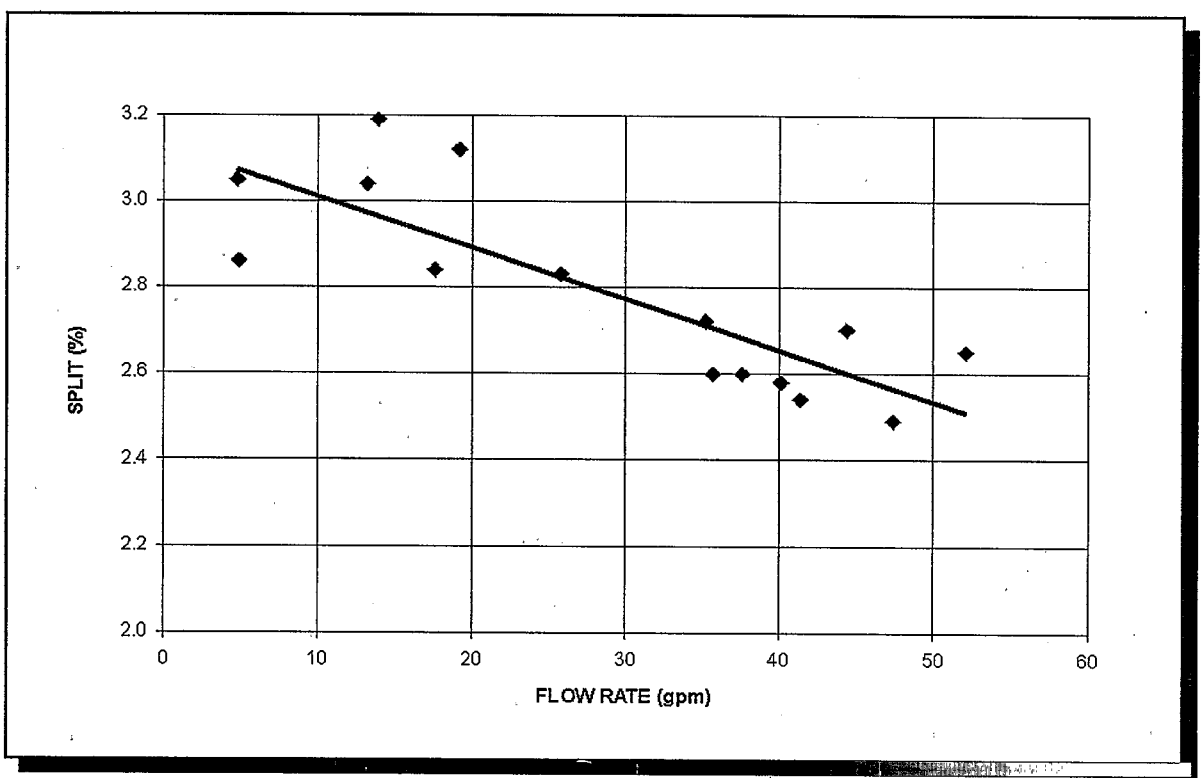


Figure 4-18. Split versus flow rate.

Table 4-30. Regression analysis of runoff sampler calibration data.

	Coefficients	Standard Error	*t* Statistic	*p* Value	Lower 95%	Upper 95%
Intercept (β_0)	3.1317	0.072914	42.950756	2.14E-15	2.97420	3.28924
Flow Rate (β_1)	-0.0119	0.002237	-5.330126	0.00014	-0.01675	-0.00709

	df	SS	MS	F	Significance F
Regression	1	0.486623	0.486623	28.410248	0.0001366
Residual	13	0.222670	0.017128		
Total	14	0.709293			

Flow Rate (gpm)	Split (%)	Predicted Split	Residual $e_i=y_i-\hat{y}_i$
52.10	2.65	2.5106	0.1394
19.20	3.12	2.9028	0.2172
4.80	3.05	3.0745	-0.0245
4.90	2.86	3.0733	-0.2133
35.20	2.72	2.7121	0.0079
44.40	2.70	2.6024	0.0976
13.20	3.04	2.9743	0.0657
25.80	2.83	2.8241	0.0059
17.60	2.84	2.9219	-0.0819
37.60	2.60	2.6835	-0.0835
41.40	2.54	2.6382	-0.0982
40.10	2.58	2.6536	-0.0736
47.40	2.49	2.5666	-0.0766
35.70	2.60	2.7061	-0.1061
13.90	3.19	2.9660	0.2240

reviewing any residual plot: If there are more points in a certain section of the residual plot, the residuals might not appear to be a uniform band of points around 0 (as suggested in Case A of Figure 4-19); instead, that section might have a somewhat wider band (Helsel and Hirsch, 1995). This is an expected result.

The normality of residuals can be assessed by examining a probability plot. Two problems with non-normal residuals are the loss of power in subsequent hypothesis tests and increased prediction intervals together with the impression of symmetry (Helsel and Hirsch, 1995).

Figure 4-20 displays all three of these plots for the split data analyzed from Table 4-29. From Figure 4-20, A and B, the split residuals appear to be independent of predicted values of *y* and time as well as having a constant variance. The regression meets assumptions 3 and 4 listed in Table 4-28. In this analysis, testing for residual independence is important since the testing apparatus was calibrated initially. The pumps or other equipment could have differed in performance over time, which in turn would affect the results. Figure 4-20C, the probability plot, suggests that the data might not rigorously follow the normality assumption, although by inspection any normality violation is believed to

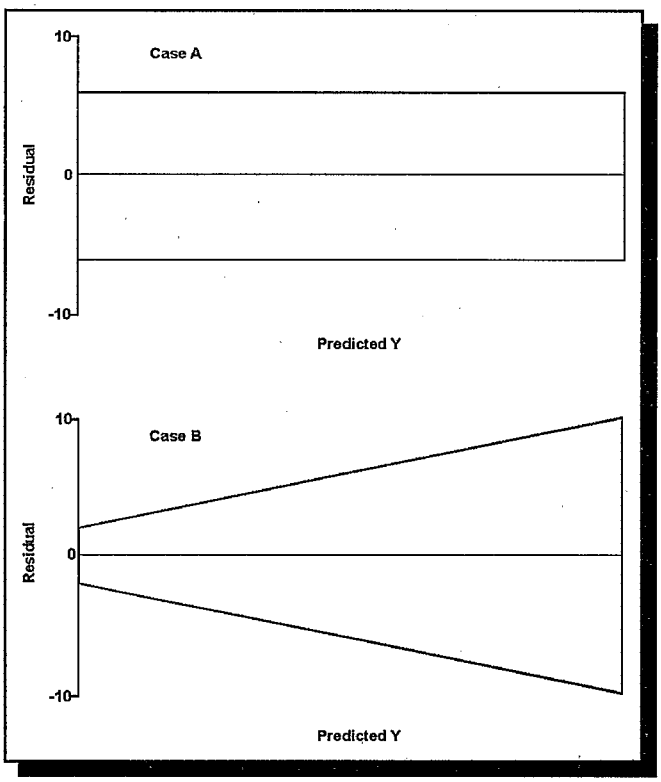


Figure 4-19. Plot of residuals versus predicted values. (Source: Ponce, 1980a)

be relatively minor. To check, the Shapiro-Wilk W statistic (see Section 4.4.1) is computed as 0.935. Comparing 0.935 to the test statistic (with $p=0.95$, $n=15$) from Table D5, 0.98, the split residuals can be accepted as being normally distributed. (Note that accepting H_o in this case might be due to small sample size and resulting lack of power.) Had this analysis violated any of these assumptions, using a different regression technique, transforming the data, or adding additional variables to the regression would have to be considered. Alternatively, the use of the regression results could be limited to those identified in Table 4-28 as restricted by the assumptions met.

Model evaluation

To determine how well the regression line fits the data, several things can be evaluated:

- Evaluate the proportion of variation in y explained by the model.
- Test whether β_0 is zero.
- Test whether $\beta 1$ is zero.
- Compute the confidence interval for β_0.
- Compute the confidence interval for β_1.

The coefficient of determination, R^2, can be used to evaluate what proportion of the variation can be explained by the model (Gaugush, 1986). R^2 can be computed as (Helsel and Hirsch, 1995)

$$R^2 = \frac{[SS_y - s^2(n-2)]}{SS_y} = 1 - \frac{SSE}{SS_y} = \frac{S_{xy}^2}{SS_x S_{xy}} \quad (4\text{-}67)$$

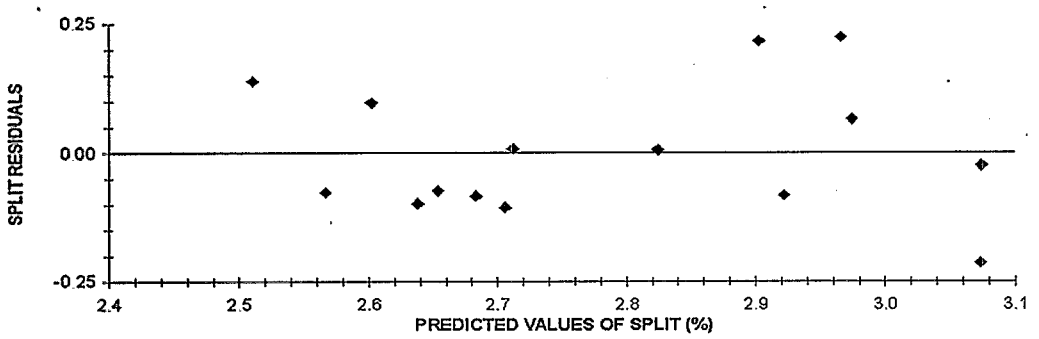
where

$$SS_y = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2 \quad (4\text{-}68)$$

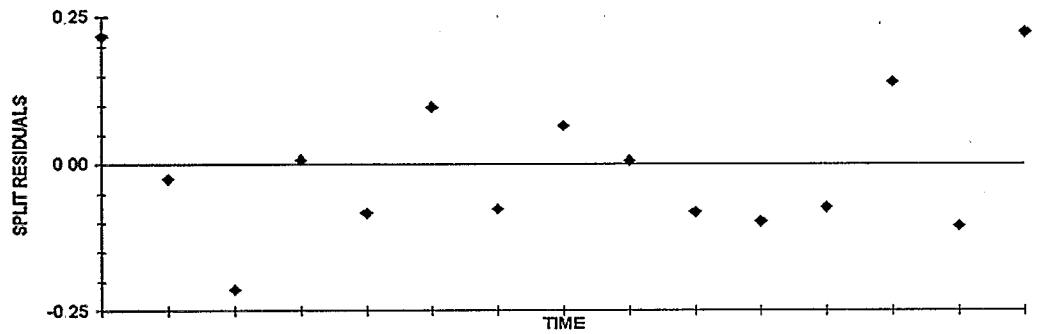
and

$$SSE = \sum_{i=1}^{n} e_i^2 \quad (4\text{-}69)$$

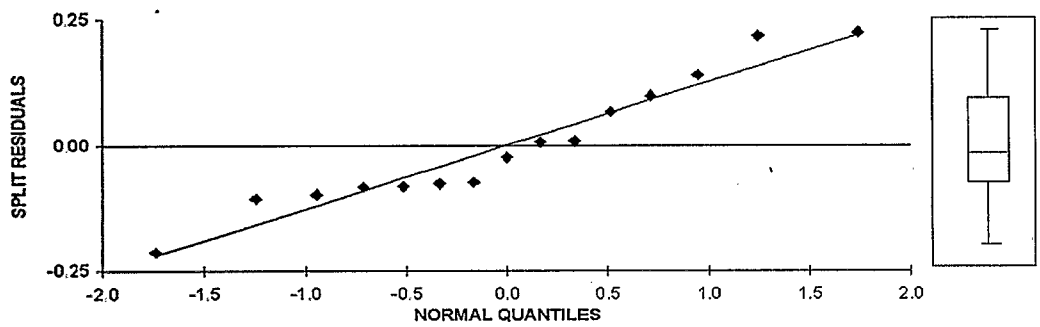
The residual, e_i, is defined as $y_i - \hat{y}_i$. S_{xy} and SS_x can be computed from Equation 4-65. Values for R^2 range between 0 and 1, with 1 representing the case where all observed y values are on the regression line. The correlation coefficient, r, measures the strength of linear relationships (Freund, 1973) and is computed as the square root of R^2. The sign of r should be the same as the sign of the slope. It ranges from -1 to 1, with the extreme values representing the strongest association and 0 representing no correlation.



A) Split residuals as function of predicted values of split.



B) Split residuals as function of time.



C) Probability plot of split residuals.

Figure 4-20. Plot of split residuals.

Using the split data from above, the sum of residuals-squared (*SSE*) is equal to 0.2227; thus, R^2 is equal to 1 - (0.2227/0.7093) = 0.686, or 68.6 percent of the variance is explained by the model. (The 0.7093 is from Table 4-29.) The overall model can also be evaluated with the *F* statistic (28.41), which is computed in Table 4-30. The *F* statistic is a measure of the variability in the data set that is explained by the regression equation in comparison to the variability that is not explained by the regression equation. Since the *p* value of 0.0001366 is less than 0.05, the overall model is significant at the 95 percent confidence level.

Are β_0 and β_1 significantly different from zero? The standard error for β_0 and β_1 in the top portion of Table 4-30 can be calculated as (Helsel and Hirsch, 1995)

$$SE(\beta_0) = s\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{SS_x}} \quad (4\text{-}70)$$

$$SE(\beta_1) = \frac{s}{\sqrt{SS_x}} \quad (4\text{-}71)$$

where

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2} \quad (4\text{-}72)$$

The value *s* is equal to the standard error of the regression (which is the same as the standard deviation of the residuals). The corresponding *t* statistics (with *n* - 2 degrees of freedom) for β_0 and β_1 are then equal to β_0 and β_1 divided by their respective standard error. The *t* statistic can then be compared to values from the *t* distribution to determine whether β_0 or β_1 are significantly different from zero. In this case β_0 and β_1 are both significantly different from zero based on inspection of their associated *p* values in Table 4-30. The overall model can also be evaluated with the *F* statistic computed in the middle portion of Table 4-30. This portion of Table 4-30 has the same format as the ANOVA tables described in the previous section. The values in this table are computed using the equations summarized in Table 4-31. Verification of the results in Table 4-30 is left to the reader.

The confidence intervals for β_0 and β_1 can be computed using the following formulas (Helsel and Hirsch, 1995):

$$\beta_0 \pm t_{\alpha/2,n-2}\, SE(\beta_0) \quad (4\text{-}73)$$

$$\beta_1 \pm t_{\alpha/2,n-2}\, SE(\beta_1) \quad (4\text{-}74)$$

where $t_{\alpha/2,n-2}$ is from Table D2. The lower and upper 95 percent confidence limits for β_0 and β_1 are provided in the top portion of Table 4-30, from which $t_{\alpha/2,n-2}$ was obtained as 2.1604.

The correlation coefficient, *r*, calculated from sample data, is an estimate of the corresponding population parameter, ρ, referred to as the population correlation coefficient. Establishing a confidence interval for ρ requires that *x* also be a normally distributed random variable (Freund, 1973). The Shapiro-Wilk *W* statistic for *x* (the flow rate data in Table 4-29) is 0.931. Comparing 0.931 to the test statistic of 0.98, obtained earlier, the data can be accepted as normally distributed. Using Table D9 (Remington and Schork, 1970) the 95 percent and 99 percent confidence limits for ρ can be obtained knowing *n* and *r*. For the data in Table 4-29, *n* is 15 and *r* is -0.828. So, the 99 percent confidence limits from Table D9 are approximately -0.95 to -0.50.

A *t* test can also be used to test H_o that ρ is zero. The *t* statistic (with *n* - 2 degrees of freedom) for this test is (Freund, 1973)

Table 4-31. Common ANOVA output format for linear regression.

Source of Variation	SS	df	MS	F	Significance F
Regression	$SSR = (S_{xy})^2/SS_x$	1	MSR = SSR/1	MSR/MSE	p
Residual	SSE	n-2	MSE = SSE/(n-2)		
Total	SSR + SSE	n-1			

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (4\text{-}75)$$

For the above data t would be -5.33. From Table D2 the two-sided t value for 95 percent significance (df = 13) is -2.1604. Therefore, H_o ($\rho = 0$) is rejected and H_a that ρ is not zero is accepted.

The Fisher Z transformation can be used to test H_o of ρ equal to values other than zero (Freund, 1973). For this test, r is changed into a Z value using (Freund, 1973)

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \qquad (4\text{-}76)$$

Freund (1973) provides a table of Z values to simplify this procedure. The test statistic is (Freund, 1973)

$$z = (Z-\mu_z)\sqrt{n-3} \qquad (4\text{-}77)$$

where μ_z corresponds to the Z value for the nonzero value of ρ being tested for. For illustration, H_o that ρ is equal to -0.8 for the regression performed can be tested using the data from Table 4-29. Equation 4-76 yields -1.1827 for r = -0.828 and -1.0986 for r = -0.8. Substituting these values in Equation 4-77 yields

$$z = (-1.1827+1.0986)\sqrt{15-3} = -0.2913$$

The two-sided z statistic at 95 percent significance (Table D1) is -1.96, H_o is accepted, $\rho = -0.8$.

A confidence interval for ρ can be determined by calculating an interval for μ_z, and then retransforming the confidence interval from Z values to ρ (Freund, 1973). The formula for the confidence interval for μ_z is (Freund, 1973)

$$Z - \frac{z_{\alpha/2}}{\sqrt{n-3}} \le \mu_z \le Z + \frac{z_{\alpha/2}}{\sqrt{n-3}} \qquad (4\text{-}78)$$

Again using the sample data, the 95 percent confidence interval for μ_z becomes

$$-1.1827 - \frac{1.96}{\sqrt{15-3}} \le \mu_z \le -1.1827 + \frac{1.96}{\sqrt{15-3}}$$

$$-1.7485 \le \mu_z \le -0.6169$$

Solving for ρ,

$$\textit{Lower limit of } \rho = -1.7485 = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$\textit{Upper limit of } \rho = -0.6169 = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$-0.94 \leq \rho \leq -0.55$$

Using the regression line

The most obvious use of the regression line is to predict y values for selected values of x. For example, using the regression equation

Split = 3.1317 - 0.0119 x Flow Rate,

the split for any flow rate can be estimated. (It is not good practice, however, to predict values beyond the range of test conditions.) For a flow rate of 10 gpm, the predicted split is 3.01 percent; for a flow rate of 50 gpm, the predicted split is 2.53 percent.

Since in most cases the regression line will not fit the data perfectly, the uncertainty associated with the predicted values should be quantified. The regression line can be used either to establish the confidence interval for the population mean of y or to determine the prediction interval for a single value of y. The limits for the single value of y are wider than the corresponding limits on the mean of y (Remington and Schork, 1970) because single observations vary more than means.

The equation for the confidence interval for the population mean y at $x = x_0$ is (Helsel and Hirsch, 1995)

$$\hat{y} \pm t_{\alpha/2,n-2}\, s \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_x}} \qquad (4\text{-}79)$$

This interval is most narrow at $\bar{x}$ and widens as x_0 moves farther from $\bar{x}$. By calculating the interval at each point along the regression line, a curve such as the dashed line in Figure 4-21 for the example data can be plotted. The equation for the prediction interval for individual values of y at $x = x_0$ is (Helsel and Hirsch, 1995)

$$\hat{y} \pm t_{\alpha/2,n-2}\, s \sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_x}} \qquad (4\text{-}80)$$

Figure 4-21 also shows this interval for the example data.

One of the simplest (in theory) nonpoint source control applications of linear regression is the regression of a water quality indicator against an implementation indicator. For example, flow-adjusted total suspended solids (TSS) concentration could be regressed against a sediment control variable such as the total combined erosion rate of all cropland for which delivery to the stream is likely to be 50 percent or greater. A significant negative slope would suggest (but not prove) that water quality has improved because of implementation of sediment control practices.

Another possible use of simple linear regression is to model a water quality parameter versus time. In this application a significant slope would indicate change over time. The sign of the slope would indicate either improvement or degradation depending on the parameter used. For nonpoint source studies, a simple regression versus time will most likely be confounded by the variability in precipitation and flows. Thus, considerable data manipulation (transformations, stratification, etc.) might be required before regression analysis can be successfully applied. In these cases, it might be more appropriate to apply one of the alternatives to regression described by Helsel and Hirsch (1995).

In many cases water quality parameters are regressed against flow. This is particularly relevant in nonpoint source studies. In analysis of covariance, regressions against flow are often performed prior to an ANOVA (Spooner et al., 1985). One of the implicit goals of nonpoint source control is to change the relationship

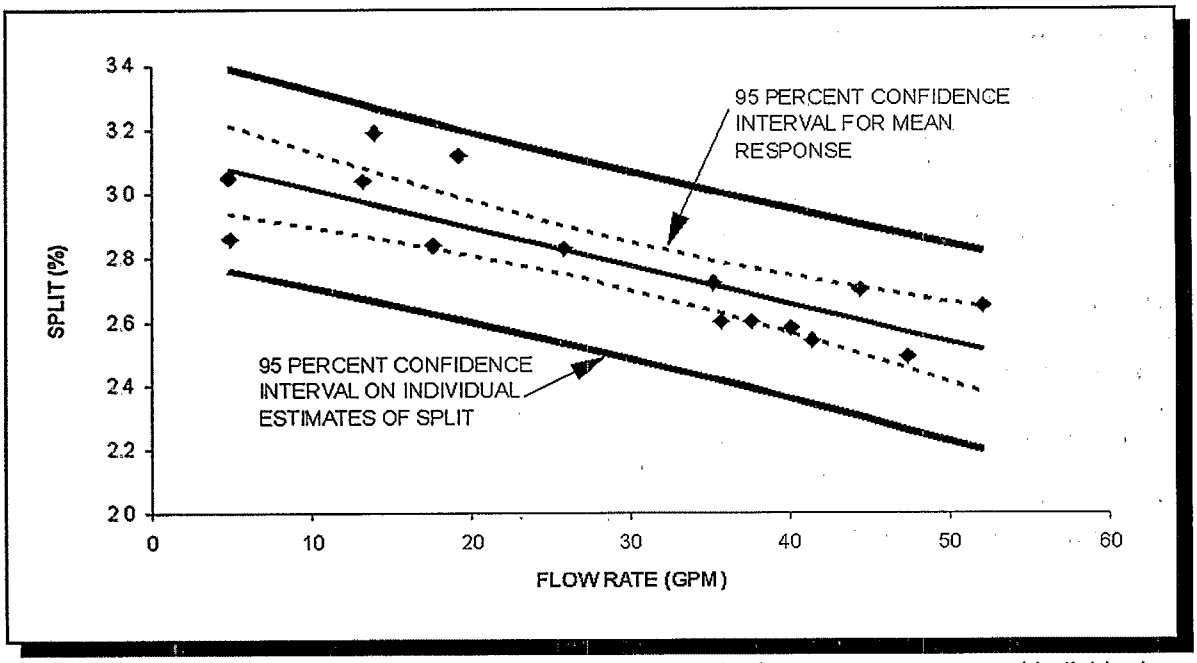


Figure 4-21. Plot of split versus flow rate with confidence limits for mean response and individual estimates.

between flow and pollutant concentration. This will be discussed in greater detail under analysis of covariance.

In paired watershed studies, measured parameters from paired samples are often regressed against each other to compare the watersheds. These regression lines can be compared over time to test for the impact of nonpoint source control efforts (Spooner et al., 1985). This will be discussed in greater detail under analysis of covariance.

4.7.3 Nonlinear Regression and Transformations

The discussion of nonlinear or curvilinear regression is limited to cases where the nonlinear relationship can be transformed into a linear relationship for which simple linear regression can be performed. Data inspection should indicate to the analyst the nature of the relationship between the dependent and independent variables. Possible curvilinear relationships include exponential curves (semi-log), power functions (log-log), and parabolas, among others (Freund, 1973).

Nonlinear regression (as discussed here) involves transformation to linear equations, followed by simple linear regression. Helsel and Hirsch (1995) provide a detailed discussion on transformations using the "bulging rule" described by Mosteller and Tukey (1977), which can be used to select appropriate transformations. Crawford et al. (1983) list the numerous regression models most often applied by the U.S. Geological Survey for flow-adjusting concentrations. The selection of which transformation to use is ultimately based on an inspection of the residuals and whether the assumptions described earlier are met. Typical transformations include x^2, x^3, $\ln x$, $1/x$, $x^{0.5}$, etc.

When the residuals do not exhibit constant variance (heteroscedasticity), one of several common transformations should be used. Logarithmic transformations are used when the standard deviation in the original scale is proportional to the mean of y. Square root transformations are used when the variance is proportional to the mean of y. In many instances, the right transformation will "fix" the nonlinear and heteroscedastic problem. With data that are percentages or proportions (between the values of 0 and 1), the variances at 0 and 1 are small. The arcsin of the square root of the individual values is a common transformation that helps spread out the values near 0 and 1 to increase their variance (Snedecor and Cochran, 1980).

There are several disadvantages when applying transformations to regression applications. The most important issue is that the regression line and confidence intervals are symmetric in the transformed form of the variables. When these lines are transformed back to their normal units, the lines will no longer be symmetrical. The most notable time in hydrology when this creates a problem is when estimating mass loading. To estimate the mass, the means for short time periods are regressed and summed to estimate the total mass over a longer period. This approach is acceptable if no transformations are used—the analyst is summing the means. However, if a log transformation is used, summing the mass over the back-transformed values results in summing the median, which will result in an estimate that is biased low for the total mass (Helsel and Hirsch, 1995).

As an example of nonlinear regression, consider a common relationship that is used to describe load (L) as a function of discharge (Q):

$$L = aQ^{b} \tag{4-81}$$

Taking the logarithms of both sides yields

$$\ln(L) = \ln(a) + b \ln(Q) \tag{4-82}$$

which has the same form as Equation 4-64, introduced at the beginning of this section, where $\ln(L)$ corresponds to y, $\ln(a)$ corresponds to β_0, b corresponds to β_1, and $\ln(Q)$ corresponds to x. By taking the logarithms of both sides, the nonlinear problem has been reduced to a simple linear model. The only additional step that the analyst must perform is to convert L and Q to $\ln(L)$ and $\ln(Q)$ before using standard software. The analyst should be aware that all of the confidence limits are in transformed units; when they are plotted in normal units, the confidence intervals will not be symmetric.

Figure 4-22 demonstrates how transforming the data may improve the regression analysis. In Figure 4-22A, sulfate concentrations (in milligrams per liter) are plotted as a function of stream flow (in cubic feet per second). The apparent downward trend is typical of a stream dilution effect; however, the trend is clearly nonlinear. The trend line plotted in this figure, as well as the residuals plotted in Figure 4-22C, demonstrate that a linear model would tend to over- and underestimate sulfate concentrations depending on the flow. Figure 4-22B displays the same data after computing the logarithms (base 10) of the sulfate and flow data. A trend line fitted to these data and the residual plot (Figure 4-22D) clearly demonstrate that applying linear regression after log transformation would be appropriate for these data.

4.7.4 Multiple Regression

Multiple regression is applied to quantify a relationship between a dependent variable and more than one independent variable (Gaugush, 1986). The assumptions made for simple linear regression also apply to multiple regression (Ponce, 1980a). The method of least squares is also used to determine the best multiple

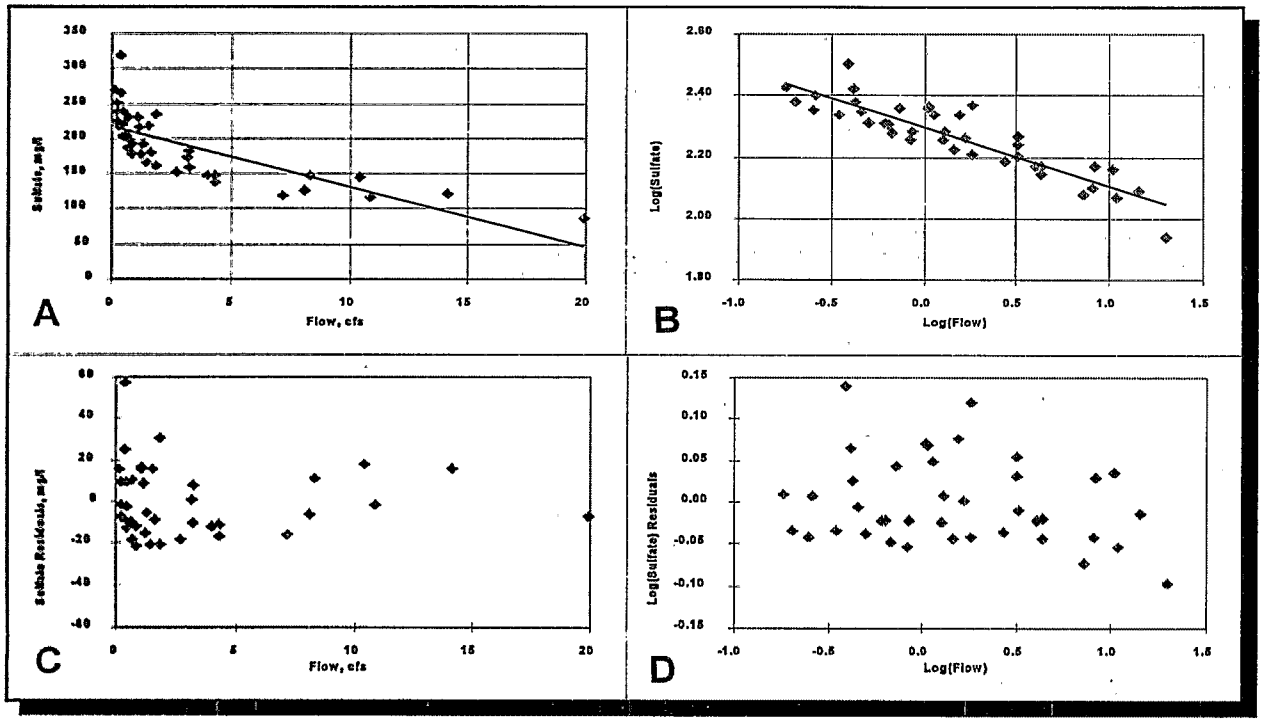


Figure 4-22. Comparison of regression analyses using raw and log-transformed data.

regression line. The general linear model to consider is (Ponce, 1980a)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon \qquad (4\text{-}83)$$

The corresponding normal equations are presented below (Ponce, 1980a).

After solving for the $\beta_1, \ldots, \beta_n$, β_0 can be calculated from (Ponce, 1980a):

$$\left(\sum x_1^2\right)\beta_1 + \left(\sum x_1 x_2\right)\beta_2 + \left(\sum x_1 x_3\right)\beta_3 + \ldots + \left(\sum x_1 x_n\right)\beta_n = \sum x_1 y \qquad (4\text{-}84)$$

$$\left(\sum x_1 x_2\right)\beta_1 + \left(\sum x_2^2\right)\beta_2 + \left(\sum x_2 x_3\right)\beta_3 + \ldots + \left(\sum x_2 x_n\right)\beta_n = \sum x_2 y \qquad (4\text{-}85)$$

$$\left(\sum x_1 x_3\right)\beta_1 + \left(\sum x_2 x_3\right)\beta_2 + \left(\sum x_3^2\right)\beta_3 + \ldots + \left(\sum x_3 x_n\right)\beta_n = \sum x_3 y \qquad (4\text{-}86)$$

. . .

$$\left(\sum x_1 x_n\right)\beta_1 + \left(\sum x_2 x_n\right)\beta_2 + \left(\sum x_3 x_n\right)\beta_3 + \ldots + \left(\sum x_n^2\right)\beta_n = \sum x_n y \qquad (4\text{-}87)$$

$$\beta_0 = \bar{y} - \beta_1\bar{x_1} - \beta_2\bar{x_2} - \ldots - \beta_n\bar{x_n} \quad (4\text{-}88)$$

Ponce (1980a) presents a hand-computed example of multiple regression using three independent variables. The reader is encouraged to follow through that example to develop an understanding of multiple regression before using computerized procedures. Gaugush (1986) states that multiple regression with two independent variables can be performed using textbook formulas, but that matrix algebra is required for broader applications. Winer (1971) provides a matrix algebra approach to multiple regression, but the discussion is complicated and probably not critical to appropriate use of multiple regression techniques (especially when the analyst consults a statistician).

Gaugush (1986) also provides an example of multiple regression in which the SAS procedure GLM (SAS Institute, Inc., 1985b) is used. This example relates pollutant level to three independent variables—distance from source, temperature, and discharge. An interpretation of the SAS output is also provided.

Key points made in the examples above include:

- An F test indicates the significance of the regression.
- The coefficient of multiple determination (R^2), which is calculated as in simple linear regression, shows the proportion of variation in y explained by the model.
- Computerized output such as that from SAS can be used to refine the model for subsequent runs.

As a further note regarding use of SAS, the RSQUARE procedure (SAS Institute, Inc., 1985b) can be used in an exploratory fashion to perform all possible multiple regressions for subsets of independent variables, listing the models in decreasing order of R^2 magnitude. Thus, the model with the largest R^2 value will be listed first. The STEPWISE procedure allows five approaches to stepwise regression for users who wish to determine which variables should be included in a regression model (SAS Institute, Inc., 1985b). However, this procedure is not guaranteed to identify the model with the largest R^2. Other computer software packages, such as SPSS (Statistical Package for the Social Sciences), can also be used for multiple regression (Ingwersen, 1980).

The following discussion of R^2, taken largely from a technical nonpoint source newsletter (Spooner, 1984), emphasizes proper interpretation of R^2 values.

The purpose of regressing a response variable (y) on one or more independent variables (x) is to "explain" some of the variation observed in the measured values in y. The F tests for each individual x variable can be used to determine whether they are individually important to the regression on y. R^2 is a measure of the fraction of variation in y explained by the linear regression on $x_1, x_2, \ldots, x_n$ variables in the model. Specifically, R^2 is the fraction of the sum of squares (SS) of the deviations of y from its mean that is attributed to the regression. R^2 values range from 0 (model useless) to 1 (model perfect)(Equation 4-87).

H_o that

$$R^2 = 0 \text{ (i.e., } \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0)$$

can be tested using the F statistic to determine whether the regression model explains any of the variation in Y. The F statistic is $(n\text{-}k\text{-}1)\ R^2/(k\text{-}1)(1\text{-}R^2)$ with $(k\text{-}1)$ and $(n\text{-}k\text{-}1)$ degrees of freedom. It should be noted that $(k\text{-}1)$ is the degrees of freedom for the regression

model SS and (n-k-1) is the degrees of freedom for the error SS.

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - F)^2}{\sum_{i=1}^{n} (Y_i - F)^2} = \frac{SS\ Regression}{SS\ Total}$$
$$= 1 - \frac{SS\ Error}{SS\ Total} \qquad (4\text{-}89)$$

A small R^2 might be significantly different from zero if n is large. Conversely, a large R^2 might be insignificant if n is small compared to the number of x's in the model.

If R^2 is small, most of the variation in Y is unexplained by the linear regression model. This remaining "noise" might be random variation, or it might be due to other independent variables not considered in the regression. If these other variables are added to the regression, the relationships among the x's already included might change.

When new variables are added to the model, R^2 always increases although the adjusted R^2 might not increase. This explains why a large R^2 might not be meaningful when the sample size is small. Also, it is not legitimate to compare two models with different numbers of x's solely by their R^2 values. However, R^2, adjusted for the degrees of freedom, may be used to compare models, where adjusted R^2 is

$$R_a^2 = (1 - R^2)\ (n - 1)\ /\ (n-k-1) \qquad (4\text{-}90)$$

How does one test whether a new variable added to a model adds significant information to explain further the variation in y (i.e., is the increase in R^2 significant)? In SAS, for example, the "type III SS or IV SS" (also known as the partial sum of squares) and their associated F tests can be used. These statistics measure the amount of variation in y explained by the addition of an individual x after all other x's are in the model. An equivalent method is to compare the SSE (sum of squares due to error) from "full" and "reduced" models (i.e., SSE from models with and without, respectively, the extra term in question). If the SSE is reduced significantly by the addition of a new variable to the model, the variable is important. The F statistic is

$$F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \qquad (4\text{-}91)$$

where df_R and df_F are the degrees of freedom for the reduced model SS and full model SS, respectively.

4.7.5 Multivariate Regression

Multivariate regression can be a very useful technique in nonpoint source monitoring and evaluation efforts. It involves the development of a linear model to relate two or more dependent variables to two or more independent variables. A detailed discussion of the theory behind multivariate regression is beyond the scope of this document. Readers are referred to statistics texts (e.g., Srivastava and Khatri, 1979) for more on multivariate regression. Multivariate regressions are designed to take into account the correlation structure of the x's and y's to reduce the overall variance.

Users of SAS (SAS Institute, Inc., 1985b) can use the REG procedure for multivariate regression. An example of the MODEL statement used in this procedure is the following (SAS Institute, Inc., 1985b):

MODEL Y1 Y2 = X1 X2 X3

where

Y1 and Y2 are the dependent variables and X1, X2, and X3 are the independent variables.

Within this procedure the MTEST statement can be used to test hypotheses regarding the multivariate regression model. F values are calculated for the following procedures (SAS Institute, Inc., 1985b):

- Wilks' lambda
- Pillai's trace
- Hotelling-Lawley trace
- Roy's maximum root

4.8 Analysis of Covariance

Suppose an analyst is interested in evaluating BMPs by comparing data collected from a paired watershed design. Data are collected from two watersheds during two periods—calibration and treatment. During calibration, neither watershed has a BMP in place, while during the later period, one of the two watersheds has a BMP installed. A natural extension of the regression techniques described in Section 4.7 is to compare regression equations between the treatment watershed and the control watershed, with one regression equation developed during the calibration phase and the second regression equation developed during the treatment phase. The analysis of covariance (ANCOVA), a procedure that combines features of ANOVA and regression, can be used to evaluate this situation. ANCOVA can also be used to test for differences in the average value for a dependent variable (e.g., sediment concentration) between the levels of a group variable (e.g., seasons or years) after adjusting for an independent variable (e.g., flow or upstream concentration).

A typical ANCOVA model in which the slopes and intercepts for the two groups are suspected to be different can be represented as (Helsel and Hirsch, 1995)

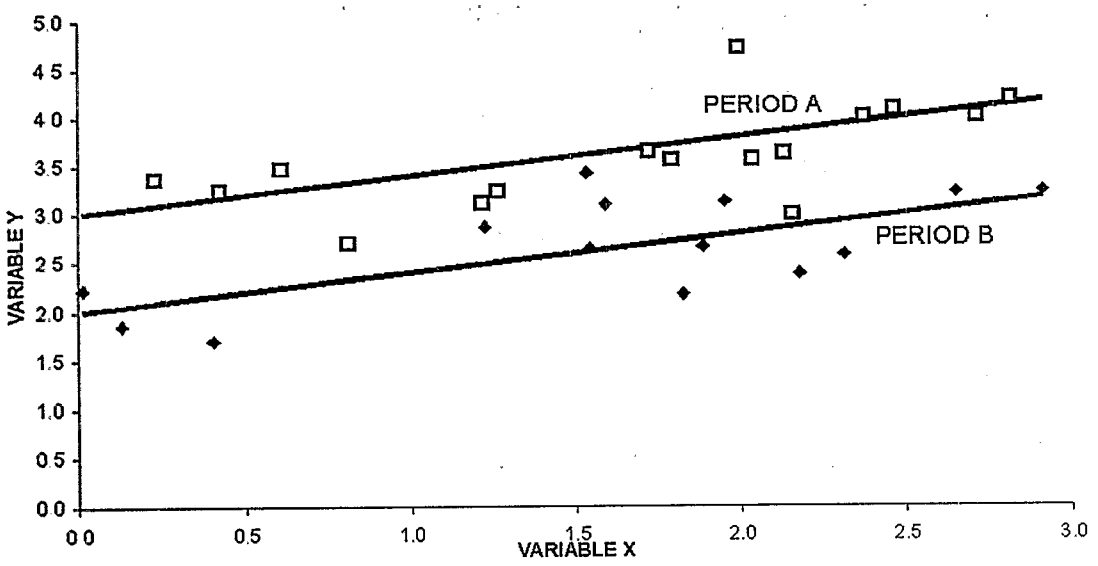
$$y = (\beta_0 + \beta_2 Z) + (\beta_1 + \beta_3 Z)x + \epsilon \qquad (4\text{-}92)$$

where Z is a binary variable that is equal to 0 or 1 depending on which group x and y are from. For example, Z could be 0 during calibration and 1 during treatment of a paired watershed analysis. In this case, β_0 and $\beta_0+\beta_2$ are the intercepts during the calibration and treatment periods, respectively. β_1 and $\beta_1+\beta_3$ are the slopes during the calibration and treatment periods, respectively. If β_2 is nonzero and β_3 is zero, the regression produced by Equation 4-92 would be a pair of parallel lines (Figure 4-23A). If β_2 and β_3 are nonzero, the regression produced by Equation 4-92 would be a pair of lines like those presented in Figure 4-23B.

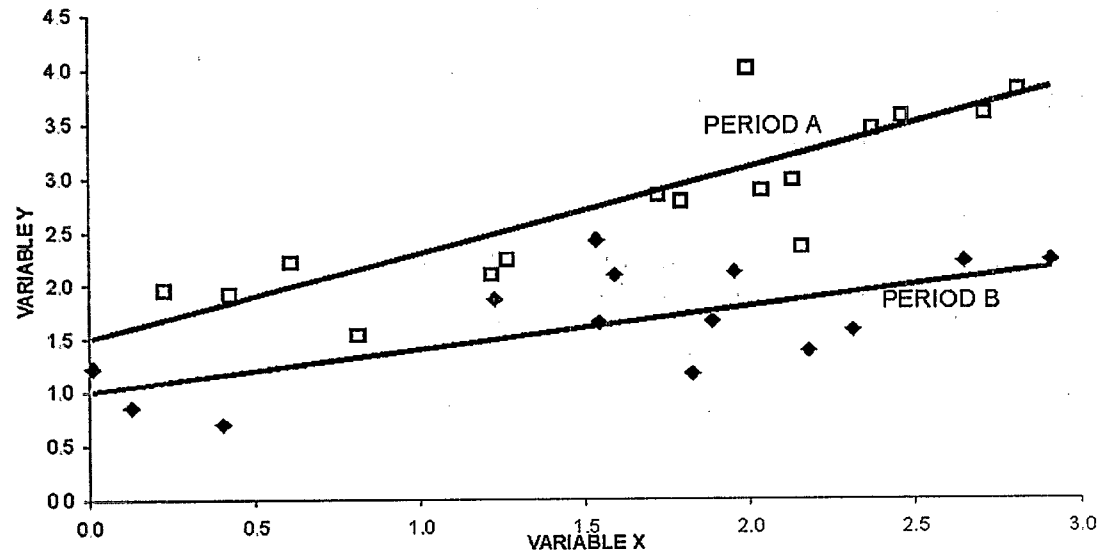
The remainder of this discussion follows an analysis performed for field runoff (cm) during the conversion from conventional to conservation tillage in Vermont (USEPA, 1993c). Two watersheds were monitored during a calibration period during which 49 (n_1) paired observations of runoff were made. Figure 4-24A is a bivariate log-log plot of storm runoff for the treatment watershed as a function of storm runoff for the control watershed. Based on an inspection of this plot, it seems reasonable to perform the analyses using log-transformed (base 10) data.

A regression analysis was performed on these data to determine whether there was a significant relationship between the watersheds, whether enough data had been collected during calibration, and whether the residual errors were smaller than the expected BMP effect. A summary of the regression ANOVA is provided in Table 4-32 (with $n_1 = 49$, $SS_y = 148.441$, $SS_x = 70.933$, and $S_{xy} = 78.463$). (Equations 4-65 through 4-67 and Table 4-31 can be used to hand-check the table entries.) The p value associated with the resulting F statistic indicates that the model explains a significant proportion of the variation.

To determine whether enough calibration data have been collected, the ratio of the MSE to the

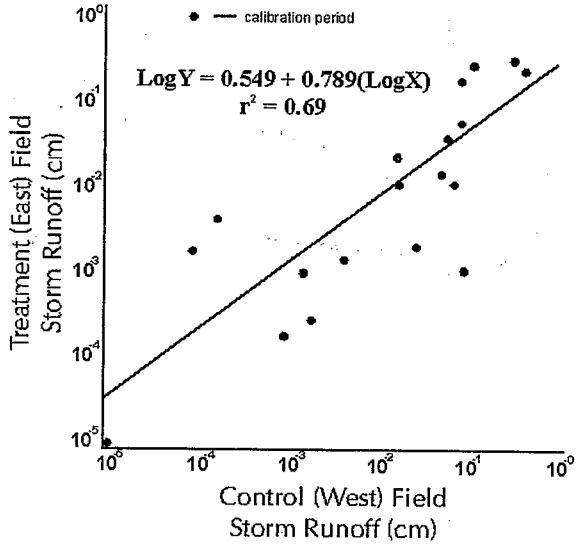


A) Variable y versus x for two periods (constant slope).

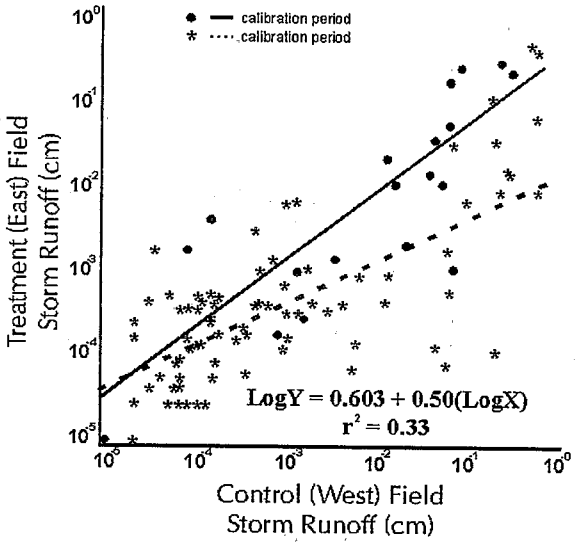


B) Variable y versus x for two periods (variable slope).

Figure 4-23. Comparison of regression equations for data from two periods.



Calibration Period Regression



Treatment and Calibration Period Regression

Figure 4-24. Storm runoff from calibration and treatment periods in Vermont. (Source: EPA, 1993c).

Table 4-32. ANOVA for regression of treatment watershed runoff on control watershed runoff during calibration.

Source of Variation	SS	df	MS	F	Significance F
Regression	86.792	1	86.79	66.17	0.0001
Residual	61.649	47	1.31		
Total	148.441	48			

smallest worthwhile difference (d) can be compared using the following formula (EPA, 1993c):

$$\frac{MSE}{d^2} = \frac{n_1 n_2}{n_1 + n_2}\left(\frac{1}{F\left(1 + \frac{F}{n_1 + n_2 - 2}\right)}\right) \qquad (4\text{-}93)$$

where n_1 and n_2 are the number of observations in the calibration and treatment periods, respectively, and F is from Table D6 with 1 and $n_1 + n_2$ - 3 degrees of freedom. If the treatment period has not been initiated, assume that $n_1 = n_2$. Using the example data where $\bar{x}$ of the log-transformed data is -2.518, the number of observations necessary to detect a 20 percent change can be estimated. The left side of the above equation would be equal to $1.31/(0.2 \times -2.518)^2$ or 5.2. With $n_1 = n_2 = 49$ and $F = 3.94$ ($p = 0.95$, 1 and 95 df), the right side of the above equation can be evaluated as 6.0. Since the left side of the equation is less than the right side, there would be enough samples to detect a 20 percent change in discharge. Equation 4-79 can be used to determine the confidence bands for the regression equation, which allow determining the level of change needed to have a significant treatment effect.

Once the treatment period data have been collected, the same type of regression analysis is performed. Following this step, the significance of an overall regression (which combines calibration and treatment data) can be evaluated and the difference between the individual slopes and intercepts can be evaluated. Continuing with the example, a summary of the regression ANOVA for the treatment period is provided in Table 4-33 (with $n_2 = 114$, $SS_y = 135.0$, $SS_x = 227.43$, and $S_{xy} = 101.32$). The p value associated with the resulting F statistic indicates that the model explains a significant proportion of the variation.

The ANCOVA can be performed by combining the results from Tables 4-32 and 4-33. Table 4-34 demonstrates the general format for performing ANCOVA hand calculations. Note that Σ indicates summation of terms. This approach is applied to the example data with the results presented in Table 4-35. Table 4-36 presents the same calculations performed with SAS. (An appropriate SAS program is provided below.) The ANCOVA indicates that the overall treatment and calibration regressions were significantly different and that the slopes and intercepts of the equations were also different. The difference in slopes is evident from Figure 4-24B. The small differences between the calculations in Tables 4-35 and 4-36 are due to rounding errors. If there are not significant

Table 4-33. ANOVA for regression of treatment watershed runoff on control watershed runoff during treatment.

Source of Variation	SS	df	MS	F	Significance F
Regression	45.13	1	45.13	56.25	0.0001
Residual	89.87	112	0.80		
Total	135.00	113			

Table 4-34. ANCOVA for comparing regression lines.

Source	df	Ss_x	S_{xy}	Ss_y	β_1	df	SS (res.)	MS	F
Within									
Calibration	n_1-1	Eq. 4-65	Eq. 4-65	Eq. 4-68	S_{xy}/SS_x	n_1-2	$SS_y-(S_{xy})^2/SS_x$	SS/df	--
Treatment	n_2-1	Eq. 4-65	Eq. 4-65	Eq. 4-68	S_{xy}/SS_x	n_2-2	$SS_y-(S_{xy})^2/SS_x$	SS/df	--
					Error:	Σ	Σ	SS/df	
Slopes	n_1+n_2-2	Σ	Σ	Σ	S_{xy}/SS_x	n_1+n_2-3	$SS_y-(S_{xy})^2/SS_x$	SS/df	
					Slope difference:	1	Slope SS-Error SS	SS/df	MS/Error MS
						1	Comb. SS- Slope SS	SS/df	MS/Slope MS
Intercepts	n_1+n_2-1	combined data				n_1+n_2-2	$SS_y-(S_{xy})^2/SS_x$		

Table 4-35. ANCOVA for comparing regression lines from calibration and treatment (hand calculations).

Source	df	Ss_x	S_{xy}	Ss_y	β_1	df	SS (res.)	MS	F
Within									
Calibration	48	70.933	78.463	148.441	1.106	47	61.650	1.3117	--
Treatment	113	227.430	101.315	135.00	0.445	112	89.866	0.8024	--
					Error:	159	151.516	0.9529	
Slopes	161	298.363	179.778	283.441	0.603	160	175.116	1.0945	
					Slope difference:	1	23.600	23.600	24.77*
						1	5.8453	5.8453	5.34†
Intercepts	162	311.671	178.762	283.492	--	161	180.961		

* Significant at p = 0.001
† Significant at p = 0.05

Table 4-36. ANCOVA for comparing regression lines from calibration and treatment (computerized software).

Source of Variation	df	MS	F	Significance F
Model	3	43.99	46.17	0.001
Error	159	0.95		
Overall	1	103.09	108.18	0.0001
Intercept	1	5.47	5.74	0.0178
Slope	1	23.42	24.58	0.0001

SAS Program to Compare Regression Lines

```
PROC GLM;
 CLASS PERIOD;
 MODEL LOGFLOW2 = LOGFLOW1 PERIOD
   LOGFLOW1*PERIOD;
 RUN;

/* LOGFLOW1 = log-transformed data from control
watershed */
/* LOGFLOW2 = log-transformed data from
treatment watershed */
/* PERIOD = indicator for whether paired data were
from the calibration or treatment period */
```

differences between the slopes, all of the periods can be represented by a common slope and the relationship between y and x is constant over the tested period.

4.9 Evaluation of Time Series

In nonpoint source data analysis, we often want to know whether there is a tendency for a pollutant concentration to increase or decrease over time. If such a tendency exists, we say there is a trend. Trend analysis is often used to determine whether the implementation of a BMP actually reduces the pollutants in a stream, or whether the development of an urban area is causing the deterioration of water quality downstream, as well as maintaining a status of ambient water quality conditions. A trend can be visually examined by plotting the observed data versus time. A statistical test is required to analyze the trend. This section describes statistical procedures for detecting and evaluating monotonic (continuously nonincreasing or nondecreasing) trends in a single time series (e.g., 10 years of monthly TSS at a single station) and presents several methods for evaluating temporal correlation.

The first issue to consider is when a monotonic trend test should be used. The most important factor before beginning the analysis is to assess whether any interventions or activities led to the hypothesis that a shift in water quality might have occurred. For example, suppose a BMP to reduce sediment loadings was installed during the course of the monitoring program. A shift in TSS concentration (hopefully downward) after BMP installation would be expected. In this case, it is more appropriate to divide the data into "before" and "after" groups and analyze the data using the two independent random sample procedures described in Section 4.5. On the other hand, if a series of BMPs are being implemented across a watershed over several years and monitoring is being performed in a downstream estuary, the changes would be expected to be gradual. In this case a monotonic trend test might be more appropriate. If there is no hypothesis to naturally divide the data, it is also best to use a monotonic trend test. Concentration data should not be used to determine data groupings for the purposes of developing hypotheses or selecting between a two-sample or monotonic trend test.

The second issue to consider is the case where sampling was interrupted for several years in the middle of a 10-year monitoring effort. It is suggested that if the data gap is greater than one-third of the total data record, it is better to use a two-sample test (Helsel and Hirsch, 1995). A similar issue to consider is the case where several data records will be examined, but they have different starting and stopping points. Helsel and Hirsch (1995) suggest that the analyst divide the data record into three periods of equal length; if any third of the record has more than 20 percent missing values, that record should not be used.

The final issue to consider is whether to account for exogenous variables (e.g., flow, temperature, rainfall) before testing for trends. A common example is the approach used by the USGS to account for flow variability in its National Stream Quality Accounting Network (NASQAN) stations. In USGS analyses, water quality variable concentrations are adjusted to account for flow.

The flow-adjusted concentrations are then evaluated for trends. These adjustments can be made using simple linear regression analyses, as discussed in Section 4.7 or the nonparametric procedures (e.g., locally weighted scatter plot smoothing) discussed by Helsel and Hirsch (1995). The purpose of adjusting the data for an exogenous variable is to reduce the background noise so that the detection of time trends is more powerful.

There are several methods to detect monotonic trends in time series. Regression analyses have already been discussed in Section 4.7. To apply linear regression where time is the independent variable, all of the assumptions listed in Table 4-28 are necessary. If these assumptions are met, linear regression is an acceptable approach. Significant trends are declared when the slope term, β_1, is significantly different from zero. Multivariate regression procedures that model the water quality variable as a function of an exogenous variable (e.g., flow) and time simultaneously can also be used to detect trends if the regression assumptions are met. When evaluating several data sets for a single report, these assumptions are rarely met for all of the data sets. In these cases, nonparametric procedures are recommended. This is not to say that data transformations for nonparametric tests are not desirable, as will be discussed later. Since simple linear and multivariate regression have been discussed, this section is limited to discussing the Mann-Kendall τ and the Seasonal Kendall tests. Both are nonparametric.

Following the monotonic trend discussion, procedures for computing the autocorrelation coefficient and Spearman's rho are provided. These procedures are useful for evaluating whether the data are truly independent, one of the fundamental assumptions in the procedures described next. If the data are serially correlated, it is possible to systematically sample from the data set, to group the data into time periods and use a summary statistic (e.g., time- or volume-weighted mean or median), or to use more advanced time series analysis procedures (Helsel and Hirsch, 1995) to analyze these data.

4.9.1 Monotonic Trends

Regression

Refer to Section 4.7 for a discussion on simple linear and multivariate regression.

Mann-Kendall τ Test

The Mann-Kendall τ test analyzes the sign of the difference between later-measured data and the earlier-measured data. Each later-measured datum is compared to all data measured earlier. This approach results in a total of n(n-1)/2 possible pairs of data, where n is the total number of observations in the time series. The Mann-Kendall τ test assumptions include the typical requirements that the data be independent and that one value can be declared larger than, smaller than, or equal to another value. The third assumption is similar to the regression requirements that the residuals must have a constant variance, but no distribution requirements are necessary.

The usual hypotheses for a Mann-Kendall τ test is whether y tends to increase or decrease with time (Helsel and Hirsch, 1995):

Mann-Kendall τ Test Assumptions

- The random variables $y_1, y_2, \ldots, y_i, \ldots, y_j, \ldots, y_n$ are mutually independent.
- The measurement scale of the data is at least ordinal (i.e., y_i can be declared as $<$, $>$, or $= y_j$).
- The data are identically distributed with only a shift in the central location if there is a trend.

Two-sided test

H_o: Prob $[y_j > y_i] = 0.5$ where $t_j > t_i$
H_a: Prob $[y_j > y_i] \neq 0.5$ (two-sided test)

One-sided test

H_o: Prob $[y_j > y_i] = 0.5$ where $t_j > t_i$
H_a: Prob $[y_j > y_i] > 0.5$ (one-sided test, increasing trend)

The next step is to compute the difference between the later-measured value and all earlier-measured values, $(y_j - y_i)$, where $j > i$ and assign the integer value of 1, 0, -1 to positive differences, no differences, and negative differences, respectively. The test statistic, S, is then computed as the sum of the integers:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(y_j - y_i) \qquad (4\text{-}94)$$

where sign(•) is equal to 1, 0, or -1 as indicated above. This task is most easily accomplished assuming the data are ordered in increasing time order. When S is a large positive number, later-measured values tend to be larger than earlier-measured values and there might be an upward trend. When S is a large negative number, later-measured values tend to be smaller than earlier-measured values and there might be a downward trend. When the absolute value of S is small, there might be no trend. The test statistic, τ, can be computed as

$$\tau = \frac{S}{n(n-1)/2} \qquad (4\text{-}95)$$

which has a range of -1 to 1 and is analogous to the correlation coefficient in regression analyses. Computing S or τ becomes tedious when n is large. Gilbert (1987) provides a FORTRAN program to alleviate the computation effort. S and τ are invariant to transformations such as logs (i.e., S and τ will be the same value whether the raw or log-transformed data are used).

For sample sizes greater than 10, the large sample approximation can be used to compute a test statistic that can be compared to a normal distribution using the following equation:

$$Z_S = \begin{cases} \dfrac{S-1}{\sigma_S} & \text{if } S>0 \\ \dfrac{S+1}{\sigma_S} & \text{if } S<0 \end{cases} \qquad (4\text{-}96)$$

where

$$\sigma_S = \sqrt{\frac{n}{18}(n-1)(2n+5)} \qquad (4\text{-}97)$$

for when there are no ties or

$$\sigma_S = \sqrt{\frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i(i)(i-1)(2i+5)}{18}} \qquad (4\text{-}98)$$

for when there are ties, where Z_S is zero if S is zero and t_i is equal to the number of ties of extent i. Z_S is compared to the critical z value from Table D1. For a two-sided 95 percent confidence level, the critical z value would be ± 1.96. If Z_S is not contained within this range, reject H_o. See Helsel and Hirsch (1995) for sample sizes of 10 or less. To determine t_i, consider the following 20 observations that are in ascending order:

<1, <1, <1, 4, 4, 6, 6, 8, 8, 10, 11, 11, 11, 11, 16, 19, 20, 22, 32, 45

In this example there are seven ties of extent 1 (i.e., no ties), three ties of extent 2 (4, 4, 6, 6, 8, 8), one tie of extent 3 (<1, <1, <1), one tie of extent 4 (11, 11, 11, 11), and zero ties of extent 5 and greater. Thus, the summation term that includes t_i from above can be evaluated as

7x1x0x7 + 3x2x1x9 + 1x3x2x11 + 1x4x3x13 = 276

Table 4-37 presents a list of annual rainfall for 21 years. Table 4-38 presents the intermediate calculations for computing S. The top portion of Table 4-38 is a table of the differences $y_j - y_i$, for example $y_2 - y_1 = 13.2$. Observations y_8 through y_{16} were omitted from Table 4-38 for presen-tation purposes. The bottom portion of Table 4-38 presents the intermediate calculations for sign$(y_j - y_i)$. Summing these values (including those not presented in this table) yields a value of 12. Since there were no ties, $Z_S = (12\text{-}1)/(1096.7)^{0.5}$ or 0.33, H_o is accepted—there is no trend in the rainfall data.

Table 4-37. Annual total rainfall for 21 years.

Year	Rainfall (in.)	Year	Rainfall (in.)
1	40.2	12	51.2
2	53.4	13	54.3
3	43.5	14	41.5
4	37.7	15	44.8
5	50.2	16	46.7
6	38.7	17	51.8
7	47.8	18	49.5
8	39.5	19	34.1
9	44.9	20	33.2
10	41.7	21	53.7
11	36.4		

Had there been a significant trend in the data, the Sen slope estimator could be estimated as (Helsel and Hirsch, 1995)

$$\beta_1 = median\left(\frac{y_j - y_i}{x_j - x_i}\right) \qquad (4\text{-}99)$$

for all $i < j$ and $i = 1, 2, \ldots, n\text{-}1$ and $j = 2, 3, \ldots, n$; in other words, computing the slope for all pairs of data that were used to compute S. The median of these slopes is the Sen slope estimator. Using the rainfall data as an example, the slope between y_4 and y_2 is equal to (37.7-53.4)/(4-2) or -7.9. Had there been a significant trend, this process would have been carried out for the remaining pairs of observations and the median slope selected as the Sen slope estimator.

As might be expected, any linear slope estimator is a poor choice when the apparent slope is exponential. In Section 4.7.3, transformations to reduce the analysis to a linear problem were discussed. These same approaches are also appropriate here. So while it does not matter for computing S or τ that the trend be linear, transforming the data prior to computing the slope estimator might be useful. For example, if the data were transformed using natural logs, the percentage change from year to year in the above example would be estimated as $(e^{\beta_1} - 1) \times 100$ (Helsel and Hirsch, 1995).

Seasonal Kendall test

In the nonpoint source area, many data follow seasonal patterns. The decision to use a seasonal Kendall test (Hirsch et al., 1982) can usually be made by examining boxplots by season. The test statistic is computed by performing a Mann-Kendall calculation for each season and then combining the results for each season. That is, if sampling is monthly, January observations are compared only to other January observations, etc. Thus S_k is computed as the sum of the S from each season (Helsel and Hirsch, 1995):

$$S_k = \sum_{i=1}^{m} S_i \qquad (4\text{-}100)$$

Table 4-38. Analysis of rainfall data using Mann-Kendall τ test.

	y_j	y_1	y_2	y_3	y_4	y_5	y_6	y_7	…	y_{17}	y_{18}	y_{19}	y_{20}	y_{21}
	y_i	40.2	53.4	43.5	37.7	50.2	38.7	47.8	…	51.8	49.5	34.1	33.2	53.7
y_1	40.2		13.2	3.3	-2.5	10.0	-1.5	7.6	…	11.6	9.3	-6.1	-7.0	13.5
y_2	53.4			-9.9	-15.7	-3.2	-14.7	-5.6	…	-1.6	-3.9	-19.3	-20.2	0.3
y_3	43.5				-5.8	6.7	-4.8	4.3	…	8.3	6.0	-9.4	-10.3	10.2
y_4	37.7					12.5	1.0	10.1	…	14.1	11.8	-3.6	-4.5	16.0
y_5	50.2						-11.5	-2.4	…	1.6	-0.7	-16.1	-17.0	3.5
y_6	38.7							9.1	…	13.1	10.8	-4.6	-5.5	15.0
y_7	47.8									4.0	1.7	-13.7	-14.6	5.9
…	…										…	…	…	…
y_{17}	51.8										-2.3	-17.7	-18.6	1.9
y_{18}	49.5											-15.4	-16.3	4.2
y_{19}	34.1												-0.9	19.6
y_{20}	33.2													20.5
y_{21}	53.7													
		40.2	53.4	43.5	37.7	50.2	38.7	47.8	…	51.8	49.5	34.1	33.2	53.7
y_1	40.2		1	1	-1	1	-1	1	…	1	1	-1	-1	1
y_2	53.4			-1	-1	-1	-1	-1	…	-1	-1	-1	-1	1
y_3	43.5				-1	1	-1	1	…	1	1	-1	-1	1
y_4	37.7					1	1	1	…	1	1	-1	-1	1
y_5	50.2						-1	-1	…	1	-1	-1	-1	1
y_6	38.7							1	…	1	1	-1	-1	1
y_7	47.8									1	1	-1	-1	1
…	…										…	…	…	…
y_{17}	51.8										-1	-1	-1	1
y_{18}	49.5											-1	-1	1
y_{19}	34.1												-1	1
y_{20}	33.2													1
y_{21}	53.7													

where S_i is S from the i^{th} season and m is the number of seasons. Z_{Sk} is estimated as

$$Z_{Sk} = \begin{cases} \dfrac{S_k - 1}{\sigma_{Sk}} & \text{if } S_k > 0 \\ \dfrac{S_k + 1}{\sigma_{Sk}} & \text{if } S_k < 0 \end{cases} \qquad (4\text{-}101)$$

or Z_{Sk} is zero if S_k is zero and

$$\sigma_{Sk} = \sqrt{\sum_{i=1}^{m} \frac{n_i}{18}(n_i - 1)(2n_i + 5)} \qquad (4\text{-}102)$$

where n_i is the number of observations in the i^{th} season.

4.9.2 Correlation Coefficients

Spearman's rho

Spearman's rho test is used to detect whether there is a correlation between paired data. Spearman's rho is computed as (Conover, 1980)

$$\rho = \frac{\sum_{i=1}^{n}\left[R(x_i)-\frac{n+1}{2}\right]\left[R(y_i)-\frac{n+1}{2}\right]}{n(n^2-1)/12} \qquad (4\text{-}103)$$

where R(•) represents the rank of the observation and n is the number of observations. If there are ties, Equation 4-104 may be used.

The resulting value of ρ is then compared to critical values in Table D10. Spearman's rho can be used in the same manner as the τ statistic computed in Section 4.9.1. Spearman's rho can also be used to evaluate serial correlation by setting $y_i = x_{i+k}$ to determine the lag-k autocorrelation. For $k = 1$, the first observation is compared to the second observation, the second observation to the third observation, and so on.

Using the rainfall data, Table 4-39 presents the intermediate calculations for Spearman's rho for $k = 1$. Notice that $y_i = x_{i+1}$ and that there are only 20 observations in this analysis. The third and fourth represent the ranks of x_i and y_i, respectively. The remaining three columns are intermediate calculations for the numerator of the above equation. Finally, ρ is equal to -126/[(20(20²-1)/12] or -0.19. Assuming a two-sided hypothesis, the critical value from Table D10 (with $n = 20$ and $\alpha = 0.05$) is ±0.4451; the rainfall data are not correlated at lag-1. This result cannot be compared with the previous example. In the previous example the correlation between annual rainfall and time was evaluated. In this example, "this year's annual rainfall" is compared to "next year's annual rainfall."

Autocorrelation coefficient

The analyst may also use the correlation coefficient, r. Salas et al. (1980) provided the formula for the lag-k autocorrelation coefficient as:

$$r = \frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{n-k}(x_t-\bar{x})^2} \qquad (4\text{-}105)$$

Anderson (1941) gave the limit

$$r_k = \frac{-1\pm 1.96(n-k-1)^{0.5}}{n-k} \qquad (4\text{-}106)$$

for the 95 percent probability levels for the lag-k autocorrelation coefficient where n is the sample size.

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}\left(\sum_{i=1}^{n} R(y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}} \qquad (4\text{-}104)$$

Table 4-39. Analysis of rainfall data using Spearman's rho.

x_i	y_i	$R(x_i)$	$R(y_i)$	$R(x_i)-(n+1)/2$	$R(x_i)-(n+1)/2$	Numer.
40.2	53.4	7	18	-3.5	7.5	-26.25
53.4	43.5	19	9	8.5	-1.5	-12.75
43.5	37.7	10	4	-0.5	-6.5	3.25
37.7	50.2	4	15	-6.5	4.5	-29.25
50.2	38.7	16	5	5.5	-5.5	-30.25
38.7	47.8	5	13	-5.5	2.5	-13.75
47.8	39.5	14	6	3.5	-4.5	-15.75
39.5	44.9	6	11	-4.5	0.5	-2.25
44.9	41.7	12	8	1.5	-2.5	-3.75
41.7	36.4	9	3	-1.5	-7.5	11.25
36.4	51.2	3	16	-7.5	5.5	-41.25
51.2	54.3	17	20	6.5	9.5	61.75
54.3	41.5	20	7	9.5	-3.5	-33.25
41.5	44.8	8	10	-2.5	-0.5	1.25
44.8	46.7	11	12	0.5	1.5	0.75
46.7	51.8	13	17	2.5	6.5	16.25
51.8	49.5	18	14	7.5	3.5	26.25
49.5	34.1	15	2	4.5	-8.5	-38.25
34.1	33.2	2	1	-8.5	-9.5	80.75
33.2	53.7	1	19	-9.5	8.5	-80.75
33.2	-	-	-	-	-	-
					Sum	-126.00

4.10 *MULTIVARIATE ANALYSES*

There are several multivariate procedures in addition to the multivariate regression discussed in 4.7.4. Mathematical descriptions of these procedures are beyond the scope of this guidance, but researchers should consult a statistician to assess the opportunities for using these procedures. In general, the multivariate procedures described in this section have not found wide usage in day-to-day applications.

With the current availability of computerized statistical procedures (e.g., SAS, SPSS), it is possible to perform multivariate analyses with ease, requiring of the researcher only that he or she understands and meets the assumptions of the particular test and knows how to interpret correctly the results of the test. It is extremely important that a qualified statistician be consulted regarding the assumptions involved and the appropriate interpretation of test results. Without such precautions, our current computer technology will only facilitate the proliferation of misguided analyses and misinterpreted results.

The multivariate analyses described briefly in this guidance include canonical correlation, cluster analysis, principal components and factor analysis, and discriminant analysis. These procedures were selected for discussion based on the work of Gaugush (1986), which should be

reviewed in addition to the detailed discussions provided in statistics texts for a better understanding of these multivariate analyses.

4.10.1 Canonical Correlation

Canonical correlation is a technique for analyzing the relationship between two sets of variables, with each set able to contain several variables (SAS Institute, Inc., 1985b). It follows that simple and multiple correlation are special cases of canonical correlation in which one or both sets of variables contain only one variable (SAS Institute, Inc., 1985b).

Gaugush (1986) states that "[c]anonical correlation is used to identify and estimate a linear function (called a canonical variate) of one set of variables that is maximally correlated with a linear function of a second set of variables." The SAS CANCORR procedure (SAS Institute, Inc., 1985b) finds as many canonical variates as there are variables in the smaller set of variables. The first and subsequent canonical variates are uncorrelated, with the first having the highest correlation coefficient, followed by the second-highest correlation coefficient for the second canonical variate, etc. It should be noted that "the first canonical correlation is at least as large as the multiple correlation between any variable and the opposite set of variables" (SAS Institute, Inc., 1985b).

Gaugush (1986) notes that the information resulting from canonical correlation is largely descriptive and therefore the procedure has not been used as much as other multivariate procedures that support hypothesis testing and/or prediction.

Gaugush (1986) promotes the use of canonical correlation to, for example, "describe the strength of a relationship between a linear combination of nutrient variables and a linear combination of biomass-related variables." The strength of such a relationship is estimated by the canonical correlation coefficient.

Another use of canonical correlation is in determining how many "common elements" are contained within two sets of variables (Gaugush, 1986). The percent overlapping variance (i.e., the squared canonical correlation coefficient) can be used to indicate the relative importance of each canonical variate (Gaugush, 1986).

To use canonical correlation in hypothesis testing, it is important that the assumption of multivariate normality is satisfied (Gaugush, 1986). Snedecor and Cochran (1980) discuss the multivariate normal distribution briefly and state its property that "any variable has a linear regression on the other variables (or on any subset of the other variables), with deviations that are normally distributed." Gaugush (1986) notes that the assumption of multivariate normality is often satisfied by "creating data distributions that are approximately normal."

To satisfy the assumptions of canonical correlation, Gaugush (1986) recommends:

- Use transformations if needed to create roughly symmetric univariate data distributions.
- Carefully examine the validity of outliers and run analyses with and without outliers to document their impact on the correlations.
- Transform data if necessary to create linear relationships among the variables in each set of variables.

Finally, Gaugush (1986) gives an example application of canonical correlation using the SAS CANCORR procedure described above.

4.10.2 Cluster Analysis

Cluster analysis is a classification method for placing "objects into groups or clusters suggested by the data, not defined a priori, such that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar" (SAS Institute, Inc., 1985b). SAS offers several clustering options under the CLUSTER procedure (SAS Institute, Inc., 1985b). It is important to recognize that numerous methods come under the heading of cluster analysis and these methods will give different results. The types of cluster analysis include the following (SAS Institute, Inc., 1985b):

- Disjoint clusters, which place each object in one and only one cluster.
- Hierarchical clusters, in which one cluster may be contained entirely within another cluster, but for which no other kind of overlap is allowed.
- Overlapping clusters with or without constraints placed on the number of objects that belong to two clusters.
- Fuzzy clusters, which are defined by a probability of membership of each object in each cluster. (These can be disjoint, hierarchical, or overlapping.)

Example analyses include the following:

- Gaugush (1986) used Ward's method of cluster analysis to group reservoirs based on similarity in log total phosphorus concentration, log total nitrogen concentration, log Secchi disk depth, and log chlorophyll *a* concentration.
- Kimball (1986) used cluster analysis to group wells based on mean nitrate, well depth, maximum nitrate, coefficient of variation of nitrate, and variance of nitrate. Mean nitrate and coefficient of variation of nitrate yielded the most information. A major conclusion made from this investigation of wells in South Dakota was that "classification of ground water sample locations by geologic environment and depth is crucial to understanding the system."

4.10.3 Principal Components and Factor Analysis

Principal component analysis (PCA) is a multivariate procedure for examining relationships among several quantitative variables (SAS Institute, Inc., 1985b). PCA is used with factor analysis to "create a relatively small number of new variables (called 'factors') from a larger number of original variables" (Gaugush, 1986). The primary use of these procedures is exploratory analysis; that is, hypothesis testing is not normally performed (SAS Institute, Inc., 1985b).

Gaugush (1986) notes that PCA is usually performed before factor analysis. Principal components are linear combinations of the original variables. The first principal component explains the most variability associated with the data, while the second principal component explains the second-most variability associated with the data and is not correlated to the first principal component. As an example, Gaugush (1986) describes how PCA can be used to develop a trophic state index from biological, nutrient, and physical data. It is sometimes helpful to prepare a scatter plot of the data using the first two principal components for exploratory analysis.

Factor analysis is then used to enhance the scientific interpretation of the principal components developed. Factor analysis can then be used to redefine the factors (i.e., the linear functions of one or more of the original variables)

so that they can be interpreted in more scientific terms. That is, factor analysis can be used to reshape a principal component such that the factors match more closely a researcher's intuitive (or research-based) model of the relationships among the variables.

Although hypothesis testing is not normally performed on the results of PCA and factor analysis, Gaugush (1986) recommends that data distributions be approximately symmetric with no outliers. As in other cases, data transformations might be needed to meet these recommendations. Because of problems of scale, Gaugush (1986) recommends that PCA and factor analysis be based on the correlation matrix unless the variables are all of approximately the same magnitude. In cases where the variables are of the same magnitude, the covariance matrix can be used.

This discussion of PCA and factor analysis is intended only to familiarize the water quality researcher with the general use of these techniques. Gaugush (1986) goes several steps farther in describing these procedures, including an illustrative example. SAS gives a fairly detailed mathematical description of PCA and factor analysis (SAS Institute, Inc., 1985b) and offers procedures for performing both (PRINCOMP and FACTOR procedures).

4.10.4 Discriminant Analysis

Discriminant analysis resembles regression analysis, but with a major difference in that the dependent variable in discriminant analysis is categorical, whereas the dependent variable in regression analysis is often continuous (Gaugush, 1986). An example application of discriminant analysis might be to predict the presence or absence of brook trout based on pH and aluminum concentration. Researchers are encouraged to follow the descriptions of discriminant analysis offered by SAS (SAS Institute, Inc., 1985b) and Gaugush (1986) before using the procedure. The following are some of the uses for discriminant analysis (SAS Institute, Inc., 1985b):

- To find a mathematical rule (or "discriminant function") for predicting to which class an observation belongs, given data for the independent quantitative variables.
- To find linear combinations of the independent quantitative variables that best reveal the differences between the classes.
- To find a subset of the independent quantitative variables that best shows the differences between the classes.

Discriminant analysis requires prior knowledge of all classes (e.g., a sample), whereas cluster analysis has no such requirement (SAS Institute, Inc., 1985b). In fact, cluster analysis is used to *define* the classes. Gaugush (1986) also cautions that outliers can adversely affect the results of discriminant analysis and that the predictor variables should follow a multivariate normal distribution within each group, with variance-covariance matrices that are constant across groups. There is, however, at least one procedure (NEIGHBOR procedure) that can be used for non-normal data (SAS Institute, Inc., 1985b).

4.11 Extreme Events

One of the key characteristics that separate environmental, and in particular nonpoint source-influenced data, is the presence of extreme events. The majority of nonpoint source pollution entering streams occurs during runoff from precipitation events. This section presents an approach for estimating annual precipitation and storm events, describes the approach used by EPA's DESCON model for estimating design flows, and concludes with statistical methods

appropriate for evaluating water quality extreme events. Earlier sections describe methods for summarizing average conditions and determining changes. This section also describes methods for evaluating extreme conditions in water quality variables. This is important for evaluating standard violations or evaluating peak concentrations to determine if a BMP was effective.

4.11.1 Rainfall Analyses

Annual precipitation

Chow (1951) presents a method for computing annual precipitation for a variety of return periods. This method is outlined below assuming that the annual rainfall is available for n years.

- Compute the mean and standard deviation for the n years of data. Also compute the coefficient of variation (CV).
- Use CV to estimate the log-probability frequency factor, K, for a given return period (Table 4-40).
- Compute the annual precipitation (X_c) for different return periods using Equation 4-107.

$$X_c = \bar{x}[1 + (CV)(K)] \quad (4\text{-}107)$$

For the rainfall data presented in Table 4-37, $\bar{x}$ and CV are equal to 44.5 inches and 0.15, respectively. From Table 4-40, the value of K corresponding to a 2 year return period is -0.09. Substituting this value into the above equation yields X_c equal to 44.5(1+(0.15)(-0.09)) or 43.9 inches. The 100 year annual precipitation would be equal to 44.5(1+(0.15)(2.70)) = 62.5 inches. The adequacy of the record length can be evaluated using (Mockus, 1960):

$$Y = (4.30 t \log_{10} R)^2 + 6 \quad (4\text{-}108)$$

where Y is the minimum record length in years, t is the Student's t quantile (Table D2) at the 90% level with Y-6 degrees of freedom, and R is the ratio of the 100 year event to the 2 year event.

To solve the above equation, an iterative approach is necessary. Using an initial guess of Y equal to 15 years, t is equal to 1.8331, while R is equal to 62.5/43.9 or 1.42. Substituting these values into the above equation yields Y = $[(4.3)(1.8331)(.1534)]^2$+6 or 7.5. Adjusting our guess of Y to 9 years, t is equal to 2.3534 and Y = $[(4.3)(2.3534)(.1534)]^2$+6 or 8.4 years (which is close enough to our initial guess). Since the actual length of record is 21 years, our 100 year return annual precipitation estimate of 62.5 inches can be expected to be reasonable.

Storm return period

The method developed by Hershfield (1961) is the most usually applied method in the field today and is commonly referred to as "TP40." The method is based on interpolating the design storm from four figures (Figures 4-25 through 4-28) and applying the following equation (Weiss, 1962):

$$I = 0.0256(C-A)x + 0.000256[(D-C)-(B-A)]xy + 0.01(B-A)y + A \quad (4\text{-}109)$$

where I is the rainfall amount (in inches); A is the 2-year, 1-hour rainfall (in inches) interpolated from Figure 4-25; B is the 2-year, 24-hour rainfall (in inches) interpolated from Figure 4-26; C is the 100-year, 1-hour rainfall (in inches) interpolated from Figure 4-27; and D is the 100-year, 24-hour rainfall (in inches) interpolated from Figure 4-28. The return period, x, and duration, y, are taken from Table 4-41 and 4-42, respectively.

Table 4-40. Theoretical log-probability frequency factors.

	Return Period (years)					
	1.01	2	5	20	100	
	Probability (%) equal to or greater than the given variate					
C_s	99	50	20	5	1	C_v
0.0	-2.33	0.0	0.84	1.64	2.33	0.0
0.5	-1.98	-0.09	0.80	1.77	2.70	0.166
1.0	-1.68	-0.15	0.75	1.85	3.03	0.324
1.139	-1.61	-0.16	0.73	1.86	3.11	0.363
1.4	-1.49	-0.19	0.69	1.88	3.26	0.436
1.5	-1.45	-0.20	0.68	1.89	3.31	0.462
2.0	-1.28	-0.24	0.61	1.89	3.52	0.596
3.0	-1.04	-0.28	0.51	1.85	3.78	0.818
4.0	-0.90	-0.29	0.42	1.78	3.91	1.000

Source: Chow, 1951

Table 4-41. Linearized rainfall frequency variate for equation 4-109.

Return Period (in years)	1	2	5	10	25	50	100
Linearized Variate (x)	-6.93	0	9.2	16.1	25.3	32.1	39.1

Source: Weiss, 1962

Table 4-42. Linearized rainfall duration variate for equation 4-109.

Duration (hours)	0.17	.033	0.5	0.67	1
Linearized Variate (y)	-37	-24	-15.6	-9.4	0
Duration (hours)	2	3	6	12	24
Linearized Variate (y)	17.6	28.8	49.9	73.4	100.0

Source: Weiss, 1962

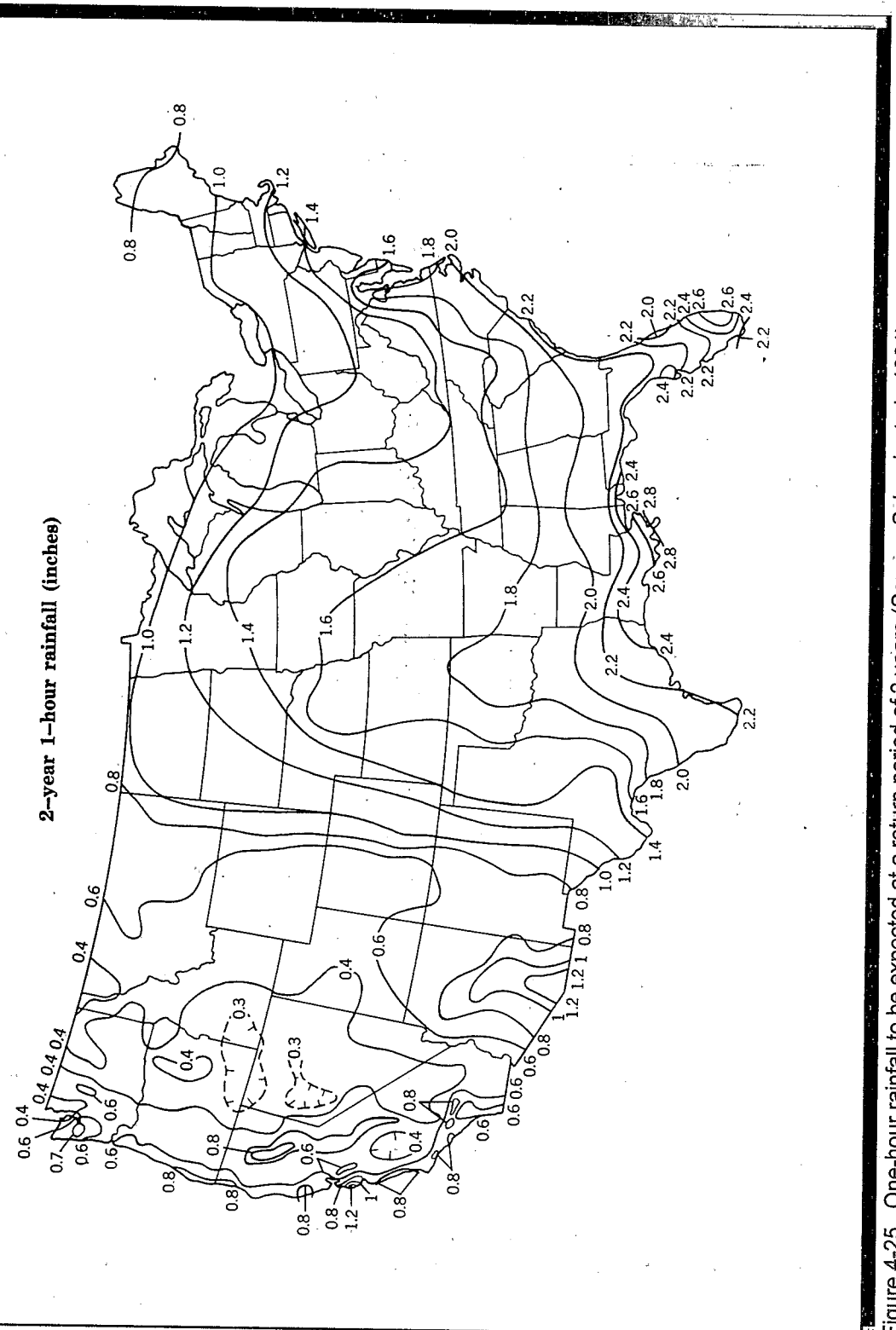


Figure 4-25. One-hour rainfall to be expected at a return period of 2 years (Source: Schwab et al., 1981)

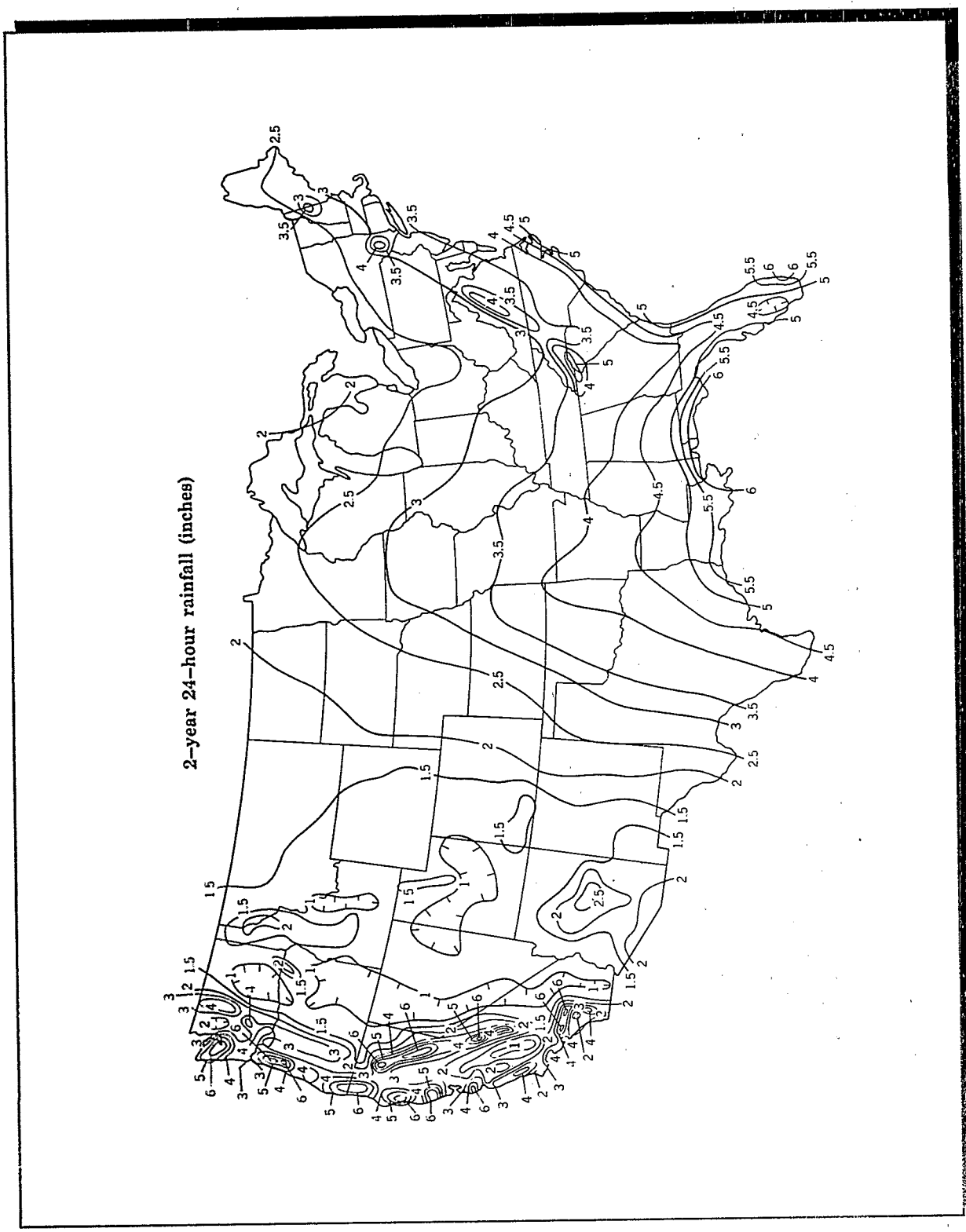


Figure 4-26. 24-hour rainfall to be expected at a return period of 2 years (Source: Schwab et al., 1981)

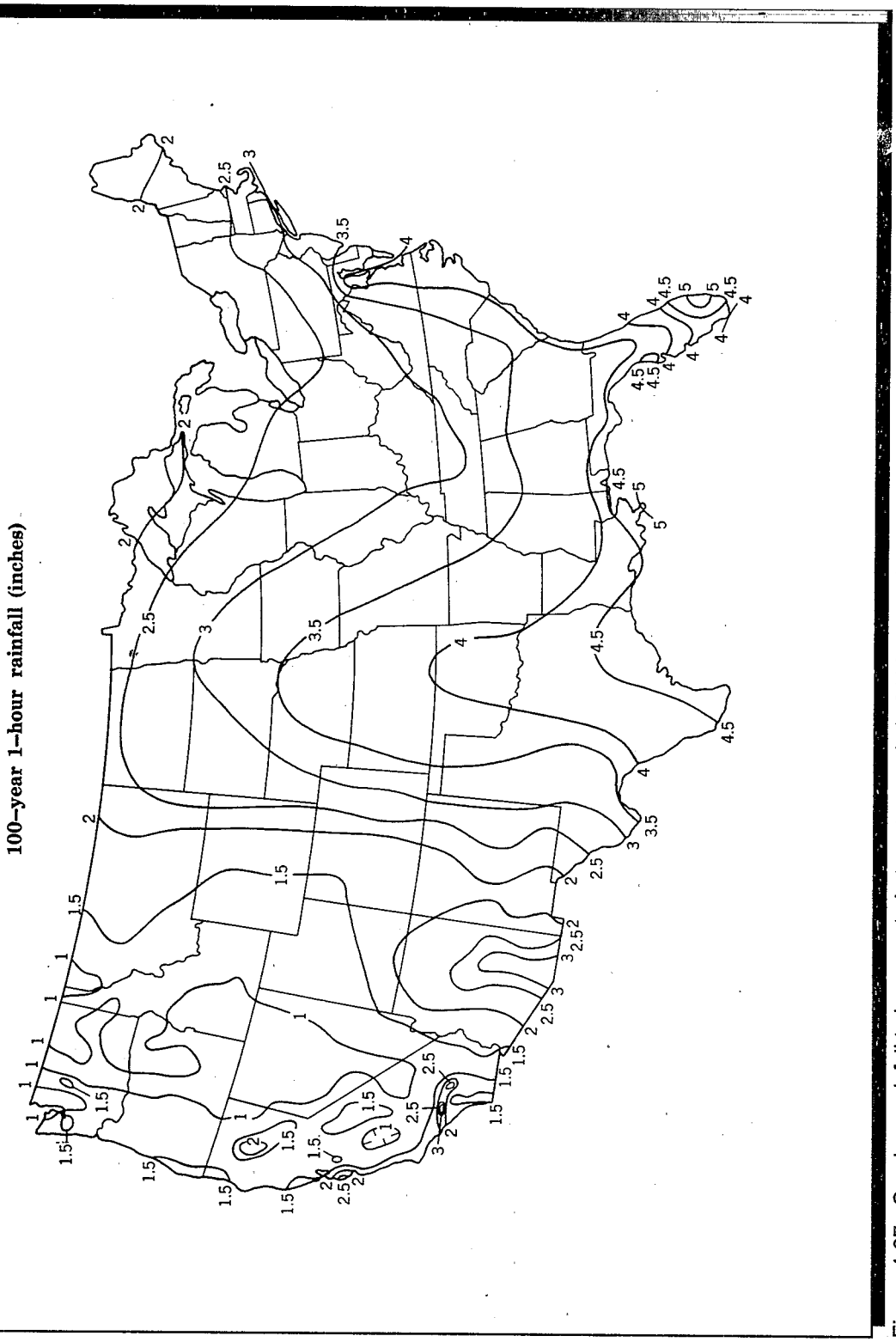


Figure 4-27. One-hour rainfall to be expected at a return period of 100 years (Source: Schwab et al., 1981)

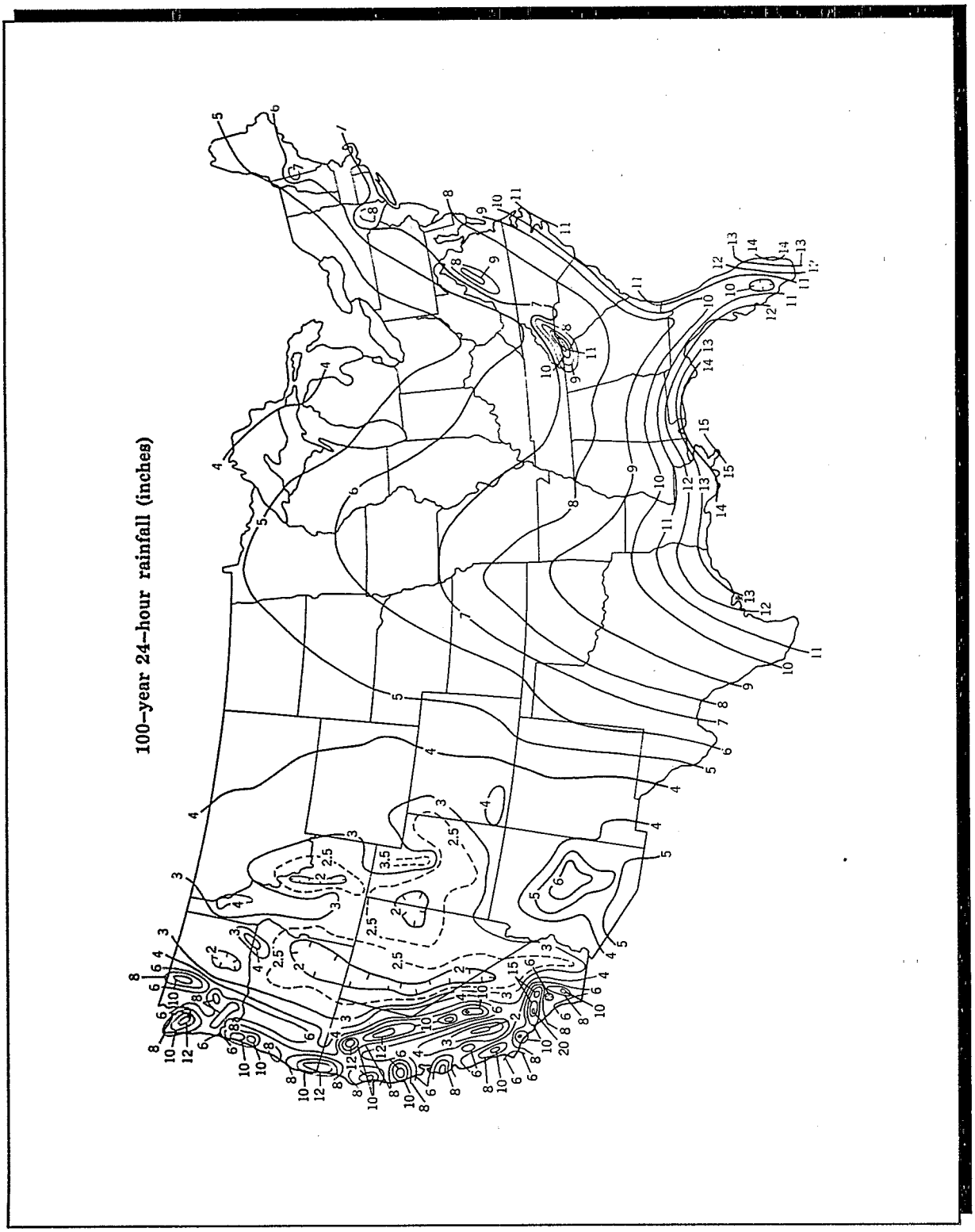


Figure 4-28. 24-hour rainfall to be expected at a return period of 100 years (Source: Schwab et al., 1981)

Suppose the analyst is interested in estimating the 1-year, 2-hour storm in El Paso, Texas. From Figures 4-25 through 4-28, *A*, *B*, *C*, and *D* are estimated as 0.8, 1.5, 2.0, and 4.0, respectively. From Table 4-41, *x* is equal to -6.93 and from Table 4-42, *y* is equal to 17.6. Substituting these values into Equation 4-109 yields a 1-year, 2-hour storm equal to 0.7 inches.

4.11.2 Design Flows

This section describes the computational steps employed by DFLOW and DESCON for each of the three types of design flows considered and has been extracted and adapted from Rossman (1990). It begins with the extreme value design flow, since this type of design flow also serves as a starting point in computing the biologically-based design flow.

Extreme value design flow (low flows)

The extreme value design flow is computed from the sample of lowest *m*-day average flows for each year of record, where "*m*" is the user-supplied flow averaging period. Established practice uses arithmetic averaging to calculate these *m*-day average flows. A log Pearson Type III probability distribution is fitted to the sample of annual minimum *m*-day flows. The design flow is the value from the distribution whose probability of not being exceeded is $1/R$, where R is the user-supplied return period. The procedure is modified slightly to accommodate situations where some annual low flows are zero.

STEP 1. Initialize each element of a vector X of daily flow values to UNKNOWN (i.e., a very large number such as 1×10^{20}).

STEP 2. Read in daily flow values from the retrieved STORET flow file into X, where $X(1)$ corresponds to the first day of record. (Note: February 29th of leap years is ignored.)

STEP 3. Create *m*-day running arithmetic averages from the daily flows in X, and replace the daily flows of X with these values. The running average of $X(i)$, $X(i+1)$, ..., $X(i+m-1)$ is placed in $X(i)$.

STEP 4. Find the lowest *m*-day running average value for each water year recorded in X (where a water year begins on April 1) and store the resulting values in vector Y. Let NY denote the number of entries in Y.

STEP 5. Let N be the number of non-zero entries in Y. Assume that these Y-values are a sample drawn from a log Pearson Type III probability distribution. The design flow is the value from this distribution whose probability of not being exceeded is $1/R$, where R is the user-supplied return period. Use the following procedure to find the design flow:

STEP 5a. Find the mean (U), standard deviation (S), and skewness coefficient (G) of the natural logarithms of the non-zero entries in Y.

STEP 5b. Let $F0$ be the fraction of entries in Y that are zero:

$$F0 = (NY - N)/NY \qquad (4\text{-}110)$$

Let P be the cumulative probability corresponding to the user-supplied return period of R years, adjusted for the presence of zero-flow years:

$$P = (1/R - F0)/(1 - F0) \qquad (4\text{-}111)$$

In other words, if $F0$ is the probability of having a year with zero stream flow, and $1/R$ is the allowed probability of a year with an excursion below the design flow, then P is the corresponding excursion probability in years with non-zero flows.

STEP 5c. Let Z be the standard normal deviate corresponding to cumulative probability P. Z can

be computed using the following formula (Joiner and Rosenblatt, 1971):

$$Z = 4.91(P^{0.14} - (1-P)^{0.14}) \quad (4\text{-}112)$$

STEP 5d. Compute the gamma deviate, K, corresponding to the standard normal deviate Z and skewness G using the Wilson-Hilferty transformation (Loucks et al., 1981):

$$K = (2/g)((1 + GZ/6 - G^2/36)^3 \quad (4\text{-}113)$$

STEP 5e. Compute the design flow as

$$\exp(U + KS) \quad (4\text{-}114)$$

Biologically-based design flow

Biologically-based design flows are computed by starting with a trial design flow, then counting how often this flow is not exceeded by m-day average flows in the historical record. (In contrast with the traditional method of computing extreme value design flows, the m-day flow averages are harmonic means, not arithmetic ones. This count is compared to the allowed number of such occurrences, and the trial design flow is adjusted accordingly. The specific computational steps involved are as follows:

STEP 1. Initialize each element of a vector X of daily flow values to UNKNOWN (i.e., a very large number such as 1×10^{20}).

STEP 2. Read in daily flow values from the retrieved STORET flow file into X, where $X(1)$ corresponds to the first day of record. (Note: February 29th of leap years is ignored.)

STEP 3. Create m-day running harmonic averages from the daily flows in X, and replace the daily flows of X with these values. The running average of $X(i)$, $X(i+1)$, ..., $X(i+m-1)$ is placed in $X(i)$ and is computed as follows:

Define $B(j)$ as $1/X(i+j-1)$ if $X(i+j-1 > 0$, and 0 otherwise, for $j = 1$ to m. Let DSUM be the sum of $B(j)$ for $j = 1$ to m and $m0$ be the number of $B(j)$ values that equal 0. Then replace $X(i)$ with $X(i) = (m-m0)/\text{DSUM}*(m-m0)/m$.

Note that this procedure takes into account the possibility of zero flows when forming a harmonic average.

STEP 4. Compute an extreme value m-day average trial design flow (DFLOW) using the biologically-based average number of years between flow excursions (R) as the return period.

STEP 5. Compute the allowed number of flow excursions, A, (i.e., the number of distinct m-day average flows allowed to be below the design flow) over the NDAYS of stream flow record: $A = \text{NDAYS}/365/R$.

STEP 6. Use the procedure described below to compute the number of biologically-based flow excursions resulting under the trial design flow DFLOW. Because the trial flow was computed as an extreme value flow, the resulting number of biologically-based excursions will most likely be larger than the allowed number, A. If it is not, then keep increasing the trial design flow by some fixed increment until the resulting number of excursions exceeds A.

STEP 7. Use the Method of False Position (Carnahan et al., 1969) to successively refine the estimate of the biologically-based design flow as follows:

STEP 7a. Set lower and upper bounds on the design flow with their corresponding excursion counts:

FL = 0; XL = 0.

FU = DFLOW; XU = number of excursions under DFLOW.

STEP 7b. Check on convergence of the bounds. If FU - FL is within 0.5 percent of FL, then end with DFLOW = FU. If XL is within 0.5 percent of A, then end with DFLOW = FL. If XU is within 0.5 percent of A, then end with DFLOW = FU. Otherwise proceed to the next step.

STEP 7c. Interpolate between the bounds to find a new trial design flow, FT:

$$FT = FL + (FU - FL)(A - XL)/(\qquad (4\text{-}115)$$

and compute the number of excursions (XT) occurring for this flow (see procedure described below).

STEP 7d. Update the bounds based on the value of XT: If XT $\leq A$, then set FL = FT and XL = XT. Otherwise set FU = FT and XU = XT. Then return to the convergence check of step 7b.

The process used to count the number of flow excursions for a given design flow proceeds in two phases. The first phase identifies all excursion periods in the period of record. An excursion period is a sequence of consecutive days where each day belongs to an m-day running average flow that is below the given design flow. Recall that "m" is the flow averaging period set by the user. Phase two groups these excursion periods into excursion clusters and counts up the total number of excursions occurring within all clusters. An excursion cluster consists of all excursion periods falling within a prescribed length of time from the start of the first period in the cluster (120 days is the default cluster length). The number of excursions counted per cluster is subject to an upper limit whose default value is 5.

Before describing the detailed procedures for each of these phases a simple numerical example will be used to illustrate the method. Suppose that the design flow under consideration is 100 cfs and that the period of record yields a sequence of 4-day running average flows as detailed in Box 1.

The first flow excursion period for this record consists of the 4-day averages occurring on days 1,2 and 3. Thus the period extends from day 1 to day 6 (days 4, 5 and 6 belong to the averaging period that begins on day 3). There are two other excursion periods consisting of days 13 to 18 and 513 to 548. Under the default clustering parameters, there are 2 excursion clusters; cluster 1 contains periods 1 and 2, and cluster 2 contains period 3. The number of excursions in each cluster is detailed in Box 2.

Note that the number of excursions in each period equals the period length divided by the averaging period. The nominal number of excursions in cluster 2 is 9, and since this exceeds the limit of 5, only 5 are counted. The total number of excursions for the design flow of 100 cfs in this example is 3 + 5 = 8.

The detailed procedure for counting biologically-based flow excursions under a specified design flow is as follows:

PHASE 1

Define:

P1(i) = day which begins excursion period i,
P2(i) = day which ends excursion period i,
XP(i) = number of excursions in period i,
XKLmax = maximum cluster length (e.g., 120 days).
t = current day of record.

STEP 1. Set i = 0, P2(0) = 0, and t = 1.

STEP 2. If the m-day running average beginning on day t is greater or equal to the specified design flow then proceed to Step 5.

Day	4-Day Average Flow (cfs)	Day	4-Day Average Flow (cfs)
1	34	513-545	< 100
2	65		
3	25	546-end	> 100
4-12	> 100		
13	57		
14	34		
15	26		
16-512	> 100		

Box 1

Cluster	Period	Start Day	Length (days)	No. of Excursions in Period	No. of Excursions in Cluster
1	1	4	6	6/4 = 1.5	3.0
	2	13	6	6/4 = 1.5	
2	3	513	36	36/4 = 9.0	5.0

Box 2

STEP 3. If the current day t is more than a day beyond the end of the current excursion period ($t > P2(i) + 1$), or if the length of the current excursion period equals XKLmax then begin a new excursion period by setting:

$i = i + 1$
$P1(i) = t$
$P2(i) = m - 1$
$XP(i) = 0.$

STEP 4. Update the ending day of the current excursion period and the excursion count for this period:

$P2(i) = P2(i) + 1$
$XP(i) = (P2(i) - P1(i)) / m.$

STEP 5. Proceed to the next day of record ($t = t + 1$). If not at the end of the record then return to Step 2. Otherwise proceed to phase 2.

PHASE 2

Define:

i = current excursion period,
k = current excursion cluster,
K1 = day of record which begins cluster k,
XK(k) = number of excursions in cluster k,
Xkmax = maximum number of excursions counted per cluster (e.g., 5),

STEP 1. Set $i = 1$, $k = 0$, and K1 = a large negative number.

STEP 2. If the length of the current cluster is greater than the maximum length (i.e., P2(i) - K1 > XKLmax) then begin a new cluster with excursion period i, i.e.,

$k = k + 1$
K1 = P1(k)
XK(k) = 0.

STEP 3. Update the excursion count for the current cluster,

XK(k) = minimum(XK(k) + XP(i), XKmax).

STEP 4. Proceed to the next excursion period ($i = i + 1$) and return to Step 2. If no more excursion periods remain, then total up the number of excursions in each cluster (XK(1) +XK(2) + ... + XK(k)) to determine the total number of excursions.

4.11.3 Frequency of Extreme Events

This section describes methods for evaluating extreme conditions in water quality variables. This is an important consideration for evaluating standard violations or evaluating peak concentrations to determine if a BMP was effective. Gilbert (1987) presents an approach for evaluating proportions. The method is based on computing the number of observations exceeding a threshold value X_c. The proportion of observations, p, exceeding X_c can be computed as

$$p = u/n \tag{4-116}$$

where u is the number of observations exceeding X_c and n is the number of observations. For $n \leq 30$, Table D11 can be used to develop nonparametric 90th or 95th percentile confidence limits. For $n > 30$, Equations 4-117 and 4-118 may be used. The lower limit is equal to 0 if u is 0 and the upper limit is 1 if the u is equal to n.

If np and $n(1-p)$ are greater than 5 (some authors suggest a value of 10), then Gilbert (1987) suggests that the normal approximation can be used to compute the upper and lower limits with the following equation:

$$p \pm Z_{1-\alpha/2}\left[\frac{p(1-p)}{n}\right]^{1/2} \tag{4-119}$$

The confidence intervals can be used to evaluate one-sample hypotheses such as

H_o: $p = 0.10$
H_a: $p \neq 0.10$

$$\text{Lower limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{(u-0.5) + \frac{Z_{1-\alpha/2}^2}{2} - Z_{1-\alpha/2}\left[(u-0.5) - \frac{(u-0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4}\right]^{1/2}\right\} \tag{4-117}$$

$$\text{Upper limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{(u+0.5) + \frac{Z_{1-\alpha/2}^2}{2} + Z_{1-\alpha/2}\left[(u+0.5) - \frac{(u+0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4}\right]^{1/2}\right\} \tag{4-118}$$

If the 95 percent confidence intervals include 0.10, we accept the null hypothesis. Otherwise the null hypothesis is rejected.

An evaluation of proportions can also be used to determine the necessary sample size to ensure that q percent of the population is less than the largest randomly sampled observation. This approach provided by Conover (1980) is demonstrated with the next example.

Example:

Determine the number of random samples that would be required to ensure with a 95 percent probability (α=0.05) that 90 percent of the population is less than the largest observation.

Solution:

Enter Table D11 with q equal to 0.9 and 1-α equal to 0.95 and directly read a sample size of 29. Therefore, it would require 29 samples to ensure that the largest observation is greater than 90 percent of the population.

Application of this example is similar to quality control processes. In this case, once 29 samples have been collected, the upper bound is set equal to the largest observation. From then on, we would expect that only 10 percent of the future samples would exceed the upper bound with 95 percent confidence. If more than 10 percent of future observations exceeded the upper bound, we would infer that some change has occurred (Ward et al., 1990).

It is also possible to compare the proportions p_1 and p_2 between two samples with sample sizes equal to n_1 and n_2. For example, it may be appropriate to compare the percent of standard violations from before and after. In this case, the null and two-sided alternative hypothesis are

H_o: $p_1 = p_2$
H_a: $p_1 \neq p_2$

Moore and McCabe (1989) provide the test statistics as

$$z = \frac{p_1 - p_2}{s_p} \qquad (4\text{-}120)$$

where s_p and p are given by

$$s_p = \sqrt{p\ (1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad (4\text{-}121)$$

$$p = \frac{u_1 + u_2}{n_1 + n_2} \qquad (4\text{-}122)$$

Moore and McCabe (1989) suggest that n_1p, $n_1(1 - p)$, n_2p, and $n_2(1 - p)$ all be greater than or equal to 5 for application. If the absolute value of z is greater than the associated normal deviate (e.g., 1.96 for a two-sided test with α equal to 0.05), then H_o is rejected.

5. QUALITY ASSURANCE AND QUALITY CONTROL

5.1 INTRODUCTION

Quality assurance (QA) and quality control (QC) are commonly thought of as procedures used in the laboratory to ensure that all analytical measurements made are accurate. Yet QA and QC extend beyond the laboratory and are essential components of all phases and all activities within each phase of a nonpoint source (NPS) monitoring project. This section defines QA and QC, discusses their value in NPS monitoring programs, and explains EPA's policy on these topics. The following sections provide detailed information and recent references for planning and ensuring quality data and deliverables that can be used to support specific decisions involving nonpoint source pollution.

5.1.1 Definitions of Quality Assurance and Quality Control

Quality assurance:

An integrated system of management procedures and activities used to verify that the quality control system is operating within acceptable limits and to evaluate the quality of data (Taylor, 1993; USEPA, 1994c).

Quality control:

A system of technical procedures and activities developed and implemented to produce measurements of requisite quality (Taylor, 1993; USEPA, 1994c).

QC procedures include the collection and analysis of blank, duplicate, and spiked samples and standard reference materials to ensure the integrity of analyses and regular inspection of equipment to ensure it is operating properly. QA activities are more managerial in nature and include assignment of roles and responsibilities to project staff, staff training, development of data quality objectives, data validation, and laboratory audits. Table 5-1 lists some common activities that fall under the headings of QA and QC. Such procedures and activities are planned and executed by diverse organizations through carefully designed quality management programs that reflect the importance of the work and the degree of confidence needed in the quality of the results.

5.1.2 Importance of QA/QC Programs

Although the value of a QA/QC program might seem questionable while a project is under way, its value should be quite clear after a project is completed. If the objectives of the project were used to design an appropriate data collection and analysis plan, all QA/QC procedures were followed for all project activities, and accurate and complete records were kept throughout the project, the data and information collected from the project will be adequate to support a choice from among alternative courses of action. In addition, the course of action chosen will be defensible based on the data and information collected. Development and implementation of a QA/QC program can require up to 10 to 20 percent of project resources (Cross-Smiecinski and Stetzenback, 1994), but this cost can be recaptured in lower overall costs due to the project's being well planned and executed. Likely problems are anticipated and accounted for before they arise, eliminating the need to spend countless hours and dollars resampling, reanalyzing data, or mentally reconstructing portions of the project to determine where an error was introduced. QA/QC procedures and activities are cost-effective measures used to determine how to allocate project energies and resources toward improving the quality of research and the usefulness of project results (Erickson et al., 1991).

Table 5-1. Common QA and QC activities.

QA Activities

- Organization of project into component parts
- Assignment of roles and responsibilities to project staff
- Use of statistics to determine the number of samples and sampling sites needed to obtain data of a required confidence level
- Tracking of sample custody from field collection through final analysis
- Development and use of data quality objectives to guide data collection efforts
- Audits of field and laboratory operations
- Maintenance of accurate and complete records of all project activities
- Personnel training to ensure consistency of sample collection techniques and equipment use

QC Activities

- Collection of duplicate samples for analysis
- Analysis of blank and spike samples
- Replicate sample analysis
- Regular inspection and calibration of analytical equipment
- Regular inspection of reagents and water for contamination
- Regular inspection of refrigerators, ovens, etc. for proper operation

Adapted from Drouse et al., 1986, and Erickson et al., 1991.

This chapter discusses many elements and aspects of QA/QC programs that do not differ significantly from one type of program to another—for instance, from a point source permit compliance sampling program to an NPS best management practice effectiveness monitoring program. Therefore, much of the following discussion is not specific to NPS projects. This does not, however, mean that a well-designed and well-implemented QA/QC program is not necessary for an NPS project. It is hoped that the following discussion will convey to the reader the importance of QA and QC to the success of every project involving the collection and analysis of environmental data.

5.1.3 EPA Quality Policy

EPA has established a QA/QC program to ensure that data used in research and monitoring projects are of known and documented quality to satisfy project objectives. The use of different methodologies, lack of data comparability, unknown data quality, and poor coordination of sampling and analysis efforts can delay the progress of a project or render the data and information collected from it insufficient for decision making. QA/QC practices should be used as an integral part of the development, design, and implementation of an NPS monitoring project to minimize or eliminate these problems (Erikson et al., 1991; Pritt and Raese, 1992; USEPA, 1994d).

EPA's mandatory agency-wide Quality System policy requires each office or laboratory generating data to implement minimum procedures to ensure that precision, accuracy, completeness, comparability, and representativeness of data are known and documented (Erickson et al., 1991; USEPA, 1984c). This policy is now based on the quality system standard developed by the American Society of Quality Control (ASQC, 1994). Each office or laboratory is required to specify the quality levels that data must meet to be acceptable and satisfy project objectives. This requirement applies to all environmental monitoring and measurement efforts mandated or supported by EPA through regulations, grants, contracts, or other formal agreements. To ensure that this responsibility is met uniformly across EPA, each organization performing work for EPA must document in a Quality Management Plan (QMP) that is approved by its senior management how it will plan, implement, and assess the effectiveness of QA and QC operations applied to environmental programs (USEPA, 1994d). In addition, each non-EPA organization must have a well-documented Quality Assurance Project Plan (QAPP) that covers each monitoring or measurement activity associated with a project (Erickson et al., 1991; USEPA, 1983c, 1994).

The purpose of writing a QAPP prior to undertaking an NPS monitoring project is to establish clear objectives for the program, including the types of data needed and the quality of the data generated (accuracy, precision, completeness, representativeness, and comparability). This information is used to design the program to meet these objectives. Developing a QAPP prior to undertaking the NPS monitoring project also establishes the boundaries of the project, in terms of the time allotted to it and the decisions that can realistically be made from the data and information that will be collected.

The QAPP should specify the policies, organization, objectives, functional activities, QA procedures, and QC activities designed to achieve the data quality goals of the project. It should be distributed to all project personnel, and they should be familiar with the policies and objectives outlined in the QAPP to ensure proper interaction of the sampling and laboratory operations and data management. All persons involved in an NPS monitoring project who either perform or supervise the work done under the project are responsible for ensuring that the QA/QC procedures and activities established in the QAPP are adhered to.

The QMP and each QAPP must be submitted for review to the EPA organization responsible for the work to be performed, and they must be approved by EPA or its designee (e.g., federal or state agency) as part of the contracting or assistance agreement process before the work can begin. In addition, it is important to note that the QMP and QAPP are "live" documents and programs in the sense that once they have been developed they cannot be placed on a shelf for the remainder of the project. All QA/QC procedures should be evaluated and plans updated as often as necessary during the course of a project to ensure that they are in accordance with the present project direction and efforts (Knapton and Nimick, 1991; USEPA, 1994c).

5.2 Data Quality Objectives (DQOs)

Before collecting environmental data in support of an NPS project, it is important to determine the type, quantity, and quality of data needed to meet the project objectives and support a specific decision based on the results of the project. Not doing so creates the risk of expending too much effort on data collection (i.e., more data are collected than necessary), not expending enough effort on data collection (i.e., more data are necessary than were collected), or expending the wrong effort (i.e., the wrong data were

collected). Proper planning and execution of a data collection effort can prevent these problems. EPA has developed the Data Quality Objectives Process as a flexible planning tool that should be used to prepare for a data collection activity. The information compiled in this effort is then used to develop the QAPP (USEPA, 1994e).

5.2.1 The Data Quality Objectives Process

The Data Quality Objectives (DQO) process takes into consideration the factors that will depend on the data (most importantly, the decision(s) to be made) or that will influence the type and amount of data to be collected (e.g., the problem being addressed, existing information, information needed before a decision can be made, and available resources). From these factors the qualitative and quantitative data needs are determined. The purpose of the DQO process is to improve the effectiveness, efficiency, and defensibility of decisions made based on the data collected, and to do so in a resource-effective manner (USEPA, 1994e).

DQOs are qualitative and quantitative statements that clarify the study objective, define the most appropriate type of data to collect, and determine the most appropriate conditions under which to collect them. DQOs also specify the minimum quantity and quality of data needed by a decision maker to make any decisions that will be based on the results of the project. By using the DQO process, investigators can ensure that the type, quantity, and quality of data collected and used in decision making will be appropriate for the intended use. Similarly, efforts will not be expended to collect information that does not support defensible decisions. The products of the DQO process are criteria for data quality and a data collection design that ensures that data will meet the criteria.

The DQO process consists of seven steps, described below. The process is iterative. As one step of the process is completed, its outputs might lead to reconsideration of previous steps. The previous steps should then be repeated. Optimization of the design (the last step) should begin only when all previous steps have been completed. When the optimization step is reached, as at any time during the DQO process, it might be necessary to reconsider earlier steps (i.e., to reiterate part or all of the process) to determine the optimum design.

A brief description of each step of the DQO process and a list of activities that are part of each step follow. For a detailed discussion of the DQO development process, refer to EPA's *Guidance for the Data Quality Objectives Process* (USEPA, 1994e), from which the following information was taken. This reference contains a case study example of the DQO process. A computer program, *Data Quality Objectives Decision Error Feasibility Trials* (EPA QA/G-4D), is also available to help the planning process by generating cost information about several simple sampling designs based on the DQO constraints before the sampling and analysis design team begins developing a final sampling design in the last step of the DQO process. (Contact EPA's Quality Assurance Management Staff, 202 260-9464).

(1) State the problem

In this first step the problem to be studied is described concisely. A review of prior studies and existing information is important during this step to gain a sufficient understanding of the problem in order to define it. The specific activities to be completed during this step (outputs) are:

- Identify members of the planning team.

- Identify the primary decision maker of the planning team and define each member's role and responsibilities during the DQO process.
- Develop a concise description of the problem.
- Specify the available resources and relevant deadlines for the study.

(2) Identify the decision

Identify what questions the study will attempt to resolve and what actions might be taken based on the study. This information is used to prepare a "decision statement" that will link the principal study question to one or more possible actions that should solve the problem. Possible options include take no action, take action, or modify an action. A decision statement might be phrased as follows: *Determine whether [or which] NPS impacts require taking [one of the alternative actions]*. For example, if the question to be addressed is "Are nutrients from agricultural runoff contributing to the growth of algal mats in the river?" and the alternative actions are "require vegetation buffers along streams" or "take no action," the decision statement is "Determine whether nutrients from agricultural runoff are contributing to algal growth and require regulation." The specific activities to be completed during this step are:

- Identify the principal study question.
- Define the alternative actions that could result from resolution of the principal study question.
- Combine the principal study question and the alternative actions into a decision statement.
- If applicable, organize multiple decisions to be made by priority.

(3) Identify the inputs to the decision

Identify the information that needs to be obtained and the measurements that need to be taken to resolve the decision statement. The specific activities to be completed during this step are:

- Identify the information that will be required to resolve the decision statement.
- Determine the sources for each item of information identified above.
- Identify the information that is needed to establish the threshold value that will be the basis of choosing among alternative actions.
- Confirm that appropriate measurement methods exist to provide the necessary data.

(4) Define the study boundaries

Specify the time periods and spatial area to which decisions will apply and determine when and where data should be collected. This information is used to define the population(s) of interest. The term *population* refers to the total collection or universe of objects from which samples will be drawn. The population could be the concentration of a pollutant in sediment, a water quality parameter, algae in the river, or bass in the lake. It is important to define the study boundaries to ensure that data collected are representative of the population being studied (since every member of a population cannot be sampled) and will be collected during the time period and from the place that will be targeted in the decision to be made. The specific activities to be completed during this step are:

- Specify the characteristics that define the population of interest.
- Identify the geographic area to which the decision statement applies (such as a county)

and any strata within that area that have homogeneous characteristics (e.g., recreational waters, dairy farms).

- Define the time frame to which the decision applies.
- Determine when to collect data.
- Define the scale of decision making, or the actual areas that will be affected by the decision (e.g., first-order streams, dairy farms with streams running through them, a county).
- Identify any practical constraints on data collection.

(5) Develop a decision rule

Define the statistical parameter of interest, specify the threshold at which action will be taken, and integrate the previous DQO outputs into a single statement that describes the logical basis for choosing among alternative actions. This statement is known as a *decision rule*. It is often phrased as an "If...then..." statement. For example, "If the mean concentration of contaminant X in the water downstream from farm Y exceeds 0.5 μg/L, then vegetation will be planted; otherwise, no action will be taken." The specific activities to be completed during this step are:

- Specify the statistical parameter that characterizes the population (the parameter of interest), such as the mean, median, or percentile.
- Specify the numerical value of the parameter of interest that would cause a decision maker to take action, i.e., the threshold value.
- Develop a decision rule in the form of an "if...then..." statement that incorporates the parameter of interest, the scale of decision making, the threshold level, and the actions that would be taken.

(6) Specify limits on decision errors

Define the decision maker's tolerable limits of making an incorrect decision (or decision error) due to incorrect information (i.e., measurement and sampling error) introduced during the study. These limits are used to establish performance goals for the data collection design. Base the limits on a consideration of the consequences of making an incorrect decision. The decision maker cannot know the true value of a population parameter because the population of interest almost always varies over time and space and it is usually impractical or impossible to measure every point (sampling design error). In addition, analytical methods and instruments are never absolutely perfect (measurement error). Thus, although it is impossible to eliminate these two errors, the combined total study error can be controlled to reduce the probability of making a decision error. The specific activities to be completed during this step are:

- Determine the possible range (likely upper and lower bounds) of the parameter of interest.
- Identify the decision errors and choose the null hypothesis. Decision errors for NPS pollution problems might take the general form of deciding there is no impact when there is [a false positive, or type I error], or deciding there is an impact when there is none [a false negative, or type II error].
- Specify the likely consequences of each decision error. Evaluate their potential severity in terms of ecological effects, human health, economic and social costs, political and legal ramifications, and other factors.

- Specify a range of possible parameter values where the consequences of decision errors are relatively minor (gray region). The boundaries of the gray region are the threshold level and the value of the parameter of interest where the consequences of making a false negative decision begin to be significant.

- Assign probability limits to point above and below the gray region that reflect the tolerable probability for the occurrence of decision errors.

(7) Optimize the design

Evaluate information from the previous steps and generate alternative data collection designs. The designs should specify in detail the monitoring that is required to meet the DQOs, including the types and quantity of samples to be collected; where, when, and under what conditions they should be collected; what variables will be measured; and the QA/QC procedures that will ensure that the DQOs are met. The QA/QC procedures are fully developed when the QAPP is written (see below). Choose the most resource-effective design that meets all of the DQOs. The specific activities to be completed during this step are:

- Review the DQO outputs and existing environmental data.

- Develop general data collection design alternatives.

- Formulate the mathematical expressions needed to solve the design problem for each data collection design alternative. This involves selecting a statistical test method (e.g., Student's *t* test), developing a statistical model that relates the measured value to the "true" value, and developing a cost function that relates the number of samples to the total cost of sampling and analysis.

- Select the optimal sample size that satisfies the DQOs for each data collection design alternative.

- Select the most resource-effective data collection design that satisfies all of the DQOs.

- Document the selected design's key features and the statistical assumptions of the selected design. It is particularly important that the statistical assumptions be documented to ensure that, if any changes in analytical methods or sampling procedures are introduced during the project, these assumptions are not violated.

The DQO process should be used during the planning stage of any study that requires data collection, and before the data are collected. EPA's policy is to use the DQO process to plan all data collection efforts that will require or result in a substantial commitment of resources. The DQO process is applicable to all studies, regardless of size; however, the depth and detail of the DQO development effort depends on the complexity of the study. In general, more complex studies benefit more from more detailed DQO development.

5.2.2 Data Quality Objectives and the QA/QC Program

The DQOs and the quality objectives for measurement data that will be specified in the QAPP are interdependent. The DQOs identify project objectives; evaluate the underlying hypotheses, experiments, and tests to be performed; and then establish guidelines for the data collection effort needed to obtain data of the quality necessary to achieve these objectives (Erickson et al., 1991; USEPA, 1994e). The

QAPP presents the policies, organization, and objectives of the data collection effort and explains how particular QA and QC activities will be implemented to achieve the DQOs of the project, as well as to determine what future research directions might be taken (Erickson et al., 1991; USEPA, 1994e). At the completion of data collection and analysis, the data are validated according to the provisions of the QAPP and a Data Quality Assessment (DQA), using statistical tools, is conducted to determine:

- Whether the data meet the assumptions under which the DQOs and the data collection design were developed.

- Whether the total error in the data is small enough to allow the decision maker to use the data to support the decision within the tolerable decision error rates expressed by the decision maker (USEPA, 1994e).

Thus, the entire process is designed to assist the decision maker by planning and obtaining environmental data of sufficient quantity and quality to satisfy the project objectives and allow decisions to be made (USEPA, 1994c, 1994e). The DQO process is the part of the quality system that provides the basis for linking the intended use of the data to the QA/QC requirements for data collection and analysis (USEPA, 1994e).

5.3 Elements of a Quality Assurance Project Plan

QAPPs must be prepared according to guidance provided in *EPA Requirements for Quality Assurance Project Plans for Environmental Data Objectives* (USEPA, 1994c). EPA requires that four types of elements be discussed in a Quality Assurance Project Plan (QAPP). These elements are listed in Table 5-2 and discussed briefly below. (For complete descriptions and requirements, be sure to see USEPA (1994c)). Additional information on the contents of a QAPP is contained in Drouse et al. (1986), Erickson et al. (1991), and Cross-Smiecinski and Stetzenback (1994). Drouse et al. (1986) and Erickson et al. (1991) are examples of EPA QAPPs prepared under previous guidance.

The elements described below should always be addressed in the QAPP, unless otherwise directed by the overseeing or sponsoring EPA organization(s). The types, quantity, and quality of environmental data collected for each project could be quite different. As noted in USEPA (1994c), "the content and level of detail in each QAPP will vary according to the nature of the work being performed and the intended use of the data." If an element is not applicable or required, then this should be stated in the QAPP. For some complex projects, it might be necessary to add special requirements to the QAPP. Again, the QAPP must be approved by the sponsoring EPA organization before work can begin, and it should be reviewed annually (for multiyear projects) and updated and reapproved as often as necessary during the project.

5.3.1 Group A: Project Management

These elements cover basic project management, including project history and objectives, roles and responsibilities of participants, and other factors to ensure that the project has a defined goal understood by all the participants and that all planning activities have been documented.

A1 Title and Approval Sheet

Provide the title of the plan; name of organization(s) implementing the project; and names, titles, and signatures of the appropriate approving officials and their approval dates.

A2 Table of Contents

List sections, figures, tables, references, and appendices. If document control format is

Table 5-2. Elements required in an EPA Quality Assurance Project Plan. (USEPA, 1994b)

	QAPP Element
A1	Title and Approval Sheet
A2	Table of Contents
A3	Distribution List
A4	Project/Task Organization
A5	Problem Definition/Background
A6	Project/Task Description
A7	Quality Objectives and Criteria for Measurement Data
A8	Project Narrative (ORD only)
A9	Special Training Requirements/Certification
B1	Sampling Process Design
B2	Sampling Methods Requirements
B3	Sampling Handling and Custody Requirements
B4	Analytical Methods Requirements
B5	Quality Control Requirements
B6	Instrument/Equipment Testing, Inspection, Maintenance Requirements
B7	Instrument Calibration and Frequency
B8	Inspection/Acceptance Requirements for Supplies and Consumables
B9	Data Acquisition Requirements (Nondirect Measurements)
B10	Data Management
C1	Assessments and Response Action
C2	Reports to Management
D1	Data Review, Validation, and Verification Requirements
D2	Validation and Verification Methods
D3	Reconciliation and User Requirements

required, see Cross-Smiecinski and Stetzenback (1994) and USEPA (1994e).

A3 Distribution List

List all individuals and organizations who will receive copies of the approved QAPP and subsequent revisions.

A4 Project/Task Organization

Discuss the specific roles and responsibilities of all individuals or organizations participating in the project.

A flow chart or box diagram is useful for depicting project organization and responsibilities (Figure 5-1). Using the diagram, explain the rationale for the organization (e.g., to maximize

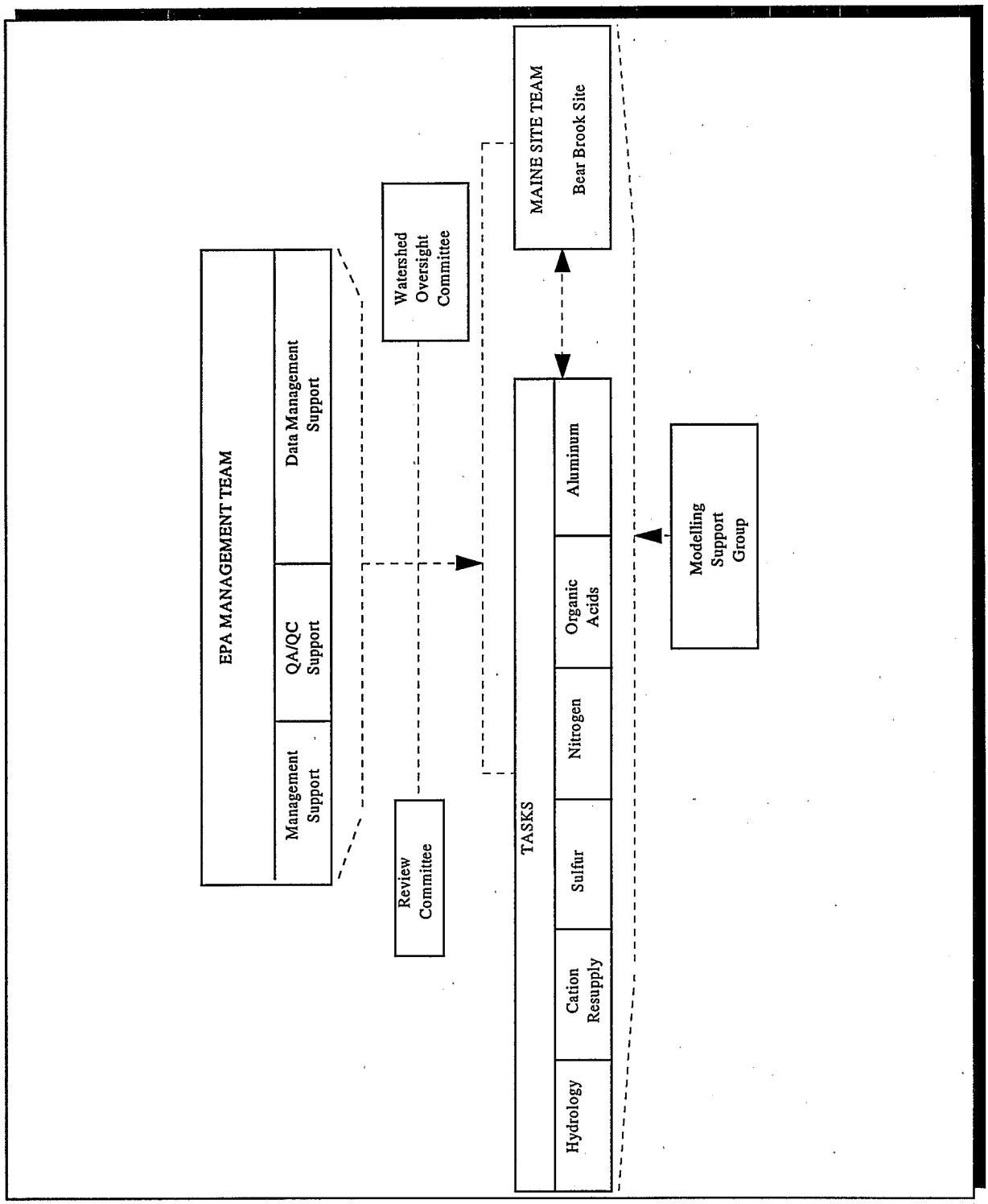


Figure 5-1. Sample organization chart for a quality assurance project plan. (Erickson et al., 1991)

the interaction of site and task leaders). This section provides details on the division of the project into teams, support teams, review committees, and other groups and identifies the persons and entities that will be involved in the project. All members of each project team should be listed along with their affiliations with participating organizations. The program manager, managers or coordinators of any specific tasks, directors of technical tasks to be conducted, and any organizations or agencies that will be involved in the project should be identified. Also identify the specific roles and responsibilities (such as field sampling, laboratory analyses, and report preparation) that will be conducted by each person and organization involved in the project.

A5 Problem Definition/Background

State the problem to be solved or the decision to be made and describe its history for this particular project.

A6 Project/Task Description

Describe the work to be performed (measurements to be made, applicable quality standards, any special personnel or equipment requirements, assessment tools needed, records and reports needed) and the schedule for its implementation.

A7 Quality Objectives and Criteria for Measurement Data

The DQO process will provide this information, or state the project quality objectives and measurement performance criteria that are necessary to support the management decision(s) to be made based on the result(s) of the project. State quality objectives in terms of project requirements, preferably in quantitative terms, rather than in terms of analytical or sampling method capabilities. Then, with the quality objectives stated, select the appropriate methods to achieve the requirements (Cross-Smiecinski and Stetzenback, 1994). The quality of data should be expressed in terms of precision, accuracy, comparability, representativeness, and completeness (defined below). A table of quality objectives, like that in Figure 5-2, is helpful.

Definitions of data quality terms

Precision (reproducibility)

(a) Precision is a measure of mutual agreement among individual measurements of the same property. The coefficient of variation (CV), also known as the percent relative standard deviation (RSD), is used to express precision (Erickson et al., 1991).

$$CV = \left(\frac{s}{x}\right) 100$$

where

s = sample standard deviation and
x = arithmetic mean.

(b) Precision is an expression of mutual agreement of multiple measurement values of the same property conducted under prescribed similar conditions. It is evaluated by recording and comparing multiple measurements of the same parameter on the same exact sample under the same conditions. Relative percent difference (RPD) is a measure of precision and is calculated with the following formula (Cross-Smiecinski and Stetzenback, 1994):

$$RPD = \frac{2\,(x_1 - x_2)}{x_1 + x_2}\,(100)$$

where

Parameter	Units	Expected Range	Accuracy	Precision	Completeness
Particulate NO_3/SO_4					
SSI[a]	$\mu g/m^3$	10-1000	10%	20%	70%
47 mm TF/PC[b]	$\mu g/m^3$	10-1000	10%	20%	90%
NO_3					
47 mm TF/PC	$\mu g/m^3$	1 to 25	20%	20%	90%
SO_2					
47 mm TF/PC	$\mu g/m^3$	1 to 75	20%	20%	90%
Meteorological					
Wind speed	m/s	0 to 75	2%	2%	90%
Wind direction	deg	0 to 360	2%	2%	90%
Dew point	0°C	-30 to 70	2%	5%	90%
Solar radiation	watts/m^2	-	-	-	-
Ambient temperature	°C	-20 to 50	1°C	2°C	90%

[a] PM_{10} Size Selective Inlet High Volume Sampler
[b] Teflon/Polycarbonate Filter

Figure 5-2. Sample quality assurance objectives. (Erickson et al., 1991).

x_1 = analyte concentration of first duplicate and
x_2 = analyte concentration of second duplicate.

Accuracy (bias)

(a) Accuracy is the degree of agreement of a measurement (or an average of measurements), X, with an accepted reference or true value, T. Accuracy is expressed as the percent difference from the true value {100 [(X-T)/T]} unless spiking materials are used and percent recovery is calculated (Erickson et al., 1991).

(b) Accuracy is the correctness of the value obtained from analysis of a sample. It is determined by analyzing a sample and its corresponding matrix spike. Accuracy can be expressed as percent recovery and calculated using the following formula (Air National Guard, 1993):

$$\%R = \frac{A - B}{C}(100)$$

where

A = spiked sample result;
B = sample result; and
C = spike added.

Comparability

(a) Comparability is defined as the confidence with which one data set can be compared to another (Erickson et al., 1991).

(b) Comparability is the quality that makes data obtained from one study comparable to data from other studies. Consistent sampling methodology, handling, and analyses are necessary to ensure comparability. Also, assurance that equipment has been calibrated properly and analytical solutions prepared identically is necessary to attain data comparability (Air National Guard, 1993).

Representativeness

(a) Representativeness can be defined both qualitatively and quantitatively; it depends on the experimental design and choice of sampling methods. The desired degree of representativeness is important in planning for the collection of samples and the subsequent uses of the data. A relevant sampling design issue, for example, is to determine how a sample will be collected to ensure it is representative of the desired characteristic (Erickson et al., 1991).

(b) Representativeness is a measure of how representative the data obtained for each parameter is compared with the value the same parameter has within the population being measured. Since the total population cannot be measured, sampling must be designed to ensure that the samples are representative of the population being sampled (Air National Guard, 1993).

Completeness

(a) Completeness is defined as the amount of valid data obtained from a measurement system compared to the amount that was expected to be obtained under anticipated sampling/analytical conditions (Erickson et al., 1991).

(b) Completeness is the amount of valid data obtained from the measurement system (field and laboratory) versus the amount of data expected from the system. An assessment of the completeness of data is performed at the end of each sampling event, and if any omissions are apparent an attempt is made to resample the parameter in question, if feasible. Data completeness should also be assessed prior to the preparation of data reports that check the correctness of all data. An example of a formula used for this purpose is

$$\%\,C = 100\left[\frac{V}{n}\right]$$

where

%C = percent complete;
V = number of measurements judged valid; and
n = total number of measurements necessary to achieve a specified level of confidence in decision making (Cross-Smiecinski and Stetzenback, 1994).

A8 Project Narrative

This is a narrative description of work to be performed that will demonstrate to technical or QA reviewers that the project or task will achieve its quality objectives. See USEPA (1994e) for complete details of what should be included in a project narrative.

A9 Special Training Requirements/Certification

If personnel will require any specialized training or certification to successfully complete the

project, discuss how this training will be obtained and documented.

A10 Documentation and Records

Itemize all of the information and records (e.g., raw data, field logs, instrument printouts, results of calibration and QC checks, analytical laboratory case narratives) that must be included in a data report package, and describe the desired report format and final disposition of records and documents.

5.3.2 Group B: Measurements and Acquisition

The Project/Task Description element (A6) contains a summary of this information, which should be provided in detail in this section. Methods that have been well documented and are available to all participants can merely be cited; for those not well documented, detailed copies of the methods and/or Standard Operating Procedures (SOPs) must be provided in the QAPP.

B1 Sampling Process Design (Experimental Design)

Explain the experimental design or data collection design, including types and numbers of samples required, sampling locations and frequencies, sampling screening criteria (if applicable), sample matrices, measurement parameters of interest, and the rationale for the design. As with all information contained in a QAPP, recording information such as the reasoning behind decisions will make the data more defensible in the future. Statistics can play an important part in determining the sampling strategy. Therefore, record all statistical procedures that will be used to determine the sampling strategy. Two basic sampling decisions that must be made are the types and numbers of quality control samples to be collected (Keith, 1988). See USEPA (1994e) for additional details on what to include in this element of the QAPP.

B2 Sampling Methods Requirements

Identify and describe all procedures for collecting samples for each sampling method, as well as what should be done when a sampling or measuring failure occurs and who is responsible for taking corrective action. Other aspects pertinent to sampling, such as record keeping, sample storage, and transport to laboratories, should also be described in this section (Cross-Smiecinski and Stetzenback, 1994).

B3 Sample Handling and Custody Requirements

Describe all aspects of sample handling and custody. Sample custody is a documentation of where and with whom samples are at all times from the moment they are collected in the field to when they are analyzed in the laboratory. A sample is considered to be under custody if (1) it is in your actual possession; (2) it is in your view, after being in your physical possession; (3) it was in your physical possession and then you locked it up to prevent tampering; or (4) it is in a designated and identified secure area (Air National Guard, 1993). Special tracking procedures called "chain-of-custody" procedures are used whenever samples are collected for use in an enforcement action or when demonstrating compliance with a regulatory requirement (e.g., NPDES). Chain-of-custody forms should be printed on multipart carbonless paper for tracking custody and should have, at a minimum, space for recording date, time, name of person accepting samples, sample numbers, and remarks (Figure 5-3). Copies of the form must be completed in the field, and signed by the fieldteam when they transfer custody of the samples to the shipper. Upon receipt in the laboratory, the laboratory signs the remaining copies, indicating they have accepted custody of the samples. Each time the form is signed, the person signing the form

CHAIN-OF-CUSTODY RECORD

PROJ NO	PROJECT NAME	No. of Containers	TAG NUMBERS	PARAMETERS				REMARKS
SAMPLERS (*Signature*)								
DATE	TIME	DESCRIPTION						

Relinquished by: (*Signature*)	Date	Time	Received by: (*Sign.*)	Relinquished by (*Sign.*)	Date	Time	Received by (*Sign.*)
Print Name:			Print Name	Print Name			Print Name
Relinquished by: (*Signature*)	Date	Time	Received by: (*Sign.*)	Date	Time	Remarks	
Print Name:			Print Name				

Figure 5-3. Sample custody chart. (After Cross-Smiecinski and Stetzenback, 1994)

retains the bottom copy and passes the remaining copies along with the samples. The laboratory should return at least one copy of the completed chain-of-custody record to the client, or proceed as directed in the QAPP.

B4 Analytical Methods Requirements

Describe the analytical methods and equipment required for both field and laboratory activities, waste disposal requirements, and specific performance requirements, as well as what should be done when a failure in the analytical system occurs and who is responsible for taking the corrective action. Also include information on any supporting methods or documents used to collect field or laboratory data. For instance, if identifications of benthic invertebrates are made, include information on the source(s) used to verify identifications; if the amount of riparian vegetation cover is estimated, describe the method used to arrive at the estimate.

Analytical methods: Describe the methods that will be used for the project. If the methods to be used are published (e.g., by the U.S. Geological Survey, EPA, or ASTM) it is sufficient to indicate what methods will be used and where descriptions of them can be found. If the best methods to be used cannot be completely ascertained until some samples have been analyzed, indicate the order of preference for use of the methods. Any modifications to published or standard methods or variations of them must be documented, and the variations must be verified as providing data of acceptable quality.

Method validation: Method validation accounts for and documents, at a minimum, the following characteristics: known and possible interferences; method precision; method accuracy, bias, and recovery; method detection level, and method comparability to superseded methods, if any (Pritt and Raese, 1992). All methods chosen for use in the project must be validated.

Generally, laboratories with their own QA/QC procedures will be used for sample analyses. The methods to be used in the laboratory must be acceptable to project managers. All potential laboratory facilities to be used in the project should be extensively evaluated before their selection and throughout their participation in the project.

B5 Quality Control Requirements

Identify the QC procedures (types, frequency, and control limits of QC checks) needed for each sampling, analysis, or measurement technique. (They might have to be modified to suit each project.) Also state what corrective action is required when control limits are exceeded. Data collected as part of field sampling and laboratory measurements must be verified as accurate. Thus, some samples are taken or measurements made to check for accuracy rather than to collect additional data. Specify what means will be used to check the accuracy of samples and measurements. Field blanks, duplicate samples, replicate samples, spiked samples, and spiked blanks are commonly used methods. Describe precisely how these control samples will be prepared for analysis.

Standard reference materials (SRMs) should be used periodically in any measurement system to monitor for changes to the system that might go unnoticed. SRMs should be used when a measurement change is noted to verify that the change is not due to a change in the measurement system. The optimum frequency of use of SRMs and also of replicates of actual test samples depends on the integrity of the measurement system and the magnitude of the errors involved when the system ceases to give predictable results. All measurements from last-known-in-control sample to first-known-out-of-control sample are suspect, so the length of the period between these two samples must be calibrated to be appropriate to the measurements being made (Taylor, 1993).

B6 Instrument/Equipment Testing, Inspection, and Maintenance Requirements

This section should include descriptions of the types of preventive maintenance for equipment that will be used to ensure that research schedules are adhered to and project objectives are completed on schedule. The section should include the following: a schedule of preventive maintenance, an inventory of critical spare parts and supplies, maintenance contract information, location of important manuals and instructions, record keeping requirements, and training of instrument and equipment operators (Cross-Smiecinski and Stetzenback, 1994). Some aspects of training can be considered a part of preventive maintenance. Describe in this section general safety precautions that will be part of project operations. Examples include materials handling, transportation of chemicals, hazardous waste

disposal procedures, emergency procedures, standard safety operations, chemical hygiene, hazard communication, hazardous waste management, waste disposal, location of safety equipment, tour of facilities, and annual classes in cardiopulmonary resuscitation and standard first aid (Pritt and Raese, 1992).

B7 Instrument Calibration and Frequency

Describe the procedures used for equipment calibration, the frequency of calibration of each piece of equipment, and the results of calibration procedures. Record any problems encountered and corrective actions taken. This section should identify each tool, gauge, instrument, or other sampling, measuring, and test equipment used for data collection activities for which quality must be controlled and which must be calibrated to maintain performance within specified limits.

B8 Inspection/Acceptance Requirements for Supplies and Consumables

Supplies and consumables to be used in the project must be inspected and accepted, according to specified criteria, for use in the project. Identify who will perform the inspections and how they will be conducted.

B9 Data Acquisition Requirements (Nondirect Measurement)

Data obtained from noninstrument sources such as computer databases, spreadsheets, and programs and literature files need to be identified and acceptance criteria established for the use of the data. Also discuss any limitations resulting from uncertainty in the quality of the data and the impact of adding more error to the results.

B10 Data Management

This section should describe all aspects of data management, from their generation in the field or laboratory to final use or storage. Discuss the control mechanisms (and provide examples of forms or checklists) for detecting and correcting errors and for preventing loss of data during data reduction. This discussion should also include all data handling equipment and procedures that will be used to process, compile, and analyze the data (hardware and software).

5.3.3 Group C: Assessment/Oversight

The purpose of these elements is to ensure that the QAPP will be implemented as prescribed; they describe the activities for assessing the effectiveness of the implementation of the QAPP and its associated QA/QC program.

C1 Assessments and Response Actions

Assessments can include a variety of activities, such as surveillence, peer review, management systems review, technical systems audit, or performance evaluation. Audits are assessments of the extent to which QA procedures and QC activities are being adhered to. They may be performed by an internal (i.e., within the project structure) but independent audit team or by an external audit team. Audits may be performed before, during, and/or after the project is performed. Audit frequency, intensity, and type should be determined, and the audit(s) should be scheduled as part of the overall program QA effort. This section of the QAPP should describe the audits to be performed and the process and procedures for responding to problems raised during audits (Cross-Smiecinski and Stetzenback, 1994).

This section should also describe actions to be taken if and when unexpected problems arise during the course of the study. Problems that can be foreseen, such as running low on commonly used laboratory supplies, should be addressed as SOPs. Many problems, however, are encountered so infrequently or are unpredictable

enough that SOPs will not be prepared for them. Special or emergency procedures address these types of problems. It is difficult to address unanticipated problems before they arise, but the QAPP should specify who is responsible for handling problems that arise from different aspects of the project (e.g., field sampling, laboratory analysis, audits). It is helpful to categorize problems based on their impact on the project (e.g., critical, important, noncritical, unimportant) and to specify the type of corrective action necessary based on the problem's category. A critical problem, for instance, would be one that would affect obtaining data of the necessary quality or quantity. If a critical problem arises, a critical-problem response by project staff would be required. This should be specified in the QAPP.

C2 Reports to Management

This section specifies the type and frequency of reports to be prepared and submitted to project management, as well as the chain of responsibility for ensuring that reports are prepared and submitted. The preparer of the reports and recipients of each report should be identified. Any required report contents and format should also be specified.

5.3.4 Group D: Data Validation and Usability

After the data collection has been completed, the data must be examined to determine whether they conform to the specified criteria and will satisfy project objectives.

D1 Data Review, Validation, and Verification Requirements

The requirements used to review and accept, reject, or qualify data should be identified, including any project-specific calculations or algorithms.

D2 Validation and Verification Methods

This section should describe each of the elements defined below in enough detail to support use of the data for their intended purpose and for comparability to past, present, and future studies (Cross-Smiecinski and Stetzenback, 1994). If computer software is used in data manipulations, record which software is to be used. Software that performs complex manipulations might have to be verified before its use to ensure that it functions properly (Cross-Smiecinski and Stetzenback, 1994).

Data reduction: The transformation of raw data into a more useful form, calculations.

Data verification: A routine activity conducted by technical, laboratory, and clerical personnel on small sets of the data to determine whether data have been accurately quantified, recorded, and transcribed; whether data have been collected and analyzed in accordance with prescribed, approved procedures; whether the data appear suitably complete; and whether the data appear to be reasonable and consistent, based on prior knowledge of the research.

For example, it is a good practice to enter data into the database twice and scan them for outlying values. This helps to detect and eliminate transcription errors. Range checks, internal consistency checks, and quality assurance evaluations should also be included for data certification (Drouse et al., 1986).

Data validation: The process by which a sample, measurement method, or datum is deemed useful for a specified purpose; an independent, timely review of a body of verified data against a predetermined set of qualitative and quantitative criteria to evaluate their adequacy for their intended use.

Data reporting: Specify any special forms or formats (e.g., tables and figures) that are to be used, as well as who is responsible for data reporting, due dates, etc.

D3 Reconciliation with User Requirements

The precision, accuracy, completeness, representativenesss, and comparability of data must be assessed using appropriate techniques. This section should give details of the formulas, statistical techniques, and procedures that will be used to assess the data. The methods used to assess the data must be in agreement with the DQOs. The terms *precision, accuracy, completeness, representativeness*, and *comparability* are defined on page 5-11, and some sample data assessment formulas are given.

The following sections provide more specific information for preparing QAPPs with respect to field and laboratory operations, and data and reporting requirements.

5.4 Field Operations

Field operations are an important activity in an NPS monitoring program. Field operations involve the organization and design of the field operation, selection of sampling sites, selection of sampling equipment, sample collection, sample handling and transport, and safety and training issues. For the purposes of QA and QC, the process of conducting field operations should be broken down into as many separate steps as are necessary to ensure complete consideration of all of the elements and processes that are a part of field activities. Field operations described in this section have been broken down into the phases mentioned above, but individual monitoring programs might require the use of more or fewer phases. For example, if the sample collection phase is very complex or if it is anticipated that sample collection will often be done under inclement weather conditions when field personnel might experience discomfort and feel rushed, it is advisable to break sample collection into separate preparation, sampling, and termination phases and discuss QA and QC for each of the phases separately. This will ensure that no details are omitted. Table 5-3 summarizes many important items that should be considered in the field operations portion of a QA/QC program.

5.4.1 Field Design

Adherence to the procedures specified in the QAPP for field operations and documentation of their use for all aspects of field operations are extremely important if the data obtained from the project are to be useful for decision making, supportable if questioned, and comparable for use by future researchers (Knapton and Nimick, 1991). Data sheets prepared beforehand, with quality reminders included where appropriate, will help ensure that all data are collected and QA/QC procedures are followed during all field activities.

General information that should be included in the documentation of the design for field operations includes the scale of the operations (laboratory, plot, hillslope, watershed); size of plots/data collection sites; designation of control sites; basin characteristics; soil and vegetation types; maps with the location of plots/data collection sites within the basin/catchment; weather conditions under which sampling is conducted; equipment and methods used; problems that might be encountered during sampling; dates of commencement and suspension of data collection; temporal gaps in data collection; frequency of data collection; intensity of data collection; and sources of any outside information (e.g., soil types, vegetation identifications) (Erickson et al., 1991). Some of these aspects are discussed in greater detail in the following sections.

Table 5-3. Checklist of items that should be considered in the field operations section of a QA/QC program.

Field Operations			
Element	**Specifics**	**Check-off**	**Responsibility**
Organization	Field organization chart created	_____	
	Staff duties and responsibilities defined	_____	
	Communication lines within and with other units established	_____	
	Project documents made available to all staff	_____	
	Staff qualifications established	_____	
Field Logistics	Sampling sites investigated and selected	_____	
	Means of access to sampling sites determined	_____	
	Sample transport and shipping procedures specified	_____	
	Field sample handling areas selected	_____	
	Chain-of-custody for samples established	_____	
	Field equipment selected and supplied	_____	
	Procedures for decontamination of sampling equipment established	_____	
Monitoring Equipment	Equipment installation procedures specified	_____	
	Equipment maintenance and control schedules established	_____	
	Equipment maintenance manual updated and distributed		
	Trouble shooting and corrective action manual updated and distributed	_____	
Quality Control Samples	Type(s) of control samples (blanks, duplicates, spikes, analytical standards, reference materials) to be used have been determined	_____	
	Frequency of control sample use has been determined	_____	
Field Audits	QA field auditor designated	_____	
	Aspects of field operations that will undergo quality assessments as part of field audits have been determined	_____	
	Acceptance criteria for compliance with SOPs and the QAP set for field events and activities	_____	
	Field audit forms, with investigations to be conducted and data to be collected, prepared	_____	
	Person(s) to review field audit records designated	_____	

Table 5-3. (continued)

Health and Safety	Field personnel properly trained ..... Proper field gear and clothing issued to field personnel	_____ _____	
SOPs	Sample management ..... Sample collection procedures ..... Reagent preparation ..... Decontamination ..... Equipment calibration and maintenance ..... Corrective action ..... Waste disposal ..... Health and safety ..... Field measurements ..... Reagent/standard preparation ..... Equipment calibration and maintenance ..... Data reduction and validation ..... Reporting ..... Corrective action ..... Waste disposal ..... Health and safety ..... Records management ..... Project-specific records ..... Field operations records	_____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____	

Source: USEPA, 1990b.

5.4.2 Sampling Site Selection

The selection of sampling sites is important to the validity of the results. Sites must be selected to provide data to meet the goals/objectives of the project. The QAPP should provide detailed information on sampling site locations (e.g., latitude and longitude); characteristics that might be important to data interpretation (e.g., percent riparian cover, stream order); and the rationale for selecting the sites used (Knapton and Nimick, 1991). Sites from other studies can be convenient to use due to their familiarity and the availability of historical data, but such sites should be scrutinized.

carefully to be certain that data obtained from them will serve the objectives of the project. If during the course of the project it is found that one or more sampling sites are not providing quality data, alternative sites might be selected and the project schedule adjusted accordingly. The adequacy of the sampling locations and the sampling program should be reviewed periodically by project managers, as determined by data needs (Knapton and Nimick, 1991).

Sampling sites should be visited before sampling begins. It is important to verify that the sites are accessible and are suitable for collection of the data needed. Consideration should be given to accessibility in wet or inclement weather if samples will be taken during such conditions. The sites should be visited, if possible, in the type(s) of weather during which sampling will occur. Plastic- laminated pictures of each sampling site with an arrow pointing to each monitoring location can assist field personnel in

finding the sites during inclement weather when the sites might appear different.

If permission to access a site is needed (for instance, if one or more sites are on or require passage through private property), such permission must be obtained before sampling begins. The person(s) granting the permission should be fully informed about the number of persons who will be visiting during each sampling event, frequency of sampling, equipment that will have to be transported to the sampling site(s), any hazardous or dangerous materials that will be used during sampling, and any other details that might affect the decision of the person(s) to grant access permission. A lack of full disclosure of information to gain access permission creates a risk of the permission's being revoked at some point during the project.

5.4.3 Sampling Equipment

Equipment for field operations includes field-resident equipment such as automatic samplers and stage-level recorders and nonresident sampling equipment such as flow, pH, and conductivity meters; equipment needed to gain access to sampling sites such as boats; and equipment for field personnel health and safety, such as waders, gloves, and life vests. The condition and manner of use of the field equipment determines the reliability of the collected data and the success of each sampling event. Therefore, operation and maintenance of the equipment are important elements of field QA and QC. All measurement equipment must be routinely checked and calibrated to verify that it is operating properly and generating reliable results (Spooner, 1994), and all access and health and safety equipment should be routinely checked to be certain that it will function properly under all expected field conditions.

A manual with complete descriptions of all field equipment to be used should be available to all field personnel. The manual should include such information as model numbers for all measurement equipment, operating instructions, routine repair and adjustment instructions, decontamination techniques, sampling preparation instructions (e.g., washing with deionized water), and use limitations (e.g., operating temperature range). If any samples are to be analyzed in the field, the techniques to be used should be thoroughly described in the manual.

5.4.4 Sample Collection

The process of sample collection should be described with the same amount of detail as the equipment descriptions. A thorough description of the sample collection process includes when the sampling is to be done (e.g., time of day, month, or year; before and/or after storms); the frequency with which each type of sample will be collected; the location at which samples are to be taken (i.e., depth, distance from shore, etc.); the time between samples (if sampling is done repetitively during a single sampling site visit); and how samples are to be labeled. Each field person must be thoroughly familiar with the sampling techniques (and equipment) prior to the first sampling event. Holding practice sampling events prior to the commencement of actual sampling is an excellent way to prepare all field personnel and will help to identify potential problems with the sampling sites (access, difficulty under different weather conditions), sampling equipment, and sampling techniques.

Quality control activities for field operations must ensure that all field operations are conducted so that sampling is done in a consistent manner and that all generated information is traceable and of known and comparable quality. Each field activity should be standardized. Standard operating procedures (SOPs) for field sampling have been developed and might be required depending on the agency for which the sampling is being conducted. Elements of the field

operations section of a QAPP should include clear statements of the regulatory requirements applicable to the project (Spooner, 1994). Any SOPs that are part of regulatory requirements should be followed precisely. The pictures taken of each sampling site to aid in locating the sampling sites also help ensure consistency of field monitoring across time and personnel by ensuring that the same spot is used at each sampling event (Spooner, 1994).

Depending on the DQOs and data requirements of the program (type of data and frequency of collection), additional quality control samples might be needed to monitor the performance of various field (as well as laboratory) operations including sampling, sample handling, transportation, and storage.

As the samples are collected, they must be labeled and packaged for transport to a laboratory for analysis (or other facility for nonchemical analyses). Computer-generated sample bottle labels prepared before the sampling event and securely attached to each bottle help minimize mistakes. Sampling location and preservation, filtration, and laboratory procedures to be used for each sample should be recorded on each label (Spooner, 1994). Be sure these labels are printed with waterproof ink on waterproof paper, and use a No. 2 pencil or waterproof/solvent-resistant marker to record information.

5.4.5 Sample Handling and Transport

Once samples have been collected, they must be analyzed, usually in a laboratory. Handling and transport of sampling containers and custody of sample suites is also a part of field operations. Sample transport, handling, and preservation must be performed according to well-defined procedures. The various persons involved in sample handling and transport should follow SOPs for this phase of the project. This will help ensure that samples are handled properly, comply with holding time and preservation requirements, and are not subject to potential spoilage, cross-contamination, or misidentification.

The chain of custody and communication between the field operations and other units such as the analytical laboratory also need to be established so that the status of the samples is always known and can be checked by project personnel at any time. The chain of custody states who the person(s) responsible for the samples are at all times. It is important that chain of custody be established and adhered to so that if any problem with the samples occurs, such as loss, the occurrence can be traced and possibly rectified, or it can be determined how serious the problem is and what corrective action needs to be taken. Field data custody sheets are essential for this effort (Cross-Smiecinski and Stetzenback, 1994; Spooner, 1994). Chain-of-custody seals must be applied to sample containers and shipping containers.

5.4.6 Safety and Training

When dealing with NPS monitoring, sampling activities often occur during difficult weather and field conditions. It is necessary to assess these difficulties and establish a program to ensure the safety of the sampling personnel. The following types of safety issues, at a minimum, should be considered and included in training and preparation activities for sampling: exposure, flood waters, debris in rivers and streams, nighttime collecting, criminal activity, and first aid for minor injuries. The trade-off between the need for data quality and the safety of personnel is a factor that project staff should consider collectively.

Finally, the QAPP for the field operations should include provisions for dealing with any foreseeable problems such as droughts, floods, frozen water, missing samples, replacement personnel during sickness or vacation, lost

samples, broken sample containers, need for equipment spare parts, and other concerns (Spooner, 1994).

5.5 LABORATORY OPERATIONS

Laboratory operations should be conducted with the same attention to detail as field operations. Often, an independent laboratory conducts sample analyses, so QA and QC for the laboratory are not under the direct control of project personnel. However, it is important that project personnel are certain that the laboratory chosen to do analyses follows acceptable QA/QC procedures so that the data produced meet the DQOs established for the project. Laboratories should be selected based on quality assurance criteria established early in the project. The Quality Assurance Officer for the project should be certain that these criteria are used for selecting a laboratory to perform any necessary analyses for the project and that any laboratories selected meet all criteria. Laboratories can be evaluated through the following measures (Air National Guard, 1993):

- Performing proficiency testing through analysis of samples similar to those which will be collected during the project.
- Performing inspections and audits.
- Reviewing laboratory QA/QC plans.
- One or more of these measures should be used by the project manager, and the laboratories should be visited before entering into a contract for sample analyses.

5.5.1 General Laboratory QA and QC

Numerous references are available on laboratory QA/QC procedures, and one or more should be consulted to gain an understanding of laboratory QA and QC requirements if project personnel are not familiar with them already. The details of a laboratory's QA/QC procedures must be included in the QAPP for the NPS monitoring project. Some elements to look for in a laboratory QA/QC plan include (Cross-Smiecinski and Stetzenback, 1994):

- How samples are received
- Proper documentation of their receipt
- Sample handling
- Sample analysis
- QC requirements (procedures and frequencies of QC checks, criteria for reference materials, types of QC samples analyzed and frequencies)
- Waste disposal
- Cleanliness and contamination
- Staff training and safety
- Data entry and reporting
- Confidentiality

This section provides some information on laboratory QA/QC procedures to which managers of monitoring programs should pay particular attention when deciding to use a particular laboratory for sample analysis (Table 5-4). More detailed references on laboratory QA and QC should be consulted for further information.

5.5.2 Instrumentation and Materials for Laboratory Operations

The laboratory chosen to do chemical analyses should have all equipment necessary to perform the analyses required, including organic analysis, inorganic analysis, and assessments of precision

Table 5-4. Checklist of items that should be considered in the laboratory operations section of a QA/QC program.

Laboratory Operations			
Element	**Specifics**	**Check-off**	**Responsibility**
Sample Management	Sample receipt Sample storage Sample handling Sample scheduling	_____ _____ _____ _____	
Equipment	Equipment calibration and maintenance	_____	
SOPs	Sample management Analytical methods Sample preparation and analysis procedures Reagent/standard preparation Raw data requirements Data reduction and validation Precision, accuracy, and method detection/reporting limits Reporting Corrective actions	_____ _____ _____ _____ _____ _____ _____ _____ _____	
Records Management	Project-specific records Laboratory operations records	_____ _____	
QC Procedures	Control samples Method blanks Matrix spikes Matrix duplication/matrix spike duplicates	_____ _____ _____ _____	
Audits	Laboratory audits schedule	_____	
Health and Safety	Fire and emergency equipment Fire and emergency equipment inspection Health equipment (masks, gloves, ...) Waste disposal	_____ _____ _____ _____	

Source: USEPA, 1990b.

and accuracy. If any specialized analyses are required (e.g., microbiology, histopathology, toxicology), be certain that the laboratory has the appropriate equipment and that laboratory staff are adequately trained to perform the desired analyses. As noted in the elements of the QAPP, periodic calibration checks that are conducted to ensure that measurement systems (instruments, devices, techniques) are operating properly should

be described in the QAPP, including procedures and frequency (Cross-Smiecinski and Stetzenback, 1994).

5.5.3 Analytical Methods

The laboratory chosen for sample analysis should use analytical methods approved by the agency for which the sampling is being conducted or by project personnel, as appropriate. Standard methods include those published by the U.S. Geological Survey, the U.S. Environmental Protection Agency, and the American Society for Testing and Materials, or those published in *Standard Methods for the Analysis of Water and Wastes* (Clesceri et al., 1989). If any methods to be used are not published, they should first be validated and verified as acceptable for the project. Each approved and published method should be accompanied by an SOP that is followed rigorously by the laboratory (Pritt and Raese, 1992).

5.5.4 Method Validation

The laboratory chosen for sample analysis should have well-developed procedures for method validation. Method validation should account for and document the following (at a minimum): Known and possible interferences; method precision; method accuracy, bias, and recovery; method detection level; and method comparability to superseded methods, if applicable (Pritt and Raese, 1992).

5.5.5 Training and Safety

An analytical laboratory should be able to ensure its customers that its personnel are adequately trained to perform the necessary analyses. Individual laboratory staff should be independently certified for each of the analyses they will be allowed to perform in the laboratory. Selection of a laboratory for sample analysis should be based on queries about how often training is conducted, whether employees are limited to using equipment for which they have been adequately trained, whether the training program is independently certified, who conducts the training, how the staff's competence with individual instruments is measured, and other factors (Pritt and Raese, 1992).

Safety for staff is an important consideration when choosing a laboratory because, aside from the paramount concern for human well-being, accidents can seriously delay sample analyses or create a need for resampling. Prospective laboratories should be inspected for their attention to safety procedures, including the availability of safety equipment such as fire extinguishers, safety showers and eyewashes, fume hoods, and ventilation systems; use and disposal practices for hazardous materials; and compliance with environmental regulations. Safety equipment should be tested on a regular basis (Pritt and Raese, 1992).

Additionally, laboratory safety includes procedures for ensuring that the laboratory is accessible only to authorized personnel to ensure confidentiality of the data. The laboratory should have a system for accounting for and limiting (or denying) laboratory access to all visitors, including persons affiliated with projects for which the laboratory is analyzing samples (Pritt and Raese, 1992).

5.5.6 Procedural Checks and Audits

A laboratory should have established procedures (SOPs) for conducting internal checks on its analyses and taking corrective action when necessary. If more than one laboratory is used for sample analyses, it will be important to know that the data obtained from the two are of the same quality and consistency. A protocol for conducting interlaboratory comparisons should also be an element of a laboratory's QA/QC plan. For many projects occasional samples are

analyzed by a second laboratory to determine whether there is any bias in the data associated with the primary laboratory's analyses.

Laboratory audits by independent auditors are normally conducted on a prescribed basis to ensure that laboratory operations are conducted according to accepted and acceptable procedures (Cross-Smiecinski and Stetzenback, 1994). Determination that a laboratory undergoes such audits and reviews audit results might be sufficient to determine that a laboratory will be adequate for conducting analyses of samples generated by the NPS monitoring project.

5.6 Data and Reports

It is essential during the conduct of an NPS monitoring project to document all data collected and used, to document all methods and procedures followed, and to produce clear, concise, and readable reports that will provide decision makers with the information they need to choose among alternative actions, as described in the DQOs.

5.6.1 Generation of New Data

All data generated during the project, whether in the field, laboratory, or some other facility, should be recorded. Include with the data any reference materials or citations to materials used for data analyses. These include computer programs, and all computer programs used for data reduction should be validated prior to use and verified on a regular basis. Calculations should be detailed enough to allow for their reconstruction at a later date if they need to be verified (Cross-Smiecinski and Stetzenback, 1994). Data generated by a laboratory should be accompanied by pertinent information about the laboratory, such as its name, address, and phone number, and names of the staff who worked directly with the project samples.

5.6.2 Use of Historical Data

Historical data are data collected for previous projects that concerned the same resource in the same area as the project to be implemented. Historical data sometimes contain valuable information, and their use can save time and effort in the implementation and/or data analysis phases of a new project. Before new data are collected, all historical data available should be obtained and their validity and usability should be assessed. *Data validity* implies that individual data points are considered accurate and precise because the field and laboratory methods used to generate the data points are known. *Data usability* implies that a database demonstrates an overall temporal or spatial pattern, though no judgment of the accuracy or precision of any individual data point is made (Spreizer et al., 1992). The validity of historical data can be difficult to ascertain, but data usability can be assessed through a combination of graphical and statistical techniques (Spreizer et al., 1992).

Specifically, historical data that can be shown to be either valid or usable can be applied to a new project in the following ways (Coffey, 1993; Spreizer et al., 1992; USEPA, 1994c):

- If the quality (i.e., accuracy and precision) of historical data is sufficiently documented, the data can be used alone or in combination with new data. The quality of historical data must be determined absolutely, generally with the help of a statistician.

- Characteristics derived from the historical data, such as the variability or mean of data, can be used in the development or selection of a data collection design. Knowledge of expected variability assists in determining the number of samples needed to attain a desired confidence level, the length of monitoring program necessary to obtain the necessary data, and the required sampling frequency.

- Spatial analysis of historical data can indicate which sampling locations are most likely to provide the desired data.
- Historical data can provide insights about past impacts and water quality that can be useful in defining an NPS pollution problem.
- Past trends can be ascertained, and the present tendency of water quality characteristics (degrading, stable, or improving) can be established for trend analysis.

5.6.3 Documentation and Record Keeping

All information and records related to the NPS monitoring project should be kept on file and kept current. This documentation should include:

- A record of decisions made regarding the monitoring project design
- Records of all personnel, with their qualifications, who participated in the project
- Intended and actual implementation schedules, and explanations for any differences
- A description of all sampling sites
- Field records of all sampling events, including any sampling problems and corrective actions taken
- Copies of all field and laboratory SOPs
- Equipment manuals and maintenance schedules (intended and actual, with explanations for any discrepancies)
- Printouts from any equipment
- Sample management and custody records
- Laboratory procedures
- Copy of the laboratory QA/QC plan
- Personnel training sessions and procedures, including any training manuals or other materials
- All data generated during the project in hard copy and electronic forms
- All correspondence related to the project
- Project interim and final reports

5.6.4 Report Preparation

The original project description should include a schedule and required format for required reports, including the final report. Adherence to this schedule is important to provide information and documentation of project progress, problems encountered, and corrective actions taken. Reports are also valuable for supporting continuation of a project if at any point during the project its continuation is scrutinized or if additional funding must be secured to ensure its completion. Reports can also become the primary sources of historical information on projects if there are changes in project personnel during the project. Project managers should decide on the necessary content and format of all reports prior to commencement of the project, and these will differ depending on funding and intended audience.

REFERENCES

Air National Guard. 1993. *Draft final remedial investigation/reasibility study work plan.* Wyoming Air National Guard 153rd Tactical Airlift Group, Cheyenne Municipal Airport, Cheyenne, Wyoming. Prepared by Aepco, Inc. and Tetra Tech, Inc. March.

Alberts, E.E., G.E. Schuman, and R.E. Burwell. 1978. Seasonal runoff losses of nitrogen and phosphorus from Missouri Valley loess watersheds. *J. Environ. Qual.* 7(2):203-207.

Allen , D.M., S.K. Service and M.V. Ogburn-Matthews. 1992. Factors influencing the collection efficiency of estuarine fishes. *Trans. Am. Fish. Soc.* 121(2):234-244.

Aller, L., T. Bennett, J.H. Lehr, and R.J. Petty. 1985. *DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeologic settings.* EPA/600/2-85/018. U.S. Environmental Protection Agency, Washington, DC.

Anderson, R.L. 1941. Distribution of the serial correlation coefficients. *Ann. Math. Stat.* 8(1):1-13.

Anderson, V.L., and R.A. McLean. 1974. *Design of experiments, a realistic approach.* Marcel Dekker, Inc., New York, New York.

ASQC. 1994. *Specifications and guidelines for quality systems for environmental data collection and environmental technology programs.* ANSI/ASQC E4-1994. American Society for Quality Control, ASQC Quality Press, Wisconsin.

Bahls, L. 1993. *Periphyton bioassessment methods for Montana streams.* Water Quality Bureau, Department of Health and Environmental Science, Helena, Montana.

Bailey, G.W. and T.E. Waddell. 1979. Best management practices for agriculture and silviculture: An integrated review. In *Best management practices for agriculture and silviculture.* Ann Arbor Science, Ann Arbor, Michigan.

Baker, D.B., K.A. Krieger, R.P. Richards, and J.W. Kramer. 1985. Effects of intensive agricultural land use on regional water quality in northwestern Ohio. In *Perspectives on nonpoint source pollution.* Proceedings of a National Conference, May 19-22, 1985, Kansas City, Missouri. EPA 440/5-85-001. U.S. Environmental Protection Agency, Washington, DC.

Ball, J. 1982. *Stream classification guidelines for Wisconsin.* Wisconsin Department of Natural Resources Technical Bulletin. Wisconsin Department of Natural Resources, Madison, Wisconsin.

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. *Environ. Toxicol. Chem.* 11(4):437-449.

References

Barbour, M.T., and J.B. Stribling. 1991. Use of habitat assessment in evaluating the biological integrity of stream communities. In *Biological criteria: Research and regulation proceedings of a symposium.* EPA-440/5-91-005. U.S. Environmental Protection Agency, Office of Water. July.

Barbour, M.T., J.B. Stribling and J.R. Karr. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. In *Biological assessment and criteria. Tools for water resource planning and decision making*, ed. W.D. Davis and T. Simon, pp. 63-76. Lewis Publishers, Boca Raton, Florida.

Berryman, D., B. Bernard, D. Cluis and J. Haemmerli. 1988. Nonparametric tests for trend detection in water quality time series. *Water Resour. Bull.* 24:545-556.

Berryman, D., B. Bobée, D. Cluis, and J. Haemmerli. 1988. Nonparametric tests for trend detection in water quality time series. *Water Resour. Bull.* 24(3):545-556.

Brady, N.C. 1984. *The nature and properties of soils.* MacMillan Publishing Company, New York.

Brakensiek, D.L., H.B. Osborn, and W.J. Rawls. 1979. *Field manual for research in agricultural hydrology.* Agriculture Handbook No. 224, U.S. Department of Agriculture, Science and Education Administration, Washington, DC.

Brinson, M.M., B.L. Swift, R.C. Plantico, and J.S. Barclay. 1981. *Riparian ecosystems: Their ecology and status.* FWS/OBS-81/17. Eastern Energy and Land Use Team, National Water Resources Analysis Group, U.S. Fish and Wildlife Service, Kearneysville, West Virginia. September.

Brookes, A. 1990. Restoration and enhancement of engineered river channels: Some European experiences. *Reg. Rivers: Res. Manage.* 5:45-56.

Brown, R.G. 1985. Effects of an urban wetland on sediment and nutrient loads in runoff. *Wetlands* 4:147-158.

Brown, M.P., P. Longabucco, and M.R. Rafferty. 1986. *Nonpoint source control of phosphorus - A watershed evaluation volume 5, The eutrophication of the Cannonsville Reservoir.* New York State Department of Environmental Conservation, Bureau of Technical Services and Research, Albany, New York.

Buttle, J.M., and F. Xu. 1988. Snowmelt runoff in suburban environments. *Nordic Hydrol.* 19:19-40.

Canning, D.J. 1988. *Construction erosion control.* Washington Department of Ecology, Shorelenads and Coastal Zone Management Program. Shorelands Technical Advisory Paper No. 3.

Carnahan, B., H.A. Luther, and J.O. Wilkes. 1969. *Applied numerical methods.* Wiley, New York.

CARWQCB. 1989. *Nonpoint source experience in the Santa Clara Valley.* California Regional Water Quality Control Board, Santa Clara Valley Water District, Santa Clara, California.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. *Graphical methods for data analysis.* Duxbury Press, Boston, Massachusetts.

Chesapeake Bay Program. 1995. *Chesapeake Bay: Introduction to an ecosystem.* Chesapeake Bay Program, Annapolis, Maryland.

Chow, V.T. 1954. *The log-probability law and its engineering applications.* Separate No. 536, 80. American Society of Civil Engineering. November.

Clark, W.H. 1986. *Rock Creek rural clean water program comprehensive water quality monitoring report 1981-1986.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, Idaho.

Clarkson, C.C., D.E. Lehnig, S.V. Plante, R.S. Taylor, and W.M. Williams. 1984. *Hydrologic basis for suspended sediment criteria.* Prepared by Camp, Dresser & McKee for U.S. Environmental Protection Agency, Washington, DC.

Clausen, J.C. 1985. The St. Albans Bay watershed RCWP: A case study of monitoring and assessment. In *Perspectives on nonpoint source pollution.* Proceedings of a national conference, May 19-22, 1985, Kansas City, Missouri. EPA 440/5-85-001, U.S. Environmental Protection Agency, Washington, DC.

Clausen, J.C. 1991. Developing a monitoring system for rural surface waters: Individual BMPs. In *Nonpoint source watershed workshop: Seminar publication.* EPA/625/4-91/027. U.S. Environmental Protection Agency, Office of Research and Development, Office of Water, Washington, DC., pp. 99-102. September.

Clausen, J.C., Wayland, K.G., K.A. Saldi, and K. Guillard. 1993. Movement of nitrogen through an agricultural riparian zone: 1. Field studies. *Water Sci. Tech.* 28(3-5):605-612.

Clesceri, L.S., A.E. Greenberg, and R.R. Trussell, eds. 1989. *Standard methods for the examination of water and wastewater.* 17th ed. American Public Health Association, American Water Works Association, and Water Pollution Control Federation, Washington, DC.

Cochran, W.G. 1977. *Sampling techniques.* 3rd ed. John Wiley and Sons, New York, New York.

Coffey, S. 1993. *The nonpoint source manager's guide to water quality and land treatment monitoring.* North Carolina Cooperative Extension Service, North Carolina State University, Raleigh, North Carolina..

Coffey, S.W., J. Spooner, and M.D. Smolen. 1993. *The nonpoint source manager's guide to water quality and land treatment monitoring.* NCSU Water Quality Group, Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina.

Cohen, A.C., Jr. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* 1(3):217-237.

Cohen, A.N. 1990. *An introduction to the ecology of the San Francisco estuary.* San Francisco Estuary Project, San Francisco, California.

Colby, B.R. 1956. *Relationship of sediment discharge to streamflow.* U.S.G.S. Open File Report. U.S. Geological Survey, Reston, Virginia.

Connin, S. [USEPA, 1991] 1991. *Characteristics of successful riparian restoration projects in the Pacific Northwest.* EPA910/9-91-033. Prepared for U.S. Environmental Protection Agency, Region 10, Water Division, Seattle, Washington. August.

Conover, W.J. 1980. *Practical nonparametric statistics*, 2nd ed. Wiley, New York.

Conquest, L. L., S. C. Ralph, and R. J. Naiman. 1994. Implementation of large-scale stream monitoring efforts: sampling design and data analysis issues. In *Biological monitoring of aquatic systems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Boca Raton, Florida.

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe. 1979. *Classification of wetlands and deepwater habitats of the United States.* U.S. Fish and Wildlife Service, Biological Services Program, Department of the Interior, Washingtion, D.C. FWS/OBS-79/31.

Cowie, G.M., J.L. Cooley and A. Dutt. 1991. *Use of modified benthic bioassessment protocols for evaluation of water quality trends in Georgia.* USDI/GS Project 14-08-0001-G1556 (03); ERC-06-91; USGS/G-1556-03. Technical Completion Report, Institute of Community and Area Development, University of Georgia, Athens, Georgia and Environmental Resources Center, Georgia Institute of Technology, Atlanta, Georgia. July.

Crawford, C.G., J.R. Slack, and R.M. Hirsch. 1983. *Nonparametric tests for trends in water-quality data using the statistical analysis system.* U.S.G.S. Open File Report 83-550. U.S. Geological Survey, Reston, Virginia.

Cross-Smiecinski, A., and L.D. Stetzenback. 1994. *Quality planning for the life science researcher: Meeting quality assurance requirements.* CRC Press, Boca Raton, Florida.

Cummins, K.W. 1994. Bioassessment and Analysis of Functional Organization of Running Water Ecosystems. In *Biological monitoring of aquatic ecosystems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Ann Arbor, Michigan.

Cupp, C. E. 1989. *Stream corridor classification for forested lands of Washington.* Report for Washington Forest Protection Association, Olympia, Washington.

Dakins, M.E., P.S. Porter, M. West, and S.T. Rao. 1996. Using uncensored trace-level measurements to detect trends in ground water contaminants. *Water Resour. Bull.* 32(4):799-805.

Dauble, D.D., and R.H. Gray. 1980. Comparison of a small seine and a backpack electroshocker to evaluate near shore fish populations in rivers. *Prog. Fish Cult.* 42:93-95.

Davenport, T.E. and M.H. Kelly. 1984a. *Water resource data and preliminary trend analysis for the Highland Silver Lake monitoring and evaluation project, Madison County, Illinois.* IEPA/WPC/84-030, Illinois Environmental Protection Agency, Springfield, Illinois.

Davenport, T.E., and M.H. Kelly. 1984b. *Soil erosion and sediment transport dynamics in Blue Creek watershed, Pike County, Illinois.* IEPA/WPC/83-004. Planning Section, Water Pollution Control, Illinois Environmental Protection Agency, Springfield, Illinois.

Davis, W.S., B.D. Snyder, J.B. Stribling, and C. Stoughton. 1996. *Summary of state biological assessment programs for streams and wadeable rivers.* EPA 230-R-96-007. U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Washington, DC.

Day J.W., Jr., C.A.S. Hall, W.M. Kemp, and A. Yáñnez-Arancibia. 1989. *Estuarine ecology.* John Wiley and Sons, Inc., New York, New York.

Dean, J.D., P.P. Jowise, and A.S. Donigian, Jr. 1984. *Leaching evaluation of agricultural chemicals (leach handbook).* PB8-4-236413, U.S. Environmental Protection Agency, Athens, Georgia.

Dewey, M.R., L.E. Holland-Bartel and S.T. Zigler. 1989. Comparison of fish catches with buoyant pop nets and seines in vegetated and nonvegetated habitats. *North. Am. J. Fish. Manage.* 9:249-253.

Dressing, S.A., J.M. Kreglow, R.P. Maas, F.A. Koehler, W.K. Snyder, and F.J. Humenik. 1982. *Best management practices for agricultural nonpoint source control: II. Commercial fertilizer.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Dressing, S.A., J. Spooner, J.M. Kreglow, E.O. Beasley, and P.W. Westerman. 1987. Water and sediment sampler for plot and field studies. *J. Environ. Qual.* 16(1):59-64.

Drouse, S.K., D.C. Hillman, J.L. Engles, L.W. Creelman and S.J. Simon. 1986. *National surface water survey. National stream survey (Phase 1 - Pilot, mid-Atlantic Phase 1 southeast screening, and episodes pilot) quality assurance plan.* EPA/600/4-86/044. NTIS No. PB87-145819. Prepared by Lockheed Engineering and Management Services Co., Inc., Las Vegas, Nevada, for U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. December.

El-Shaarawi, A.H., and E. Damsleth. 1988. Parametric and nonparametric tests for dependent data. *Water Resour. Bull.* 24:513-519.

Erickson, H.E., M. Morrison, J. Kern, L. Hughes, J. Malcolm and K. Thornton. 1991. *Watershed manipulation project: Quality assurance implementation plan for 1986-1989.* EPA/600/3-91/008. NTIS No. PB91-148395. Prepared by NSI Technology Services Corporation, Corvallis, Oregon for Corvallis Environmental Research Laboratory, Oregon. January.

Fisher, J.S., R.R. Perdue, M.F. Overton, M.D. Sobsey, and B.L. Sill. 1987. *A comparison of water quality at two recreational marinas during a peak-use period.* North Carolina State University, Raleigh, North Carolina. Prepared for the National Oceanic and Atmospheric Administration.

Freund, J.E. 1973. *Modern elementary statistics.* Prentice-Hall, Englewood Cliffs, New Jersey.

Freund, R.J., and R.C. Littell. 1981. *SAS for linear models: A guide to the ANOVA and GLM procedures.* SAS Institute, Inc., Cary, North Carolina.

Frissell, C.A., W.J. Liss, C.E. Warren, and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environ. Manage.* 10(2):199-214.

Gaines, A.G., Jr., and A.R. Solow. 1990. *The distribution of fecal coliform bacteria in surface waters of the Edgartown Harbor coastal complex and management implications.* Edgartown Harbor Project, Interim Report No. 5. Edgartown Harbor Association, Inc.

GAO. 1986. *The Nation's water: Key unanswered questions about the quality of rivers and streams.* Report No. GAO/PEMD-86-6. U.S. General Accounting Office, Washington, DC.

Gaugush, R.F., ed. 1986. *Statistical methods for reservoir water quality investigations.* Instruction Report E-86-2, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi.

Gerritsen, J., J.S. White, E.W. Leppo, J.B. Stribling, and M.T. Barbour. 1996. *Examination of the effects of habitat quality, land use, and acidification on the macroinvertebrate communities of coastal plain streams.* CBWP-MANTA-TR-95-2. Chesapeake Bay Research and Monitoring Division, Maryland Department of Natural Resources, Annapolis, Maryland.

Gibson, G.R., M.T. Barbour, J.B. Stribling, J. Gerritsen and J.R. Karr. 1994. *Biological criteria: Technical guidance for streams and small rivers.* EPA 822-B-94-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC. September.

Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring.* Van Nostrand Reinhold Company, New York.

Gilliom, R.J., and D.R. Helsel. 1986. Estimation of distributional parameters for censored trace level water quality data, 1, estimation techniques. *Water Resour. Res.* 22(2):135-146.

Goodman, J. 1991. Evaluation of site-specific ground-water quality data. In *Nonpoint source watershed workshop: Seminar publication*, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Pp. 145-151.

Goodman, J., D. German, and J. Bishoff. 1996. *Ground water monitoring. A guide to monitoring for agricultural nonpoint source pollution projects.* South Dakota Department of Environment and Natural Resources, Pierre, South Dakota.

Gray, J.S. 1989. Effects of environmental stress on species-rich assemblages. *Biol. J. Linn. Soc.* 37:19-32.

Green, R.H. 1993. Application of repeated measures designs in environmental impact and monitoring studies. *Austr. J. Ecol.* 18:81-98.

Griffin, C.B. 1995. *Introduction to GIS for public agencies*. State University of New York, College of Environmental Science and Forestry, Syracuse, New York. June.

Gulland, J.A. 1983. *Fish stock assessment: A manual of basic methods.* FAO/Wiley Series, Vol. 1. Wiley & Sons, New York.

Gustin, M.S., G.E. Taylor, Jr., and T.L. Leonard. 1995. Atmospheric mercury concentrations above mercury contaminated mill tailings in the Carson River Drainage Basin, NV. In D. Porcella, J. Huckabee, B. Wheatley (eds.), *Third international conference on mercury as a global pollutant*, Whistler, British Columbia, Canada, July 10-14. Pp. 217-220.

Hallberg, G.R., B.E. Hoyer, E.A. Bettis III, and R.D. Libra. 1983. *Hydrogeology, water quality, and land management in the Big Spring basin, Clayton County, Iowa.* Iowa Geologic Survey Open-File Report 83-3.

Hallberg, G.R., R.D. Libra, E. A. Bettis, and B.E. Hoyer. 1984. *Hydrogeologic and water quality investigations in the Big Spring basin, Clayton County, Iowa.* Iowa Geological Survey, Open File Report 84-4.

Harcum, J.B. 1990. Water-quality data analysis protocol development. Ph.D. diss., Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, Colorado.

Harcum, J.B., J.C. Loftis, and R.C. Ward. 1992. Selecting trend tests for water quality series with serial correlation and missing values. *Water Resour. Bull.* 28(3):469-478.

Harris, J., J.C. Loftis, and R.H. Montgomery. 1987. Statistical methods for characterizing groundwater quality. *Ground Water* 25(2):185:193.

Hawkins, C.P., J.L. Kershner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman and M.K. Young. 1993. A hierarchical approach to classifying stream habitat features. *Fisheries* 18(6):3-12.

Hayes, M.L. 1983. Active fish capture methods. In *Fisheries techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 123-145. American Fisheries Society, Bethesda, Maryland.

Hayslip, G.A. 1993. *EPA Region 10 in-stream biological monitoring handbook (for wadable streams in the Pacific Northwest).* EPA-910-9-92-013. U.S. Environmental Protection Agency, Region 10, Environmental Services Division, Seattle, Washington.

Heath, R.C. 1989. *Basic ground-water hydrology.* Water-supply Paper 2220. U.S. Geological Survey, Washington, D.C.

Helsel, D.R., and T.A. Cohn. 1988. Estimation of descriptive statistics for multiply-censored water-quality data. *Water Resour. Res.* 24(12):1997-2004.

Helsel, D.R., and R.M. Hirsch. 1995. *Statistical methods in water resources*. Elsevier. Amsterdam.

Hershfield, D.N. 1961. *Rainfall frequency atlas of the United States*. U.S. Weather Bureau Technical Paper 40. May.

Hipel, K.W. 1988. Nonparametric approaches to environmental impact assessment. *Water Resour. Bull.* 24(3).

Hirsch, R.M., and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resour. Res.* 20:727-732.

Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* 18(1):107-121.

Hoaglin, D.C., F. Mosteller, and J.W. Tukey. 1983. *Understanding robust and exploratory data analysis.* Wiley, New York.

Hodges, J.L., and E.L. Lehmann. 1963. Estimates of location based on rank tests. *Ann. Math. Stat.* 34:598-611.

Horowitz, A.J., K.A. Elrick, J.A. Robbins, and R.B. Cook. 1993. *The effect of mining and related activities on the sediment-trace element geochemistry of Lake Coeur D'Alene, Idaho, USA - Part II: Subsurface sediments.* U.S. Geological Survey, Denver, Colorado. Open-File Report 93-656.

Hubert, W.A. 1983. Passive capture techniques. In *Fisheries techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 95-122. American Fisheries Society, Bethesda, Maryland.

Hughes, R.M., and J.R. Gammon. 1987. Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. *Trans. Am. Fish. Soc.* 116(2):196-209.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54(2):187-211.

IEPA. 1979. *Water quality management plan, Volume III.* Illinois Environmental Protection Agency, Springfield, Illinois.

Ingwersen, J.B. 1980. *Statistical analysis using SPSS at the USDA-Fort Collins computer center.* WSDG-AD-00002. U.S. Department of Agriculture, Forest Service, Washington, DC.

ITFM. 1995a. *The strategy for improving water-quality monitoring in the United States*. Final report of the Intergovernmental Task Force on Monitoring Water Quality. Intergovernmental Task Force on Monitoring Water Quality. February.

ITFM. 1995b. *The strategy for improving water-quality monitoring in the United States. Technical appendixes.* Final report of the Intergovernmental Task Force on Monitoring Water Quality. Intergovernmental Task Force on Monitoring Water Quality, Washington, DC. February.

Jamieson, C.A., ed. 1986. *Summary of the 1986 RCWP data analysis workshop.* U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

Jerald, A., Jr. 1983. Age determination. In *Fisheries techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 301-324. American Fisheries Society, Bethesda, Maryland.

Johengen, T.H., and A.M. Beeton. 1992. The effects of temporal and spatial variability on monitoring agricultural nonpoint source pollution. In *The national rural clean water program symposium: 10 years of controlling agricultural nonpoint source pollution: The RCWP experience. Seminar publication.* EPA/625/R-92/006. U.S. Environmental Protection Agency, Office of Research and Development, Office of Water, Washington, DC., pp. 89-95. August.

Joiner, B.L., and J.R. Rosenblatt. 1971. Some properties of the range in samples from Tukey's symmetric lambda distributions. *J. Amer. Stat. Assoc.* 66:394-399.

Karr, J.R., and D.R. Dudley. 1981. Ecological perspective on water quality goals. *Environ. Manage.* 5:55-68.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant and I.J. Schlosser. 1986. *Assessing biological integrity in running waters: A method and its rationale.* Special Publication 5. Illinois Natural History Survey, Urbana, Illinois.

Kaufmann, P.R., A.T. Herlihy, and L.A. Baker. 1992. Sources of acidity in lakes and streams of the United States. *Environ. Pollut.* 77(2/3):115-122.

Keith, L.H., ed. 1988. Principles of environmental sampling. American Chemical Society, Washington, DC.

Kendall, M.G. 1975. *Rank correlation methods, 4th edition.* Charles Griffin, London.

Kentucky Department of Environmental Protection. 1993. *Methods for assessing biological integrity of surface waters.* Division of Water, Kentucky Department of Environmental Protection, Frankfort, Kentucky.

Kerans, B.L., J.R. Karr, and S.A. Ahlstedt. 1992. Aquatic invertebrate assemblages: Spatial and temporal differences among sampling protocols. *J. North Am. Benthol. Soc.* 11:377-390.

Khaleel, R., K.R. Reddy, and M.R. Overcash. 1980. Transport of potential pollutants in runoff water from land areas receiving animal waste: a review. *Water Res.* 14:421-436.

Kimball, C.G. 1986. *1986 annual RCWP progress report - Project 20.* South Dakota Dept. of Water and Natural Resources, Pierre, South Dakota.

Klemm, D.J., P.A. Lewis, F. Fulk, and J.M. Lazorchak. 1990. *Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters.* EPA-600-4-90-030. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, Ohio.

Klemm, D.J., Q.J. Stober and J.M. Lazorschak. 1992. *Fish field and laboratory methods for evaluating the biological integrity of surface waters.* EPA/600/R-92/111. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, Ohio. March.

Knapton, J.R., and D.A. Nimick. 1991. *Quality assurance for water-quality activities of the U.S. Geological Survey in Montana.* U.S. Geological Survey Open File Report 91-216. U.S. Geological Survey, Helena, Montana. August.

Koterba, M.T., F.D. Wilde, and W.W. Lapham. 1995. *Ground water data collection protocols and procedures for the National Water Quality Assessment Program: Collection and documentation of water quality samples and related data.* Open File Report 95-399. U.S. Geological Survey, Reston, Virginia.

Kreglow, J.M., S.A. Dressing, R.P. Maas, F.A. Koehler, L. Christensen, W.K. Snyder, and F.J. Humenik. 1982. *Best management practices for agricultural nonpoint source control: I. Animal waste.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Lapham, W.W., F.D. Wilde, and M.T. Korterba. 1995. *Ground water data collection protocols and procedures for the National Water Quality Assessment Program: Selection, installation, and documentation of wells, and collection of related data.* Open File Report 95-398. U.S. Geological Survey, Reston, Virginia.

Larsen, R.J., and M.L. Marx. 1981. *An introduction to mathematical statistics and its applications.* Prentice Hall, Englewood Cliffs, New Jersey.

Leigh, D.S. 1994. Mercury contamination and floodplain sedimentation from former gold mines in north Georgia. *Wat. Resour. Bull.* 30(4):739-748.

Leopold, L.B. 1968. *Hydrology for urban land planning.* Circular 559. U.S. Geological Survey, Reston, Virginia.

Lettenmaier, D.P. 1988. Multivariate nonparametric tests for trend in water quality. *Water Resour. Res.* 24(3):505-512.

Loftis, J.C., P.S. Porter, and G. Settembre. 1987. Statistical analysis of industrial wastewater monitoring data. *J. Water Pollut. Control* 59(3):145-151.

Loucks, D.P., J.R. Stedinger, and D.A. Haith. 1981. *Water resource systems planning and analysis.* Prentice-Hall, Englewood Cliffs, New Jersey.

Luchetti, G., K. Thornburgh, and C. Rawson. 1987. *A water quality monitoring program for major agricultural areas of the Snohomish River basin.* Tulalip Technical Report, Shohomish Co., Washington.

Lyons, J. 1992. *Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin.* General Technical Report NC-149. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota.

Maas, R.P., J.M. Kreglow, S.A. Dressing, F.A. Koehler, L. Christensen, W.K. Snyder, and F.J. Humenik. 1982. *Best management practices for agricultural nonpoint source control: III. Sediment.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Maas, R.P., S.A. Dressing, J. Spooner, M.D. Smolen, and F.J. Humenik. 1984. *Best management practices for agricultural nonpoint source control: IV. Pesticides.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska.* EPA/910/9-91-001. U.S. Environmental Protection Agency, Region 10, Seattle, Washington.

Mann, H.B. 1945. Non-parametric tests against trend. *Econometrica* 13:245-259.

Mann, H.B., and D.R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* 18:50-60.

Marcus, J.M., and T.P. Stokes. 1985. Polynuclear aromatic hydrocarbons in oyster tissue and around three coastal marinas. *Bull. Environ. Contam. Toxicol.* 35:833-844.

Maxted, J. R. and E. L. Dickey. 1993. *Statistical analysis of replicate data - Fall 1991 biological survey.* Preliminary technical report. Delaware Department of Natural Resources and Environmental Control, Division of Water Resources. Dover, Delaware. August.

McMahon, P.J.T. 1989. The impact of marinas on water quality. *Water Sci. Tech.* 21(2):39-43.

Meador, M.R., T.F. Cuffney, and M.E. Gurtz. 1993. *Methods for sampling fish communities as part of the National Water Assessment Program.* Open File Report 93-104. U.S. Geological Survey, Raleigh, North Carolina.

Meals, DW. 1991. Surface water trends and land-use treatment. In *Nonpoint source watershed workshop: Seminar publication*, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Pp. 136-142.

Meals, DW. 1992. Relating land use and water quality in the St. Albans Bay watershed, Vermont. In *Seminar publication: The National Rural Clean Water Program symposium.* 10 years of controlling agricultural nonpoint source pollution: The RCWP experience. EPA/625/R-92/006. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Pp. 131-143.

Mendenhall, W. 1971. *Introduction to probability and statistics.* Duxbury Press, Belmont, California. Cited in MacDonald et al., 1991.

Miller, D.L., P.M. Leonard, R.M. Hughes, J.R. Karr, P.B. Moyle, L.H. Schrader, B.A. Thompson, R.A. Daniels, K.D. Fausch, G.A. Fitzhugh, J.R. Gammon, D.B. Halliwell, P.L. Angermeier, and D.J. Orth. 1988. Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 5:12-20.

Milliken, A., and V. Lee. 1990. *Pollution impacts from recreational boating: A bibliography and summary review.* Rhode Island Sea Grant, Narragansett, Rhode Island.

Mitsch, W.J., R.J. Stevenson, J.R. Taylor, and M.A. Cardamone. 1983. Water contamination in the Ohio River basin in Kentucky - a summary report. Systems Science Institute, University of Louisville, Kentucky.

Mockus, V. 1960. Selecting a flood-frequency method. *Am. Soc. Eng. Trans.* 3:48-51,54.

Montgomery, R.H., and J.C. Loftis. 1987. Applicability of the *t* test for detecting trends in water quality variables. *Water Resour. Bull.* 23(4):653-662.

Moore, D.S., and G.P. McCabe. 1989. *Introduction to the practice of statistics.* W.H. Freeman and Company, New York.

Mosteller, F., and R.E.K. Rourke. 1973. *Sturdy statistics.* Addison-Wesley, Reading, Massachusetts.

Mosteller, F., and J.W. Tukey. 1977. *Data analysis and regression.* Addison-Wesley Publishers, Menlo Park, California.

National Research Council. 1986. *Ecological knowledge and environmental problem-solving. Concepts and case studies.* National Academy Press, Washington, DC.

NCDEM. 1990. *North Carolina coastal marinas: Water quality assessment.* North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, North Carolina.

NCDEM. 1991. *Coastal marinas: Field survey of contaminants and literature review.* Report No. 91-03. North Carolina Division of Environmental Management, Raleigh, North Carolina.

Netter, J., W. Wasserman, and M.H. Kutner. 1985. *Applied linear statistical models.* 2nd ed. Irwin Publishers, Homewood, Illinois.

Nixon, S.W., C.A. Oviatt, and S.L. Northby. 1973. *Ecology of small boat marinas.* Rhode Island Sea Grant, Narragansett, Rhode Island.

NOAA. 1990. *Estuaries of the United States. Vital statistics of a national resource base.* National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

Novotny, V., and G. Chesters. 1989. Delivery of sediment and pollutants from nonpoint sources: A water quality perspective. *J. Soil Water Cons.* 44:568-576.

Oak Ridge National Laboratory. 1993. *Public health assessment for Carson River mercury site, Moundhouse, Lyon County, Nevada. Preliminary report.* Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Ohio EPA. 1987. *Biological criteria for the protection of aquatic life: Volumes I-III.* Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio.

Ohio EPA. 1992. Executive Summary and Vol I: Summary, Status, and Trends. In *Ohio water resource inventory.* Ohio Environmental Protection Agency, Division of Surface Water, Ecological Assessment Section, Columbus, Ohio.

Omernik, J.M. 1987. Ecoregions of the coterminous United States. *Ann. Assoc. Am. Geogr.* 77(1):118-125.

Omernik, J.M. 1995. Ecoregions: A spatial framework for environmental management. In *Biological assessment and criteria: Tools for water resource planning and decision making*, ed. W.S. Davis and T.P. Simon, pp. 49-62. Lewis Publishers, Boca Raton, Florida.

Pankratz, A. 1983. *Forecasting with univariate box-Jenkins models.* Wiley and Sons, Inc., New York.

Pennsylvania RCWP Coordinating Committee. 1984. *The Conestoga headwaters rural clean water program, pn19, 1984 annual progress report.* U.S. Department of Agriculture, Soil Conservation Service, Harrisburg, Pennsylvania.

Pitt, R., and J. McLean. 1992. Stormwater, baseflow, and snowmelt pollutant contributions from an industrial area. *Water Environment Federation 65th annual conference & exposition, surface water quality & ecology symposia, volume VII.* September 20-24, New Orleans, Louisiana.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross and R.M. Hughes. 1989. *Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish.* EPA/440/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Platts, W.S., W.F. Megahan, and G.W. Minshall. 1983. *Methods for evaluating stream, riparian, and biotic conditions*. Gen. Tech. Rept. INT-138. U.S. Department of Agriculture, Forest Service, Ogden, Utah.

Ponce, S.L. 1980a. *Statistical methods commonly used in water quality data analysis.* WSDG Technical Paper WSDG-TP-00001. US Department of Agriculture, Forest Service, Washington, DC.

Ponce, S.L. 1980b. *Water quality monitoring programs*. WSDG Technical Paper WSDG-TP-00002. U.S. Department of Agriculture, U.S. Forest Service, Washington, DC.

Porter, P.S., R.C. Ward, and H.F. Bell. 1988. The detection limit, water quality monitoring data are plagued with levels of chemicals that are too low to be measure precisely. *Environ. Sci. Tech.* 22:856-861.

Pritt, J.W., and J.W. Raese, ed. 1992. *Quality assurance/quality control manual.* Open File Report 92-495. U.S. Geological Survey, National Water Quality Laboratory, Reston, Virginia.

Reckhow, K.H. 1979. Sampling designs for lake phosphorus budgets. In eds. Everett, L.G. and K.D. Schmidt, *Establishment of water quality monitoring programs*. Proceedings of a symposium held in San Francisco, California, June 12-14, 1978, AWRA, Minneapolis, Minnesota.

Reckhow, K.H., K. Kepford and W.W. Hicks. 1993. *Methods for the analysis of lake water quality trends.* EPA 841-R-93-0003. U.S. Environmental Protection Agency, Washington, DC.

Remington, R.D., and M.A. Schork. 1970. *Statistics with applications to the biological and health sciences.* Prentice-Hall, Englewood Cliffs, New Jersey.

Renard, K. G.; Foster, G. R.; Weesies, G. A.; Porter, J. P. 1991. RUSLE: Revised universal soil loss equation. *J. Soil Water Cons.* 46(1):30-33.

Richards, R.P. 1986. *Monte carlo studies of sampling strategies for estimating tributary loads, II: Effects on bias and precision due to differences among watershed sizes and the transported materials monitored.* Water Quality Laboratory, Heidelberg College, Tiffin, Ohio.

Richards, C., and G.W. Minshall. 1992. Spatial and temporal trends in stream macroinvertebrate communities: the influence of catchment disturbance. *Hydrobiologia* 241(3):173-184.

Rosen, B.H. 1995. Use of periphyton in the development of biocriteria. In W.S. Davis and T.P. Simon (ed.), *Biological assessment and criteria. Tools for water resource planning and decision making.* Lewis Publishers, Boca Raton, Florida. Pp. 209-215.

Rossman, L.A. 1990. *DFLOW user's manual.* EPA 600/8-90/051. U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, Ohio.

Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane. 1980. *Applied modeling of hydrologic time series.* Water Resources Publications, Ft. Collins, Colorado.

SAS Institute, Inc. 1985a. *SAS user's guide: Basics.* Version 5 edition. SAS Institute, Inc., Cary, North Carolina.

SAS Institute, Inc. 1985b. *SAS user's guide: Statistics.* Version 5 edition.SAS Institute, Inc., Cary, North Carolina.

SAS Institute, Inc. 1990. *SAS/STAT user's guide. Volume 1, CECLUS-FREQ.* Version 6, 4th edition. SAS Institute, Inc., Cary, North Carolina.

Sawyer, C.M., and A.F. Golding. 1990. Marina pollution abatement. In *1990 Environmental management for marinas conference*. International Marina Institute, Wickford, Rhode Island.

Scalf, M.R., J.F. McNabb, W.J. Dunlap, R.L. Cosby, and J. Fryberger. 1981. *Manual of ground-water sampling procedures.* NWWA/EPA Series.

SCDHEC. 1987. *Heavy metals and extractable organic chemicals from the coastal toxics monitoring network 1984-1986.* South Carolina Department of Health and Environmental Control. Technical Report No. 007-87.

Schlosser, I.J., and J.R. Karr. 1980. *Determinants of water quality in agricultural watersheds.* UILU-WRC-80-0147, University of Illinois Water Resources Center, Urbana, Illinois.

SCRIP. 1996. *How coal mining produces pollution in the upper Conemaugh River basin.* Internet Home Page.

Seabloom, R.A., G. Plews, F. Cox, and F. Kramer. 1989. *The effect of sewage discharges from pleasure craft on Puget Sound waters and shellfish quality.* Washington State Department of Health, Shellfish Section. Olympia, Washington.

Sen, P.K. 1968. Estimates of the regression coefficient based on Kendalls tau. *J. Amer. Stat. Assoc.* 63:1379-1389.

Shackleford, B. 1988. *Rapid bioassessments of lotic macroinvertebrate communities: Biocriteria development.* Arkansas Department of Pollution Control and Ecology, Little Rock, Arkansas.

Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611.

Shelley, P.E. 1979. *Monitoring requirments, methods, and costs for the nationwide urban runoff program.* U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

Shelton, L.R. 1994. *Field guide for collecting and processing stream-water samples for the National Water Quality Assessment Program.* Open File Report 94-455. U.S. Geological Survey, Reston, Virginia.

Shelton, L.R., and P.D. Capel. 1994. *Guidelines for collecting and processing samples of stream bed sediment for analysis of trace elements and organic contaminants for the National Water Quality Assessment Program.* Open File Report 94-458. U.S. Geological Survey, Reston, Virginia.

Sherwani, J.K. and D.H. Moreau. 1975. *Strategies for water quality monitoring.* Report No. 107, Water Resources Research Institute of the University Raleigh, North Carolina.

Sherwood, C.R., D.A. Jay, R.B. Harvey, P. Hamilton, and C.A. Simenstad. 1990. Historical changes in the Columbia River Estuary. *Prog. Oceanog.* 25:299-352.

Simons, D.B., R. Li, T.P. Ballestero, and K.G. Eggert. 1981. *Workshop in sampling requirements for monitoring and evaluaiton of watershed management practices, May 20-21, 1981.* U.S.EPA Environmental Research Laboratory, Athens, Georgia.

Singh, V.P., and M. Fiorentino, eds. 1996. *Geographical information systems in hydrology.* Kluwer Academic Publishers, Dordrecht, Netherlands.

Skalski, J.R., and D.H. McKenzie. 1982. A design for aquatic monitoring programs. *J. Environ. Manage.* 14:237-251.

Smith, R.A., R.B. Alexander, M.G. Wolman. 1987. Water-quality trends in the Nation's rivers. *Science* (Mar 27):1607-1615.

Smith, F., S. Kulkami, L.E. Myers, and M.J. Messner. 1988. Evaluating and presenting quality assurance data. In *Principles of environmental sampling*, ed. L.H. Keith, pp. 157-168. American Chemical Society, Washington, DC.

Smolen, M. 1984. *Water quality models.* National Water Quality Evaluation Project Notes, No. 9, Biological and Agricultural Engineering Deptartment, North Carolina State University, Raleigh, North Carolina.

Smolen, M.D., R.P. Maas, J. Spooner, C.A. Jamieson, S.A. Dressing, L.C. Stanley, and F.J. Humenik. 1986. *NWQEP 1986 annual report, status of agricultural NPS projects - Draft.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical methods.* 7th ed. The Iowa State University Press, Ames, Iowa.

Sokal, R.R., and F.J. Rohlf. 1981. *Biometry.* W.H. Fremman and Co. San Francisco, California. Cited in MacDonald et al., 1991.

Southerland, M., and J.B. Stribling. 1995. Status of biological criteria development and implementation. In *Biological criteria tools for water resources planning and decision making*, ed. W.D. Davis and T. Simon, pp. 79-94. Lewis Publishers, Boca Raton, Florida.

Spooner, J. 1984. *Use and misuse of r-square in regression analysis.* National Water Quality Evaluation Project Notes, Technical Supplement to No. 12, Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Spooner, J. 1986a. *Analysis of covariance, Part I.* National Water Quality Evaluation Project Notes, Technical Supplement to No. 18, Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Spooner, J. 1986b. *Analysis of covariance, Part II.* National Water Quality Evaluation Project Notes, Technical Supplement to No. 19, Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Spooner, J. 1994. *Nonpoint source water quality monitoring designs and data analysis techniques: A review.* Review draft. North Carolina Cooperative Extension Service, North Carolina State University.

Spooner, J. 1994a. Comparisons between parametric and nonparametric statistical trend tests. Second National Nonpoint Source Watershed Monitoring Conference.

Spooner, J. and R.P. Maas. 1984. *Regression-basic concepts and definitions.* National Water Quality Evaluation Project Notes, Technical Supplement to No. 11, Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Spooner, J., R.P. Maas, S.A. Dressing, M.D. Smolen, and F.J. Humenik. 1985. Appropriate designs for documenting water quality improvements from agricultural nps control programs. In *Perspectives on nonpoint source pollution.* Proceedings of a national conference, May 19-22, Kansas City, Missouri. EPA 440/5-85-001. Washington, DC.

Spooner, J., C.A. Jamieson, R.P. Maas, S.A. Dressing, M.D. Smolen, and F.J. Humenik. 1986. *Rural clean water program status report on the CM&E projects 1985 - supplemental report: Analysis methods.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

Spreizer, G.M., T.J. Calabrese and R.S. Weidner. 1992. Assessing the usability of historical water-quality data for current and future applications. In *Current practices in ground water and vadose zone investigations*, ASTM STP 1118. D.M. Nielsen and M.N. Sara, ed. Pp. 377-390. American Society for Testing and Materials, Philadelphia, Pennsylvania.

Srivastava, M.S., and C.G. Khatri. 1979. *An introduction to multivariate statistics.* North Holland, New York.

Steedman, R.J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. *Can. J. Fish. Aquat. Sci.* 45:492-501.

Storch, T.A. 1986. A comparison of water quality in Chautauqua Lake, New York: 1937-1983. In *Lake and reservoir management, volume II.* Proceedings of the Fifth Annual Conference and International Symposium on Applied Lake & Watershed Management, November 13-16, 1985, Lake Geneva, Wisconsin. North American Lake Management Society, Merrifield, Virginia.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Amer. Geophys. Union Trans.* 38:913-920.

Taylor, J.K. 1993. *Standard reference materials handbook for SRM users.* NIST Special Publication 260-100. National Institute of Standards and Technology, Standards and Reference Materials Program, U.S. Department of Commerce, Technology Administration, Washington, DC. February.

Tchernia, P. 1980. *Descriptive regional oceanography.* Pergamon Marine Series, No. 3. Pergamon Press, Great Britain.

Thomas, B.J. 1970. Sedimentation activities in the north Pacific division, In *Proceedings of a seminar on sediment transport in rivers and reservoirs.* April 7-9. U.S. ACE, Hydrologic Engineering Center.

Thornton, K.W., B.L. Kimmer, and F.E. Payne (editors). 1990. *Reservoir limnology: Ecological perspectives*. John Wiley and Sons, Inc., New York, New York.

Travis, C.C., and M.L. Land. 1990. Estimating the mean of data sets with nondetectable values. *Environ. Sci. Tech.* 24:961-962.

USDA. 1981. *Forest management for water quality.* Forest Service, Area Planning and Development.

USDA-NRCS. 1996. *National handbook on water quality monitoring.* National Water Quality Handbook, Part 600. U.S. Department of Agriculture, Natural Resources Conservation Service. December.

USDOI. (n.d). *Appalachian clean stream initiative.* U.S. Department of the Interior, Office of Surface Mining Reclamation and Enforcement, Washington, DC. Leaflet.

USDOI. 1991. *Riparian-wetland initiative for the 1990's.* BLM/WO/GI-91/001+4340. U.S. Department of the Interior, Bureau of Land Management, Washington, DC. September.

USEPA. 1975. *Model state water monitoring program.* National Water Monitoring Panel. EPA-440/9-74-002. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1977. *Nonpoint source—stream nutrient level relationships: A nationwide study.* U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1978a. *An assessment of major nonpoint pollution sources: relative magnitudes, cost-effective controls, institutional constraints.* Booz, Allen, and Hamilton, Inc., Bethesda, Maryland.

USEPA. 1978b. *Microbiological methods for monitoring the environment, water, and wastes.* EPA-600/8-78-017. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1980a. *An approach to water resources evaluation of nonpoint silvicultural sources (a procedural handbook).* EPA-600/8-80-012. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1980b. *Water quality management needs assessment FY '80-'84.* Draft. U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1981. *Guidelines for evaluation of agricultural nonpoint source water quality projects.* U.S. Environmental Protection Agency, Interagency Taskforce, Washington, DC. April.

USEPA. 1982b. *Data collection and quality assurance for the nationwide urban runoff program.* U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1982c. *Results of the Nationwide Urban Runoff Program.* Volume II—Appendices. U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1983a. *Results of the Nationwide Urban Runoff Program, volume I - Final report.* U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1983b. *Chesapeake Bay Program: Findings and recommendations.* U.S.Environmental Protection Agency, Region 3, Philadelphia, Pennsylvania.

USEPA. 1983c. *Interim guidelines and specifications for preparing quality assurance project plans.* EPA-600/4-83-004. QAMS-005/80. U.S. Environmental Protection Agency, Office of Monitoring Systems and Quality Assurance, Office of Research and Development, Washington, DC. February.

USEPA. 1984a. *Report to Congress: Nonpoint source pollution in the United States.* U.S. Environmental Protection Agency, Office of Water Program Operations, Washington, DC.

USEPA. 1984b. *A probabilistic methodology for analyzing water quality effects of urban runoff on rivers and streams, first draft.* U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1984c. *Policy and program requirements to implement the mandatory quality assurance program.* EPA Order 5360.1. U.S. Environmental Protection Agency, Washington, D.C. April.

USEPA. 1987a. *Handbook, ground water.* EPA/625/6-87/016. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1987. *An overview of sediment quality in the United States.* EPA-905/9-88-002. U.S. Environmental Protection Agency, Region 5, Chicago, Illinois.

USEPA. 1990. *Surface water monitoring program guidance.* Draft. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Assessment and Watershed Protection Division. Washington, D.C. September.

USEPA. 1990a. *The lake and reservoir restoration guidance manual.* EPA-440/4-90-006. U.S. Environmental Protection Agency, Office of Water, Washington, DC. August.

USEPA. 1991. *Characteristics of successful riparian restoration projects in the Pacific Northwest.* EPA 910/9-91-033. U.S. Environmental Protection Agency, Region 10, Water Division, Nonpoint Sources Section, Seattle, Washington. August.

USEPA. 1993a. *Guidance specifying management measures for sources of nonpoint pollution in coastal waters.* EPA 840-B-92-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1993b. *Guide to federal water quality programs and information.* EPA-230-B-93-001. U.S. Environmental Protection Agency, Washington, DC. February.

USEPA. 1993c. *Paired watershed study design..* EPA 841-F-93-009. U.S. Environmental Protection Agency, Washington, DC. September.

References

USEPA. 1994a. *Biological criteria: Technical guidance for streams and small rivers.* EPA 822B94001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1994b. *Volunteer stream monitoring: A methods manual.* EPA-841-D-95-001 (Draft). U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Assessment and Watershed Protection Division, Washington, DC. December.

USEPA. 1994c. *EPA requirements for quality management plans.* Interim final. EPA QA/R-2. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC.

USEPA. 1994d. *EPA requirements for quality assurance project plans for environmental data operations.* Interim Final. EPA QA/R-5. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC. August.

USEPA. 1994e. *Guidance for the data quality objectives process.* EPA QA/G-4. Final. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC. September.

USEPA. 1995. *National water quality inventory 1994 Report to Congress.* EPA 841-R-95-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1997. *Field and laboratory methods for macroinvertebrate and habitat assessment for low gradient non-tidal streams.* Mid-Atlantic Coastal Streams Workgroup, U.S. Environmental Protection Agency, Environmental Services Division, Region 3, Wheeling, West Virginia.

USEPA/USDOI. 1995. *1995 progress report. Statement of mutual intent strategic plan for the restoration and protection of streams and watersheds polluted by acid mine drainage from abandoned coal mines.* U.S. Environmental Protection Agency, Region 3 and U.S. Department of the Interior, Office of Surface Mining, Appalachian Regional Coordinating Center, Pittsburgh, Pennsylvania. November

USFWS. 1982. *Mitigation and enhancement techniques for the upper Mississippi River system and other large river systems.* Resource Publication 149. U.S. Department of the Interior, U.S. Fish and Wildlife Service, Washington, DC.

USFWS. 1991. *Principles and techniques of electrofishing.* U.S. Fish and Wildlife Service, Fisheries Academy, Office of Technical Fisheries Training, Kearneysville, West Virginia.

USGS. 1977. *National handbook of recommended methods for water-data acquisition.* Chapter 2: Groundwater. U.S. Department of the Interior, Reston, Virginia.

USGS. 1984. *National water summary 1983 - hydrologic events and issues.* Water Supply Paper 2250. U.S. Government Printing Office, Washington, DC.

USGS. 1985. *National water summary 1984, hydrologic events, selected water-quality trends, and ground-water resources.* Water-Supply Paper 2275, U.S. Government Printing Office, Washington, D.C..

U.S. House Committee on Public Works and Transportation. 1980. *Report on implementation of the FWPCA.* House Report No. 97-71. Washington, DC.

van Belle, G., and J.P. Hughes. 1984. Nonparametric tests for trend in water quality. *Water Resour. Res.* 20(1):127-136.

Vermont RCWP Coordinating Committee. 1984. *St. Albans Bay rural clean water program summary report 1984.* U.S. Department of Agriculture and Vermont Water Resources Research Center, Burlington, Vermont.

Vermont RCWP Coordinating Committee. 1986. *St. Albans Bay rural clean water program annual progress report 1986.* U.S. Department of Agriculture and Vermont Water Resources Research Center, Burlington, Vermont.

Viessman, W., J.W. Knapp, G.L. Lewis, and T.E. Harbough, 1977. *Introduction to hydrology.* Hurpert Row, New York, New York.

Walburg, C.H., J.F. Novotny, K.E. Jacobs, and W.D. Swink. 1983. *Effects of reservoir releases on water quality, macro-invertebrates, and fish in tailwaters: Field study results.* U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi.

Ward, R.C., and J.C. Loftis. 1986. Establishing statistical design criteria for water quality monitoring systems, review and synthesis. *Water Resour. Bull.* 22(5):759-767.

Ward, R.C., J.C. Loftis, and G.B. McBride. 1986. The "data-rich but information-poor syndrome" in water quality monitoring. *Environ. Manag.* 10(3):291-297.

Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. *Design of water quality monitoring systems.* Van Nostrand Reinhold Company, New York.

Washington State Department of Ecology. 1991. *Stormwater management manual for the Puget Sound basin - public review draft.* Publication No. 90-73.

Weatherley, A.H. 1972. *Growth and ecology of fish populations.* Academic Press, New York, New York.

Weisberg, S. B. and 11 co-authors. 1993. *EMAP-estuaries Virginian Province 1990 demonstration project report.* EPA 600/R-92/100. United States Environmental Protection Agency, Environmental Research Laboratory, Narragansett, Rhode Island.

Weiss, L.L. 1962. A general relation between frequency and duration of precipitation. *Monthly Weather Rev.* 87-88.

Wendt, P.H., R.F. van Dolah, M.Y. Bobo, and J.J. Manzi. 1990. *The effects of a marina on certain aspects of the biology of oysters and other benthic macrofauna in a South Carolina estuary.* Unpublished draft manuscript. South Carolina Department of Health and Environmental Control, Columbia, South Carolina.

usability 5-27
validity 5-27
water quality 2-10
data management 5-17
data quality objectives 5-3
development steps 5-4
data set, balanced 4-61
data stratification 4-8
decision rule 5-6
decision statement 5-5
design flow
biologically based 4-102
extreme event 4-101
detection limit 4-15
Diatom Bioassessment Index 3-26
discriminant analysis 4-94
DQO see *data quality objectives*
ecoregions 3-16
EMAP see *Environmental Monitoring and Assessment Program*
EMC see *event mean concentration*
Environmental Monitoring and Assessment Program 3-29
error
decision 5-6
gross 2-13
sampling, random 2-13
Type I 4-2, 4-43
Type II 4-2, 4-43
estuaries 1-3, 1-16
characteristics 1-17
estuarine drainage area 1-17
event mean concentration 4-11
extreme events 4-94
design flow 4-101
frequency 4-105
extreme values 4-8
factor analysis 4-93
fertilizers 1-4, 1-10, 1-21
field operations 5-19
Fisher Z transformation 4-72
flow alteration 1-8
flow-weighted mean concentration 4-11
fluvial drainage area 1-17
forest harvesting 1-10
Friedman test 4-52, 4-63
FWMC see *flow-weighted mean concentration*
geographic information systems 2-11
geology 1-21
geometric mean 4-11
see also *statistic*
GLM 4-77
Great Lakes 1-1, 1-3, 1-6, 1-10
ground water 1-19
characteristics 1-20
habitat
and biological condition 3-6
assessment 3-4, 3-6
quality 3-2, 3-4
riparian 1-8
structure 3-4
histogram 4-21
Hodges-Lehman estimator 4-8, 4-52
hydraulic residence time 1-13
hydromodification 1-8
hypothesis
alternative 4-2
null 4-2
testing 4-2
hypothesis testing 2-12, 4-2
multiple comparisons 4-63
see also *test*
imperviousness 1-7
Intergovernmental Task Force on Monitoring 2-1
interquartile range 4-1, 4-16
see also *statistic*
interval estimate 4-1
Invertebrate Community Index 3-38
ITFM see *Intergovernmental Task Force on Monitoring*
Kendall 4-8
seasonal 4-86, 4-88
slope estimator 4-8
Kolmogorov-Smirnov 4-31
Kruskal-Wallis 4-7, 4-52, 4-56
kurtosis 4-18, 4-30
laboratory operations 5-24
lakes 1-13
characteristics 1-14
stratification 1-14
least significant difference 4-55
livestock
grazing 1-7
wastes 1-4
loading rate 4-14
log transformations 4-26, 4-75
log-log plot 4-23
Mann-Kendall τ 4-86
one-sided 4-87
two-sided 4-87
Mann-Whitney 4-5, 4-7, 4-47, 4-50, 4-56
marinas 1-11
mean, Winsorized 4-15
mercury 1-9
metrics, biological 3-11, 3-20
mines, abandoned 1-10
mining 1-9, 2-25
coal 1-9
ore 1-9
underwater 1-8
monitoring
biological 1-16, 3-1
chemical 2-31
evaluation 2-6
ground water 1-20
objectives 2-4, 2-5
physical 2-31
trend 2-24
monitoring plan
development 2-1
monitoring programs
design 2-12
design, probabilistic 2-12, 2-14
design, targeted 2-12
geographic scale 2-3
purposes 2-1
responsibilities 2-28
temporal scale 2-3

NASQAN see *National Stream Quality Accounting Network*
National Stream Quality Accounting Network 2-25
nonpoint source pollution *1-1*
  categories 1-1, 1-6
  effects of 1-4
  extent of problem 1-1
ocean shoreline waters 1-3
open coastal waters 1-18
  characteristics 1-19
PCA see *principal component analysis*
periphyton *3-25*
pesticides 1-6, 1-7, 1-10, 1-21
plot of residuals 4-66
point estimate *4-1*, 4-10
point source pollution *1-1*
pollution, natural sources 1-2
ponds 1-13
  characteristics 1-14
population *5-5*
port construction 1-11
power, analytical 4-5
precision *5-11*
Prince George's County, MD 3-32, 3-38
principal component analysis 4-93
QA/QC
  EPA policy 5-2
  importance 5-1
  see also *quality assurance and quality control*
QAPP see *quality assurance project plan*
QMP see *quality management plan*
quality assurance 2-31, *5-1*
quality assurance project plan 2-31, 5-3
  elements 5-8
  optimizing 5-8
quality control 2-31, *5-1*
quality management plan 5-3
quartile 4-19
range 4-16
rapid bioassessment protocols 3-1, 3-16
RBP see *rapid bioassessment protocol*
RCWP see *Rural Clean Water Program*
record keeping 5-28
reference conditions, establishment 3-16
regression
  evaluation 4-69
  independence of residuals 4-66
  linear 4-64
  multiple 4-75
  multivariate 4-78
  nonlinear 4-74
  normality of residuals 4-68
  plot of residuals 4-66
  simple linear 4-65
report preparation 5-28
representativeness *5-13*
reproducibility *5-11*
reservoirs 1-13
  characteristics 1-14
resource extraction 1-9
rivers 1-12
  characteristics 1-12
rivers and stream 1-12
Rural Clean Water Program
  Idaho 2-26
  Vermont 2-24
safety 5-23
salts 1-6, 1-7
sample
  composite 2-31
  continuous 2-31
  grab 2-31
  integrated 2-31
  mean 4-18
  type 2-31
sample custody 5-14
sample handling 5-14, 5-23
sampling
  cluster 2-19
  cluster, two-stage 2-20
  double 2-21
  equipment 2-32
  fish 3-24
  frequency 2-13
  macroinvertebrate 3-21
  optimizing 2-18, 5-7
  record keeping 2-32
  simple random 2-14
  site locations 2-25
  site selection 3-29
  station 2-32
  stratified random 2-17
  systematic 2-18
  systematic stratified 2-19
  time 2-21
  two-stage 2-19, 2-20
  two-step 2-15
SAS 4-77
scatter plot 4-25
Sen slope estimator 4-88
Shapiro-Wilk *W* 4-31, 4-38
silviculture 1-10
skewness 4-18
software
  data analysis 4-10
  multivariate 4-91
soils 1-21
Spearman's rho 4-90
standard deviation 4-1, 4-16, 4-19
  see also *statistic*
statistic
  average conditions 4-7
  central tendency 4-4, 4-10
  changing conditions 4-7
  coefficient of variation 4-16
  descriptive 4-1
  geometric mean 4-1, 4-11
  interquartile range 4-1, 4-16
  mean 4-1
  median 4-1
  mode 4-1
  nonparametric 4-4
  parametric 4-4
  standard deviation 4-1, 4-16, 4-19
  variability 4-4

variance 4-16
weighted mean 4-11
stem-and-leaf plot 4-21
storm return period 4-95
storm water 1-2, 1-3, 1-4, 1-7, 1-11, 2-10, 4-8, 4-11, 4-94
agriculture 1-1
stream
characteristics 1-12
order 2-3
segmentation 1-13
temperature 1-8, 1-9, 1-10
streambank
destabilization 1-1, 1-2, 1-4
erosion 1-6, 1-7
scouring 1-7
streamside vegetation
removal 1-7, 1-9
Student's *t* 2-15, 4-7, 4-18, 4-35, 4-46
paired 4-8
two-sample 4-47
test
Duncan's multiple range 4-63
equal variances 4-32
Gabriel's multiple-comparison 4-63
graphical 4-26
kurtosis 4-30
normality 4-26
one sample 4-34
one-sided 4-34
paired data 4-34
randomness 4-33
Ryan-Einot-Gabriel-Welsch multiple *F* 4-63
sign 4-44, 4-63
skewness 4-27
two-sample 4-47
two-sided 4-34
Waller-Duncan *k*-ratio 4-63
time series plot 4-20
time-weighted mean concentration 4-11
topography 1-21
slope 1-21
trend, monotonic 4-85, 4-86
Tukey's method 4-63
TWMC see *time-weighted mean concentration*
Universal Soil Loss Equation 2-11, 2-12
Revised 2-11
urban runoff 1-7
urbanization 1-7
variability
hydrologic 2-11
spatial 2-11
see also *statistic*
variables
selection 2-7
surrogate measures 2-9
variance 4-16
see also *statistic*
Water Erosion Prediction Project 2-11
water quality, correlation to land treatment 2-11
water resource types 1-12
watershed sampling
multiple 2-24
nested paired approach 2-22
paired approach 2-22, 3-3, 4-74
single 2-24
weighted mean 4-11
see also *statistic*
Wilcoxon signed rank 4-8, 4-34, 4-43, 4-46
Wilcoxon's rank sum
see *Mann-Whitney*

[Note: Italicized page numbers indicate pages on which definitions of terms are found.]

APPENDIX A: REVIEW OF AVAILABLE MONITORING GUIDANCES

A.1 INTRODUCTION

The monitoring guidances and papers reviewed in this appendix are grouped under the following headings:

- General Monitoring Guidances and References
- Guidances for Stream and River Monitoring
- Guidances for Lake and Reservoir Monitoring
- Guidances for Watershed Monitoring
- Guidances for Ground Water Monitoring
- Guidances for Biological Monitoring
- Program-Specific Monitoring Guidances

The reviews have been organized primarily according to the type of waterbody or waterbody system to which the guidance is most applicable—streams and rivers, lakes and reservoirs, or watersheds. Guidances that are not specific to one type of waterbody are grouped under the remaining headings. For example, a guidance for biological monitoring in streams and rivers is listed under "Guidances for Stream and River Monitoring," whereas a biological monitoring guidance that is not specific to waterbody type is listed under "Guidances for Biological Monitoring." Numerous guidances are not specific to either a waterbody type or a particular type of monitoring, and these are listed under "General Monitoring Guidances and References." Guidances specific to programs mandated under the Clean Water Act, such as the National Pollutant Discharge Elimination System (NPDES) program or the National Estuary Program, are listed separately.

A complete listing of the guidances reviewed in this appendix is presented first, followed by the individual reviews. A table at the end of the appendix indicates at a glance the aspects of monitoring addressed in each guidance. The numbers at the top of the table correspond to the numbers preceding the guidance titles in this listing and in the reviews.

A.2 CATEGORICAL LISTING OF MONITORING GUIDANCES REVIEWED

A.2.1 General Monitoring Guidances and References

1. *National Handbook on Water Quality Monitoring*. USDA-NRCS 1996.

2. *Water Quality Indicators Guide*. USDA-NRCS 1997.

3. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring*. Coffey, Spooner, and Smolen 1993.

4. *Coordinated Nonpoint Source Water Quality Monitoring Program for Idaho*. Idaho Department of Health and Welfare 1990.

5. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document.* Minnesota Pollution Control Agency 1989.

6. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards.* Dissmeyer 1994.

7. *Designing Effective Nonpoint Source Water Quality Monitoring Programs.* Maas 1989.

8. *Volunteer Estuary Monitoring: A Methods Manual.* USEPA 1993.

9. *Volunteer Water Monitoring: A Guide for State Managers.* USEPA 1990.

10. *The Volunteer Monitor.* The National Newsletter of Volunteer Water Quality Monitoring.

11. *The Strategy for Improving Water-Quality Monitoring in the United States.* Intergovernmental Task Force on Monitoring Water Quality 1995.

12. *The Strategy for Improving Water-Quality Monitoring in the United States.* Technical Appendixes. Intergovernmental Task Force on Monitoring Water Quality 1995.

13. *Guidance for State Water Monitoring and Wasteload Allocation Programs.* USEPA 1985.

14. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects.* USEPA 1981.

A.2.2 Guidances for Stream and River Monitoring

15. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* MacDonald, Smart and Wissmar 1991.

16. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams.* USEPA 1993.

17. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* USEPA 1989.

18. *An Improved Biotic Index of Organic Stream Pollution.* Hilsenhoff 1987.

19. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin.* Lyons 1992.

20. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects.* Simonson and Lyons 1992.

21. *1996 Water Body Assessment Guidance: A Stream to Standards Process*. Idaho Division of Environmental Quality 1996.

22. *1996 Beneficial Use Reconnaissance Project Workplan*. Idaho Division of Environmental Quality 1996.

23. *Handbook: Stream Sampling for Waste Load Allocation Applications*. USEPA 1986.

24. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects*. Simonson and Lyons 1992.

25. *Techniques for Detecting Effects of Urban and Rural Land-Use Practices on Stream-Water Chemistry in Selected Watersheds in Texas, Minnesota, and Illinois*. Walker 1993.

A.2.3 Guidances for Lake and Reservoir Monitoring

26. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Volumes I-III. USEPA 1983-1984.

27. *Volunteer Lake Monitoring: A Methods Manual*. USEPA 1991.

28. *Statistical Methods for the Analysis of Lake Water Quality Trends*. USEPA 1993.

29. *Monitoring Lake and Reservoir Restoration*. USEPA 1990.

A.2.4 Guidances for Watershed Monitoring

30. *Monitoring Primer for Range Watersheds*. Bedill and Buckhouse Draft.

31. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 years of controlling agricultural nonpoint source pollution: The RCWP experience*. USEPA 1992.

32. *Watershed Monitoring Manual*. Kansas Biological Survey 1993.

A.2.5 Guidances for Ground Water Monitoring

33. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water*. USEPA 1991.

34. *Ground Water Monitoring: A Guide to Monitoring for Agricultural Nonpoint Source Pollution Projects*. Goodman, German, and Bishoff 1995.

35. *Response to Ground Water Contaminants Detected through Idaho's Statewide Ambient Ground Water Quality Monitoring Program*. Idaho Department of Health and Welfare 1995.

A.2.6 Guidances for Biological Monitoring

36. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters.* USEPA 1993.

37. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* Volume I. USEPA 1987.

38. *Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents.* USEPA 1973.

A.2.7 Program-specific Monitoring Guidances

39. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects.* USEPA 1991.

40. *NPDES Storm Water Sampling Guidance Document.* USEPA 1992.

41. *Monitoring Guidance for the National Estuary Program.* USEPA 1991.

42. *Water Quality Standards Handbook.* USEPA 1983.

43. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* USEPA 1989.

44. *Environmental Monitoring and Assessment Program: Ecological Indicators.* USEPA 1990.

45. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document.* USEPA 1986.

46. *Summary of U.S. EPA-Approved Methods, Standard Methods, and Other Guidance for 301(h) Monitoring Variables.* USEPA 1985.

47. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance.* USEPA 1989.

48. *CWA Section 403: Procedural and Monitoring Guidance.* USEPA 1994.

49. *Methods for Collecting Benthic Invertebrate Samples as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

50. *Methods for Collecting Algal Samples as Part of the National Water Quality Assessment Program.* Porter, Cuffney, Gurtz, and Meador 1993.

51. *Methods for Characterizing Stream Habitat as Part of the National Water Quality Assessment Program.* Meador, Hupp, Cuffney, and Gurtz 1993.

52. *Guidelines for the Processing and Quality Assurance of Benthic Invertebrate Samples Collected as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

53. *Methods for Sampling Fish Communities as a Part of the National Water Quality Assessment Program.* Meador, Cuffney, and Gurtz. 1993.

54. *Guidelines for Studies of Contaminants in Biological Tissues for the National Water Quality Assessment Program.* Crawford and Luoma 1994.

A.3 GUIDANCE REVIEWS

A.3.1 General Monitoring Guidances and References

1. *National Handbook on Water Quality Monitoring*. USDA-NRCS 1996.

> **REFERENCE:** USDA-NRCS. 1996. *National Handbook on Water Quality Monitoring*. U.S. Department of Agriculture, Natural Resources Conservation Service. December.
>
> **CONTENTS:** Water quality problem; Objectives; Statistical designs; Scale of study; Variable selection; Sample type; Sampling location; Sampling frequency and duration; Station type; Sample collection and analysis; Land use and management monitoring; Data management.
>
> **MAIN FOCUS:** Primarily useful for designing a monitoring program. Thoroughly discusses all aspects of monitoring program design (site selection, parameter selection, sampling program design) with numerous examples and illustrations to help the reader determine what is most appropriate for particular applications.

This document describes methods for monitoring water quality responses to land use, land management activities, and conservation practices in streams, lakes, and ground water. It is intended to be a comprehensive guidance for water quality managers who have little experience in monitoring study design and implementation. The many purposes that monitoring studies can serve are discussed to help water quality managers define the scope of their monitoring programs. Examples of programs meant to serve the different purposes are included.

Specific guidance is provided on designing a monitoring study, setting up a monitoring station, and analyzing water quality data. Worksheets are provided to facilitate rapid and complete monitoring study design. The document contains 12 chapters, each of which covers a specific step of monitoring study design: (1) define the water quality problem, (2) define the monitoring objectives, (3) describe statistical design, (4) determine the study scale, (5) select water quality variables, (6) select sample type, (7) determine sampling locations, (8) determine sampling frequency and duration,(9) design station type, (10) describe sample collection and analysis methods, (11) describe land use monitoring, and (12) describe data management.

2. *Water Quality Indicators Guide*. USDA-NRCS.

> **REFERENCE:** USDA-NRCS. 1997. *Water Quality Indicators Guide*. U.S. Department of Agriculture, Natural Resources Conservation Service.
>
> **CONTENTS:** Pollution related to agriculture; Water quality field analysis; Ecology of freshwater streams; Sediment; Nutrients; Pesticides; Animal Wastes; Salts. Appendices: Water quality procedures; Aquatic organisms; Glossary; Conservation and best management practices; Field sheets.
>
> **MAIN FOCUS:** Field analysis of water pollution. Indicators for water pollution due to sediment, nutrients, pesticides, animal wastes, and salts are explained and field sheets are provided for tabulating water quality conditions.

The core of the *Water Quality Indicators Guide* is the field sheets and list of associated practices to remedy or abate agricultural nonpoint source pollution. The field sheets are arranged in matrix format with environmental indicators given for sediment, animal wastes, nutrients, pesticides, and salts. Each indicator is divided into descriptions of the environment from excellent to poor, and each description is given a weighted numerical ranking. The user matches the individual description with what is observed in the water or on the land. By totaling the individual rankings, a score is obtained indicating the potential for agricultural nonpoint source problems. Practices can be selected from the list to alleviate problem situations. *(From document preface.)*

3. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring*. Coffey, Spooner, and Smolen 1993.

REFERENCE: Coffey, S.W., J. Spooner, and M.D. Smolen. 1993. *The Nonpoint Source Manager's Guide to Land Treatment and Water Quality Monitoring*. NCSU Water Quality Group, Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina.

CONTENTS: Overview of monitoring program; Management objectives and problem development; Monitoring program objectives; Monitoring program design; Data collection; Data analysis; Program evaluation.

MAIN FOCUS: Specifically addresses monitoring to track land treatments. Distinguishes between two levels of monitoring: Level I—basic, low-cost monitoring, and Level II—more intensive and comprehensive monitoring, at higher cost. Water quality and land treatment monitoring parameters are discussed separately for each level of monitoring.

This guide discusses the objectives and design of monitoring program used to evaluate present conditions, identify water quality problems, detect trends and impacts, and document water quality improvements associated with the implementation of land treatments. Variations in monitoring program objectives and designs to make them appropriate for streams, lakes, wetlands, and reservoirs are indicated. Two levels of detail, applicable to different monitoring objectives, are discussed separately for each of three types of monitoring programs: land use and land treatment monitoring, water quality monitoring, and loading rate monitoring. For each level of detail and each type of monitoring, appropriate biological, habitat, physical, and chemical variables are discussed.

A chapter on the design of monitoring programs emphasizes the importance of monitoring both land treatments and water quality to provide feedback on the impact of land management activities on water quality. Guidance on the use of existing data in monitoring program design is provided, and estimated costs for some procedures are included to help water quality managers estimate monitoring program budgets. Only brief discussions of data collection and analysis and program evaluation are presented. While this guidance primarily addresses agricultural land treatments, it is equally applicable to designing and implementing programs to monitor the effects of other types of land treatments, or general status and trend monitoring programs.

4. *Coordinated Nonpoint Source Water Quality Monitoring Program for Idaho.* Idaho Department of Health and Welfare 1990.

REFERENCE: Clark, W.H. 1990. *Coordinated nonpoint source water quality monitoring program for Idaho.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, Idaho. January.

CONTENTS: Introduction; Program development; Interagency monitoring responsibilities; Surface water quality program; Monitoring program implementation; Summary. Appendices (partial listing): Idaho monitoring sampling site listing, suggested monitoring parameters and protocols, and a monitoring plan checklist.

MAIN FOCUS: Statewide, integrated nonpoint source monitoring program for Idaho, addressing monitoring for agriculture, forestry, and mining.

The document serves as a guide for a coordinated statewide monitoring program established in Idaho. It describes basin and watershed trend monitoring, beneficial use monitoring, and best management practice effectiveness monitoring. The three main nonpoint source activities that occur in Idaho—agriculture, forestry, and mining—are addressed by the statewide monitoring program and in the guide. For each of these activities, there is an introduction and objectives section, a description of the current program, and a description of the recommended program. The monitoring program described in the guide addresses trends in major river basins and watersheds, beneficial use support status, and best management practice effectiveness. The guide also includes a listing of appropriate variables and protocols and a checklist of major items to be included in a nonpoint source water quality monitoring plan. [From document abstract]

5. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document.* Clean Water Partnership 1989.

REFERENCE: Clean Water Partnership. 1989. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document.* Minnesota Pollution Control Agency, Division of Water Quality. March.

CONTENTS: Purpose; Steps in developing a monitoring plan; Water quality parameters; Sampling site selection; Sampling strategies; Hydrology; Quality assurance/quality control.

MAIN FOCUS: Brief discussion of each monitoring parameter covering reason for its measurement, considerations for measuring the parameter, and sampling and analysis techniques.

This guidance covers all of the technical steps in developing and implementing monitoring programs for the evaluation of the effect of projects on water quality. Brief descriptions of each of the parameters to be monitored constitute most of the document. The monitoring program described is one that meets the requirements for Clean Water Partnership projects.

6. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards*. Dissmeyer 1994.

REFERENCE: Dissmeyer, G.E. 1994. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards*. USDA Forest Service, Southern Region, Atlanta, Georgia. Miscellaneous Publication 1520. June.

CONTENTS: Planning monitoring projects; Quality assurance/quality control; Statistical considerations; Selecting the appropriate BMP effectiveness monitoring level; Monitoring methods; Deciding on BMP effectiveness: Some case histories.

MAIN FOCUS: Discussion of four levels of monitoring for different monitoring program objectives. Focuses on monitoring program design for determination of BMP effectiveness, management activity water quality impact, choice of BMPs for best water quality protection, and longevity of BMP effectiveness. Discussions of numerous monitoring protocols, with references to detailed descriptions for further information.

This manual is intended to assist managers and staff in developing water quality monitoring plans to evaluate the effectiveness of forestry BMPs in meeting water quality goals or standards for streams, including chemical, physical, biological, and habitat integrity. The focus is on monitoring project design and parameters and methods selection. The methods discussed in the document are appropriate for fourth order streams and smaller. Monitoring project design is discussed with the intention of separating the NPS impact of management activities from other NPS impacts. The document also addresses assessments of how long BMPs remain effective and which BMPs best protect water quality. Effectiveness monitoring is separated into four levels, depending on program needs. Monitoring methods for chemical, physical, biological, and habitat integrity are recommended to achieve each of the four monitoring levels. Generally, the reader is referred to the original literature or to other documents for procedural details of the monitoring methods, and this guide serves as a manual for monitoring program design.

7. *Designing Effective Nonpoint Source Water Quality Monitoring Programs*. Maas 1989.

REFERENCE: Maas, R.P. 1989. *Designing Effective Nonpoint Source Water Quality Monitoring Programs*. Prepared for the Tennessee Valley Authority. November.

CONTENTS: I. General guidance for developing nonpoint source monitoring programs: Background; Purposes and objectives of water quality monitoring in NPS projects; Source inventories; Water sampling in NPS control projects; NPS water quality monitoring designs; Biological monitoring; Using preliminary data to get the most from your water quality monitoring program; Determining the minimum detectable change in water quality required to be observable from an NPS water quality monitoring program; Matching purposes and design of nonpoint source water quality monitoring programs; II. Case study: Oostanaula Creek, Tennessee.

MAIN FOCUS: Focuses primarily on monitoring program design for documenting the benefit of NPS control programs; Provides details (i.e., monitoring program design, statistical analyses) about time-trend and paired watershed designs; Includes a case study of implementing a cost-effective NPS monitoring program.

The purpose of this document is to provide guidance to those involved in NPS pollution control in designing and carrying out successful and cost-efficient water quality monitoring programs. It is intended as a technical

document to provide NPS monitoring professionals with the necessary resource materials to make informed decisions regarding the design of water quality monitoring programs. It should also provide NPS program managers with sufficient background to evaluate technical decisions on NPS monitoring strategies for their consistency with an overall purpose and their probability of documenting water quality improvement. This document focuses primarily on experience gained through state and federal agricultural and urban NPS projects, such as the Model Implementation Program, initiated in 1978, and the Rural Clean Water Program, initiated in 1980. (from author's preface)

A particularly useful application of NPS monitoring information is to document the beneficial effect of an NPS control program on water quality. This, however, can be a particularly difficult monitoring objective to fulfill. This document examines monitoring design characteristics that maximize the ability to achieve this objective.

8. *Volunteer Estuary Monitoring: A Methods Manual.* USEPA 1993.

REFERENCE: USEPA. 1993. *Volunteer Estuary Monitoring: A Methods Manual.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 842-B-93-004. December.

CONTENTS: Our troubled estuaries; Setting the stage; Monitoring dissolved oxygen; Monitoring nutrients and phytoplankton; Monitoring submerged aquatic vegetation; Monitoring bacteria; Monitoring other estuarine conditions; Training volunteers; Presenting monitoring results; Appendices.

MAIN FOCUS: Clearly written, intended to serve a lay audience. Step-by-step descriptions of common estuary monitoring methods.

This manual is meant to be a companion to three other EPA documents: *Volunteer Water Monitoring: A Guide for State Managers*, *Volunteer Lake Monitoring: A Methods Manual*, and *Volunteer Stream Monitoring: A Methods Manual*. The manual reviews those water quality parameters considered most important to monitor to determine an estuary's water quality: dissolved oxygen, bacteria, nutrients, phytoplankton, and submerged aquatic vegetation. Each chapter discusses why it is important to monitor the particular parameter, the role of the parameter in estuarine ecology, and sampling equipment for taking measurements of the parameter. Methods for sampling the parameter are set out in easy-to-follow steps. Two introductory chapters discuss the state of the Nation's estuaries, basic estuarine ecology, basic monitoring equipment, and gross conditions to note while monitoring, such as temperature and shoreline condition. Other chapters discuss volunteer training, the importance of credible data, and data presentation techniques. Supply houses are listed in one appendix, and quality assurance is addressed in another.

9. *Volunteer Water Monitoring: A Guide for State Managers.* USEPA 1990.

This guide is directed toward those contemplating setting up a volunteer monitoring program. Nearly every aspect of creating a successful monitoring program is discussed, including attracting volunteers, keeping volunteers interested and motivated, finding funding sources, quality assurance and quality control, and ensuring that the collected data are put to use. The techniques and methods of monitoring are discussed in other manuals (e.g., *Volunteer Lake Monitoring: A Methods Manual*).

Descriptions of five state volunteer monitoring programs are provided: Illinois, Kentucky, New York, Ohio, and the Chesapeake Bay. Each discussion provides details of program objectives, how volunteers were recruited and trained, sampling protocols, data management, program administration, volunteer recognition, and expenses and funding. State contact names, addresses, and phone numbers are provided for each program.

REFERENCE: USEPA. 1990. *Volunteer Water Monitoring: A Guide for State Managers*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 440/4-90-010.

CONTENTS: Volunteers in water monitoring; Planning a volunteer monitoring program; Implementing a volunteer monitoring program; Providing credible information; Costs and funding. Appendix: Descriptions of five successful programs.

MAIN FOCUS: A guide to the administrative details of a volunteer monitoring program, including training volunteers, funding, program design, and media relationships. Presentations of volunteer monitoring programs in the appendix provide good examples of program implementation to meet differing needs and to deal with different local and state requirements. State contacts are provided.

10. *The Volunteer Monitor*. Rhode Island Sea Grant.

REFERENCE: *The Volunteer Monitor*. Published by the Coastal Resources Center, the University of Rhode Island, Rhode Island Sea Grant Program, Narragansett, Rhode Island.

CONTENTS: Example issue content: Special topic: biological monitoring; Monitoring groups need a national association; EPA Lakes Methods manual; Wetlands field guide; Lake monitoring network; Technical tips; Doing your own lab analysis for fecal coliform; Monitoring aquatic plants; Monitoring diseased eelgrass; Local bank sponsors River Rescue; Third National Volunteer Monitoring Conference set.

MAIN FOCUS: Typically provides information on monitoring programs throughout the country, new publications, materials and technical methods, contacts for equipment supplies, school programs, and information on monitoring particular plants or animals.

The Volunteer Monitor is a newsletter published by Rhode Island Sea Grant that is devoted entirely to topics related to volunteer monitoring. Issues contain articles on a diversity of topics, including chemical, physical, and biological monitoring, reviews of documents related to monitoring, EPA activities, monitoring conferences, monitoring associations, equipment and methods, data collection and analysis, public education, and technology transfer. Information on purchasing monitoring supplies and articles of special interest, such as on methods to monitor particular species or habitats, are regularly published.

11. *The Strategy for Improving Water-Quality Monitoring in the United States*. Intergovernmental Task Force on Monitoring Water Quality 1995.

The Intergovernmental Task Force on Monitoring Water Quality was established to address a need for greater coordination among agencies collecting water-quality information to make that information more meaningful to all users. This is the final report of the ITFM. It describes the strategy of the ITFM for nationwide water quality monitoring and its implementation. Highlights of the strategy and recommendations, discussed in detail in the final report and technical appendixes, include a focus on goal-oriented monitoring and the use of indicators; gathering and using existing information; ensuring institutional collaboration; striving for methods comparability; improving information automation, accessibility, and utility; adhering to quality assurance and quality control protocols; evaluating monitoring activities; identifying needs for research and development; promotion of training; and provision of adequate funding.

REFERENCE: ITFM. 1995. *The Strategy for Improving Water-Quality Monitoring in the United States.* Final report of the Intergovernmental Task Force on Monitoring Water Quality. Intergovernmental Task Force on Monitoring Water Quality. February.

CONTENTS: Major conclusions and recommendations; Summary; General intent; Background; Water quality questions; Nationwide strategy for improving water quality monitoring; Implementation; Initial agency actions to improve monitoring; Conclusion.

MAIN FOCUS: Recommends a strategy for nationwide water quality monitoring and technical monitoring improvements to support sound water quality decision-making at all levels of government and in the private sector.

12. *The Strategy for Improving Water-Quality Monitoring in the United States. Technical Appendixes.* Intergovernmental Task Force on Monitoring Water Quality 1995.

This companion document to the ITFM 1995 final report provides detailed discussions of the more technical aspects of the national monitoring program. In addition, general information on monitoring programs, such as program objectives and design, are discussed in the context of the national program. Numerous tables related to the technical topics discussed consolidate much information contained in many other technical monitoring documents in a single place.

REFERENCE: ITFM. 1995. *The Strategy for Improving Water-Quality Monitoring in the United States. Technical Appendixes.* Final report of the Intergovernmental Task Force on Monitoring Water Quality. Intergovernmental Task Force on Monitoring Water Quality. February.

CONTENTS: Glossary; Framework for a water-quality monitoring program; Terms of reference—National Water-quality Monitoring Council; Indicators for meeting management objectives—summary and rationale; Indicator selection criteria; Ecoregions, reference conditions, and index calibration; Multimetric approach for describing ecological conditions; Terms of reference—Interagency Methods and Data Comparability Board; Data comparability and performance-based methods policy paper—comparability of data-collection methods; Target audiences, monitoring objectives, and format considerations for reporting water-quality information; Annotated bibliography of selected outstanding water-quality reports; Ground water quality monitoring framework; Date elements glossary; Evaluation of a performance-based methods system approach to field and prelaboratory methods; Performance-based methods system for biological collection methods—A framework for examining method comparability.

MAIN FOCUS: Discussions of numerous technical points related to the national monitoring program, including indicators, ecoregions, and multimetrics. Provides useful charts and graphs related to these topics.

13. *Guidance for State Water Monitoring and Wasteload Allocation Programs*. USEPA 1985.

REFERENCE: U.S. EPA. 1985. *Guidance for State Water Monitoring and Wasteload Allocation Programs*. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC. EPA 440/4-85-031. October.

CONTENTS: Overview of water quality program monitoring; Monitoring for water quality-based controls; Monitoring for compliance and enforcement; Water quality assessments; Quality assurance; Data reporting; Total maximum daily loads and wasteload allocations.

MAIN FOCUS: Provides a general overview of state water quality monitoring responsibilities under the Clean Water Act (pre-1987).

This guidance is oriented toward program management and does not contain extensive technical information on monitoring programs. The guidance is intended to be used by states and EPA regional offices for developing the monitoring and wasteload allocation portions of annual state 106 and 205(j) work programs. Monitoring and wasteload allocation activities are defined in accordance with EPA regulations. Two principal areas are covered by the guidance. The first is an outline of the objectives of the water monitoring program, and the second is a description of the process for calculating total maximum daily loads (TMDLs), waste load allocations for point sources, and load allocations for nonpoint sources of pollution.

Separate chapters of the guidance address different aspects of state programs (conducting water quality assessments, developing water quality-based controls, and assessing compliance), and for each of these program aspects the guidance discusses the types of data needed, who is responsible for doing the work, the methods to be used, data reporting requirements, parameters, and uses of the data. References to technical guidance and sample program checklists are provided in an appendix.

14. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects*. USEPA 1981.

REFERENCE: USEPA. 1981. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects*. U.S. Environmental Protection Agency. L.R. Shuyler, EPA project officer. Prepared under the Rural Nonpoint Source Control Water Quality Evaluation and Technical Assistance project.

CONTENTS: Evaluation procedures for nonpoint source control measures; Evaluation and sampling needs; Streams; Lakes; Ground water; Socioeconomic evaluation.

MAIN FOCUS: The sections on streams and lakes contain an assessment of evaluation alternatives and specific monitoring recommendations relevant to the type of receiving water, including guidance for identification of impacted beneficial uses. The important characteristics of each receiving water type that should be monitored are discussed.

This guidance was developed under EPA's Rural Nonpoint Source Control Water Quality Evaluation and Technical Assistance project as a joint effort of EPA and USDA. It is intended to provide basic guidelines for measuring water quality changes and estimating socioeconomic impacts resulting from NPS pollution control programs. Monitoring requirements for a basic, level I monitoring program are discussed, as well as more

intensive requirements for a detailed monitoring program. The guidance provided in this document is basic in the sense that the parameters that should be measured for minimum level I and level II assessments are mentioned and discussed briefly, but no guidance for taking measurements, identifying monitoring stations, designing a monitoring program, or analyzing results is provided. Emphasis is placed on the collection of historical data prior to monitoring and on monitoring physical, chemical, and biological characteristics that will indicate present condition and changes in those characteristics due to BMP implementation.

A separate section on socioeconomic evaluation of NPS pollution control programs is provided. Questions of impacts on farmer income, overall project value with respect to alternative community needs, and choices of alternative environmental projects are mentioned as important considerations prior to the implementation of an NPS pollution control project. Two levels of socioeconomic evaluations are discussed. A level I evaluation considers the effects of an NPS pollution control project on land use, crop production, income, pollution control effectiveness, and project efficiency in a cost/benefit sense. A level II evaluation includes the elements of a level I evaluation as well as an estimation of community and off-farm impacts and an evaluation of alternative options. Tables at the end of the paper summarize the recommendations presented in the text.

A.3.2 Guidances for Stream and River Monitoring

15. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* MacDonald, Smart, and Wissmar 1991.

> **REFERENCE:** MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* U.S. Environmental Protection Agency, Region 10, Water Division, Seattle, Washington. EPA/910/9-91-001. May.
>
> **CONTENTS:** Part I: Context and structure of monitoring projects; Statistical considerations in water quality monitoring; Principles of developing a monitoring plan and selecting the monitoring parameters; parameter recommendations and interactions. Part II: Physical and chemical constituents; Changes in flow; sediment; Channel characteristics; Riparian monitoring; Aquatic organisms.
>
> **MAIN FOCUS:** Provides a good review of the importance of proper statistical design in a monitoring program. Tabulates monitoring parameters according to their usefulness for monitoring different land treatments. Specific to monitoring stream conditions in the Pacific Northwest and Alaska.

The focus of these guidelines is on monitoring water quality in streams, and it does not directly address monitoring in lakes, reservoirs, or other downstream areas. The discussions are also limited to conditions in the Pacific Northwest and Alaska, which reduces the scope of conditions and activities considered in the document. However, the information provided on monitoring objectives and monitoring parameters is generally applicable. Information on monitoring to detect the water quality impacts of grazing, mining, and recreation is also provided because these activities occur on or near lands where forestry activities are conducted and it can be difficult to separate the water quality impacts of these activities from those of forestry operations.

The document has two parts. Part 1 discusses seven purposes of monitoring, legal requirements for NPS pollution monitoring, statistical considerations in water quality monitoring, monitoring plan development, and the selection of monitoring parameters. Monitoring parameters are recommended for different forestry-related activities (e.g., forest harvest, road construction). Part 2 is a comprehensive discussion of individual monitoring parameters and is intended to facilitate the selection of the most appropriate monitoring parameters for specific monitoring

objectives. This is not a technical guidance on sampling procedures or statistical analyses used for monitoring programs, but rather a comprehensive discussion of the various elements of a monitoring program to help water resource or forestry operations managers make informed decisions concerning monitoring programs. References are provided to direct the reader to the appropriate technical guidance where needed.

16. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams*. USEPA 1993.

REFERENCE: USEPA. 1993. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams*. U.S. Environmental Protection Agency, Region 10, Water Division, Surface Water Branch. EPA 910/R-93-017. October.

CONTENTS: Impacts of grazing on water quality and beneficial uses; Monitoring plan procedure; Stratification, reconnaissance, and classification of rangeland riverine riparian areas; Evaluation/recommendation of monitoring methods: Stream temperature and shade; Nutrients; Bacterial indicators; Stream channel morphology; Streambank stability; Substrate fine sediment; Pool quality; Streamside vegetation.

MAIN FOCUS: Describes a methodology to classify a stream and riparian vegetation prior to selecting monitoring sites. Provides tables of monitoring parameter attributes (sampling frequency, collection time necessary, equipment requirements, lab costs, level of expertise needed). Detailed information on recommended methods is provided.

This document describes a monitoring system to assess the impacts of grazing on water quality in streams of the western United States and protocols used to assess changes in water quality that result from stream restoration projects. Protocols that are easy to use and cost-effective (i.e., have reduced sampling frequency, minimized need for specialized equipment, and a reduced requirement for laboratory analyses) were selected. The document focuses on monitoring attributes of the stream channel, stream bank, and streamside vegetation that are important to the support of aquatic life and that are impacted by grazing. A discussion of the impacts of grazing on stream ecosystems provides a basis for selecting monitoring parameters, and a procedure to develop a monitoring plan is recommended. Methods to stratify and classify stream reaches based on stream type, dominant soils, and riparian vegetation communities are described, and they provide a basis for selecting monitoring sites and reference areas. Forms for recording data and a list of equipment needed for each protocol are provided.

Monitoring methods that are commonly used to assess the effects of grazing on water quality are described, and the attributes of each are tabulated. The advantages and disadvantages of the methods are described and tabulated for ease of comparison, and recommendations for specific protocols are made based on ease of use and cost-effectiveness. These recommended protocols, including data collection and analysis procedures, are then thoroughly described. The recommended protocols are stream temperature and shade, nutrients, bacterial indicators, stream channel morphology, streambank stability, substrate fine sediment, pool quality, and streamside vegetation. A list of references pertaining to each protocol is provided at the end of the discussion of each one. The usefulness of the recommended protocols is not limited to monitoring the impacts of grazing, and the reader should find the discussions of the protocols valuable if the same protocols are being considered for monitoring the water quality effects of other land use activities. Other references should be consulted for specific guidance on sampling and statistical analysis of data.

17. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* USEPA 1989.

REFERENCE: U.S. EPA. 1989. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA/440/4-89/001. May.

CONTENTS: The concept of biomonitoring; Overview of protocols and summary of components; Quality assurance/quality control; Habitat assessment and physicochemical parameters; Benthic macroinvertebrate biosurvey and data analysis; Fish biosurvey and data analysis; Integration of habitat, water quality, and biosurvey data. Appendices: Guidance for use of field and laboratory data sheets; Rapid bioassessment subsampling methods for benthic protocols; Family and species-level macroinvertebrate tolerance classifications; Tolerance, trophic guilds, and origins of selected fish species.

MAIN FOCUS: Provides both an introduction to the concept of biomonitoring and detailed methods sections for conducting rapid bioassessment protocols. Sample data sheets for all protocols are provided, and data analysis and interpretation is thoroughly discussed.

This is a comprehensive technical reference for bioassessment procedures for streams and rivers. Three macroinvertebrate protocols and two fish protocols are presented. The macroinvertebrate protocols were tested in wadable freshwater streams, though they should be applicable to large freshwater rivers as well. They were developed by consolidating various procedures in use by a variety of state water quality agencies, and they require levels of effort ranging from fairly simple to rigorous. The fish protocols were validated in freshwater streams and large rivers and are therefore equally applicable to both. They were developed based on previous work by other researchers and on standard fish population assessment models.

The document contains an introduction to the biomonitoring approach of detecting aquatic life impairments and estimating their severity. Procedures for assessing the habitat where the sampling is done are explained, and the physical and chemical parameters relevant to biological survey data interpretation are discussed. Complete instructions for conducting benthic biosurveys and fish surveys are provided, laboratory methods and data analysis techniques are explained, and quality assurance and quality control are addressed. Field data forms and guidance for their use are provided.

18. *An Improved Biotic Index of Organic Stream Pollution.* Hilsenhoff 1987.

REFERENCE: Hilsenhoff, W.L. 1987. An Improved Biotic Index of Organic Stream Pollution. *Great Lakes Entomologist* 20:31-39.

CONTENTS: Abstract; Reassignment of tolerance values; Identification; Collection and evaluation of samples; Literature cited; Appendix 1: Tolerance values for stream arthropods.

MAIN FOCUS: Provides a discussion of identification problems specific to insect orders. Includes a detailed discussion of proper stream arthropod collection techniques. An appendix provides revised tolerance values for stream arthropods.

Hilsenhoff introduced a biotic index for evaluating the water quality of streams in 1977 and offers improvements to it in this paper. The initial index assigned tolerance values of 0-5 to species, but after evaluation of over 1,000 samples the index has been revised to accommodate tolerance values of 0-10 to provide greater precision. Tolerance values of species are provided in an appendix. Some orders of arthropods are difficult to identify, and the problems associated with those orders are discussed. A revised procedure for collecting, sorting, and evaluating samples using the biotic index presented in the paper is provided.

19. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin*. Lyons 1992.

REFERENCE: Lyons, J. 1992. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin.* U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. General Technical Report NC-149.

CONTENTS: General considerations; Applying the IBI in Wisconsin warmwater streams: Collecting and processing the field data; Analyzing the data; Interpreting IBI scores.

MAIN FOCUS: Complete discussion of data analysis. Provides Maximum Species Richness plots for data interpretation. Applicable to similar streams in nearby states.

This paper summarizes the results of a 4-year fish collection and data analysis effort aimed at developing a version of the IBI for warmwater Wisconsin streams. The paper is designed primarily as a "how to" manual and therefore contains little discussion of the principles of the IBI. Discussion focuses on collection of fish samples for analysis, and analysis and interpretation of the data. Maximum Species Richness (MSR) plots for data interpretation are provided.

Because of the similarity in stream characteristics and fish fauna between Wisconsin and parts of adjacent states, the Wisconsin version of the IBI described in this paper should be useful in southeastern and northeastern Minnesota, the entire Upper Peninsula and the northern Lower Peninsula of Michigan, extreme northwestern Illinois, and extreme northeastern Iowa.

20. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects*. Simonson and Lyons 1992.

REFERENCE: Simonson, T., and J. Lyons. 1992a. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects*. Wisconsin Department of Natural Resources. May.

CONTENTS: Station summary data sheet; Station map data sheet; Station flow data sheet; Transect data sheet; Appendix: Gear used for habitat sampling.

MAIN FOCUS: Each data element to be recorded is explained separately. Data recording forms are provided.

This paper describes all of the data elements to be recorded during evaluation phase monitoring of fish habitat for WDNR Priority Watershed Projects. The purpose of the monitoring is to document changes in fish communities and fish habitat that occur in streams where improved land use practices are implemented to reduce NPS pollution. The level of detail given for each data element make this paper a manual for conducting an evaluation,

not just a discussion of data elements. Two sets of sample data entry sheets are provided, one blank and one filled out with example data. The gear needed to conduct an evaluation is described in an appendix, and the names and addresses of suppliers are provided.

21. *1996 Water Body Assessment Guidance: A Stream to Standards Process.* Idaho Division of Environmental Quality 1996.

REFERENCE: Idaho Division of Environmental Quality. 1996. *1996 water body assessment guidance: A stream to standards process.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise Idaho. August.

CONTENTS: Introduction; Water body initialization; Aquatic life beneficial uses status determinations; Recreational beneficial uses status determinations; Water supply beneficial uses status determinations; Wildlife habitat and aesthetics beneficial uses status determinations; Idaho water quality standards and wastewater treatment requirements narrative criteria; Criterion evaluation process; Beneficial use attainability; Beneficial uses status determinations appeals process; Listing water quality-limited water bodies. Appendices (partial listing): Macroinvertebrate biotic index; Habitat index; Reconnaissance index of biotic integrity; Index of biotic integrity; Algal biotic index; Water body assessment form; Water body assessment guidance assumptions.

MAIN FOCUS: An analytical tool for determining if a water body is or is not supporting a beneficial use.

The guidance presents the assessment method that Idaho uses to list water quality-limited water bodies and to designate beneficial uses. The guidance incorporates a beneficial use attainability designation process for unclassified waters. The method used by Idaho uses ecological indicators to make water quality assessments. The document is easy to follow, owing to its flow chart type of organization. In this manner, an assessment is broken down into many easily-answered questions.

22. *1996 Beneficial Use Reconnaissance Project Workplan.* Idaho Division of Environmental Quality 1996.

REFERENCE: Idaho Division of Environmental Quality. 1996. *1996 Beneficial Use Reconnaissance Project workplan.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise Idaho. May.

CONTENTS: Introduction: Beneficial uses, Purpose, Objectives, Scope, Rationale for stream selection; Methods: Stream site selection, Core parameters, Pilot investigations to validate procedures, Rationale for parameter selection and summary of procedures, Recommended procedure sequence for site evaluation; Quality assurance; Safety training and certification; Data analyses and interpretation; Appendices: Streams proposed for monitoring in 1996, Field equipment check list, Field forms, Audit forms.

MAIN FOCUS: Descriptive overview of IDDEQ's Beneficial Use Reconnaissance Project.

The workplan lists the parameters to monitor, with references for each method where detailed information on them can be obtained, and monitoring level-of-intensity information for each parameter. The quality assurance (QA) section summarizes the QA approach for the project and references QA manuals to be consulted for detailed information. The workplan is modified annually to incorporate changes in methods and protocol. This workplan is

to be used as a guide for training stream monitoring field crews. It describes the Beneficial Use Reconnaissance Project, but is not a detailed "how-to" manual.

The protocols used by IDDEQ to measure water quality, beneficial use attainability, beneficial use status, and general stream health are described in it. The protocols are meant to provide a reconnaissance level screen of stream conditions, and as reconnaissance level protocols they are only intended to differentiate between impaired and non-impaired streams. The protocols are meant to be applicable to any wadable stream regardless of location in Idaho. Another objective of the protocols is to identify the principal measures that are likely to provide insight into stream ecology, biology, and water quality, and to determine their relationships to beneficial uses. Time constraints, staff limitations, and cost effectiveness were considered in selecting the protocols.

23. *Handbook: Stream Sampling for Waste Load Allocation Applications*. USEPA 1986.

REFERENCE: USEPA. 1986. *Handbook: Stream Sampling for Waste Load Allocation Applications*. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/625/6-86/013. September.

CONTENTS: Sampling requirements for wasteload allocation modeling; Sampling requirements for conventional pollutants; Sampling requirements for toxic pollutants; Whole effluent approach; Example application.

MAIN FOCUS: Summarizes data requirements for modeling applications and discusses the applicability and advantages of numerous models. Flow charts for modeling approaches and parameter formulas are provided.

This report discusses sampling requirements in support of wasteload allocation studies in rivers and streams. Two approaches to wasteload allocation are addressed: the chemical-specific approach and the whole effluent approach. Numerical or analytical toxicant fate models are used to implement the chemical-specific approach. Modeling requirements and sampling guidelines are delineated for this method.

For the whole effluent approach, the method is summarized and then instream dye study requirements are presented. The report concludes with example applications of the chemical-specific approach for conventional and toxic pollutants.

This guidance does not discuss equipment requirements, personnel requirements, sample collection, stream characterization, or laboratory analytical techniques. The primary purpose of the document is to assist water quality specialists in designing stream surveys to support modeling applications for wasteload applications. The data collection process required to calibrate, verify, and apply models used for wasteload allocations to critical design conditions is described.

24. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects*. Simonson and Lyons 1992.

This paper, the companion to Simonson and Lyons' paper *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects*, describes the data elements to be recorded during evaluation phase monitoring of fish communities for WDNR Priority Watershed Projects. This paper is similar in content to that paper, except that only one set (blank) of data sheets is provided.

REFERENCE: Simonson, T., and J. Lyons. 1992b. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects*. Wisconsin Department of Natural Resources. May.

CONTENTS: Station summary data sheet; Catch summary data sheet; Individual fish data sheet.

MAIN FOCUS: Each data element to be recorded is explained separately. Data recording forms are provided.

25. *Techniques for Detecting Effects of Urban and Rural Land-Use Practices on Stream-Water Chemistry in Selected Watersheds in Texas, Minnesota, and Illinois*. USGS 1993.

REFERENCE: Walker, J.F. 1993. *Techniques for Detecting Effects of Urban and Rural Land-Use Practices on Stream-Water Chemistry in Selected Watersheds in Texas, Minnesota, and Illinois*. U.S. Geological Survey and Wisconsin Department of Natural Resources. USGS Open-file Report 93-130.

CONTENTS: Techniques for detecting effects of land-use practices on water chemistry; Application of techniques to selected watersheds; Summary and conclusions.

MAIN FOCUS: Statistical techniques for the detection of effects of land-use practices on water chemistry are applied to selected watersheds. Alternative procedures for assessing the effects of land-use practices are compared.

There is little information available about the effectiveness of best management practices at the watershed scale. This report presents a discussion of several parametric and nonparametric statistical techniques for detecting changes in water-chemistry data. The use of storm load data is discussed as an alternative to using fixed-frequency instantaneous concentration data. Statistical techniques were applied to three urban watersheds in Texas and Minnesota and three rural watersheds in Illinois. For the urban watersheds, single- and paired-site data collection strategies were considered. For the rural watersheds, the selected techniques were found not to be effective at identifying changes. The use of regressions improved the ability to detect changes. (from author's abstract)

A.3.3 Guidances for Lake and Reservoir Monitoring

26. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Volumes I-III. USEPA 1983-1984.

These documents contain EPA guidance to assist states in implementing the revised Water Quality Standards Regulation that appeared in the *Federal Register* in November 1983. Consideration of the suitability of a waterbody for attaining a given use is an integral part of the water quality standards review and revision process. This guidance is intended to assist states in determining the uses currently being achieved, the potential uses of the waterbodies, and the causes of any impairment of the uses. A framework for determining the attainable aquatic protection use is described, and parameters to be used to make the determinations mentioned above are provided. Methods and approaches that can be used by states for conducting use attainability analyses are discussed. Volume I discusses rivers and streams, Volume II discusses estuarine systems, and Volume III discusses lake systems.

REFERENCE: USEPA. 1983. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Volume I. U.S. Environmental Protection Agency, Office of Water, Washington, DC. November.

USEPA. 1984. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Volume II: Estuarine systems. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1984. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Volume III: Lake systems. U.S. Environmental Protection Agency, Office of Water, Washington, DC. November.

CONTENTS: Volume I: Physical evaluations; Chemical evaluations; Biological evaluations; Interpretation. Volume II: Physical and chemical characteristics; Characteristics of plant and animal communities; Synthesis and interpretation. Volume III: Physical and chemical characteristics; Biological characteristics; Synthesis and interpretation.

MAIN FOCUS: A general discussion of conducting physical, chemical, and biological analyses of river and stream, estuarine, and lake systems.

27. *Volunteer Lake Monitoring: A Methods Manual.* USEPA 1991.

REFERENCE: USEPA. 1991. *Volunteer Lake Monitoring: A Methods Manual.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 440/4-91-002. December.

CONTENTS: Focusing on a lake condition; Monitoring algae; Monitoring aquatic plants; Monitoring dissolved oxygen; Monitoring other lake conditions; Training citizen volunteers; Presenting monitoring results. Appendix: Scientific supply houses.

MAIN FOCUS: Specifically addresses lake monitoring. Provides step-by-step instructions for setting up and implementing a volunteer lake monitoring program. Sample data collection forms are provided, and a discussion on lake ecology provides background for volunteer monitors. Well-illustrated.

This manual presents specific information on volunteer lake water quality monitoring methods. It is intended for organizers of volunteer lake monitoring programs and for the volunteers who actually sample lake conditions. It summarizes the steps necessary to plan and manage a volunteer monitoring program, including setting general goals, identifying the uses and users of collected data, and establishing sound quality assurance procedures. The document concentrates special attention on three of the most common lake pollution problems: increased algal growth, increased growth of rooted aquatic plants, and lowered or fluctuating levels of dissolved oxygen. It also briefly discusses other lake pollution problems: sedimentation, turbidity, lake acidification, and bacteriological conditions. Appropriate parameters to monitor and specific steps for each selected monitoring method are identified, and example sampling forms are provided.

28. *Statistical Methods for the Analysis of Lake Water Quality Trends.* Reckhow, Kepford, and Hicks 1993.

REFERENCE: Reckhow, K.H., K. Kepford, and W.W. Hicks. 1993. *Statistical Methods for the Analysis of Lake Water Quality Trends.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 841-R-93-003. December.

CONTENTS: Basic statistics and statistical concepts—Descriptive statistics; Robustness, resistance, and influence; Hypothesis testing; Statistical methods for trend detection. Individual lake analysis—Summary statistics; Graphical analyses; Normality; Seasonality; Independence; Trend detection in total phosphorus; Total nitrogen; Conclusions; Regional and statewide lake analysis. Appendices—Basic descriptive statistics; Introduction to SAS macros; SAS tables; List of SAS programs and files on disk.

MAIN FOCUS: Guides lake program managers on the application and interpretation of methods of trend detection in lake water quality. Provides step-by-step descriptions of statistical methods, as well as graphs and figures demonstrating analyses.

This manual is a technical supplement to *The Lake and Reservoir Restoration Guidance Manual.* This document and the accompanying software in the SAS system present nonparametric statistical methods for trend assessment in water quality, emphasizing lakes. The purpose of the document is to provide lake program managers with guidance on the application and interpretation of methods for the detection of trends in lake water quality. This manual discusses basic statistical concepts and approaches in applied statistics that are relevant to trend detection, before introducing the procedures and tests. Numerous graphs and figures demonstrate analyses done with actual data.

29. *Monitoring Lake and Reservoir Restoration.* USEPA 1990.

REFERENCE: USEPA. 1990. *Monitoring Lake and Reservoir Restoration.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 440/4-90-007. August.

CONTENTS: Planning the monitoring program; Monitoring methods; Watershed monitoring; In-lake restoration techniques and monitoring; A long-term monitoring protocol; Case study: Detection of trends and sampling strategy evaluations.

MAIN FOCUS: Three levels of watershed monitoring are discussed. Complete discussions of lake and reservoir restoration objectives and the methods used to achieve the objectives are provided.

This manual is a technical supplement to *The Lake and Reservoir Restoration Guidance Manual* (USEPA, 1988). It provides guidance for both design and implementation of a monitoring program by outlining specific standards for specific types of lake restoration and protection projects. It is intended to guide monitoring carried out under the Clean Lakes Program in connection with the Phase II or implementation portion of a lake restoration project. Phase I or diagnostic/feasibility monitoring is more exploratory in nature and more generic in terms of parameters used, and it is not directly addressed in this guidance. The primary users of this guidance are expected to be Regional EPA Clean Lakes project officers, state and local project managers, and project sponsors and consultants.

Procedures for performing in-lake sampling, measuring stream flow, handling and preserving samples, and analyzing data are described. Three levels of watershed monitoring are described: watershed inventories, limited stream monitoring, and comprehensive watershed monitoring. Control techniques for four lake restoration objectives are described. The four objectives are the control of nuisance algae, an increase in depth, the control of nuisance plants, and the mitigation of acidic conditions. Where appropriate, monitoring parameters for both during and after the treatments are tabulated and discussed. Formulas for the calculation of monitoring parameters are provided. Guidance is also provided for implementing a long-term monitoring protocol.

A.3.4 Guidances for Watershed Monitoring

30. *Monitoring Primer for Range Watersheds*. Bedill and Buckhouse Draft.

REFERENCE: Not available.

CONTENTS: Basis and rationale for monitoring in range watersheds; Water quality in the context of range watersheds; Some new concepts to go with the current ones; Environmental factors and range watersheds; Management of range watersheds which directly affects water quality; Monitoring methods and measurements.

MAIN FOCUS: Provides a good introduction to monitoring terms and concepts. Although the document is directed toward range monitoring, the terms and concepts are generally applicable to all monitoring.

This short document provides a general introduction to monitoring in the context of rangeland management. Watersheds are defined, their characteristics and functions are described, and the rationale behind their being the unit for rangeland management and monitoring is discussed. The focus of the document is on monitoring riparian areas and vegetation, not upland areas or streams. A general discussion of the concepts and steps involved in designing and implementing a rangeland monitoring program is provided. Detailed information on designing a monitoring program, selecting monitoring parameters and protocols, sampling procedures, and data analysis is not provided.

31. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 Years of Controlling Agricultural Nonpoint Source Pollution: The RCWP Experience*. USEPA 1992.

REFERENCE: USEPA. 1992. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 Years of Controlling Agricultural Nonpoint Source Pollution: The RCWP Experience*. September 13-17, 1992. U.S. Environmental Protection Agency, Office of Research and Development, Office of Water, Washington, DC. EPA/625/R-92/006. August.

CONTENTS: Water quality and land treatment monitoring; Relating water quality to land treatment; Land treatment and operation and maintenance of BMPs; Project coordination and farmer participation; Institutional arrangements, program administration, and project spin-offs; Information and education; Socioeconomics, technology transfer, lessons learned; Research needs and future vision; Additional information.

MAIN FOCUS: Reports of hands-on experiences encountered during nonpoint source and watershed project implementation. Since the RCWP had projects in states from most regions of the United States, this document contains specific information relevant to a variety of circumstances.

This symposium proceedings is intended to provide guidance for state nonpoint source programs and local watershed projects. It is the result of 10 years of experience of the National Rural Clean Water Program, and the papers in this document address both the successes and difficulties experienced in the 21 projects that composed the program. The papers included in the proceedings were peer reviewed. Most of the papers address individual projects of the RCWP, and they provide valuable insights into the variety of approaches and solutions for addressing specific circumstances and obstacles that may be encountered during the implementation of a nonpoint source or watershed program. In this respect the information in this document goes beyond the general "how-to" information provided in most guidance documents.

32. *Watershed Monitoring Manual.* Kansas Biological Survey 1993.

REFERENCE: Kansas Biological Survey. 1993. *Watershed monitoring manual.* Ecotoxicology Program, Kansas Biological Survey, University of Kansas, Lawrence, Kansas. July.

CONTENTS: Introduction and background information; Chemical measurements: Grab samples, *In situ* water quality analyses, Local water quality analyses; Physical measurements: Habitat, Stream discharge; Biological measurements: Macroinvertebrate, Fish, Drift sampling, Primary productivity; Biological laboratory procedures; Appendices: Preparation of equipment and supplies; Water quality glassware and plasticware washing procedure; Glass fiber filter preparation.

MAIN FOCUS: Detailed instructions of methods for stream monitoring.

The manual is a guide for field crews undertaking long-term watershed monitoring. The manual describes all field procedures and local laboratory procedures to be undertaken for the measurement of chemical, physical, and biological parameters. The manual is geared toward a specific project, but the procedures can serve as a guide for monitoring streams for other projects. Each section (on chemical, physical, and biological measurements) is presented as separate procedural descriptions of the various tasks to be performed. Each procedure is presented with an objective statement, listing of necessary equipment, and detailed description of the sampling procedure to follow. Reference sections list the literature from which the procedures are taken. *[From document introduction]*

A.3.5 Guidances for Ground Water Monitoring

33. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* USEPA 1991.

REFERENCE: USEPA. 1991. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 570/9-91-010. April.

CONTENTS: Methods for measuring or estimating nonpoint source contaminated ground-water discharge to surface water; The impact of nonpoint source contaminated ground-water discharge to surface water in water quality-limited water bodies: determining total maximum daily load and waste load allocations.

MAIN FOCUS: Each method is thoroughly described, including any limitations and assumptions in its use, the expertise needed to apply it, and data inputs and data outputs. A short evaluation of the method and relevant references with full citations are then provided.

This document is a summary of methods that have been applied to measure or estimate nonpoint-source-contaminated ground-water discharge to surface water. An overview of methods is presented, but this guidance is not a manual for employing the methods described. After the review of analytical methods, a separate chapter presents an overview of the total maximum daily load assessment and waste load allocation processes and discusses the applicability of the methods described. Some of the methods reviewed are the use of seepage meters, geophysical techniques, numerical models, and isotope methods.

An annotated bibliography is available as a companion volume to this report. The papers that provided the background for this guidance are referenced throughout and are abstracted in the annotated bibliography. Full citations for each of the background papers are provided in this report.

34. *Ground Water Monitoring: A Guide to Monitoring for Agricultural Nonpoint Source Pollution Projects.* Goodman, German, and Bishoff 1995.

REFERENCE: Goodman, J., D. German, and J. Bishoff. 1995. *Ground Water Monitoring: A Guide to Monitoring for Agricultural Nonpoint Source Pollution Projects.* Draft. February.

CONTENTS: Rural Clean Water Program background; Nonpoint source project process; Vadose zone monitoring. Ground water monitoring—Monitoring project planning; Monitoring system development; Monitoring system implementation; Quality assurance/quality control; Data management and evaluation; New developments in ground water monitoring. Ground water/surface water monitoring.

MAIN FOCUS: Geared mainly toward hydrologic systems and geologic regimes of the glaciated regions of the upper midwestern area of the United States. Includes detailed considerations and steps to take in nearly all of the content areas.

This document provides a usable "how to" handbook for designing and operating a ground water monitoring network for nonpoint source pollution control projects. It is the product of a comprehensive ground water monitoring and evaluation project that was conducted in the Oakwood Lakes-Poinsett area of east-central South Dakota under the Rural Clean Water Program (RCWP). This document is not intended to be an all-encompassing guidebook, but it does document the successes and failures in the South Dakota RCWP project and other RCWP projects that conducted ground water monitoring.

35. *Response to Ground Water Contaminants Detected through Idaho's Statewide Ambient Ground Water Quality Monitoring Program.* Idaho Department of Health and Welfare 1995.

This document provides guidance to the IDDEQ for response to ground water quality contaminant detections reported by the Idaho Statewide Ambient Ground Water Quality Monitoring Program. The objective of the protocol is to describe how IDDEQ can investigate a contaminant detected through statewide monitoring and to establish consistent guidance for the development of information to be provided to IDDEQ for the determination of an applicable response. A phased approach for characterizing the nature and extent of reported contamination is taken in the protocol, with separate phases to investigate whether the contaminant detected is associated with a known problem or existing project; persistent or non-persistent; isolated, localized, or regional; and whether it can be associated with a suspected source.

REFERENCE: Cardwell, J. 1995. *Response to ground water contaminants detected through Idaho's statewide ambient ground water quality monitoring program.* Water Quality Monitoring Protocols - Report No. 9. Idaho Department of Health and Welfare, Division of Environmental Quality, Monitoring and Technical Support Bureau, Boise, Idaho. January.

CONTENTS: Introduction; Notification procedure; Protocol; Quality assurance project plans; Data management. Appendices (partial listing): Selected ground water quality information sources listing; Selected pollutants measured and their most commonly associated uses; Selected analytes measured and their most commonly associated uses.

MAIN FOCUS: Investigation and reporting protocols for reports of ground water contamination in Idaho.

A.3.6 Guidances for Biological Monitoring

36. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters.* USEPA 1993.

REFERENCE: USEPA. 1993. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters.* U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/600/R-92/111. March.

CONTENTS: Quality assurance and quality control; Safety and health; Sample collection for analysis of the structure and function of fish communities; Specimen processing techniques; Sample analysis techniques; Special techniques; Fish bioassessment protocols for use in streams and rivers; Family-level ichthyoplankton index methods; Fish health and condition assessment profile methods; Guidelines for fish sampling and tissue preparation for bioaccumulative contaminants; Fisheries bibliography.

MAIN FOCUS: Provides guidance for all methods of fish collection, from netting to electrofishing, and rapid bioassessment protocols. A sample analysis section discusses fish identification and how specimens are measured and weighed properly for data analysis purposes. Separate sections discuss fish kill investigations and marking and tagging techniques. The fisheries bibliography is comprehensive.

This document describes guidelines and standardized procedures for the use of fish in evaluating the biological integrity of surface waters, and it provides biomonitoring programs with fisheries methods for measuring the status and trends of environmental pollution. Separate chapters in the document describe a variety of fish collection methods, including the use of nets, electricity, chemicals, and hook and line; specimen handling; specimen analysis; and methods to calculate the age of fish. The use of fish for rapid bioassessments of habitat and water quality impacts to fish populations is also discussed. A section on special techniques discusses flesh tainting methodology (used to relate flavor impairment to a particular waste source), fish kill investigations, and Instream Flow Incremental Methodology (IFIM), which measures impacts to fish and other aquatic organisms resulting from changes in instream flow. An extensive bibliography that is organized by topic is provided. It includes a section on fish identification, with the references separated by geographic region of the United States and by marine and freshwater species.

37. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* USEPA 1987.

REFERENCE: USEPA. 1987. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* Volume I. U.S. Environmental Protection Agency, Office of Marine and Estuarine Protection, Washington, DC. EPA 430/9-86-005. March.

CONTENTS: Recommended target species; Additional sampling considerations; Historical data for target species; Summary of recommendations.

MAIN FOCUS: Specifically addresses monitoring fish and macroinvertebrates for bioaccumulation of toxic substances. A separate section discusses the use of historic data, and data summaries of metals and priority pollutant concentrations for target species are provided for reference.

This guidance is intended for 301(h) programs, but the information presented is applicable to bioaccumulation monitoring in general. Guidance for the selection of target species for bioaccumulation monitoring is its main focus. A compilation, evaluation, and summarization of 1987 data on concentrations of priority pollutants in the suggested target species is included. This information provides a set of data for comparative purposes, to aid the user in interpreting data. The document explains the ranking procedure and criteria used to select the target species. Selected target species are tabulated geographically. Species of fish are selected for the geographic areas of Massachusetts to Virginia, and California and Washington, and macroinvertebrates are selected for Massachusetts to Virginia; Alaska to California; Florida, the Virgin Islands, and Puerto Rico; and Hawaii. Information on the types of tissue to analyze, the time of sampling, and the use of historical data is provided as well.

38. *Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents.* USEPA 1973.

REFERENCE: USEPA. 1973. *Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents.* C.I. Weber, ed. U.S. Environmental Protection Agency, National Environmental Research Center, Office of Research and Development, Cincinnati, Ohio. Program Element 1BA027. EPA-670/4-73-001. July.

CONTENTS: Biometrics; Plankton; Periphyton; Macrophyton; Macroinvertebrates; Fish; Bioassay; Appendix.

MAIN FOCUS: A section on biometrics provides a complete discussion of data analysis. Bioassays for phytoplankton/algae, periphyton, macroinvertebrates, and fish are discussed.

This manual was published to provide pollution biologists with a methods reference guide for measuring the effects of environmental contaminants on freshwater and marine organisms. Both field and laboratory methods are discussed for fish, macroinvertebrates, plankton, periphyton, and macrophyton. A section on biometrics provides a full discussion of sampling (simple random and stratified random) and statistical analysis methods (T-test, chi square, F-test, analysis of variance, confidence intervals, and linear regression). Sections on different types of organisms to be sampled (e.g., fish, periphyton) discuss sample collection and preservation, sample preparation and analysis, sampling methods, and special techniques where appropriate. References are provided for each section of the manual. Special sections on fathead minnow and brook trout chronic tests are included in a

section on bioassay techniques. An appendix contains data recording sheets and a discussion of equipment and supplies.

A.3.7 Program-specific Monitoring Guidances

39. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects*. USEPA 1991.

> **REFERENCE:** USEPA. 1991. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. August.
>
> **CONTENTS:** Selection criteria for National Monitoring Program projects; NonPoint Source Management System (NPSMS) software; Management file; Monitoring plan file; Annual report file.
>
> **MAIN FOCUS:** Outlines the types of data that should be collected and documentation that should be kept for watershed monitoring projects; specifically addresses the requirements of the National Monitoring Program established pursuant to CWA Section 319.

Under section 319 of the Clean Water Act as amended in 1987, EPA is establishing a national program to intensively monitor and evaluate a subset of watershed projects. A nationally consistent protocol is to be followed by the projects, and EPA has developed and distributed a national framework for the National Monitoring Program. This guidance provides monitoring and reporting guidelines for the program.

Nonpoint Source Management System (NPSMS) software has been developed and distributed to states that have received grants under section 319 of the Clean Water Act. The software facilitates information tracking and reporting under the National Monitoring Program. It is menu-driven, and this document discusses the proper entry of data into the software system and provides a step-by-step guide to it. Much of the information presented is therefore not of relevance to the reader interested in designing a monitoring program, though there is limited discussion of monitoring objectives, monitoring program design, and monitoring parameters.

40. *NPDES Storm Water Sampling Guidance Document*. USEPA 1992.

> **REFERENCE:** USEPA. 1992. *NPDES Storm Water Sampling Guidance Document*. U.S. Environmental Protection Agency, Office of Water. Washington, DC. EPA 833-B-92-001. July.
>
> **CONTENTS:** Background for storm water sampling; Fundamentals of sampling; Analytical considerations; Flexibility in sampling; Health and safety.
>
> **MAIN FOCUS:** Contains a good discussion of grab and composite samples, sample collection methods, and sample handling and documentation. Addresses many points specific to the NPDES program.

This guidance is intended for operators of facilities that discharge storm water containing industrial pollutants and operators of large and medium-sized municipal separate storm sewer systems. Its purpose is to assist facility operators and/or owners in planning for and fulfilling the NPDES storm water discharge sampling requirements for NPDES permit applications. The information presented pertains to individual industrial storm water

applications, group storm water applications, and municipal storm water permit applications. The guidance was issued in support of EPA regulations and policy initiatives involving the development and implementation of a national storm water program, and serves as Agency guidance.

The legal requirements of storm water sampling under the Clean Water Act are explained, and storm water sampling methodologies, other measurements necessary for permit compliance (e.g., flow, rainfall), sample documentation, and sample analysis requirements are discussed. Modifications to standard sampling procedures, which are allowed under specific circumstances on a case-by-case basis, are explained. Acceptable techniques for manual and automatic sample collection are described. Health and safety considerations are also discussed. NPDES storm water permit requirements are the focus of this guidance, and much of the information provided in it is not directly related to monitoring the water quality and habitat effects of NPS pollution. However, the technical information that it contains on choosing sampling locations and sampling procedures should be useful to those monitoring storm water as part of an NPS monitoring program.

41. *Monitoring Guidance for the National Estuary Program.* USEPA 1991.

REFERENCE: USEPA. 1991. *Monitoring Guidance for the National Estuary Program.* U.S. Environmental Protection Agency, Office of Water. EPA 503/8-91-002. August.

CONTENTS: Develop monitoring objectives and performance objectives; Establish testable hypotheses and select statistical methods; Select analytical methods and alternative sampling designs; Evaluate monitoring program performance; Implement monitoring study and data analysis; Communicate program results; Appendices: Case studies; Methods.

MAIN FOCUS: Appendices contain case studies of the Puget Sound and Chesapeake Bay monitoring programs, and detailed methods sections. The methods sections include water column physical and chemical parameters, sediment, plankton, aquatic vegetation, benthos, fish, bioaccumulation, and bacteria and viruses. Each method section discusses monitoring design, analytical methods, QA/QC, and statistics for the parameter.

The National Estuary Program (NEP) was created by the Water Quality Act of 1987 to promote long-term planning and management in nationally significant estuaries threatened by pollution, development, or overuse. Management conferences, with representatives from EPA, the affected state(s), local governments, the scientific community, and citizens' groups, are established to develop Comprehensive Conservation and Management Plans (CCMPs) for the estuaries. The first task of a management conference is to identify and characterize the problems in the estuary. Then, based on the findings, a CCMP is developed to guide the implementation of actions undertaken to overcome the identified problems and protect the estuarine environment. A requirement of the enacting legislation is that the effectiveness of actions taken pursuant to CCMPs be monitored, and this document provides guidance on the design, implementation, and evaluation of NEP monitoring programs.

NEP monitoring programs are designed to serve two goals—to measure the effectiveness of management actions and programs implemented under CCMPs, and to provide essential information that can be used to redirect and refocus the estuarine management efforts. Because the intended audience for this document is those involved in estuary management efforts, including environmental managers, governmental agencies, and citizens, this guidance discusses both background issues and technical aspects relevant to estuarine monitoring programs.

This guidance presents a systems design approach to designing a monitoring program, with discussions of each of the steps involved in the approach: developing monitoring program objectives, designing a monitoring program, establishing hypotheses, selecting statistical methods and sampling designs, evaluating the monitoring program's

performance, and managing and analyzing data. An extensive methods section is organized according to the parameter being monitored, e.g., water column chemistry, sediment grain size, aquatic vegetation, fish community structure, viral pathogens. Numerous references to other texts providing in-depth discussions of each step in monitoring program design are provided. Two case studies, the Puget Sound Ambient Monitoring Program and the Chesapeake Bay Monitoring Program, are used to provide examples from existing estuarine monitoring programs. In addition, the case studies also address options for funding monitoring programs, how to incorporate existing monitoring studies into a coordinated basin-wide monitoring effort, and methods for determining the effectiveness and feasibility of monitoring efforts.

42. *Water Quality Standards Handbook.* USEPA 1983.

REFERENCE: USEPA. 1983. *Water Quality Standards Handbook.* U.S. Environmental Protection Agency, Office of Water Regulations and Standards. Washington, DC. December.

CONTENTS: Water quality standards review and revision process; General program guidance; Water body survey and assessment guidance for conducting use attainability analyses; guidelines for deriving site-specific water quality criteria. Appendices: Bioassay test methods; Determination of statistically significantly different LC50 values; Case studies.

MAIN FOCUS: Provides a complete discussion of the state water quality standards development process and the process for conducting waterbody surveys and use attainability analyses. Excellent background material on state water quality standards and stream classification systems.

This guidance is meant to assist states in implementing the 1983 water quality standards regulation (48 FR 51400, November 8). It is not a monitoring guidance per se, but rather a guidance for determining whether awaterbody survey and/or use attainability analysis, as required by the Clean Water Act, meets specifications set forth by EPA. Numerous case studies illustrate acceptable state approaches.

The handbook provides a general description of the standards setting process, information on program administrative policies and procedures, and a description of analyses used to determine appropriate water uses and criteria. States are to use the data and analyses set forth in this document, or similar data and analyses, to conduct use attainability analyses or to establish water quality criteria. Certain regulatory requirements to which states must adhere when they develop water quality criteria are discussed as well. EPA has determined that certain types of scientific and technical data and analyses are necessary in order for the public and EPA to conduct informed reviews of proposed water quality standards. Data and analyses required for this purpose are noted. Explanations of terms and concepts used in the regulatory language, e.g., mixing zone and antidegradation, are provided.

While this document does not provide guidance on the development or implementation of a monitoring program, it is a useful reference for those interested in the rationale behind and derivation of water quality standards. Monitoring is an essential step in setting and modifying site-specific water quality standards. Therefore, this document is valuable as a companion to documents that deal more directly with monitoring program design and implementation.

43. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* USEPA 1989.

REFERENCE: USEPA. 1989. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* W. Warren-Hicks, B.R. Parkhurst, and S.S. Baker, Jr., eds. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/600/3-89/013. March.

CONTENTS: Ecological endpoints; Assessment strategies and approaches; Field sampling design; Quality assurance and data quality objective; Toxicity tests; Biomarkers; Field assessments; Data interpretation.

MAIN FOCUS: A comprehensive discussion of all aspects of ecological risk assessment. A specific list of references follows each subsection that discusses a technical aspect of risk assessment.

This is a comprehensive field and laboratory reference document for the design, implementation, and interpretation of ecological risk assessments, and it specifically addresses assessments of hazardous waste sites. Complete discussions of ecological endpoints, assessment methods, statistical considerations, toxicity testing, field assessments, and data interpretation are provided. The section on toxicity testing is divided into subsections on aquatic, terrestrial, and microbial tests; the section on field assessments is divided into subsections on aquatic, vegetation, terrestrial vertebrates, and terrestrial invertebrates.

44. *Environmental Monitoring and Assessment Program: Ecological Indicators.* USEPA 1990.

REFERENCE: USEPA. 1990. *Environmental Monitoring and Assessment Program: Ecological Indicators.* U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/600/3-90/060. September.

CONTENTS: EMAP indicator concepts; Indicator strategy for near coastal waters; Indicator strategy for inland surface waters; Indicator strategy for wetlands; Indicator strategy for forests; Indicator strategy for arid lands; Indicator strategy for agroecosystems; Indicators relevant to multiple resource categories; Indicator strategy for atmospheric stressors; Conclusions and future directions; Appendices (indicator fact sheets).

MAIN FOCUS: EMAP indicators are related to ecological and social (e.g., recreation) variables and to monitoring objectives. The reasoning behind the choice of each indicator is fully explained.

The Environmental Monitoring and Assessment Program (EMAP) is a nationwide initiative to assess and document the status and trends in the condition of the Nation's ecological resources. The program is organized into seven resource groups that focus monitoring and assessment activities on defining the environmental condition of agricultural lands, estuaries, forests, the Great Lakes, lakes and streams, rangelands, and heterogeneous ecological resources (landscape ecology). Regional EMAP (REMAP) projects that use the same approach as the nationwide EMAP program focus on issues of particular importance in the region or a state or states within the region.

This report presents the approach proposed to describe ecological condition; defines a common strategy within the program for selecting and prioritizing; and summarizes the indicators chosen for evaluation as core indicators for major types of ecosystems. The discussions in this document of monitoring indicators for the major types of ecosystems provide valuable information on types of indicators to choose for different monitoring goals.

Indicators that were considered but not chosen for use in EMAP are also discussed, and this is valuable in that it points out the purposes for which these monitoring parameters would be inappropriate. Fact sheets for each of the indicators chosen for EMAP are included as appendices. Each fact sheet discusses the particular indicator's application, measurement, variability, and primary problems, and provides pertinent references.

45. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document*. USEPA 1986.

REFERENCE: USEPA. 1986. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document*. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC. OSWER-9950.1. September.

CONTENTS: Characterization of site hydrogeology; Placement of detection monitoring wells; Monitoring well design and construction; Sampling and analysis; Statistical analysis of detection monitoring data; Assessment monitoring.

MAIN FOCUS: Provides technical information on site characterization, well placement, and well design and construction. Sample collection, handling, and preservation techniques are thoroughly discussed.

This guidance describes in detail what EPA deems to be the essential components of a ground water monitoring system that meets the goals of the Resource Conservation and Recovery Act (RCRA). It is intended to be used by enforcement officials, permit writers, field inspectors, and attorneys at the federal and state levels to assist them in making informed decisions regarding the adequacy of existing or proposed ground water monitoring systems.

The guidance contains technical information on site characterization, well design and construction, and assessment of contamination of ground water. Hydrogeologic regimes vary widely from site to site, and this guidance does not attempt to address all possible circumstances for the purposes of ground water monitoring programs. It does provide a framework within which a decision-making process can be applied using site-specific considerations.

Ground water monitoring is a specific type of monitoring and may be beyond the scope of many monitoring programs. Although this guidance is specific to the RCRA program, the protocols presented are rigorous and could be used to provide defensible data for any ground water monitoring program.

46. *Summary of U.S. EPA-Approved Methods, Standard Methods, and Other Guidance for 301(h) Monitoring Variables*. USEPA 1985.

REFERENCE: USEPA. 1985. *Summary of U.S. EPA-Approved Methods, Standard Methods, and Other Guidance for 301(h) Monitoring Variables*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA 503/4-90-002. September.

CONTENTS: Introduction; Table 1: U.S. EPA-Approved Methods and Guidance Documents for Measuring Biological, Sediment, and Water Quality Variables in 301(h) Monitoring Programs; Water Quality Variables; Sediment Analyses; Biological Variables; References.

MAIN FOCUS: A convenient listing of where to find information on analytical techniques for many monitoring variables.

This short (16-page) document tabulates the types of analytical methods available for the water quality, sediment, and biological variables used in 301(h) monitoring programs. The methods are listed as EPA-approved, EPA-suggested, standard, or additional, and the availability of guidance for each method is noted. Following the table is a brief note specifying method numbers (e.g., USEPA method No. 150.1; additional procedure No. 413.2) for each of the variables listed. A list of references at the end of the document indicates where to find information about each of the monitoring variables and analytical methods.

47. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance.* USEPA 1989.

REFERENCE: USEPA. 1989. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance.* U.S. Environmental Protection Agency, Office of Solid Waste, Waste Management Division, Washington, DC. EPA/530-SW-89-026. NTIS PB89-151047. February.

CONTENTS: Regulatory overview; Choosing a sampling interval; Choosing a statistical method; Background well to compliance well comparisons; Comparisons with MCLs or ACLs; Control charts for intra-well comparisons; Miscellaneous topics; Appendices.

MAIN FOCUS: Detailed discussion of all aspects of statistical analyses of ground-water monitoring data. Flow charts are provided to assist the reader in choosing the proper statistical method and interpreting data.

The hazardous waste regulations under the Resource Conservation and Recovery Act (RCRA) require owners and operators of hazardous waste facilities to use design features and control measures that prevent the release of hazardous waste into ground water. This document provides guidance to RCRA facility permit applicants and writers concerning the statistical analysis of ground water monitoring data at RCRA facilities. Sections of the document provide an overview of regulations concerning the statistical analyses of ground water monitoring data, hydrogeologic parameters to consider when choosing a sampling interval, guidance on choosing an appropriate statistical method, statistical methods that may be used to evaluate ground water monitoring data, statistical procedures that are appropriate for special circumstances, and special topics. Appendices cover general statistical considerations, a glossary of statistical terms, statistical tables, and references.

48. *CWA Section 403: Procedural and Monitoring Guidance.* USEPA 1994.

REFERENCE: USEPA. 1994. *CWA Section 403: Procedural and Monitoring Guidance.* U.S. Environmental Protection Agency, Office of Water, Oceans and Coastal Protection Division, Washington, DC. EPA 842-B-94-003. March.

CONTENTS: Section 403 procedure; Options for monitoring under the basis of "no irreparable harm"; Summary of monitoring methods: physical characteristics, water chemistry, sediment chemistry, sediment grain size, benthic community structure, fish and shellfish pathology, fish populations, plankton, habitat identification methods, bioaccumulation, pathogens, effluent characterization, mesocosms and microcosms.

MAIN FOCUS: Each monitoring parameter is dealt with separately, with a separate discussion of monitoring design, analytical methods, QA/QC, statistical considerations, and use of data. An appendix contains an extensive list of monitoring methods references.

This document is designed to provide EPA Regions and NPDES-authorized states with a framework for the decision-making process for section 403 (ocean dumping) evaluations and to provide guidance on the type and level of monitoring that should be required as part of permit issuance under the "no irreparable harm" provisions of section 403. Options for monitoring under the basis of no irreparable harm, including criteria for evaluating perceived potential impact and establishing monitoring requirements to assess actual impacts, are discussed. Summaries of monitoring methods for evaluating numerous parameters (see contents listing) are provided. Each method section contains an explanation of the usefulness of the parameter of concern in a 403 monitoring program and a discussion of analytical methods, the use of data generated, and considerations of monitoring design, statistical design, and quality assurance/quality control.

49. *Methods for Collecting Benthic Invertebrate Samples as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

REFERENCE: Cuffney, T.F., M.E. Gurtz, and M.R. Meador. 1993. *Methods for Collecting Benthic Invertebrate Samples as part of the National Water Quality Assessment Program.* USGS Open-File Report 93-406. U.S. Geological Survey, Reston, Virginia.

CONTENT: National Water Quality Assessment Program sampling design; Sampling design for benthic invertebrates; Methods for collecting benthic invertebrates; Maintenance of sampling equipment; Sample processing and labeling; Field data sheets; Contract laboratories and the Biological Quality Assurance Unit; Adapting collection methods for other National Water Quality Assessment Program objectives; Safety and health.

MAIN FOCUS: Evaluating benthic invertebrate communities as part of the ecological survey component of the U.S. Geological Survey's National Water Quality Assessment Program.

This document presents a variety of sampling methods and procedures for collecting benthic invertebrates as part of the National Water Quality Assessment (NAWQA) Program. Numerous sample-collecting techniques, equipment, and data forms are presented for use at basic fixed sampling sites. Each technique or method is thoroughly explained with diagrams, flowcharts, or examples. These techniques and methods are easily adaptable for use in other components of the NAWQA Program, or where needed in other programs of the USGS's Water Resources Division. Field data sheets are provided.

The objectives of benthic invertebrate community characterizations are to develop for each site a list of taxa within the associated stream reach and determine the structure of the benthic invertebrate communities within selected habitats of that reach. This document presents the nationally consistent approach used by the USGS to achieve these objectives. It provides guidance on site, reach, and habitat selection and methods and equipment for qualitative multihabitat sampling and semi-quantitative single-habitat sampling. Appropriate quality assurance and quality control guidelines are used to maximize the ability to analyze data within and among study units.

50. *Methods for Collecting Algal Samples as Part of the National Water Quality Assessment Program.* Porter, Cuffney, Gurtz, and Meador 1993.

This document describes the sampling methods, procedures and equipment for collecting algal samples as part of the National Water Quality Assessment (NAWQA) Program. The approach used in the sampling design for algal communities provides a common spatial scale to assess biological communities and habitat characteristics. The design also addresses seasonal and hydrologic conditions that affect algal communities. The NAWQA Program provides an integrated assessment of water quality within selected environmental settings by collecting and analyzing a combination of physical, chemical, and biological data. The algal component of the data is designed

to characterize the species distribution and community structure of benthic algae and their relation to water quality.

REFERENCE: Porter, S.D., T.F. Cuffney, M.E. Gurtz, and M.R. Meador. 1993. *Methods for Collecting Algal Samples as part of the National Water Quality Assessment Program.* USGS Open-File Report 93-409. U.S. Geological Survey, Reston, Virginia.

CONTENT: National Water Quality Assessment Program sampling design; Sampling design for algal communities; Methods for collecting algal samples; Sample processing and labeling; Contract laboratories and the Biological Quality Assurance Unit.

MAIN FOCUS: Describes methods, procedures, and equipment for collecting algal samples at basic fixed sites as part of the U.S. Geological Survey's National Water Quality Assessment Program.

A variety of sample collection methods focus on qualitative multihabitat and quantitative targeted habitat periphyton samples. Forms for recording sampling data are provided. This manual also discusses preservation methods, preparation of subsamples, and labeling while processing algal samples.

51. *Methods for Characterizing Stream Habitat as Part of the National Water Quality Assessment Program.* Meador, Hupp, Cuffney, and Gurtz 1993.

REFERENCE: Meador, M.R., C.R. Hupp, T.F. Cuffney, and M.E. Gurtz. 1993. *Methods for Characterizing Stream Habitat as part of the National Water Quality Assessment Program.* USGS Open-File Report 93-408. U.S. Geological Survey, Reston, Virginia.

CONTENT: National Water Quality Assessment Program sampling design; Stream habitat sampling design; Methods for characterizing stream habitat.

MAIN FOCUS: Provides detailed procedures for characterizing stream habitat as part of the U.S. Geological Survey's National Water Quality Assessment Program.

This document provides detailed instructions for characterizing stream habitat as part of the National Water Quality Assessment (NAWQA) Program. These procedures allow for appropriate habitat descriptions and standardization of measurement techniques to create unbiased evaluations of habitat influences on water resource conditions at a variety of spatial levels. Using the methods presented, evaluation of stream habitat is based on a spatially hierarchical framework that incorporates habitat data at basin, segment, reach, and microhabitat scales. This framework provides a basis for national consistency in collection techniques while allowing flexibility in habitat assessment within individual study units. Procedures are described for collecting habitat data at basin and stream segment scales using geographic information system (GIS) databases, maps, and aerial photographs. Detailed diagrams and characterization forms are provided.

52. *Guidelines for the Processing and Quality Assurance of Benthic Invertebrate Samples Collected as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

REFERENCE: Cuffney, T.F., M.E. Gurtz, and M.R. Meador. 1993. *Guidelines for the Processing and Quality Assurance of Benthic Invertebrate Samples Collected as part of the National Water Quality Assessment Program.* USGS Open-File Report 93-407. U.S. Geological Survey, Reston, Virginia.

CONTENT: Sample processing strategy; Specifications for sample processing by contract laboratories; Biological quality assurance unit.

MAIN FOCUS: Specifies the procedures and guidelines that contract laboratories should use to identify and quantify invertebrates from the large-rare and main-body components of benthic invertebrate samples collected as part of the U.S. Geological Survey's National Water Quality Assessment Program.

This report provides nationally consistent guidelines and criteria for the processing of benthic invertebrate samples collected as part of the National Water Quality Assessment (NAWQA) Program. These guidelines include procedures for the tracking and labeling of samples, the use of standard methods and equipment for removing benthic invertebrates from the sample matrix, subsampling procedures, target levels of identification for major groups of invertebrates, quantification procedures, quality assurance/quality control procedures, and standard formats for reporting data. In addition, standards and procedures for the initial qualification and continual review of contract laboratories are discussed. Diagrams, flow charts, example data sheets, and examples of situations are used in the presentation of these guidelines.

53. *Methods for Sampling Fish Communities as a Part of the National Water Quality Assessment Program.* Meador, Cuffney, and Gurtz. 1993.

REFERENCE: Meador, M.R., T.F. Cuffney, and M.E. Gurtz. 1993. *Methods for Sampling Fish Communities as a part of the National Water Quality Assessment Program.* USGS Open-File Report 93-104. U.S. Geological Survey, Reston, Virginia.

CONTENT: National Water Quality Assessment sampling design; Fish community sampling design; Fish community sampling considerations; Methods for sampling fish communities; Biological Quality Assurance Unit; Field data sheets.

MAIN FOCUS: Provides detailed procedures for use by biologists to evaluate stream fish communities as part of the U.S. Geological Survey's National Water Quality Assessment Program.

This manual provides complete instructions for procedures used to sample and evaluate stream fish communities as part of the National Water Quality Assessment (NAWQA) Program. The methods allow standardization of collection methods and descriptions of fish communities to facilitate unbiased evaluations of the relationships between physical, chemical, and biological components of water quality conditions. The methods are established standard procedures for characterizing fish communities in streams ranging from headwaters to large rivers. The focus of the sampling procedures is electrofishing and seining techniques, but other methods are mentioned. Taxonomic identification, physical measurements, examination of fish for external anomalies, and preservation of specimens are covered in the discussion of sample processing. Forms for recording these data are provided.

54. *Guidelines for Studies of Contaminants in Biological Tissues for the National Water Quality Assessment Program.* Crawford and Luoma 1994.

REFERENCE: Crawford, J.K., and S.N. Luoma. 1994. *Guidelines for Studies of Contaminants in Biological Tissues for the National Water Quality Assessment Program.* USGS Open-File Report 92-494. U.S. Geological Survey, Reston, Virginia.

CONTENT: Tissue analysis in the National Water Quality Assessment Program. Tissue analysis activities of other agencies—USFWS, NOAA, EPA National Study of Chemical Residues in Fish, EPA EMAP, Regional and state. Approach of tissue analysis surveys—General approach; Objectives, priorities, and time line; Study strategies; Factors affecting selection of target chemicals; Target compounds for contaminant occurrence; Targeted chemical analyses following the contaminant-occurrence survey; Selection of taxa for analysis; Field procedures for collecting and processing tissue samples; Field records; Laboratory procedures for analyzing tissue samples; Quality assurance and quality control; Voucher collections and sample archival; Data management; Data interpretation.

MAIN FOCUS: Describes the concepts and field methods to be used by the U.S. Geological Survey's National Water Quality Assessment Program for evaluating contaminants in tissues of organisms.

This document presents the rationale, objectives, approach, and procedures to be used in the National Water Quality Assessment (NAWQA) Program for determining the occurrence, distribution, and trends in concentrations of trace elements and synthetic organic compounds in tissues. Part 1 of this manual describes the rationale for analyzing tissue contaminants and overviews the approach used. Part 2 explains the tissue-contaminant strategies of other agencies and compares them to the strategy used in the NAWQA Program. Part 3 discusses the approach for the use of tissue analysis as an aid to interpret water quality in NAWQA study units. Also, suggestions for interpretation of data are offered to facilitate consistency between study units.

Table A.1. Summary of guidances reviewed.

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Guidance for Monitoring:																		
Streams & Rivers	•	•	•		•	•			•	•	•	•	•		•	•	•	•
Lakes & Reservoirs	•	•	•		•				•	•	•	•	•					
Estuaries/Marine								•	•	•	•	•						
Ground Water	•				•						•	•						
Storm Water																		
Land Use	•					•	•					•						
Land Management	•					•												
Forest				•		•						•						
Range																		
Agriculture		•		•			•					•						
Watersheds	•		•				•					•						
Program Elements:																		
Objectives	•		•	•	•	•	•	•	•		•	•	•			•	•	
Design	•		•	•	•	•	•		•	•	•	•			•	•	•	•
Protocols	•			•		•				•	•	•				•	•	•
Parameters	•	•		•	•	•				•	•	•				•	•	•
Related Subjects:																		
Data Analysis	•		•			•			•							•	•	•
Data Management	•		•	•			•		•		•	•	•				•	
QA/QC			•	•	•	•		•	•		•	•	•				•	
Stream Classification																•		

(Table continued on next page)

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Samples and Sampling:																		
Types of Samples	•				•			•							•		•	•
Sampling Equipment	•									•								•
Sampling Techniques					•			•		•							•	•
Sample Handling	•							•		•							•	•
Sample Documentation								•									•	
Sample Analysis	•							•		•							•	
Sampling Locations	•		•		•			•		•					•	•	•	•
Sampling Frequency	•			•	•			•							•	•	•	
Health and Safety										•								
Monitoring Parameters:																		
Physical Constituents			•	•	•	•		•		•					•	•	•	
Chemical Constituents		•			•	•		•		•							•	
Nutrients		•	•	•	•			•		•					•	•		
Sediment	•	•	•							•		•			•	•	•	
Bacteria & Viruses		•	•		•			•		•		•			•	•		
Phytoplankton	•	•	•		•			•		•								
Zooplankton	•		•		•	•				•		•						
Macroinvertebrates	•	•	•			•				•		•			•		•	•
Fish	•	•	•			•				•					•		•	
Aquatic Vegetation	•	•	•			•		•		•		•			•		•	
Riparian Vegetation			•							•		•			•	•	•	
Habitat Quality			•			•										•	•	
Flow	•											•			•		•	
Channel Morphology						•						•			•	•	•	
Bank Stability						•										•	•	

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Guidance for Monitoring:																		
Streams & Rivers	•	•	•	•	•	•		•			•		•	•				•
Lakes & Reservoirs								•	•	•	•		•					•
Estuaries/Marine								•			•							•
Ground Water													•		•	•	•	
Storm Water																		
Land Use													•			•		
Land Management													•					
Forest																		
Range												•						
Agriculture													•			•		
Watersheds							•				•	•	•	•		•		
Program Elements:																		
Objectives		•		•	•	•			•		•	•						•
Design		•		•	•	•					•	•	•			•	•	•
Protocols	•	•				•					•					•		•
Parameters	•	•			•	•					•	•						•
Related Subjects:																		
Data Analysis	•						•			•	•							•
Data Management									•		•		•				•	•
QA/QC				•					•		•					•	•	•
Stream Classification				•														

(Table continued on next page)

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Samples and Sampling:																		
Types of Samples	•				•		•		•		•			•	•			•
Sampling Equipment														•		•		
Sampling Techniques		•				•			•		•			•	•			•
Sample Handling											•			•				•
Sample Documentation	•										•			•				•
Sample Analysis											•			•				•
Sampling Locations	•	•			•	•	•		•		•			•		•		
Sampling Frequency		•			•	•	•		•		•			•		•		
Health and Safety																		•
Monitoring Parameters:																		
Physical Constituents		•		•				•	•		•			•				•
Chemical Constituents		•						•	•		•			•				•
Nutrients											•			•				
Sediment				•				•	•		•							
Bacteria & Viruses									•									
Phytoplankton		•	•						•		•			•				
Zooplankton																		
Macroinvertebrates			•	•				•						•				
Fish	•		•	•		•		•			•			•				•
Aquatic Vegetation		•							•		•							
Riparian Vegetation								•										
Habitat Quality			•	•				•						•				•
Flow		•		•				•			•			•				
Channel Morphology		•						•										
Bank Stability		•	•															

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Guidance for Monitoring:																		
Streams & Rivers		•	•			•	•	•					•	•	•	•	•	•
Lakes & Reservoirs		•	•			•		•										
Estuaries/Marine	•	•	•		•			•		•		•						
Ground Water							•		•		•							
Storm Water				•														
Land Use																		
Land Management																		
Forest								•										
Range								•										
Agriculture								•										
Watersheds								•										
Program Elements:																		
Objectives		•	•		•		•	•	•			•	•	•	•		•	•
Design		•			•		•		•		•	•	•	•	•		•	•
Protocols	•	•					•				•	•	•	•			•	•
Parameters	•	•					•	•	•			•						•
Related Subjects:																		
Data Analysis	•	•					•		•		•	•						•
Data Management											•							•
QA/QC					•		•		•			•				•	•	•
Stream Classification															•			

(Table continued on next page)

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																	
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Samples and Sampling:																		
Types of Samples	•	•		•			•				•	•	•	•				
Sampling Equipment													•	•			•	
Sampling Techniques		•					•		•			•	•	•			•	•
Sample Handling		•		•			•		•				•	•		•	•	•
Sample Documentation		•		•					•				•	•		•	•	•
Sample Analysis		•		•			•									•		•
Sampling Locations	•							•	•		•	•	•	•	•		•	•
Sampling Frequency	•						•	•	•		•	•	•	•			•	•
Health and Safety				•					•									
Monitoring Parameters:																		
Physical Constituents					•					•		•						
Chemical Constituents					•		•			•		•						
Nutrients					•					•								
Sediment					•					•		•						
Bacteria & Viruses					•					•		•						
Phytoplankton		•			•					•		•		•				
Zooplankton		•			•					•		•						
Macroinvertebrates	•	•			•					•		•	•					•
Fish	•	•			•					•		•					•	•
Aquatic Vegetation					•													•
Riparian Vegetation																		
Habitat Quality												•						
Flow					•							•						
Channel Morphology																		
Bank Stability																		

Index

1. *Water Quality Monitoring.* USDA-SCS Draft.
2. *Water Quality Indicators Guide.* USDA-NRCS 1997.
3. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring.* Coffey, Spooner and Smolen 1993.
4. *Coordinated Nonpoint Source Water Quality Monitoring Program for Idaho.* Idaho Department of Health and Welfare 1990.
5. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document.* Minnesota Pollution Control Agency 1989.
6. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards.* Dissmeyer 1994.
7. *Designing Effective Nonpoint Source Water Quality Monitoring Programs.* Maas 1989.
8. *Volunteer Estuary Monitoring: A Methods Manual.* USEPA 1993.
9. *Volunteer Water Monitoring: A Guide for State Managers.* USEPA 1990.
10. *The Volunteer Monitor.* The National Newsletter of Volunteer Water Quality Monitoring.
11. *The Strategy for Improving Water-Quality Monitoring in the United States.* Intergovernmental Task Force on Monitoring Water Quality 1995.
12. *The Strategy for Improving Water-Quality Monitoring in the United States. Technical Appendixes.* Intergovernmental Task Force on Monitoring Water Quality 1995.
13. *Guidance for State Water Monitoring and Wasteload Allocation Programs.* USEPA 1985.
14. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects.* USEPA 1981.
15. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* MacDonald, Smart and Wissmar 1991.
16. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams.* USEPA Draft.
17. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* USEPA 1989.
18. *An Improved Biotic Index of Organic Stream Pollution.* Hilsenhoff 1987.
19. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin.* Lyons 1992.
20. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects.* Simonson and Lyons 1992.
21. *1996 Water Body Assessment Guidance: A Stream to Standards Process.* Idaho Division of Environmental Quality 1996.
22. *1996 Beneficial Use Reconnaissance Project Workplan.* Idaho Division of Environmental Quality 1996.
23. *Handbook: Stream Sampling for Waste Load Allocation Applications.* USEPA 1986.
24. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects.* Simonson and Lyons 1992.

25. *Techniques for Detecting Effects of Urban and Rural Land-Use Practices on Stream-Water Chemistry in Selected Watersheds in Texas, Minnesota, and Illinois.* Walker 1993.

26. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses.* Volumes I-III. USEPA 1983-1984.

27. *Volunteer Lake Monitoring: A Methods Manual.* USEPA 1991.

28. *Statistical Methods for the Analysis of Lake Water Quality Trends.* USEPA 1993.

29. *Monitoring Lake and Reservoir Restoration.* USEPA 1990.

30. *Monitoring Primer for Range Watersheds.* Bedill and Buckhouse Draft.

31. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 Years of Controlling Agricultural Nonpoint Source Pollution: The RCWP Experience.* USEPA 1992.

32. *Watershed Monitoring Manual.* Kansas Biological Survey 1993.

33. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* USEPA 1991.

34. *Ground Water Monitoring: A Guide to Monitoring for Agricultural Nonpoint Source Pollution Projects.* Goodman, German, and Bishoff 1995.

35. *Response to Ground Water Contaminants Detected through Idaho's Statewide Ambient Ground Water Quality Monitoring Program.* Idaho Department of Health and Welfare 1995.

36. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters.* USEPA 1993.

37. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* Volume I. USEPA 1987.

38. *Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents.* USEPA 1973.

39. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects.* USEPA 1991.

40. *NPDES Storm Water Sampling Guidance Document.* USEPA 1992.

41. *Monitoring Guidance for the National Estuary Program.* USEPA 1991.

42. *Water Quality Standards Handbook.* USEPA 1983.

43. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* USEPA 1989.

44. *Environmental Monitoring and Assessment Program: Ecological Indicators.* USEPA 1990.

45. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document.* USEPA 1986.

46. *Summary of U.S. EPA-Approved Methods, Standard Methods, and Other Guidance for 301(h) Monitoring Variables.* USEPA 1985.

47. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance.* USEPA 1989.

48. *CWA Section 403: Procedural and Monitoring Guidance.* USEPA 1994.

49. *Methods for Collecting Benthic Invertebrate Samples as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

50. *Methods for Collecting Algal Samples as Part of the National Water Quality Assessment Program.* Porter, Cuffney, Gurtz, and Meador 1993.

51. *Methods for Characterizing Stream Habitat as Part of the National Water Quality Assessment Program.* Meador, Hupp, Cuffney, and Gurtz 1993.

52. *Guidelines for the Processing and Quality Assurance of Benthic Invertebrate Samples Collected as Part of the National Water Quality Assessment Program.* Cuffney, Gurtz, and Meador 1993.

53. *Methods for Sampling Fish Communities as a Part of the National Water Quality Assessment Program.* Meador, Cuffney, and Gurtz. 1993.

54. *Guidelines for Studies of Contaminants in Biological Tissues for the National Water Quality Assessment Program.* Crawford and Luoma 1994.

APPENDIX B. DATA SOURCES

The data sources and information on the sources provided in this appendix are current as of August 1996. Availability of the data should be investigated prior to development of a project data collection plan.

1. STORAGE AND RETRIEVAL (STORET) SYSTEM — USEPA/OWOW

The Storage and Retrieval (STORET) System is one of the oldest and largest water information systems currently in use. STORET stores information on ambient, intensive-survey effluent, and biological water quality monitoring information. Although most STORET information has been added since 1975, records go back to 1899. STORET has four information areas:

- Water Quality System: WQS, the main component of STORET, contains chemical and physical information obtained during monitoring of waterways within and contiguous to the United States. This includes information for estuaries, streams, lakes, rivers, ground water, canals, and coastal and international waters.

- Biological System Field Survey File: BIOS contains information on the distribution, abundance, and physical condition of aquatic organisms in waters within and contiguous to the United States, as well as descriptions of their habitats. BIOS provides a central repository for biological information and analytical tools for data analyses.

- Daily Flow File: DFF contains daily observations of stream flow and miscellaneous water quality parameters collected at gaging stations belonging to the U.S. Geological Survey's (USGS's) national network. The DFS contains essentially the same information as the USGS Daily Values File; the DFS provides an alternative source for the information and simplifies linkages to other, non-USGS water databases.

- Fish Kill File: The Fish Kill File tracks fish kills caused by pollution that have occurred throughout the United States. The kills are a result of a variety of industrial, municipal, agricultural, and transportation-related operations.

Currently about 800 organizations have submitted information to STORET. There are over 735,000 sampling stations in STORET and more than 180 million parametric observations covering some 12,000 water quality parameters.

Many organizations, including federal, state, interstate, and international agencies, submit information to STORET. Users submit new information in the appropriate format daily. STORET data files are updated weekly. Each organization is responsible for the information it submits to STORET; STORET is a user-owned system. States submitting information follow quality assurance and control procedures. All STORET data are checked for invalid data ranges or missing mandatory fields before they are added to the system. Although STORET software edits incoming data for errors and inconsistencies, the owners of the data have the primary responsibility for its content.

Contact:

Office of Wetlands, Oceans, and Watersheds
USEPA Fairchild Building (4503F)
499 South Capitol Street, SW
Washington, DC 20024
Phone: 202-260-7166

2. WATER QUALITY ANALYSIS SYSTEM (WQAS) — USEPA/OWOW

The EPA Water Quality Analysis Branch has developed several databases and procedures that operate under the umbrella of the Water Quality Analysis System (WQAS). WQAS databases are available on the same mainframe as the STORET files, and many procedures intrinsic to one system access and manipulate data from the other. WQAS includes many databases; the following are a sample of those which would be most useful in a nonpoint source monitoring and evaluation study.

- Association of State and Interstate Water Pollution Control Administrators (ASIWPCA): Data covering water quality impairments for 1972, 1982, and 1984.
- CITY: Data on 53,000 cities in the United States and its territories.
- GAGE: Data on 36,000 stream gaging locations across the United States.
- Reach File (RF): Data on stream reaches across the United States. Information on RF3 streams is limited; 100 percent of all RF1 and RF2 streams are included.

Contact:

Office of Wetlands, Oceans, and Watersheds
USEPA Fairchild Building (4503F)
499 South Capitol Street, SW
Washington, DC 20024
Phone: 202-260-7166

3. NATIONAL WATER INFORMATION SYSTEM (NWIS) — USDOI/USGS

USGS is in the process of designing and developing a new National Water Information System (NWIS). The goal of the NWIS effort is to develop and implement a highly flexible hydrologic data management and processing system—one that can be easily changed and expanded in a rapidly changing technological environment. The NWIS will replace the following two systems:

- National Water Data Storage and Retrieval System (WATSTORE): WATSTORE consists of several files in which data are grouped and stored by common characteristics and data collection frequencies. The system also is designed to allow for the inclusion of additional data files as needed. Currently, files are maintained for the storage of (1) surface water, quality-of-water, and ground water data measured on a daily or continuous basis; (2) annual peak values for stream flow stations: (3) chemical analyses for surface and ground water sites; (4) water data parameters measured more frequently than daily; (5) geologic and inventory data for ground water sites; and (6) summary data on water use. In addition, an index file of sites for which data are stored in the system is maintained.

- National Water Data Exchange System (NAWDEX): NAWDEX is a national confederation of water-oriented organizations working together to improve access to water data. Its primary objective is to assist users of water data in the identification, location, and acquisition of needed data. NAWDEX consists of member organizations from the water data community. The members are linked so that their water data holdings can be readily exchanged for maximum use. It encompasses four major areas of operation: (1) maintaining an internal data center, including access to automated data processing facilities for maintenance and use of its information files; (2) indexing water data held by participating organizations; (3) providing facilities and personnel for responding to requests for water data; and (4) formulating recommended water data handling and exchange standards.

NWIS will be a single integrated system that will have the functionality of current data systems as well as expanded capability for processing and managing additional chemical constituent, sediment, biological, and spatial data.

Contact:

National Water Information System
U.S. Geological Survey
National Center, MS 437
12201 Sunrise Valley Drive
Reston, VA 22092
Phone: 703-648-5659

4. NATIONAL STREAM QUALITY ACCOUNTING NETWORK (NASQAN) *USDOI/USGS*

The National Stream Quality Accounting Network (NASQAN) program, started in 1972, provides a nationally uniform basis for assessing large-scale and long-term trends in the physical, chemical, and biological characteristics of the Nation's surface waters. Water quality monitoring is carried out at stations that are generally located on major rivers at the downstream end of the accounting unit.

NASQAN was redesigned in 1995 to focus on four of the largest river basins in the Nation—the Mississippi, the Columbia, the Colorado, and the Rio Grande. Monitoring stations are located at the confluence of major rivers within these basins and at the mouths of the rivers. Coastal stations are operated at a few additional large rivers (the St. Lawrence, Susquehanna, Alabama, Tombigbee, and Yukon) to more fully describe the export of chemicals from the continent to the ocean. NASQAN complements the other national water quality monitoring and assessment programs operated by the USGS by providing information on very large rivers.

NASQAN seeks to characterize the water quality of these large rivers by measuring concentration and mass transport of a wide range of dissolved and suspended constituents including:

- Major ions (salts such as sodium, potassium, and sulfate)
- Nutrients (including various forms of nitrate and phosphate that contribute to eutrophication of lakes and rivers)
- Dissolved and sediment-bound heavy metals (such as lead, cadmium, and arsenic)

- Common pesticides that can dissolve in water (including herbicides such as atrazine)
- Inorganic and organic forms of carbon

This information will be used to describe the long-term trends and changes in the concentration and transport of these constituents within the four river basins (USGS, 1996a).

Contact:

Office of Water Quality
U.S. Geological Survey
National Center, MS 412
12201 Sunrise Valley Drive
Reston, VA 22092
Phone: 703-648-6861

5. National Water-Quality Assessment Program (NAWQA) — *USGS/OWQ*

The National Water-Quality Assessment Program (NAWQA) is designed to assess historical, current, and future water quality conditions in representative river basins and aquifers nationwide. One of the primary objectives of the program is to describe relations between natural factors, human activities, and water quality conditions and to define those factors which most affect water quality in different parts of the Nation. Information from the NAWQA should be useful to water resource managers, planners, and policy makers, and for guiding research, monitoring, and regulatory activities in cost-effective ways.

When fully implemented in 1997, the NAWQA will provide consistent and comparable information on water resources in 60 important river basins and aquifers across the Nation, which combined account for 60 to 70 percent of the Nation's water use and population served by public water supplies and cover about one-half of the land area of the Nation. Each of these 60 river basins and aquifers is a study unit within NAWQA. One-third of the study units will be studied intensively at any given time, for a 3- to 5-year period, followed by a 5- to 6-year period of less intensive study and monitoring. Coinciding with the study unit investigations are national synthesis assessments. Two to four national synthesis topics will be studied at any given time. The first issues being studied—the occurrence of nutrients and pesticides in rivers and ground water—were chosen because of widespread environmental and public health concerns and because the information necessary for a national assessment of these topics is incomplete. The next topic for national synthesis will be the occurrence and distribution of volatile organic compounds (USGS, 1994).

At the study unit level, each investigation has a local liaison committee that consists of representatives with water resources responsibilities or interests from federal, state, and local agencies, universities, and the private sector. Each liaison committee is charged with the following:

- The exchange of information about water quality issues of regional and local interest.
- The identification of sources of data and information.
- Assistance in the design and scope of project products.
- The review of project planning documents and reports.

Contact:

NAWQA Program
U.S. Geological Survey
National Center, MS 413
12201 Sunrise Valley Drive
Reston, VA 22092
Phone: 703-648-

6. NATIONAL HYDROLOGIC BENCHMARK NETWORK (HBN) — USGS/OWQ

USGS's National Hydrologic Benchmark Network (HBN) was established in 1964 as a unique network of stream discharge and water quality monitoring stations in basins that have little human influence on hydrologic characteristics. The program operated a fixed-station, fixed-interval network of 50 sites for water quality and quantity data collection. Drainage areas of the HBN basins range from 2 to 2,000 square miles, with a median basin size of 57 square miles. Elevations of the watersheds range from about 100 to 14,000 feet above sea level. Many of the stations are located in national parks, wilderness areas, state parks, national forests, and specially protected areas set aside for scientific investigations. Throughout the history of the network, sampling frequency has been monthly, bimonthly, or quarterly. In 1996, all stations are scheduled to be sampled twice a year.

Constituent coverage includes:

- Field parameters, including discharge
- Dissolved and total nutrients
- Dissolved inorganics, major ions, and trace elements
- Suspended sediment
- Bacteriology

Precipitation quantity has also been measured at many stations. Depth- and width-integrating sampling techniques are used.

The principal objectives of the network are as follows:

- To document natural changes in hydrologic characteristics.
- To provide a better understanding of the hydrologic structure of natural basins.
- To provide a comparative base for studying the effects of humans on the hydrologic environment.

Data from the network are used to detect water quality trends and to describe water quality conditions. Results from the HBN have both national and regional application for identifying natural hydrologic conditions. Each station represents an integration of upstream water quality conditions with generally minimal influence from land-use activities.

All data collected for the HBN are stored in USGS's WATSTORE and NWIS databases (see data source 3 in this appendix). On a monthly basis, the data are transferred from WATSTORE to

STORET (see data source 1 in this appendix) and can be accessed through either system (USGS, 1996b).

5. NATIONAL SURFACE WATER SURVEY (NSWS) — USEPA/ERL

The National Surface Water Survey (NSWS) consists of two parts: the National Lake Survey and the National Stream Survey. The purpose of the National Lake Survey is to quantify, with known statistical confidence, the current status, extent, and chemical and biological characteristics of lakes in regions of the United States that are potentially sensitive to acidic deposition.

The purpose of the National Stream Survey (NSS) is to determine the percentage, extent, and location of streams in the United States that are presently acidic or have low acid-neutralizing capacity and might therefore be susceptible to future acidification, as well as to identify streams that represent important classes in each region for possible use in more intensive studies or long-term monitoring. The NSS provides an overview of stream water chemistry in regions of the United States that are expected, on the basis of previous alkalinity data, to contain predominantly waters with a low acid-neutralizing capacity.

Variables monitored include acid neutralizing capacity, aluminum, base cations, conductance, major ions, metals, nitrate, organics, pH, and sulfate. A randomly selected subset of lakes was sampled using appropriate methods. The sample results were then weighted to estimate the chemical compositions of lake populations with known confidence. Uncertainties related to time of sampling, spatial variability, and population definition are included in specific research projects to improve confidence in estimates. The NSS employed a randomized, systematic sample of regional stream populations and used rigorous quality assurance protocols for field sampling and laboratory chemical analysis.

Contact:

Environmental Research Laboratory
U.S. Environmental Protection Agency
200 SW 35th Street
Corvallis, OR 97333
Phone: 503-754-4423

6. NATIONAL COASTAL POLLUTANT DISCHARGE INVENTORY (NCPDI) — USDOC/NOAA

The National Coastal Pollutant Discharge Inventory (NCPDI) Program is a series of database development and analytical activities within the National Oceanic and Atmospheric Administration's (NOAA) Strategic Assessment Program for coastal and estuarine areas. The cornerstone of the program is a comprehensive database and computational framework that has been developed over the last 9 years. The database contains pollutant loading estimates for all major categories of point, nonpoint, and riverine sources located in coastal counties of the 200-mile exclusive economic zone that discharge to the estuarine, coastal, and oceanic waters of the conterminous United States (excluding the Great Lakes). The pollutant discharge estimates in the NCPDI are made for each coastal component for the following base years:

- East coast - 1982
- West coast - 1984
- Gulf coast - 1987

The estimates can be considered to approximate pollutant discharge conditions for a 5-year period around the base year. Estimates are made for nine major source categories and 17 pollutants. Source categories include:

- Point sources
- Urban nonpoint sources
- Nonurban nonpoint sources
- Irrigation return flow
- Oil and gas operations
- Marine transportation operations
- Accidental spills
- Dredging operations

Pollutant estimates can be aggregated by county, USGS hydrologic cataloging unit, or estuarine watershed. Pollutant parameters include:

- Flow (wastewater flow or surface runoff)
- Biochemical oxygen demand
- Particulate matter
- Nutrients (total nitrogen and phosphorus)
- Metals (arsenic, cadmium, chromium, copper, iron, lead, mercury, and zinc)
- Petroleum hydrocarbons (oil and grease)
- Pesticides (35 compounds)
- Pathogens (fecal coliform bacteria)
- Wastewater treatment sludges

Estimates are based on a combination of computed methodologies and actual monitored observations. Estimates are seasonal (winter, spring, summer, fall) for a base year. Updated discharge estimates for 1987 for the coastal areas of the Gulf of Mexico and for 1989 for the east coast are being prepared.

Contact:

Pollutant Sources Characterization Branch
National Oceanic and Atmospheric Administration
6001 Executive Boulevard, Room 220
Rockville, MD 20852
Phone: 301-443-0454

7. Ocean Data Evaluation System (ODES) — *USEPA/OWOW*

The Ocean Data Evaluation System (ODES) is a menu-driven system for storing and analyzing water quality and biological data from marine, estuarine, and freshwater environments. The system supports

federal, state, and local decision makers associated with marine monitoring programs. The system was designed to support managers and analysts in meeting regulatory objectives through the evaluation of marine monitoring information.

ODES contains over 2.5 million records of data from the National Estuary Program, the Great Lakes National Program Office, the Ocean Disposal Program, the 301(h) Sewage Discharge Program, the National Pollutant Discharge Elimination System (NPDES) Program, and the 403(c) Program.

Contact:

Office of Wetlands, Oceans, and Watersheds
USEPA Fairchild Building (4503F)
499 South Capitol Street, SW
Washington, DC 20024
Phone: 202-260-7028

8. WATERBODY SYSTEM (WBS) — *USEPA/OWOW*

The Waterbody System (WBS) is an automated database of state water quality assessment information. WBS facilitates collection, storage, retrieval, and analysis of water quality assessment information collected by the states to meet EPA's congressional reporting requirements under section 305(b) of the Clean Water Act.

The WBS contains information that helps program managers report accurately and quickly on the water quality status of a particular waterbody. It might also be used to target resource expenditures and to set surface water program priorities. Under the Clean Water Act, states submit information to EPA on several types of surface waters affected by point or nonpoint source pollution, lakes monitored under the Clean Lakes Program, and surface waters requiring the assigning of total maximum daily load limits to restore or maintain their water quality.

WBS serves as an inventory of each state's navigable waters that have been assessed for water quality and is used as the basis for the 305(b) report to Congress every 2 years. States assemble available monitoring information and make judgments on water quality before summary information can be entered into the system. WBS stores the components and the results of the assessment. The WBS is not designed to store, manipulate, or analyze raw monitoring data.

Participation in the WBS is voluntary. The database is currently used by approximately 40 states, territories, and river basin commissions. The database consists of assessments rather than monitoring data and includes many optional fields. The consistency of WBS information within a state is quite good. Those wishing to aggregate to a regional or national level should discuss data characteristics with the WBS coordinator.

Contact:

Office of Wetlands, Oceans, and Watersheds
USEPA Fairchild Building (4503F)
499 South Capitol Street, SW

Washington, DC 20024
Phone: 202-260-3667

9. PESTICIDE INFORMATION NETWORK (PIN) — USEPA/OPP

The Pesticide Information Network (PIN), maintained by EPA's Office of Pesticide Programs, enables pesticide monitoring data generated by a variety of sources to be routinely identified, obtained, and used. PIN also provides federal, state, and local agencies with a means for sharing information and expertise on pesticides. In addition, information in PIN is used to enhance the accuracy of pesticide risk assessments and risk/benefit regulatory decisions regarding exposure and effects of pesticides under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). PIN is composed of three files:

- The Pesticide Monitoring Inventory (PMI) is a nationwide compilation of synopses of pesticide monitoring projects conducted by federal, state, and local governments as well as private groups. PMI includes the location of the monitoring project, the pesticides involved, an abstract of the project, and the name and address of a contact person. PMI does not contain hard data or results; these can be obtained from individual contact persons for each project.

- The Restricted Use Products (RUP) File is a regulatory file that serves as an information resource for states. Information provided includes pesticide active ingredients, dates of restriction, reasons for restriction, and all products that contain the restricted active ingredients.

- The Coordination File is a cross-referencing chemical index of all synonyms for the active ingredients listed in the PMI and RUP files.

Contact:
Office of Pesticide Programs
USEPA Waterside Mall (H7507C)
401 M Street, SW
Washington, DC 20460
Phone: 202-557-5455

10. ENVIRONMENTAL CONTAMINANT DATA MANAGEMENT SYSTEM (ECDMS) — USDOI/USFWS

The Environmental Contaminant Data Management System (ECDMS) is the cataloging, sample management, and data storage system for residue data from field studies conducted by the U.S. Fish and Wildlife Service (USFWS). Data are from sample matrices consisting of animal and plant tissues, sediments, soils, and water. The system contains data on pesticides, elements, PCBs, and other compounds.

The National Contaminant Biomonitoring Program (NCBP) is maintained by USFWS to document temporal and geographic trends in concentrations of persistent environmental contaminants that might threaten fish and wildlife. NCBP data are maintained by the ECDMS. The NCBP is the USFWS segment of the National Pesticide Monitoring Program, a multiagency monitoring effort by the member

agencies of the Federal Committee on Pest Control. Since 1965, USFWS has periodically determined concentrations of potential toxic elements and selected organochlorine chemicals in fish and wildlife collected from a nationwide network of stations. The NCBP is being phased out with the implementation of the broader Biomonitoring of Environmental Status and Trends (BEST) Program.

In addition to organochlorine chemical residues, freshwater fish, starlings, and waterfowl samples are analyzed for arsenic, cadmium, copper, lead, mercury, selenium, and zinc. Composite samples of whole freshwater fish are collected in replicate from 112 stations in major rivers throughout the United States and in the Great Lakes. Starlings are collected in replicate from 139 terrestrial sites in the conterminous 48 states. Wings of mallards and black ducks shot by hunters in the continental United States are collected to assess body burden of organochlorine compounds in migratory birds. This monitoring program has continued at 2- to 4-year intervals since 1965.

Contact:

Division of Environmental Contaminants
U.S. Fish and Wildlife Service
4401 North Fairfax Drive
Suite 330
Arlington, VA 22203
Phone: 703-358-2148

11. NATIONAL SHELLFISH REGISTER (NSR) — *USDOC/NOAA*

Classified shellfishing waters are monitored as an indicator of bacterial water quality nationwide. Waters are classified for the commercial harvest of oysters, clams, and mussels based on the presence of actual or potential pollution sources and coliform bacteria levels in surface waters. Each shellfish-producing state classifies its waters in accordance with guidelines established by the national Shellfish Sanitation Program. Approximately 2,000 classified shellfishing areas are defined by:

- Name
- Location (nautical chart number, estuary, state, region)
- Classification (approved, prohibited, conditionally approved, restricted)
- Size
- Pollution sources (identified for all nonapproved areas)

Trends in classification by region from 1966 to 1990 and by selected estuaries in the Northeast, Southeast, Gulf of Mexico, and Pacific from 1971 to 1990 are available. Areas that were reclassified because of improved or diminished water quality are distinguished from those which were reclassified as a result of improved monitoring. Data also are collected on administration of state programs, including:

- Identification of state agencies responsible for monitoring waters, assigning classification, analyzing water samples, etc.
- Number of personnel
- Budgets

- Number of sampling stations
- Frequency of sampling
- Other factors that might influence classification

Data are collected by questionnaire, followed by interviews. Classifications are noted on 265 nautical charts. Data were compiled in 1966, 1971, 1974, 1980, 1985, 1990, and 1995.

Contact:

N/ORCA
National Oceanic and Atmospheric Administration
6001 Executive Boulevard
Rockville, MD 20852
Phone: 301-443-8843

12. NATIONAL STATUS AND TRENDS DATABASE (NST) — USDOC/NOAA

Beginning in 1984, NOAA undertook the task of providing information on the status and trends of environmental quality in estuarine and coastal areas. The program defines the geographic distribution of contaminant concentrations in tissues of marine organisms and in sediments. Status and trend data are available from the Mussel Watch and Benthic Surveillance for 4 major elements, 12 trace elements, DDT and its metabolites, selected chlorinated pesticides, selected PCB congeners, approximately 22 polyaromatic hydrocarbons, and ancillary sediment and tissue parameters.

Samples have been collected since 1984 at about 50 Benthic Surveillance sites and since 1986 at about 150 Mussel Watch sites. Sediment samples are collected at all sites. At Benthic Surveillance sites, benthic fish are collected and their livers excised and stored for subsequent chemical analysis. At Mussel Watch sites, bivalve mollusks are collected for analysis. Data are collected annually.

Contact:

Ocean Assessments Division
National Oceanic and Atmospheric Administration
6001 Executive Boulevard
Rockville, MD 20852
Phone: 301-443-8655

13. NATIONAL CLIMATE DATA CENTER (NCDC) — USDOC/NOAA

The National Climatic Data Center (NCDC) collects, processes, and archives meteorological and climatological data from a global network of stations. Records begin in the mid-19th century and continue to the present. Climatic variables (e.g., temperature, precipitation, solar radiation, storms, wind, and floods) are summarized for both short-term and long-term periods of record. Data are available in published form, on microfiche, or on magnetic tape. Derived values relating to growing season and heating and cooling degree days are also produced. Special statistical summaries of actual and derived values of meteorological elements over the world's oceans, as well as summaries used in the study of air pollution, are available.

For about four decades, NCDC has been receiving climatic data from across the United States and around the globe. Principal sources in the United States are the National Weather Service (NWS), the Federal Aviation Administration, the U.S. Air Force, the U.S. Navy, and the U.S. Coast Guard. The NWS's Cooperative Station Network is composed mainly of 10,000 volunteer observers and has been recording daily records since the 1800s. As aircraft began to fill the skies, information on the upper atmosphere was needed. Balloon-borne instruments radioed data; radars began to probe the clouds; rockets reached the fringes of the atmosphere; weather satellites, both geo-stationary and polar orbiting, now continuously watch and record the weather. Technical advancements led NCDC to archive some of its data on CD-ROMs so that users could look at a large amount of climatic data at one time. NCDC plans to archive new data sets using the latest technical advances available. Observations are taken at varying intervals, from every 15 minutes to once per month. Collections are daily or monthly depending on type and source of information.

Contact:
Climate Services Division
NOAA/NESDIS E/CC3
Federal Building
Asheville, NC 28801-2696
Phone: 704-259-0682

14. Synoptic Rainfall Data Analysis Program (SYNOP) — *USEPA*

EPA maintains a Synoptic Rainfall Data Analysis Program (SYNOP) as a tool for summarizing and statistically characterizing rainfall records. SYNOP summarizes hourly rainfall data by storm events, calculating for each event the volume (inches), duration (hours), average intensity (inches per hour), maximum intensity (inches per hour), time since the previous storm (hours), and antecedent rainfall (inches); the hours of missing data; and the hours that the meter did not read. SYNOP then uses these storm event statistics to determine monthly and annual means and coefficients of variation for the various storm parameters.

Contact:
SYNOP
USEPA Waterside Mall (H7507C)
401 M Street, SW
Washington, DC 20460
Phone: 202-382-7112

15. Acid Deposition System (ADS) — *USDOI/USGS*

The National Acid Deposition Program/National Trends Network (NADP/NTN) was the first, and continues to be the only, U.S. network to monitor precipitation chemistry on a national scale. The current network consists of 196 sites in the conterminous United States, Hawaii, Puerto Rico, and American Samoa. Sites are located in predominantly rural areas to avoid the localized influences of large point sources and major urban centers. Nearly 14 years of continuous data are available from the

sites with the greatest longevity; many of these sites are associated with the State Agricultural Experiment Stations.

The primary objective of the NADP/NTN is the determination of geographical patterns of temporal trends in chemical deposition. The program provides scientists, managers, and policy makers with weekly precipitation chemistry data and information on geographical patterns and temporal trends in concentrations and deposition of hydrogen, sulfate, nitrate, ammonium, calcium, magnesium, sodium, potassium and chloride, and ortho-phosphate ions in precipitation. Final quality-assured data are available to a multitude of data users upon request, within 6 months of sample collection. Principal constituents monitored in precipitation and analyzed for trends are:

- pH
- Specific conductance
- Hydrogen ions
- Sulfate and nitrate ions
- Ammonium and calcium ions
- Chloride, magnesium, sodium, and potassium ions

The NADP/NTN monitoring program has developed criteria and protocols that ensure uniformity in siting, sampling methods, analytical techniques, data handling, and overall network operations. Precipitation is collected by wet/dry precipitation collectors and rain gages. Analytical methods for the chemical variables measured are:

- pH
- Field pH
- Laboratory conductivity
- Electronic detection of hydrogen (also reported as pH)
- Automated calorimetric detection of ammonium
- Atomic absorption spectrophotometric detection of calcium, magnesium, sodium, and potassium
- Ion chromatographic detection of sulfate, nitrate, and chloride

Samples are collected weekly. Data from some sites are available from 1979. The data are maintained on the Acid Deposition System (ADS).

Contact:

U.S. Geological Survey
National Center, MS 416
12201 Sunrise Valley Drive
Reston, VA 22092
Phone: 703-648-6875

16. MAJOR LAND USES IN THE UNITED STATES (MLU) — USDA/ERS

For more than 50 years, Economic Research Services (ERS) and its predecessor agencies have estimated acreage and maintained an inventory of the major uses of land in the United States at intervals coinciding with the Census of Agriculture. Estimates are made for major land use classes:

- Cropland
- Grassland pasture and range
- Forest land
- Special use
- Unclassified use

Each major class is further classified by specified uses, and some are classified by ownership. Land uses are also designated as agricultural and nonagricultural. Agricultural land uses include:

- Cropland (cropland harvested, cropland failure, cultivate summer fallow, and idle cropland)
- Grazing lands (cropland pasture and permanent pasture and range)
- Grazed forest land
- Miscellaneous agricultural uses (farmsteads, farm roads, and farmlands)

Special uses include:

- Forest land not grazed
- Intensive uses (highways and roads, railroads, and airports)
- Extensive uses (national parks, state parks, wilderness areas, federal wildlife areas, state wildlife areas, national defense areas, and federal industrial facilities)

Other unclassified land uses include urban, special uses not inventoried, and other miscellaneous areas, such as marshes, open swamps, bare rock areas, deserts, and tundra. Data are analyzed for trends.

Data from the Bureau of the Census, agencies of the Department of Agriculture, public land management and conservation organizations, and other sources are assembled, analyzed, and synthesized to estimate state, regional, and national land use acreage. The major uses of land are inventoried every 5 years, coinciding with years in which the Census of Agriculture is completed. The inventories generally have been comparable in format and coverage since 1945. The series on "cropland used for crops" dates back to 1909.

Contact:

Economic Research Service
U.S. Department of Agriculture
1301 New York Avenue, NW, Room 408
Washington, DC 20005-4788
Phone: 202-219-0424

17. AGRICULTURAL CENSUS (AGCENSUS) — USDOC/BOC

The Agricultural Census (AgCensus) database includes about 750 variables reported at the county level for 1978, 1982, and 1987. The census is scheduled for years ending in 2 and 7. AgCensus variables include simple statistics such as the number of farms, total acreage of farms, acreage of various major crops, and total number of various livestock. It also includes such data as the number and acreage of irrigated farms, the number and acreage of farms irrigated by various sources, the number and acreage of farms by farm size, commercial fertilizer and other agricultural chemical expenditures, the number of farms and livestock by herd size, the number and acreage of vegetable farms, and a subset of these statistics for farms with sales greater than $10,000.

AgCensus does not include pesticide application rates or even the acreage to which specific pesticides are applied. Fertilizer and manure application rates and acreage are also not reported. Irrigation rates and farm management practices are likewise not included in the database.

Contact:

Agriculture Division
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233
Phone: 301-763-1113

18. FOREST SERVICE RANGE MANAGEMENT INFORMATION SYSTEM (FSRAMIS) — USDA/USFS

The Forest Service Range Management Information System (FSRAMIS) collects and analyzes data on grazing in national forests and grasslands. FSRAMIS provides grazing use statistical data. Data on the number of grazing animals (cattle, horses and burros, sheep and goats), animal unit month, and number of permittees are reported at the national level and for each type of Forest Service land, region, and state. Other variables measured include:

- Allotment condition
- Improvement inventory and activity
- Grazing capacity
- Actual use
- Authorized use
- Unauthorized use

Data are analyzed for trends in ecological potential. Data on grazing on the National Forest System lands are extracted from the grazing permits. Data on free-roaming horse and burro populations are estimated by census. Data are collected on cycles ranging from annual to once every 3 to 5 years.

Contact:

Range Management Staff
U.S. Forest Service
Department of Agriculture

P.O. Box 96090
Washington, DC 20090-6090
Phone: 202-205-1460

19. FOREST INVENTORY AND ANALYSIS PROGRAM (FIA) — USDA/USFS

The Forest Inventory and Analysis (FIA) program is responsible for making and keeping current a comprehensive inventory and analysis of the renewable forest and rangeland resources of the United States. Initial inventory efforts began in the West in 1930; by the 1960s, inventories had been completed for all of the 48 conterminous states and many of the important forested states had been reinventoried. The inventory data and analysis provide trend information on the extent, condition, ownership, and composition of the Nation's forests as well as information about wildlife habitat, forage production, and other resource characteristics needed for resource planning. At least 43 kinds of resource data are collected for sample plots during the inventory, including:

- Land use
- Land ownership
- Forest type
- Stand age
- Stand size and volume classes
- Harvest history
- Soils data
- Tree data (species, diameter at breast height, height, cull, etc.)
- Other vegetation data
- Non-timber data

These data are used to make estimates of forest land areas, species composition, timber volume, and net annual timber growth, removals, and mortality by forest type, state, region, ownership, softwood and hardwood sawtimber species, productivity class, diameter class, and other classifications. The volume of roundwood products harvested by material species group, region, and product is estimated. Estimates also are made of areas harvested or otherwise disturbed, regenerated to forest, or cleared for other uses. Additional estimates of recreation use, wildlife values, site productivity, physiographic characteristics, and other items are made. The data are maintained in three databases:

- National Resources Planning Act (RPA) Timber Database
- Eastwide Forest Inventory Database
- Forest Inventory and Analysis Database

Data are gathered using a two-phase sampling design, with the first phase involving ground measurements at sample plots, each covering an acre. Depending on the extent to which sensing is used, ground sample intensity ranges from one plot per 3,000 acres to one plot per 10,000 acres. Statewide timber inventory information has been collected continuously for about 50 years. In most regions of the United States, the third inventory cycle has been completed and some areas have been inventoried as many as five times. Each year, some 50 million acres are inventoried in the

conterminous United States. Currently this rate of coverage translates into an inventory cycle of 12 years for the Nation.

Contact:

U.S. Department of Agriculture
Forest Service
P.O. Box 96090
Washington, DC 20090-6090
Phone: 202-205-1343

20. NATIONAL LAND USE AND LAND COVER MAPS — USDOI/USGS

As part of its National Mapping Program, the USGS produces and distributes land use and land cover maps and digitized data. Land use refers to human activities that are directly related to the land. Land cover describes the vegetation, water, natural surface, and artificial constructions at the land surface. Associated maps display information on political units, hydrologic units, census county subdivision, and, in some cases, federal land ownership. Land use and land cover areas are classified into nine major classes:

- Built-up land
- Agricultural land
- Rangeland
- Forest land
- Water areas
- Wetlands
- Barren land
- Tundra
- Perennial snow or ice

Each major class is subdivided into several minor classes, for 37 minor classes total. For example, forest lands are further classified as deciduous, evergreen, or mixed forest land, and water is further classified as streams and canals, lakes, reservoirs, or bays and estuaries. Remote sensing methods are used, including satellite imagery, high-altitude imagery, medium-altitude remote sensing, and low-altitude imagery. Data were collected in the late 1970s and early 1980s.

Contact:

Office of Geographic and Cartographic Research
U.S. Geological Survey
National Center, MS 590
12201 Sunrise Valley Drive
Reston, VA 22092
Phone: 703-648-5741

21. POPULATION CENSUS DATA (CENDATA) — USDOC/BOC

The decennial census provides a comprehensive set of population statistics for the United States. Basic demographic characteristics are collected on a 100-percent basis. Social and economic characteristics are collected from a large sample of all households and persons in group quarters. The decennial census provides demographic (e.g., age, race, sex, relationship), social (e.g., education, migration, ancestry, language), and economic (e.g., occupation, industry, income, place of work) characteristics of the population of the United States, Puerto Rico, the Virgin Islands, Guam, American Samoa, the Norther Marianas, and Palau. Trend data are available from previous decennial censuses. Basic demographic data are collected from 100 percent of the population. Social and economic characteristics are collected from a large sample, approximately one in six in 1980 and 1990.

Contact:

Population Division
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233
Phone: 301-763-7890

22. NATIONAL RESOURCES INVENTORY (NRI) — USDA/NRCS

For 50 years, the Natural Resources Conservation Service (NRCS), formerly the Soil Conservation Service (SCS), has been conducting periodic inventories of the Nation's soil, water, and related resources. The National Resources Inventory (NRI), which is an extension and modification of earlier inventories, provides data on the status, condition, and trends of these resources on nonfederal land in the United States. The many types of data collected by the NRI process are organized into eight general categories:

- Soil characteristics and interpretations (including agricultural land capability)
- Land cover
- Land use (including irrigated and nonirrigated cropland, grazed and ungrazed forest land, prime farmland, etc.)
- Erosion (such as sheet and rill, wind, and ephemeral gullies)
- Land treatment (such as irrigation, tillage, and windbreaks)
- Conservation treatment needs
- Vegetative conditions (such as wetlands, rangeland condition and species, and pasture management)
- Potential for conversion to cropland

The NRI is a multiresource inventory based on soils and related resource data collected at scientifically selected random sample sites. The NRI sample design was developed by the Iowa State University Statistical Laboratory at Ames. It uses census area and point methods for data collection. Data collection involves both field investigation and remote sensing. Data are collected on a 5-year cycle. Recent surveys were conducted in 1977, 1982, and 1987.

The 1987 NRI data were collected from nearly 300,000 sample sites from all counties of the United States, except those in Alaska, and in Puerto Rico and the Virgin Islands. Most of these samples were part of the 1982 NRI, which had nearly 1 million sample sites. The 1987 NRI data have a high degree of reliability at the state level, and the 1982 NRI date provide a high degree of reliability at the multicounty level. Data estimates can be made by Major Land Resource Areas (MLRA), NRCS Administrative Areas, Water Resources Council Aggregated Subareas, and other multicounty geographic subdivisions.

Contact:

Resources Inventory and Geographic Information Systems Division
U.S. Department of Agriculture
Natural Resources Conservation Service
P.O. Box 2890
South Agricultural Building, Room 6175
Washington, DC 20013
Phone: 202-720-4530

23. ECOREGION MAPS — USEPA/ERL

Ecoregions are defined by EPA to be regions of relative homogeneity in ecological systems or in relationships between organisms and their environments. Ecoregions of the United States have been mapped to help water resource managers understand better the regional patterns of ecosystem quality and the relative importance of factors that might be determining this quality. Specifically, the ecoregion framework can establish a logical basis for characterizing ranges of ecosystem conditions or quality that are realistically attainable. A national ecoregion map has been prepared at a scale of 1:7,500,000 and regional maps are prepared at a scale of 1:2,500,000. The maps are available in an ARCINFO format for use by individual users.

Contact:

U.S. Environmental Protection Agency
Environmental Research Laboratory
200 SW 35th Street
Corvallis, OR 97333
Phone: 503-757-4601

24. NATIONAL WETLANDS INVENTORY (NWI) — USDOI/USFWS

In 1975, the U.S. Fish and Wildlife Service (USFWS) established the National Wetlands Inventory (NWI) to develop technically sound and comprehensive information on the characteristics and extent of wetland resources in the United States. Status and trends information is available for selected wetland types including estuarine wetlands, palustrine wetlands, lacustrine wetlands, and deepwater habitats in the lower 48 states. In addition, statistical data are available for coastal waters and bay bottoms, coastal marshlands and mangroves, recent changes in inland vegetated wetlands, recent changes in lacustrine deepwater habitats, estimates of current annual wetland losses, estimates of wetland losses by flyways, states with significant changes in wetland resources, indicators of development pressures on

wetland resources, and causes of wetland losses. The Emergency Wetlands Resources Act of 1986 requires that updates of the wetland status and trends be produced on 10-year cycle with reports due in 1990, 2000, 2010, etc. Data are collected continuously. The 1990 update provides trend data on wetlands losses and gains between the 1970s and the 1980s.

The wetland mapping phase of the project has produced map coverage for approximately 70 percent of the lower 48 states, 22 percent of Alaska, and all of Hawaii, Puerto Rico, and Guam. Wetland status and trends information is designed to provide statistical estimates on a national basis (lower 48 states). In addition, regional intensification studies are available for the Chesapeake Bay Region and the Central Valley of California. Other statewide status information is available for the states of Florida, Delaware, New Jersey, Illinois, Washington, Maryland, and Connecticut. Status reports covering the coastal wetlands of Alaska and the Prairie Pothole Region (North Dakota, South Dakota, Minnesota) are also available.

Contact:

National Wetlands Inventory
U.S. Fish and Wildlife Service
Suite 101 Monroe Building
9720 Executive Center Drive
St. Petersburg, FL 3702-2440
Phone: 813-893-3624

DATABASE	SUBDATABASES/ DATA TYPES	AVAILABILITY	NOTES
STORET	• Water Quality File • Biological System Field Survey File (BIOS) • Daily Flow File (DFF) • Fish Kill File (FK)	• USEPA mainframe • USGS NAWDEX	• A modernized STORET, with greater computing and data-handling capabilities, will be on-line in 1997.
NAWDEX	• Water Data Sources Directory: Information on organizations that collect water and water-related data • Master Water Data Index: Information on sites at which surface water, ground water, or water quality data are collected	USGS NAWDEX	• An interagency program to facilitate the exchange of water data and promote the improvement of water data handling • Provides access to WATSTORE and STORET
WBS	• Surface waters • Overall water quality, suitability for designated uses, causes and sources of water quality problems	USEPA mainframe	• Management tool to track state assessments of ambient water quality for surface waters • Source of 305(b) report information
NASQAN	• Concentration, mass transport • Major ions, nutrients, heavy metals, pesticides, inorganic and organic carbon	USGS NAWDEX	• Focuses on large river systems: Mississippi, Columbia, Colorado, Rio Grande, St. Lawrence, Susquehanna, Alabama, Tombigbee, Yukon
NWIS	• Spatial, chemistry, flow, biological • Surface water and ground water	USGS	• Replacing WATSTORE and NAWDEX • Designed to be a flexible hydrologic data management and processing system
PIS	• PMI (Pesticide Monitoring Inventory): synopses of pesticide monitoring projects; location, pesticides monitored, abstract, contact • RUP (Restricted Use Products file): Pesticide active ingredients, restriction dates and reasons, names of all products with restricted active ingredients • Chemical Index: Cross-references for synonyms for active ingredients listed in PMI and RUP	USEPA	• Contains up-to-date pesticide information

DATABASE	SUBDATABASES/ DATA TYPES	AVAILABILITY	NOTES
ODES	• Ocean outfall data (301(h), 403(c)) • NPDES discharge permit data • Ocean dumping data • National Estuary Program data • GLNPO data	USEPA	• For storing and analyzing marine environmental data • Contains statistical, graphical, and modeling tools
WQAS	• ASIWPCA: water quality impairments • CITY: data on U.S. cities • GAGE: stream gaging data • REACH FILE: stream reach data	USEPA mainframe	
HBN	• Stream discharge • Nutrients • Inorganics, major ions, trace elements • Suspended sediment • Bacteriology	• USGS WATSTORE • USEPA STORET	• 50 monitoring sites in basins with little human influence on hydrologic characteristics
NAWQA	• Water quality • Sediment	USGS	• Designed to assess historical, current, and future water quality conditions in representative river basins and aquifers • Covers 60 river basins and aquifers that account for 60%-70% of U.S. water use and 1/2 of U.S. land area
NSWS	• NSS (National Stream Survey): stream acidity, stream chemistry, acid neutralizing capacity aluminum, base cations, conductance, major ions, metals, nitrate, organics, pH, sulfate • NLS (National Lake Survey): lake chemistry	USEPA/ERL, Corvallis, OR	• NSS's purpose is to determine percentage, extent, and location of streams with low acid-neutralizing capacity
NCPDI	• Pollutant loading estimates for all major categories of point, nonpoint, and riverine sources in coastal counties. • Flow, particulates, BOD, nutrients, metals, petroleum hydrocarbons, pesticides, pathogens	NOAA	• Baseline years: 1982 - east coast; 1984 - west coast; 1987 - Gulf coast
ECDMS	• Residue data from USFWS field studies • Animal and plant tissues, sediments, soils, water • Pesticides, elements, PCBs, other compounds • NCBP (National Contaminant Biomonitoring Program): USFWS segment of National Pesticide Monitoring Program; temporal and geographic trends in concentrations of persistent environmental contaminants	USFWS	• NCBP being phased out with implementation of BEST Program

DATABASE	SUBDATABASES/ DATA TYPES	AVAILABILITY	NOTES
NSR	• Classified shellfishing waters; name, location, NSSP classification, size, pollution sources • State program administration information	NOAA	• Trends in classifications are available
NST	• Estuarine and coastal waters • Geographic distribution of contaminant concentrations in tissues of organisms and sediments • 4 major elements, 12 trace elements, DDT and metabolites, chlorinated pesticides, PCB congeners, 22 polyaromatic hydrocarbons, sediments, tissues	NOAA	• Data from Benthic Surveillance sites and Mussel Watch sites
NCDC	• Temperature, precipitation, solar radiation, storms, floods, wind • Derived values for growing season, heating and cooling degree days • Statistical summaries of meteorological elements over the oceans • Air pollution summaries	NOAA	• Available as published, microfiche, or magnetic tape
SYNOP	• Hourly rainfall data summarized by storm event for volume, duration, average intensity, maximum intensity, time since previous storm, antecedent rainfall, hours of missing data, hours not read • Monthly and annual means, coefficients of variation for storm parameters	USEPA	• Tool for summarizing and statistically characterizing rainfall records
ADS	• Specific conductance, pH, hydrogen ions, sulfate and nitrate, ammonium, calcium chloride, magnesium, sodium, potassium	USGS	• Only national network to monitor precipitation chemistry • Monitoring sites are predominantly rural
MLU	• Cropland, grassland pasture and range, forest land, special uses, unclassified uses • Subclassified by specific use, ownership	USDA	• Special uses: intensive (roads, airports, railroads), extensive (parks etc., defense, federal industrial)
National Land Use and Land Cover Maps	• Human land use activities • Land cover by vegetation, water, natural surface, artificial construction • Political unit, hydrologic unit, census county subdivision, federal land ownership	USGS	• Data for maps were collected from the mid-1970s to early 1980s
CENDATA	• Population, demographic, social, economic	USDOC, Bureau of the Census	• Updated every 10 years

DATABASE	SUBDATABASES/ DATA TYPES	AVAILABILITY	NOTES
AgCensus	• Total acreage of farms, acreage of major crops, total number of livestock, number and acreage of irrigated farms, number and acreage of farms irrigated by various sources, number and acreage of farms by size, commercial fertilizer and other agricultural chemical expenditures, number of farms and livestock by herd size, number and acreage of vegetable farms, statistics for farms with sales > $10,000.	USDOC, Bureau of the Census	• County-level data
FSRAMIS	• Grazing activity on national forests and grasslands; cattle, horses, burros, sheep, goats • Number of grazing animals, animal unit month, number of permittees • National level, region, state, type of forest service land • Allotment condition, improvement inventory and activity, grazing capacity, actual use, authorized use, unauthorized use	USDA, USFS	• Data are analyzed for trends in ecological potential
FIA	• RPA Timber Database • Eastwide Forest Inventory Database • Forest Inventory and Analysis Database • Land use, land ownership, forest type, stand age, stand size and volume class, harvest history, soils, trees, other vegetation, non-timber • Forest land area; species composition; timber volume; net annual timber growth, removals, mortality.; area harvested; area regenerated; area converted to non-forest use • Recreation use, wildlife value, site productivity, physiographic characteristics	USDA, USFS	• Comprehensive inventory and analysis of forest land and rangeland
NRI	• Soil characteristics, land cover, land use, erosion, land treatment, conservation treatment needs, vegetative condition, potential for conversion to cropland	USDA, NRCS	• Information on non-federal lands • Data collected at 5-year intervals
Ecoregion Maps	• Ecoregions of the U.S. • National (1:7,500,000), regional (1:2,500,000)	USEPA/ERL, Corvallis, OR	• Intended for water resource management • Available in ARCINFO format

DATABASE	SUBDATABASES/ DATA TYPES	AVAILABILITY	NOTES
NWI	• Estuarine, palustrine, lacustrine, deepwater wetlands • Coastal waters, bay bottoms, coastal marshlands, mangroves • Changes in inland vegetated wetlands, changes in lacustrine deepwater habitats • Current annual wetland loss, wetland loss by flyway, states with significant changes in wetland resources, causes of wetland loss	USDOI, USFWS	• Updated every 10 years • Regional intensive data available for Chesapeake Bay, Central Valley (CA) • Statewide information for FL, DE, NJ, IL, WA, MD, CT • Status reports for Coastal AK, Prairie Pothole Region (ND, SD, MN)

APPENDIX C. EXAMPLE MONITORING PROGRAMS

This appendix contains three examples of nonpoint source monitoring programs that are considered by many nonpoint source experts to be good programs for the objectives they address. These examples should not be copied indiscriminately for use in every other similar watershed, but should only be used as references for designing good monitoring programs for similar situations.

C.1 St. Albans Bay, Vermont Rural Clean Water Program

The St. Albans Bay, Vermont, monitoring and evaluation program was funded at $1.6 million over 11 years (1980-1990), a price well beyond reach for many watersheds of similar size (33,344 acres) (Smolen et al., 1986). The program had three objectives (Vermont RCWP Coordinating Committee, 1986):

- Document changes in the water quality of specific tributaries within the watershed resulting from implementation of manure management practices.
- Measure changes in suspended sediment and nutrients entering St. Albans Bay resulting from implementation of water quality management programs within the watershed.
- Evaluate trends in the water quality of St. Albans Bay and the surface waters within the St. Albans Bay watershed during the period of the St. Albans Bay RCWP Watershed Project.

To achieve these objectives, the monitoring strategy for the St. Albans Bay Watershed included long-term water quality monitoring, related long-term monitoring, and short-term intensive studies. The monitoring sites for all sampling are shown in Figure C-1 and listed in Table C-1.

C.1.1 Long-term Water Quality Monitoring

The long-term monitoring in St. Albans Bay included four monitoring levels. Level 1 Bay Sampling was designed to determine long-term water quality trends in St. Albans Bay over the life of the project (four monitoring stations in the bay). Level 2 Tributary Sampling was designed to determine the long-term water quality trends for the major tributaries including the Bay and the St. Albans City wastewater treatment plant (six monitoring stations located throughout the watershed). Level 3 monitoring was designed to evaluate the effect of best manure management practices on the quality of surface runoff from individual fields (two monitoring stations located in the headwaters of the Jewett Brook subwatershed), while Level 4 was designed to supplement the Level 2 monitoring by sampling additional tributaries to St. Albans Bay and to isolate subunits within the Level 2 subwatersheds (four monitoring stations) (Vermont RCWP Coordinating Committee, 1984). Table C-2 lists the variables monitored for each level of the long-term monitoring.

C.1.2 Related Long-Term Monitoring

Precipitation and other climatological data were collected in the St. Albans Bay Watershed at four locations. Biological monitoring was conducted in St. Albans Bay and in bay tributaries. Bay biological monitoring corresponded to the long-term Level 1 stations. Tributary biological monitoring

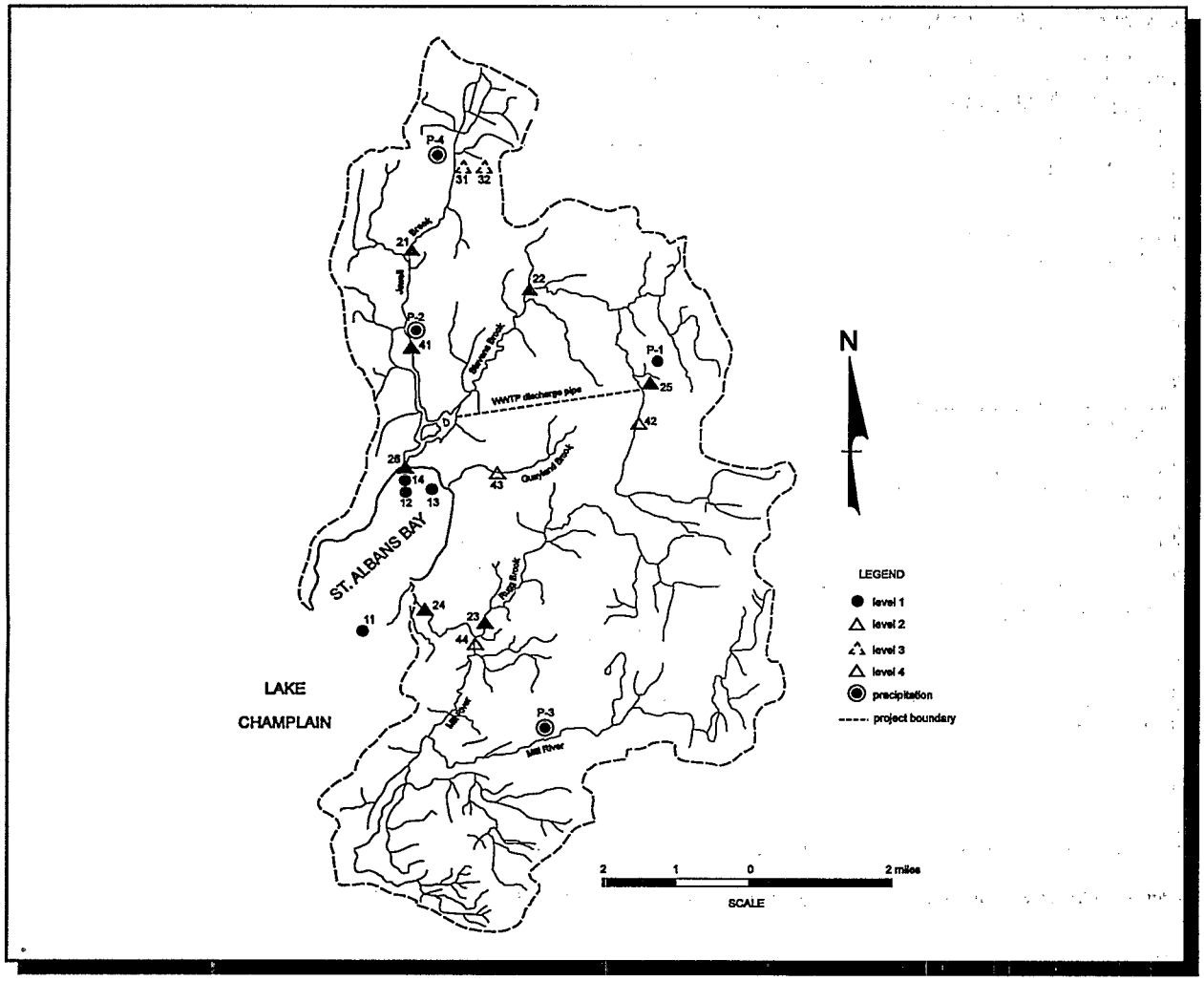


Figure C-1. St. Albans Bay Watershed, Franklin County, Vermont, sampling locations. (Source: USEPA, 1982c)

was conducted at five locations. A cooperative program was developed to collect land use and agricultural activity information from each farm in the watershed. Both baseline information and daily field log data were collected and entered into a geographic information system. Table C-3 lists the parameters monitored for each of the related long-term monitoring efforts.

C.1.3 Short-term Intensive Studies

The short-term intensive studies included three components: wetland influences, bay and wetland sediment, and bay circulation. Stevens Brook wetland was sampled to determine the effects of the

Table C-1. St. Albans Bay watershed, Franklin County, Vermont, sampling station summary.

Type of Monitoring	Station Description
Level 1: Long-term Bay and Biological	Station 11: Outer Bay Station 12: Inner Bay Station 13: Beach Station 14: Off Bridge (new)
Level 2: Long-term Tributary	Station 21: Jewett Brook Station 22: Stevens Brook Station 23: Rugg Brook Station 24: Mill River Station 25: St. Albans Wastewater Treatment Plant Station 26: Stevens Wetland
Level 3: Manure Management Evaluation	Station 31: Larose ditch - below site Station 32: Larose ditch - above site
Level 4: Additional Tributary	Station 41: Jewett Brook Station 42: Stevens Brook Station 43: Guayland Brook Station 44: Mill River
Meteorological	P-1: St. Albans Radio Station P-2: Dunsmore Farm P-3: Franklin Ford Tractor P-4: LaRose Farm
Biological	Station 21: Jewett Brook Station 22: Rugg Brook Station 23: Mill River Station 22A: Stevens Brook above STP outfall Station 22B: Stevens Brook below STP outfall

wetland on water entering St. Albans Bay from point and nonpoint sources. Fifteen sampling stations were located within the wetland along the brook channels. Station 26 served as a site for continuous monitoring of the wetland outlet. Fifteen sampling stations were located within the bay and contiguous wetland to determine the chemical and physical properties of the sediments. Six of these stations were used for within-year temporal studies. Sediment phosphorus release studies were also conducted at these sites. Wind, water current, and concentration data were collected in St. Albans Bay to determine the effect of bay circulation on water quality. A model was developed to predict phosphorus concentrations in the bay under different loading rates and meteorological conditions.

Table C-4 lists the parameters monitored for each of the short-term studies.

C.1.4 Sampling and Analytic Techniques

Table C-5 summarizes the frequency and type of sample collected for long-term water quality monitoring, although not all stations were sampled throughout the 10-year study. The methods used to determine each water quality parameter are given in Table C-6.

Table C-2. Long-term variables monitored for each level in the St. Albans Bay watershed.

Variable	Long-term Monitoring Level			
	1	2	3	4
Turbidity	●	●		●
Total Suspended Solids	●	●	●	●
Volatile Suspended Solids	●	●	●	●
Total Phosphorus	●	●	●	●
Ortho-Phosphorus	●	●	●	●
Ammonia-Nitrogen	●	●	●	●
Total Kjeldahl Nitrogen	●	●	●	●
Nitrite + Nitrate as Nitrogen	●	●	●	●
Chlorophyll *a*	●			
Fecal Coliform	●	●		●
Fecal Streptococcus	●	●		●
Temperature	●	●		●
Dissolved Oxygen	●	●		●
pH	●	●		●
Conductivity	●	●		●
Secchi Disc	●			
Flow				●

Table C-3. Monitored variables for related long-term monitoring in the St. Albans Bay watershed.

Type of Monitoring	Variables
Meteorological	Continuous precipitation and streamflow Continuous temperature (Level 3 only) Wind speed and direction (Bay only)
Biological	Fish species and abundance, benthic invertebrates, periphyton, macroinvertebrates
Land Use	Land use - field by field, activity dates Livestock - type, number, housing, dates Manure management - type, capacity, fields, dates Fertilizer - type, amount, field, dates Pesticide/agrichemicals - type, amount, field, dates Milkhouse - type, disposal Barnyard - size, paving, use schedule Drainage - type, fields Soils, topography, streams, farm and watershed boundaries

Table C-4. Monitored variables for the short-term intensive studies in the St. Albans Bay watershed.

Type of Monitoring	Variables	
Wetland Influences	Temperature Dissolved Oxygen Discharge Chloride Total Suspended Solids Volatile Suspended Solids	Total Phosphorus Ortho-phosphorus Total Kjeldahl Nitrogen Ammonia-Nitrogen Nitrate + Nitrite as Nitrogen pH
Bay and Wetland Sediments	Redox potential Porosity Grain size distribution % organic matter Iron	Total phosphorus NH_4Cl phosphorus NH_4Ac phosphorus NaOH phosphorus HCl phosphorus
Bay Circulation	Total phosphorus Chloride	Water velocity and direction Wind speed and direction

Table C-5. Water quality sampling schedule for the St. Albans Bay watershed.

Level	Variables	Sample Type	Season	Sample Frequency
1	All	Grab	October-April	Monthly
			May-July	Biweekly
			August-September	Weekly
2	Fecal coliform Fecal streptococcus	Grab	All	Weekly
	Temperature Turbidity pH Dissolved oxygen	In situ	All	Biweekly
	Solids Nitrogen Phosphorus	Composite	All	2-48 hour and 1-72 hour Weekly
3	All	Composite	All	4 hour
4	All	Grab	May-February	20 days (avg.)
			March-April	Weekly

Table C-6. Methods of water quality analysis used in the St. Albans Bay RCWP.

Variable	Storage	Preservation/Method of Analysis	Reference
pH	in situ	Orion 399A meter	Method 150.1[b]
Dissolved Oxygen	in situ	YSI 51B meter	Method 360.1[b]
Conductance/Temperature	in situ	YSI 33 S-C-T meter	Method 120.1[b]
Turbidity	cool, 4°C	Hach Ratio Turbidimeter (NTU)	Method 180.1[b]
Total Suspended Solids[a]	cool, 4°C	Gravimetric filtration	Method 160.2[b]
Volatile Suspended Solids	cool, 4°C	Filter ignition at 550 °C	Method 160.4[b]
Total Phosphorus	cool, 4°C acid (pH≤2)	Persulfate digestion	Method 365.2[b]
Orthophosphate[a]	cool, 4°C	Ascorbic acid	Method 365.2[b]
Nitrate + Nitrite-N[a]	cool, 4°C acid (pH≤2)	Cd-Cu reduction sulfanilamide	Method 353.3[b]
Ammonia-N	cool, 4°C acid (pH≤2)	Boric acid distillation	Method 350.2[b]
Total Kjeldahl Nitrogen	cool, 4°C acid (pH≤2)	Macro-digestion with sulfuric acid	Bremner 1965
Chemical Oxygen Demand	cool, 4°C acid (pH≤2)	Potassium dichromate oxidation	Oceanography International Corp.
Fecal Coliform	cool, 4°C	Membrane filtration	Method 909A[c]
Fecal Streptococci	cool, 4°C	Membrane filtration	Method 909C[c]

[a] Filtration through glass fiber filter.
[b] USEPA, 1974.
[c] American Public Health Administration, 1975.

C.2 BELLEVUE, WASHINGTON NATIONWIDE URBAN RUNOFF PROGRAM

The Nationwide Urban Runoff Program (NURP) was designed such that each project met minimum monitoring requirements to support national analyses. Thus, the monitoring programs were very similar for all projects in NURP. The Bellevue, Washington, monitoring program is one example.

Bellevue (25 square miles, 1982 population about 75,000) is located in the Puget Sound lowlands on the west side of the Cascade Mountains and immediately east of Lake Washington (Figure C-2) (USEPA, 1982c). Land use is primarily residential and mean annual precipitation, mostly rain, is about 42 inches. Bellevue is hilly, with moderate slopes predominating. Drainage is carried by a system of separate storm sewers (i.e., separate from sanitary wastes), open channels, and streams into Lake Washington through Mercer Slough (Figure C-3) (USEPA, 1982c). Some drainage flows east into Lake Sammamish through Phantom Lake and another stream.

The project objectives were (USEPA, 1982c):

- *To apply uniformly, in selected drainage basins, a variety of management practices that are available to and achievable by local units of government.*
- *To improve standard practices and operations by varying the frequency and manner of application, developing management programming methods, and altering monitoring and inspection practices for greater responsiveness to water quality needs.*
- *To test, analyze, and document the impact of local management practices on storm water quality, isolating causal factors and their impacts on water quality and evaluating and developing functional relationships between the quantity and quality of runoff and the hydrologic and cultural characteristics of the basins involved.*
- *To develop, test, and document methods of source control of common urban storm water pollutants.*
- *To document temporal changes in storm runoff and constituent concentrations within several drainage basins of differing land use.*
- *To develop and document means of incorporating best management practices into the institutional and operational framework of local government agencies.*
- *To expand the toxic metals, sediment, herbicides and pesticides, and other databases for various land use categories, contributing to the data base of storm water quality modeling efforts nationally.*
- *To develop methods for estimating storm and annual loads of water quality constituents from unsampled watersheds in each urban-study area.*
- *To evaluate methods of transferring the data to ungaged watersheds in other regions.*

C.2.1 Monitoring Approach

Monitoring activities took place in three catchment areas (Figure C-3) (USEPA, 1982c): (1) Surrey Downs, (2) Lake Hills, and (3) 148th Avenue. Three groups participated in the monitoring efforts: the U.S. Geological Survey (USGS), the City of Bellevue, and the municipality of Metropolitan Seattle (Metro). Two of the three study catchments, Surrey Downs and Lake Hills, were single-family residential areas of similar size. These two basins were used to investigate the effectiveness of street sweeping for reducing the amount of pollutants in storm runoff. The third catchment, 148th Avenue,

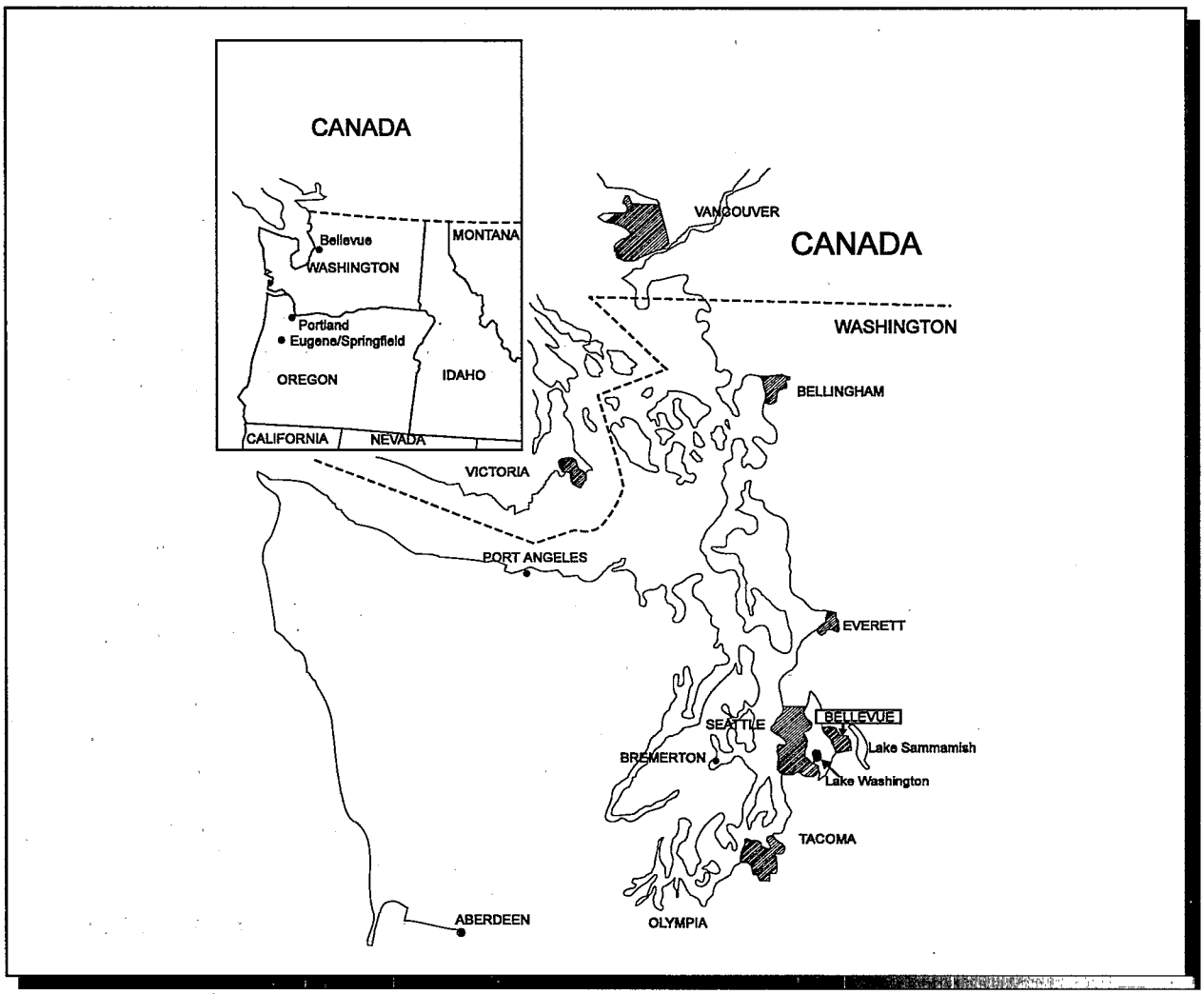


Figure C-2. State locus of Bellevue, Washington, NURP. (Source: USEPA, 1982c)

contained a divided 4-lane arterial street. The data from this site were used to investigate the effects of detention basins on the runoff quality. The monitoring activities also allowed investigators to:

- Define pollutant hydrographs for each of the three catchments during approximately 12 storms per year.
- Determine the effectiveness of street cleaning equipment for various levels of effort under the actual conditions encountered.
- Describe the quantities and characteristics of sewerage system particulates in the study area.
- Obtain a continuous mass balance relationship between total runoff yields and all the sources of urban runoff pollution.
- Analyze samples for the 129 EPA toxic or "priority" pollutants.

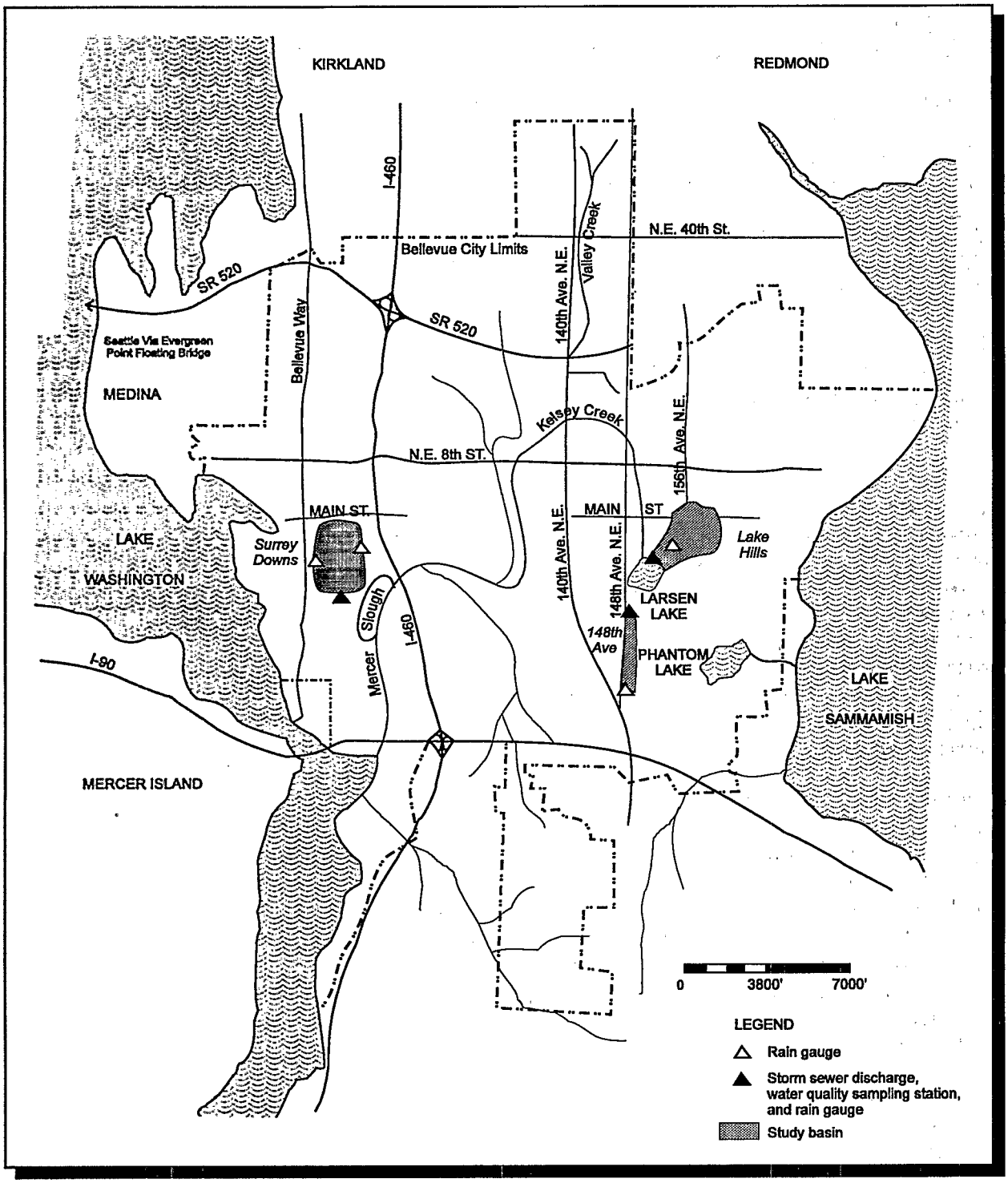


Figure C-3. Bellevue, Washington, NURP stream systems and sampling sites. (Source: USEPA, 1982c)

The following, taken from NURP project summaries (USEPA, 1982c), describes in some detail the sampling approaches and equipment used to meet the monitoring objectives.

The area comprising the Surrey Downs catchment consisted of single family homes and the Bellevue Senior High School. Slopes in the basin were generally moderate, with the exception of the steep slopes on the west side. Surrey Downs was relatively isolated from neighboring communities by the lack of easy vehicular access and convenient "short cuts" through this residential neighborhood.

The Lake Hills catchment contained single family residences and the St. Louise Parish Church and School. Although there were relatively isolated residential areas within the catchment, two through-streets, which carry more traffic than a typical residential street, cross the area.

The 148th Avenue catchment contained 4,960 feet of 148th Avenue, a four-lane, divided arterial street, and some adjacent land with sidewalks, apartments, parking lots, office buildings, and grassy swales that were used as detention basins. A little over one-fourth of the catchment area was taken up by the 148th Avenue street surface.

USGS sample collection and management procedures were essentially the same at all three sites. A digital paper punch recorder recorded: (1) clock time, (2) a number code which indicated if a sample was taken by the automatic sampler, (3) accumulated precipitation in up to three rain gages, and (4) up to two stages for computing discharge. Data were recorded at 5-minute intervals whenever the gage exceeded a [preset] threshold or whenever there was measurable precipitation. Precipitation was measured with tipping-bucket rain gages. Three gages were operated for the Surrey Downs catchment and two each were operated for the Lake Hills and 148th Avenue catchments. Rainfall and dry deposition quality samples were collected at one location in each catchment. Discrete runoff samples were taken during storms for defining the temporal variation of water quality during storm hydrographs. Samples were taken at a preset time interval (5 to 50 minutes) once the stage exceeded a preset threshold.

The Manning composite sampler [used by Bellevue to collect composite flow and proportional stormwater runoff samples] *was triggered at pre-determined increments of flow (300 and 500 cubic feet, the former to obtain more subsamples when small events were expected). The Manning flowmeters used an ultrasonic transducer to sense relative stage. Stage was converted to discharge by a programmed microprocessor in the flowmeter and presented on a circular flow chart as a percentage of maximum rated flow. The microprocessor was programmed from a stage/discharge rating developed by the USGS. Storm samples were removed from the samplers as soon as possible after storms, typically within two or three hours. Samples were kept on ice until pH, conductivity and turbidity were measured in-house. Subsamples were preserved and sent to a contract lab in Seattle for the remaining chemical analysis.*

To obtain street surface particulate samples, the City of Bellevue used the following procedures. Because the street surfaces were more likely to be dry during daylight hours (necessary for good sample collection), collection did not begin before sunrise nor continue after sunset, unless additional personnel were available for traffic control. Subsamples were collected in a narrow strip about six inches wide from one side of the street to the other (curb-to-curb). In heavily traveled streets where

traffic was a problem, some subsamples consisted of two separate half-street strips (curb-to-crown). A pick-up truck was used to carry the equipment components, consisting of a generator, tools, fire extinguisher, vacuum hose and wand, and two wet-dry vacuum units during sample collection. The truck had warning lights, including a roof-top flasher unit.

To carry out the catch basin sampling tasks, all catch basins in each study area were surveyed for location, length, size and slope of pipes, and depth of catchment. Another survey was done to record the dimensions of each catch basin. Sediment volume was then calculated from a measurement of sediment depth.

Some experimental design work was done in 1979 and early 1980 to determine the concentrations of some pollutant constituents. Grab samples of supernatant and sediment were taken from selected catch basins in each study area and submitted to a contract lab for chemical analysis. During 1980, two complete catch basin inventories were made; recording sediment depth, and thus mass loading in the system. Monthly inventories were scheduled for 1981. After December, 1980, spot checks of fifteen to twenty-five selected catch basins in each study area were made after each significant storm event. This information, along with storm and street loading data allowed characterization of flushing and deposition within the sewerage system.

For the toxicant inventory portion of the study, stormwater runoff samples were collected as flow proportioned composites using a Manning S300T automated sampler (all teflon and glass contact surfaces) activated by ultrasonic flowmeters, except for the volatile samples which were collected as grabs early in the storm events. Samplers and containers were cleaned between events according to USEPA protocols using "Micro" brand soap and nitric acid; the hydrochloric acid and methylene chloride rinses were not used. Deionized distilled water blanks were taken through each sampler before use and have proven to be completely clean of organic and metal contaminants. Street surface dust samples were collected as described above using a stainless steel vacuum and PVC flexible hose. No special cleaning protocol was applied to the vacuum. Some sample contamination could have occurred from the PVC hose, but no functional alternatives [have] been found for collecting the dust samples. Interstitial water samples from the stream-bed in Kelsey Creek were collected through aluminum standpipes set in the stream gravel, using a Manning S3000T sampler to draw the water up from the perforated base of the standpipe. This sampling was in conjunction with the "Ecological Impacts of Stormwater Runoff in Urban Streams" project of the University of Washington.

C.2.2 Equipment

For the City's street sampling task various vacuum and hose lengths were tested. Relative air flows and suction pressures in the hose were monitored for different test set-ups. Both one- and two-vacuum configurations and 1.5 inch hoses in lengths varying from 10 to 35 feet were tested, along with a Vacu-Max unit. The standard "reference" system was two vacuums and a 35-foot hose. The best suction and higher air velocities were observed with two vacuums and short hose lengths (10 feet), but the short hose length would require that the vacuums be dismounted from the truck at each subsampling location. The longer hose, with the two vacuums, was judged adequate, and resulted in great cost and time savings.

Two industrial vacuum cleaners (2-hp) with one secondary filter and a primary dacron filter bag were used. The vacuum units were heavy duty and made of stainless steel to reduce contamination of the samples. The two 2-hp vacuums were used together by using a wye connector at the end of the hose. This combination extended the useful length of the 1.5 inch hose to 35 feet and increased the suction.

C.3 CONESTOGA HEADWATERS, PENNSYLVANIA RURAL CLEAN WATER PROGRAM

The following are the key elements of the ground water monitoring program of the Pennsylvania Rural Clean Water Program (Pennsylvania RCWP Coordinating Committee, 1984):

There were four components to the monitoring strategy and which included three scales or levels of monitoring [Figure C-4]. *The first component, the regional network, consisted of general monitoring on a regional scale and included the entire 188 square-mile area. The second component involved more detailed monitoring in a small watershed area of about 5.8 square miles (Little Conestoga Creek*

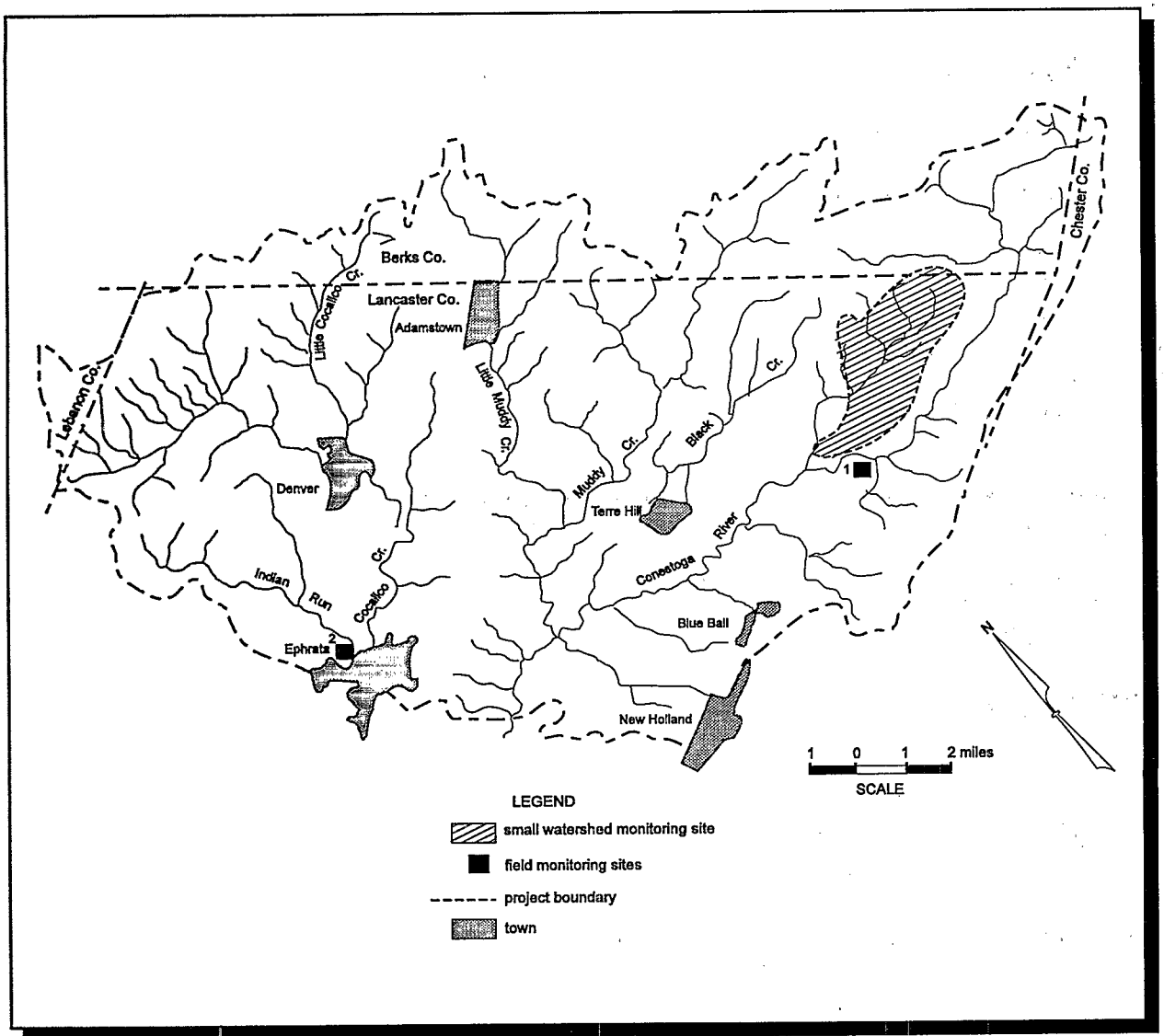


Figure C-4. Location of Conestoga River headwaters area and four monitoring components of the Pennsylvania RCWP. (Source: Pennsylvania RCWP Coordinating Committee, 1984)

basin). The third and fourth components consisted of intensive monitoring on a field scale. Both field sites were on farms.

All four components were designed to permit the comparison of water quality before and after the implementation of BMPs. As such, they included pre-BMP and post-BMP monitoring periods. Although the basic strategy was the same, each component addressed a slightly different problem [Table C-7]. *All four components addressed the problem of nutrients in ground water and surface water; three addressed the sediment problem in surface water; and two addressed the pesticide problem in ground and surface waters. Since the single most important source of water quality problems appeared to be the excess of nutrients, nutrient management BMPs were emphasized in most of the monitoring components.*

C.3.1 Regional Network:

The monitoring schedule for the regional network consisted of three one-year monitoring periods: (1) a pre-BMP period; (2) an early post-BMP period; and (3) later post-BMP period [Table C-8]. *The quality of ground and surface waters, as well as precipitation, were monitored.*

The 43 ground water sites included 42 private domestic, and farm wells and one spring. Of the 43 ground water sites, 33 were located in carbonate rocks [Figure C-5]. *This difference in geology (carbonate vs. noncarbonate areas) played an important role in the quality of water. The areas having the highest levels of nutrients, sediment, pesticides, and bacteria in water consistently coincided with areas underlain by carbonate rocks. This was probably due to the greater number of farms in the carbonate areas and the greater permeability of the carbonate rocks. Because of the greater problems*

Table C-7. Specific problems to be addressed for each monitoring component of the Pennsylvania RCWP.

REGIONAL NETWORK (188 mi²)	
Problem:	What is the composite effect of all implemented BMPs on sediment, nutrients, and pesticides in streams and nutrients and pesticides in ground water?
SMALL WATERSHED SITE-LITTLE CONESTOGA CREEK (5.8 mi²)	
Problem:	What is the effect of a typical combination of BMPs (with emphasis on nutrient management) on sediment and nutrients in the creek and nutrients in ground water?
FILED SITE 1 - DAIRY FARM (21.7 acres)	
Problem:	What is the effect of manure storage and terracing on sediment, nutrients, and pesticides in runoff and nutrients and pesticides in groundwater?
FIELD SITE 2 - HOG/STEER FARM (55 acres)	
Problem:	What is the effect of nutrient management on nutrients in runoff and ground water?

(Source: Pennsylvania RCWP Coordinating Committee, 1984)

in the carbonate areas, the regional network, as well as the other three monitoring components, were concentrated in the carbonate areas.

C.3.2 Small Watershed Site:

In addition to ground water, surface water, and precipitation data, data on nutrients in soils and manure were collected [Table C-9]. *More detailed land-use data were also collected. One gage, four additional base-flow sites, and all the ground water sites were located in the eastern end of the basin where BMP implementation was projected to be greatest* [Figure C-6].

C.3.3 Field Site 1

In addition to the same data that was collected at the small watershed site (with the exception of base-flow data), data were obtained from lysimeters [Table C-10]. *The locations of the data collection facilities at one field site are shown in* [Figure C-7].

C.3.4 Field Site 2

The planned schedule, approach, and data collection were similar to those for field site 1 [Table C-11]. *The only significant difference was that data collection concentrated on nutrients at field site 2 since nutrient management BMPs were planned for this component. The locations of the data collection*

Table C-8. Monitoring plan for regional network of the Pennsylvania RCWP.

SCHEDULE

1982 - 1983 Pre-BMP
1985 - 1986 Early Post-BMP
1988 - 1989 Later Post-BMP

APPROACH

Compare concentrations and/or loads of sediment, nutrients, and pesticides before and after BMP implementation.

WATER QUALITY DATA COLLECTION

2 Stream Gages - sediment, nutrients, and pesticides for major storms

4 Base-flow Sites - monthly sediment, nutrients, and pesticides

43 Ground Water Sites - nutrients and pesticides in spring; summer, and fall

3 Precipitation Gages

LAND USE DATA

Aerial photograph analysis for distribution of major land uses for each of the 3 time periods.

Types and locations of BMPs implemented.

(Source: Pennsylvania RCWP Coordinating Committee, 1984.)

facilities [at this field site] *are shown in* [Figure C-8]. *Field site 2 was in a carbonate area and was monitored for two years for pre-BMP monitoring.*

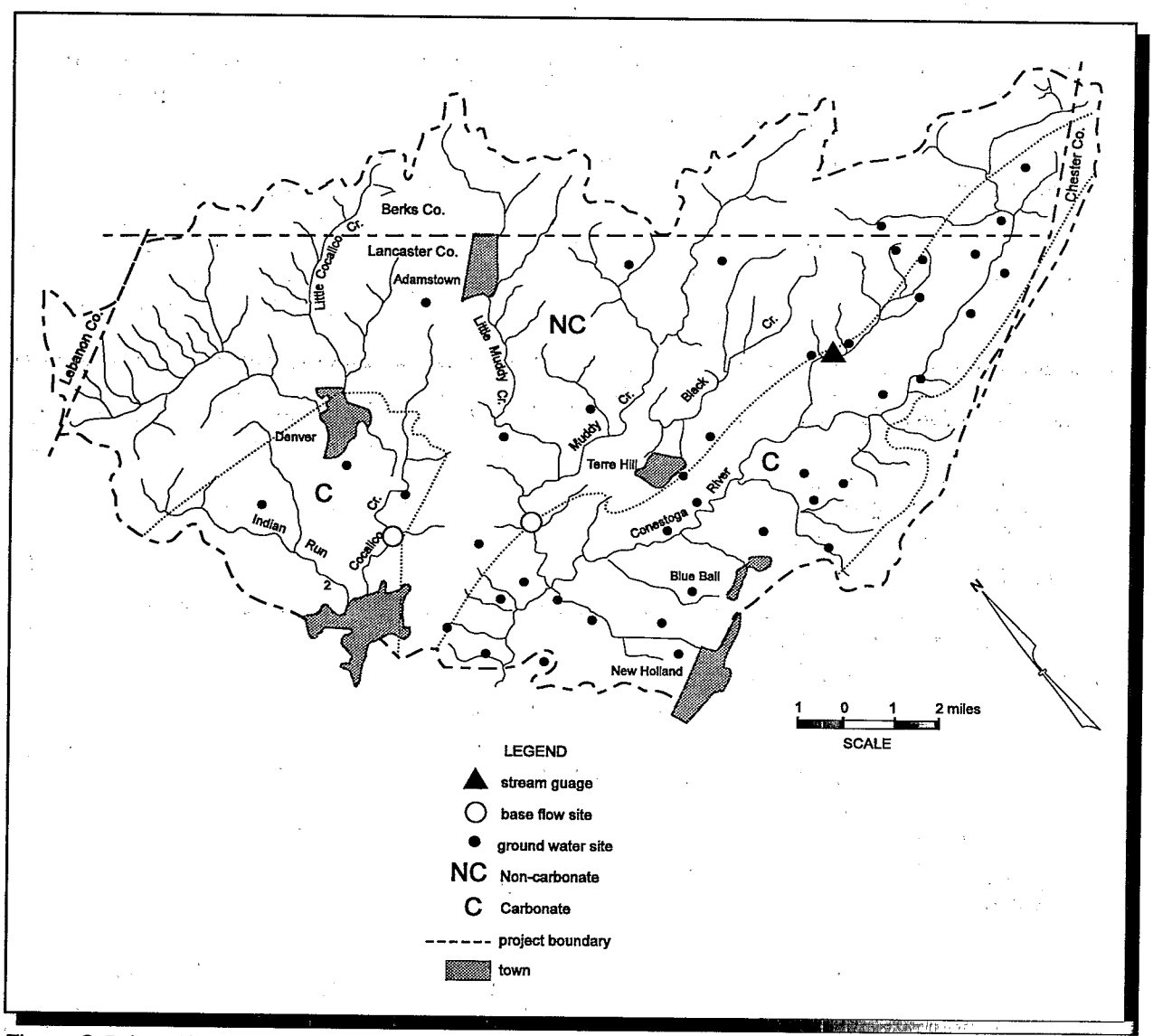


Figure C-5 Location of monitoring facilities for the regional network of the Pennsylvania RCWP. (Source: Pennsylvania RCWP Coordinating Committee, 1984)

Table C-9. Monitoring plan for small watershed site of the Pennsylvania RCWP.

SCHEDULE

1982 - 1985 Pre-BMP
1985 - 1986 BMP
1986 - 1988 Post-BMP

APPROACH

Compare concentrations and loads of sediment and nutrients before and after BMP implementation

WATER-QUALITY DATA COLLECTION

2 Stream Gages - sediment and nutrients for major storms
7 Base-flow Sites - sediment and nutrients every 3 weeks
5-10 Ground Water Sites - nutrients 4 times per year
Soil - spring, fall
Manure - spring
1 Precipitation Gage

LAND-USE DATA

Monthly report from each farmer, including information on:

Plowing	Planting
Harvesting	Field conditions
Liming	Fertilizing
Manure spreading	Pesticide application

(Source: Pennsylvania RCWP Coordinating Committee, 1984)

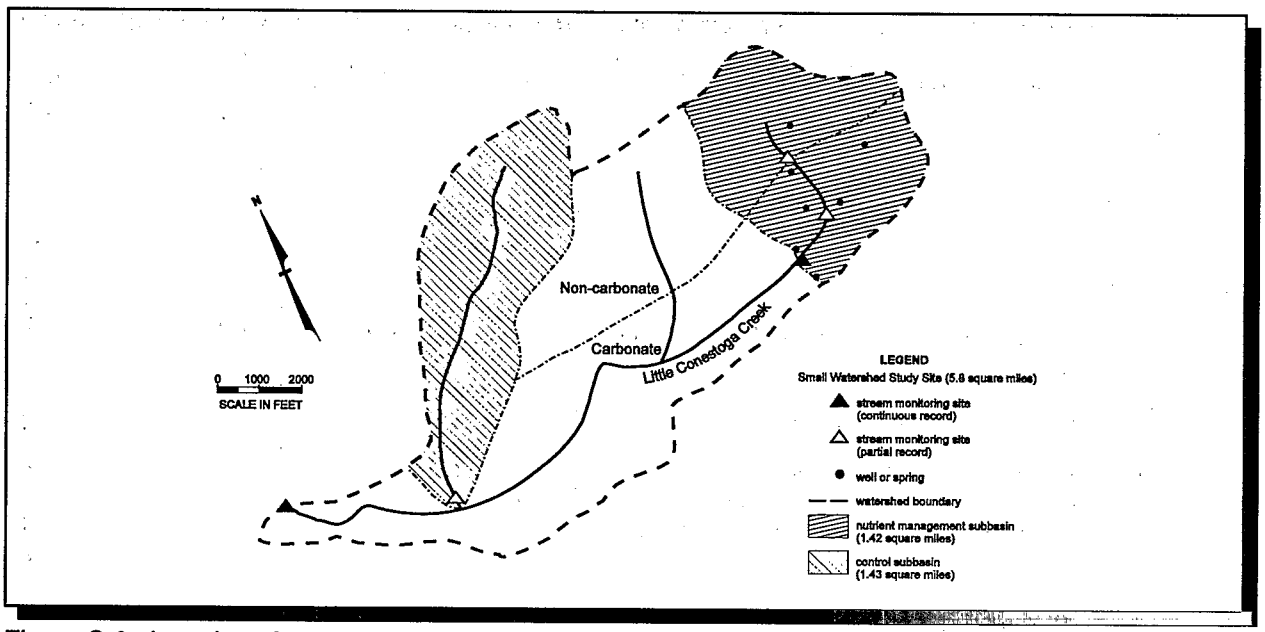


Figure C-6. Location of monitoring facilities for the small watershed site of the Pennsylvania RCWP. (Source: Pennsylvania RCWP Coordinating Committee, 1984)

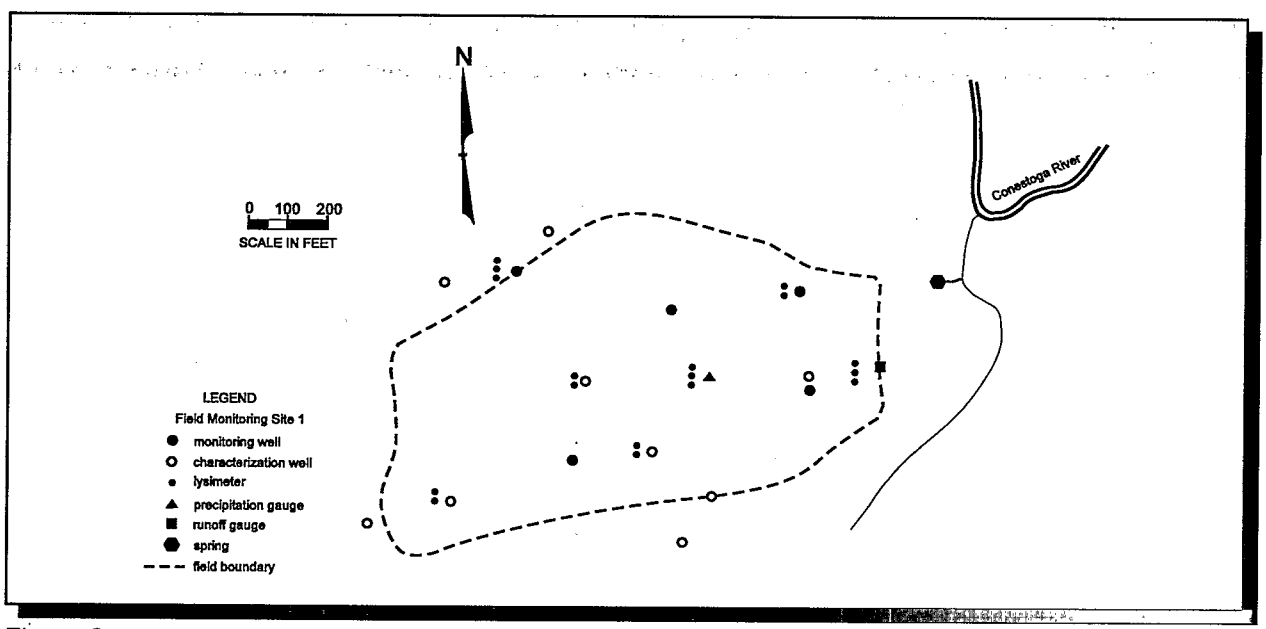


Figure C-7. Location of monitoring facilities for field site 1 of the Pennsylvania RCWP. (Source: Pennsylvania RCWP Coordinating Committee, 1984)

Table C-10. Monitoring plan for field site 1 of the Pennsylvania RCWP.

SCHEDULE

1982 - 1984	Pre-BMP
1984 - 1985	BMP
1985 - 1987	Post-BMP

APPROACH

Compare concentrations and loads of sediment, nutrients, and pesticides before and after BMP implementation.

WATER-QUALITY DATA COLLECTION

1 Runoff Gage - sediment and nutrients for major storms (pesticides for selected storms)

7 Groundwater Sites - nutrients monthly and 4 storms per year (pesticides less frequently)

5-10 Lysimeters - nutrients 4 storms per year (pesticides for selected storms)

Soil - spring, fall

Manure - spring

1 Precipitation Gage

LAND-USE DATA

Biweekly reports from farmer, including information on:

Plowing	Planting
Harvesting	Field conditions
Liming	Fertilizing
Manure spreading	Pesticide application

(Source: Pennsylvania RCWP Coordinating Committee, 1984)

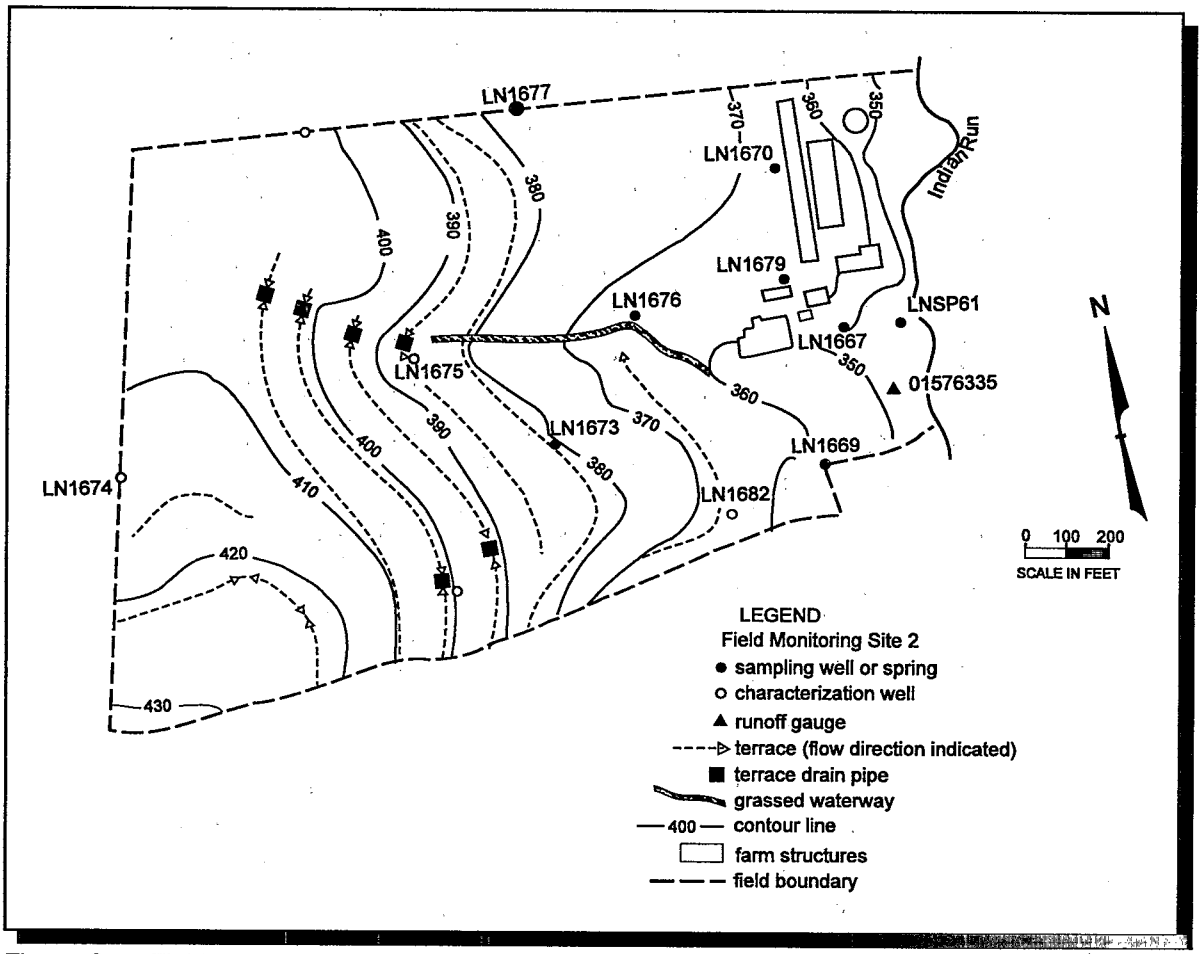


Figure C-8. Field site 2, Lancaster County, Pennsylvania, Pennsylvania RCWP. (Source: Pennsylvania RCWP Coordinating Committee, 1984)

Table C-11. Monitoring plan for field site 2 of the Pennsylvania RCWP.

SCHEDULE

1984 - 1986	Pre-BMP
1984	BMP
1986 - 1988	Post-BMP

APPROACH

Compare concentrations and loads of nutrients before and after BMP implementation.

WATER-QUALITY DATA COLLECTION

1 Runoff Gage - nutrients for major storms
5-10 Groundwater Sites - nutrients monthly and 4 storms per year
5-10 Lysimeters - nutrients 4 storms per year
Soil - spring, fall
Manure - spring
1 Precipitation Gage

LAND-USE DATA

Biweekly report from farmer, including information on:

Plowing	Planting
Harvesting	Field conditions
Liming	Fertilizing
Manure spreading	

(Source: Pennsylvania RCWP Coordinating Committee, 1984)

APPENDIX D

Statistical Tables

D1 Cumulative areas under the Normal distribution (values of p corresponding to Z_p)

D2 Percentiles of the $t_{\alpha,df}$ distribution (values of t such that $100(1-\alpha)\%$ of the distribution is less than t)

D3 Upper and lower percentiles of the Chi-square distribution

D4 Coefficients a_i for the Shapiro-Wilk test for normality

D5 Quantiles of the Shapiro-Wilk test for normality (values of W such that $100p\%$ of the distribution of W is less than W_p)

D6 Upper percentiles of the F distribution for $p = 0.995$, $p = 0.99$, $p = 0.975$, $p = 0.95$, $p = 0.90$

D7 Noncentral t distribution. Values of the noncentrality parameter for a one-sided test with $\alpha = 0.050$, $\alpha = 0.025$, $\alpha = 0.010$

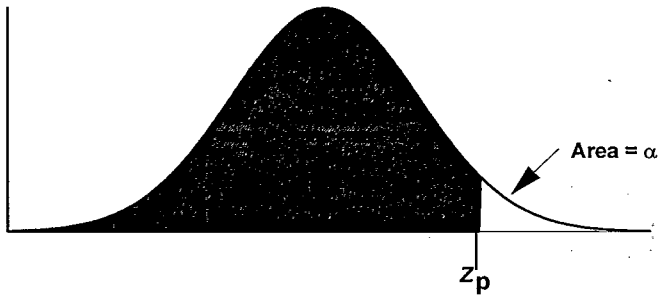
D8 Binomial distribution with y successes out of n trials

D9 Population correlation coefficients

D10 Quantiles of the Spearman test statistic

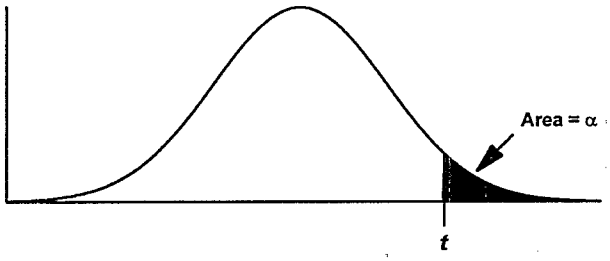
D11 Sample sizes for one-sided nonparametric tolerance limits

Table D1. Cumulative areas under the Normal distribution (values of p corresponding to Z_p)



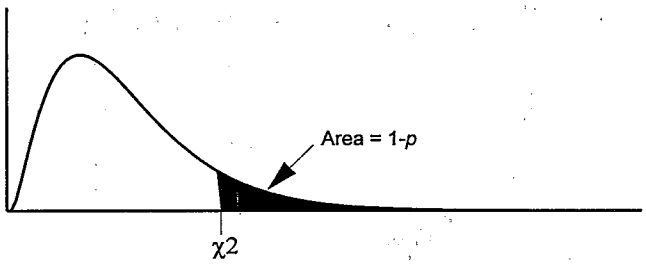
Z_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table D2. Percentiles of the $t_{\alpha,df}$ distribution (values of t such that $100(1-\alpha)\%$ of the distribution is less than t)



df	α = 0.40	α = 0.30	α = 0.20	α = 0.10	α = 0.05	α = 0.025	α = 0.010	α = 0.005
1	0.3249	0.7265	1.3764	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.2887	0.6172	1.0607	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.2707	0.5686	0.9410	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.2590	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.2582	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.2576	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.2571	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.2569	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.2567	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.2566	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.2564	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.2563	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.2562	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.2561	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.2560	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.2559	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.2558	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.2557	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.2556	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500
35	0.2553	0.5292	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238
40	0.2550	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.2547	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.2545	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603
80	0.2542	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387
100	0.2540	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259
150	0.2538	0.5255	0.8440	1.2872	1.6551	1.9759	2.3515	2.6090
200	0.2537	0.5252	0.8434	1.2858	1.6525	1.9719	2.3451	2.6006
inf.	0.2533	0.5244	0.8416	1.2816	1.6449	1.9600	2.3264	2.5758

Table D3. Upper and lower percentiles of the Chi-square distribution



	p											
df	0.001	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995	0.999
1	....	....	....	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	10.827
2	0.002	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	13.815
3	0.024	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	18.466
5	0.210	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	22.457
7	0.599	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	24.321
8	0.857	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819	34.527
14	3.041	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319	36.124
15	3.483	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801	37.698
16	3.942	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718	40.791
18	4.905	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582	43.819
20	5.921	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	45.314
21	6.447	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	46.796
22	6.983	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928	52.619
26	9.222	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290	54.051
27	9.803	11.808	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645	55.475
28	10.391	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672	59.702
35	14.688	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275	66.619
40	17.917	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766	73.403
50	24.674	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490	86.660
60	31.738	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952	99.608
70	39.036	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.43	104.21	112.32
80	46.520	51.172	53.540	57.153	60.391	64.278	96.578	101.88	106.63	112.33	116.32	124.84
90	54.156	59.196	61.754	65.647	69.126	73.291	107.57	113.15	118.14	124.12	128.30	137.21
100	61.918	67.328	70.065	74.222	77.929	82.358	118.50	124.34	129.56	135.81	140.17	149.45
200	143.84	152.24	156.43	162.73	168.28	174.84	226.02	233.99	241.06	249.45	255.26	267.54

Table D4. Coefficients a_i for the Shapiro-Wilk test for normality.
(Source: Shapiro and Wilk, 1965)

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2	-	0.0000	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3	-	-	-	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	-	-	-	-	-	0.0000	0.0561	0.0947	0.1224
5	-	-	-	-	-	-	-	0.0000	0.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	-	-	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	-	-	-	-	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	-	-	-	-	-	-	0.0000	0.0163	0.0303	0.0422
10	-	-	-	-	-	-	-	-	0.0000	0.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	-	-	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	-	-	-	-	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	-	-	-	-	-	-	0.0000	0.0084	0.0159	0.0227
15	-	-	-	-	-	-	-	-	0.0000	0.0076

Table D4. (continued)

f \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2462	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1116	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	-	-	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	-	-	-	-	0.0000	0.0057	0.0110	0.0156	0.0203	0.0244
19	-	-	-	-	-	-	0.0000	0.0053	0.0101	0.0146
20	-	-	-	-	-	-	-	-	0.0000	0.0049

f \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	-	-	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	-	-	-	-	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	-	-	-	-	-	-	0.0000	0.0037	0.0071	0.0104
25	-	-	-	-	-	-	-	-	0.0000	0.0035

Table D5. Quantiles of the Shapiro-Wilk test for normality (values of W such that $100p\%$ of the distribution of W is less than W_p). (Source: Shapiro and Wilk, 1965)

n	$W_{0.01}$	$W_{0.02}$	$W_{0.05}$	$W_{0.10}$	$W_{0.50}$
3	0.753	0.756	0.767	0.789	0.959
4	0.687	0.707	0.748	0.792	0.935
5	0.686	0.715	0.762	0.806	0.927
6	0.713	0.743	0.788	0.826	0.927
7	0.730	0.760	0.803	0.838	0.928
8	0.749	0.778	0.818	0.851	0.932
9	0.764	0.791	0.829	0.859	0.935
10	0.781	0.806	0.842	0.869	0.938
11	0.792	0.817	0.850	0.876	0.940
12	0.805	0.828	0.859	0.883	0.943
13	0.814	0.837	0.866	0.889	0.945
14	0.825	0.846	0.874	0.895	0.947
15	0.835	0.855	0.881	0.901	0.950
16	0.844	0.863	0.887	0.906	0.952
17	0.851	0.869	0.892	0.910	0.954
18	0.858	0.874	0.897	0.914	0.956
19	0.863	0.879	0.901	0.917	0.957
20	0.868	0.884	0.905	0.920	0.959
21	0.873	0.888	0.908	0.923	0.960
22	0.878	0.892	0.911	0.926	0.961
23	0.881	0.895	0.914	0.928	0.962
24	0.884	0.898	0.916	0.930	0.963
25	0.886	0.901	0.918	0.931	0.964
26	0.891	0.904	0.920	0.933	0.965
27	0.894	0.906	0.923	0.935	0.965
28	0.896	0.908	0.924	0.936	0.966
29	0.898	0.910	0.926	0.937	0.966
30	0.900	0.912	0.927	0.939	0.967
31	0.902	0.914	0.929	0.940	0.967
32	0.904	0.915	0.930	0.941	0.968
33	0.906	0.917	0.931	0.942	0.968
34	0.908	0.919	0.933	0.943	0.969
35	0.910	0.920	0.934	0.944	0.969
36	0.912	0.922	0.935	0.945	0.970
37	0.914	0.924	0.936	0.946	0.970
38	0.916	0.925	0.938	0.947	0.971
39	0.917	0.927	0.939	0.948	0.971
40	0.919	0.928	0.940	0.949	0.972
41	0.920	0.929	0.941	0.950	0.972
42	0.922	0.930	0.942	0.951	0.972
43	0.923	0.932	0.943	0.951	0.973
44	0.924	0.933	0.944	0.952	0.973
45	0.926	0.934	0.945	0.953	0.973
46	0.927	0.935	0.945	0.953	0.974
47	0.928	0.936	0.946	0.954	0.974
48	0.929	0.937	0.947	0.954	0.974
49	0.929	0.937	0.947	0.955	0.974
50	0.930	0.938	0.947	0.955	0.974

Table D6. Upper percentiles of the F distribution for $p = 0.995$

f_d \ f_n	1	2	3	4	5	6	7	8	9	10	12
											$\alpha = 0.005$
1	1.6E+4	2.0E+4	2.2E+4	2.3E+4	2.3E+4	2.3E+4	2.4E+4	2.4E+4	2.4E+4	2.4E+4	2.4E+4
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4
3	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.39
4	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.70
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03
7	16.24	12.40	10.88	10.05	9.522	9.155	8.885	8.678	8.514	8.380	8.176
8	14.69	11.04	9.597	8.805	8.302	7.952	7.694	7.496	7.339	7.211	7.015
9	13.61	10.11	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417	6.227
10	12.83	9.427	8.081	7.343	6.872	6.545	6.303	6.116	5.968	5.847	5.661
12	11.75	8.510	7.226	6.521	6.071	5.757	5.524	5.345	5.202	5.085	4.906
15	10.80	7.701	6.476	5.803	5.372	5.071	4.847	4.674	4.536	4.424	4.250
20	9.944	6.987	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.678
24	9.551	6.661	5.519	4.890	4.486	4.202	3.991	3.826	3.695	3.587	3.420
30	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.451	3.344	3.179
40	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	2.953
50	8.626	5.902	4.826	4.232	3.849	3.579	3.376	3.219	3.092	2.988	2.825
60	8.495	5.795	4.729	4.140	3.760	3.492	3.291	3.134	3.008	2.904	2.742
70	8.403	5.720	4.661	4.076	3.698	3.431	3.232	3.076	2.950	2.846	2.684
80	8.335	5.665	4.611	4.028	3.652	3.387	3.188	3.032	2.907	2.803	2.641
100	8.241	5.589	4.542	3.963	3.589	3.325	3.127	2.972	2.847	2.744	2.583
120	8.179	5.539	4.497	3.921	3.548	3.285	3.087	2.933	2.808	2.705	2.544

f_d \ f_n	15	20	24	30	40	50	60	70	80	100	120
1	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4	2.5E+4
2	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	199.5	199.5
3	43.08	42.78	42.62	42.47	42.31	42.21	42.15	42.10	42.07	42.02	41.99
4	20.44	20.17	20.03	19.89	19.75	19.67	19.61	19.57	19.54	19.50	19.47
5	13.15	12.90	12.78	12.66	12.53	12.45	12.40	12.37	12.34	12.30	12.27
6	9.814	9.589	9.474	9.358	9.241	9.170	9.122	9.088	9.062	9.026	9.001
7	7.968	7.754	7.645	7.534	7.422	7.354	7.309	7.276	7.251	7.217	7.193
8	6.814	6.608	6.503	6.396	6.288	6.222	6.177	6.145	6.121	6.087	6.065
9	6.032	5.832	5.729	5.625	5.519	5.454	5.410	5.379	5.356	5.322	5.300
10	5.471	5.274	5.173	5.071	4.966	4.902	4.859	4.828	4.805	4.772	4.750
12	4.721	4.530	4.431	4.331	4.228	4.165	4.123	4.092	4.069	4.037	4.015
15	4.070	3.883	3.786	3.687	3.585	3.523	3.480	3.450	3.427	3.394	3.372
20	3.502	3.318	3.222	3.123	3.022	2.959	2.916	2.885	2.861	2.828	2.806
24	3.246	3.062	2.967	2.868	2.765	2.702	2.658	2.627	2.603	2.569	2.546
30	3.006	2.823	2.727	2.628	2.524	2.459	2.415	2.383	2.358	2.323	2.300
40	2.781	2.598	2.502	2.401	2.296	2.230	2.184	2.150	2.125	2.088	2.064
50	2.653	2.470	2.373	2.272	2.164	2.097	2.050	2.015	1.989	1.951	1.925
60	2.570	2.387	2.290	2.187	2.079	2.010	1.962	1.927	1.900	1.861	1.834
70	2.513	2.329	2.231	2.128	2.019	1.949	1.900	1.864	1.837	1.797	1.769
80	2.470	2.286	2.188	2.084	1.974	1.903	1.854	1.817	1.789	1.748	1.720
100	2.411	2.227	2.128	2.024	1.912	1.840	1.790	1.752	1.723	1.681	1.652
120	2.373	2.188	2.089	1.984	1.871	1.798	1.747	1.709	1.679	1.636	1.606

f_n and f_d equal the degrees of freedom in the numerator and denominator

Table D6. Upper percentiles of the F distribution $p = 0.99$ (continued)

f_d \ f_n	1	2	3	4	5	6	7	8	9	10	12
										$\alpha = 0.01$	
1	4.1E+3	5.0E+3	5.4E+3	5.6E+3	5.8E+3	5.9E+3	5.9E+3	6.0E+3	6.0E+3	6.1E+3	6.1E+3
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89
6	13.75	10.92	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.843
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.563
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496
70	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585	2.450
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.415
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.368
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336

f_d \ f_n	15	20	24	30	40	50	60	70	80	100	120
1	6.2E+3	6.2E+3	6.2E+3	6.3E+3	6.3E+3	6.3E+3	6.3E+3	6.3E+3	6.3E+3	6.3E+3	6.3E+3
2	99.43	99.45	99.46	99.47	99.48	99.48	99.48	99.48	99.48	99.49	99.49
3	26.87	26.69	26.60	26.50	26.41	26.35	26.32	26.29	26.27	26.24	26.22
4	14.20	14.02	13.93	13.84	13.75	13.69	13.65	13.63	13.61	13.58	13.56
5	9.722	9.553	9.466	9.379	9.291	9.238	9.202	9.176	9.157	9.130	9.112
6	7.559	7.396	7.313	7.229	7.143	7.091	7.057	7.032	7.013	6.987	6.969
7	6.314	6.155	6.074	5.992	5.908	5.858	5.824	5.799	5.781	5.755	5.737
8	5.515	5.359	5.279	5.198	5.116	5.065	5.032	5.007	4.989	4.963	4.946
9	4.962	4.808	4.729	4.649	4.567	4.517	4.483	4.459	4.441	4.415	4.398
10	4.558	4.405	4.327	4.247	4.165	4.115	4.082	4.058	4.039	4.014	3.996
12	4.010	3.858	3.780	3.701	3.619	3.569	3.535	3.511	3.493	3.467	3.449
15	3.522	3.372	3.294	3.214	3.132	3.081	3.047	3.022	3.004	2.977	2.959
20	3.088	2.938	2.859	2.778	2.695	2.643	2.608	2.582	2.563	2.535	2.517
24	2.889	2.738	2.659	2.577	2.492	2.440	2.403	2.377	2.357	2.329	2.310
30	2.700	2.549	2.469	2.386	2.299	2.245	2.208	2.181	2.160	2.131	2.111
40	2.522	2.369	2.288	2.203	2.114	2.058	2.019	1.991	1.969	1.938	1.917
50	2.419	2.265	2.183	2.098	2.007	1.949	1.909	1.880	1.857	1.825	1.803
60	2.352	2.198	2.115	2.028	1.936	1.877	1.836	1.806	1.783	1.749	1.726
70	2.306	2.150	2.067	1.980	1.886	1.826	1.785	1.754	1.730	1.695	1.672
80	2.271	2.115	2.032	1.944	1.849	1.788	1.746	1.714	1.690	1.655	1.630
100	2.223	2.067	1.983	1.893	1.797	1.735	1.692	1.659	1.634	1.598	1.572
120	2.191	2.035	1.950	1.860	1.763	1.700	1.656	1.623	1.597	1.559	1.533

f_n and f_d equal the degrees of freedom in the numerator and denominator

Table D6. Upper percentiles of the F distribution for $p = 0.975$ (continued)

f_d \ f_n	1	2	3	4	5	6	7	8	9	10	12
											$\alpha = 0.025$
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	976.7
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34
4	12.22	10.65	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844	8.751
5	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.676
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.541
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.412
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.288
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.216
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.169
70	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237	2.136
80	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213	2.111
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.077
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	2.055

f_d \ f_n	15	20	24	30	40	50	60	70	80	100	120
1	984.9	993.1	997.3	1001	1006	1008	1010	1011	1012	1013	1014
2	39.43	39.45	39.46	39.46	39.47	39.48	39.48	39.48	39.49	39.49	39.49
3	14.25	14.17	14.12	14.08	14.04	14.01	13.99	13.98	13.97	13.96	13.95
4	8.657	8.560	8.511	8.461	8.411	8.381	8.360	8.346	8.335	8.319	8.309
5	6.428	6.329	6.278	6.227	6.175	6.144	6.123	6.107	6.096	6.080	6.069
6	5.269	5.168	5.117	5.065	5.012	4.980	4.959	4.943	4.932	4.915	4.904
7	4.568	4.467	4.415	4.362	4.309	4.276	4.254	4.239	4.227	4.210	4.199
8	4.101	3.999	3.947	3.894	3.840	3.807	3.784	3.768	3.756	3.739	3.728
9	3.769	3.667	3.614	3.560	3.505	3.472	3.449	3.433	3.421	3.403	3.392
10	3.522	3.419	3.365	3.311	3.255	3.221	3.198	3.182	3.169	3.152	3.140
12	3.177	3.073	3.019	2.963	2.906	2.871	2.848	2.831	2.818	2.800	2.787
15	2.862	2.756	2.701	2.644	2.585	2.549	2.524	2.506	2.493	2.474	2.461
20	2.573	2.464	2.408	2.349	2.287	2.249	2.223	2.205	2.190	2.170	2.156
24	2.437	2.327	2.269	2.209	2.146	2.107	2.080	2.060	2.045	2.024	2.010
30	2.307	2.195	2.136	2.074	2.009	1.968	1.940	1.920	1.904	1.882	1.866
40	2.182	2.068	2.007	1.943	1.875	1.832	1.803	1.781	1.764	1.741	1.724
50	2.109	1.993	1.931	1.866	1.796	1.752	1.721	1.698	1.681	1.656	1.639
60	2.061	1.944	1.882	1.815	1.744	1.699	1.667	1.643	1.625	1.599	1.581
70	2.028	1.910	1.847	1.779	1.707	1.660	1.628	1.604	1.585	1.558	1.539
80	2.003	1.884	1.820	1.752	1.679	1.632	1.599	1.574	1.555	1.527	1.508
100	1.968	1.849	1.784	1.715	1.640	1.592	1.558	1.532	1.512	1.483	1.463
120	1.945	1.825	1.760	1.690	1.614	1.565	1.530	1.504	1.483	1.454	1.433

f_n and f_d equal the degrees of freedom in the numerator and denominator

Table D6. Upper percentiles of the F distribution for $p = 0.95$ (continued)

f_d \ f_n	1	2	3	4	5	6	7	8	9	10	12
										$\alpha = 0.05$	
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.745
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.284
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.952
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.893
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.875
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.850
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.834

f_d \ f_n	15	20	24	30	40	50	60	70	80	100	120
1	245.9	248.0	249.1	250.1	251.1	251.8	252.2	252.5	252.7	253.0	253.3
2	19.43	19.45	19.45	19.46	19.47	19.48	19.48	19.48	19.48	19.49	19.49
3	8.703	8.660	8.638	8.617	8.594	8.581	8.572	8.566	8.561	8.554	8.549
4	5.858	5.803	5.774	5.746	5.717	5.699	5.688	5.679	5.673	5.664	5.658
5	4.619	4.558	4.527	4.496	4.464	4.444	4.431	4.422	4.415	4.405	4.398
6	3.938	3.874	3.841	3.808	3.774	3.754	3.740	3.730	3.722	3.712	3.705
7	3.511	3.445	3.410	3.376	3.340	3.319	3.304	3.294	3.286	3.275	3.267
8	3.218	3.150	3.115	3.079	3.043	3.020	3.005	2.994	2.986	2.975	2.967
9	3.006	2.936	2.900	2.864	2.826	2.803	2.787	2.776	2.768	2.756	2.748
10	2.845	2.774	2.737	2.700	2.661	2.637	2.621	2.609	2.601	2.588	2.580
12	2.617	2.544	2.505	2.466	2.426	2.401	2.384	2.372	2.363	2.350	2.341
15	2.403	2.328	2.288	2.247	2.204	2.178	2.160	2.147	2.137	2.123	2.114
20	2.203	2.124	2.082	2.039	1.994	1.966	1.946	1.932	1.922	1.907	1.896
24	2.108	2.027	1.984	1.939	1.892	1.863	1.842	1.828	1.816	1.800	1.790
30	2.015	1.932	1.887	1.841	1.792	1.761	1.740	1.724	1.712	1.695	1.683
40	1.924	1.839	1.793	1.744	1.693	1.660	1.637	1.621	1.608	1.589	1.577
50	1.871	1.784	1.737	1.687	1.634	1.599	1.576	1.558	1.544	1.525	1.511
60	1.836	1.748	1.700	1.649	1.594	1.559	1.534	1.516	1.502	1.481	1.467
70	1.812	1.722	1.674	1.622	1.566	1.530	1.505	1.486	1.471	1.450	1.435
80	1.793	1.703	1.654	1.602	1.545	1.508	1.482	1.463	1.448	1.426	1.411
100	1.768	1.676	1.627	1.573	1.515	1.477	1.450	1.430	1.415	1.392	1.376
120	1.750	1.659	1.608	1.554	1.495	1.457	1.429	1.408	1.392	1.369	1.352

f_n and f_d equal the degrees of freedom in the numerator and denominator

Table D6. Upper percentiles of the F distribution for $p = 0.90$ (continued)

$\alpha = 0.1$

f_d \ f_n	1	2	3	4	5	6	7	8	9	10	12
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.408
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.216
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.896
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.268
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.905
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.668
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.502
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.379
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.284
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.147
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.017
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.892
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.832
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.773
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.715
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.680
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.657
70	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691	1.641
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680	1.629
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.612
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.601

f_d \ f_n	15	20	24	30	40	50	60	70	80	100	120
1	61.22	61.74	62.00	62.26	62.53	62.69	62.79	62.87	62.93	63.01	63.06
2	9.425	9.441	9.450	9.458	9.466	9.471	9.475	9.477	9.479	9.481	9.483
3	5.200	5.184	5.176	5.168	5.160	5.155	5.151	5.149	5.147	5.144	5.143
4	3.870	3.844	3.831	3.817	3.804	3.795	3.790	3.786	3.782	3.778	3.775
5	3.238	3.207	3.191	3.174	3.157	3.147	3.140	3.135	3.132	3.126	3.123
6	2.871	2.836	2.818	2.800	2.781	2.770	2.762	2.756	2.752	2.746	2.742
7	2.632	2.595	2.575	2.555	2.535	2.523	2.514	2.508	2.504	2.497	2.493
8	2.464	2.425	2.404	2.383	2.361	2.348	2.339	2.333	2.328	2.321	2.316
9	2.340	2.298	2.277	2.255	2.232	2.218	2.208	2.202	2.196	2.189	2.184
10	2.244	2.201	2.178	2.155	2.132	2.117	2.107	2.100	2.095	2.087	2.082
12	2.105	2.060	2.036	2.011	1.986	1.970	1.960	1.952	1.946	1.938	1.932
15	1.972	1.924	1.899	1.873	1.845	1.828	1.817	1.808	1.802	1.793	1.787
20	1.845	1.794	1.767	1.738	1.708	1.690	1.677	1.667	1.660	1.650	1.643
24	1.783	1.730	1.702	1.672	1.641	1.621	1.607	1.597	1.590	1.579	1.571
30	1.722	1.667	1.638	1.606	1.573	1.552	1.538	1.527	1.519	1.507	1.499
40	1.662	1.605	1.574	1.541	1.506	1.483	1.467	1.455	1.447	1.434	1.425
50	1.627	1.568	1.536	1.502	1.465	1.441	1.424	1.412	1.402	1.388	1.379
60	1.603	1.543	1.511	1.476	1.437	1.413	1.395	1.382	1.372	1.358	1.348
70	1.587	1.526	1.493	1.457	1.418	1.392	1.374	1.361	1.350	1.335	1.325
80	1.574	1.513	1.479	1.443	1.403	1.377	1.358	1.344	1.334	1.318	1.307
100	1.557	1.494	1.460	1.423	1.382	1.355	1.336	1.321	1.310	1.293	1.282
120	1.545	1.482	1.447	1.409	1.368	1.340	1.320	1.305	1.294	1.277	1.265

f_n and f_d equal the degrees of freedom in the numerator and denominator

Table D7. Noncentral t distribution. Values of the noncentrality parameter for a one-sided test with $\alpha = 0.050$. (Source: Owen, 1965)

f	β = type 2 error										
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
1	16.47	12.53	10.51	8.19	6.63	5.38	4.31	3.35	2.46	1.60	.64
2	6.88	5.52	4.81	3.98	3.40	2.92	2.49	2.07	1.63	1.15	.50
3	5.47	4.46	3.93	3.30	2.85	2.48	2.13	1.79	1.43	1.02	.46
4	4.95	4.07	3.60	3.04	2.64	2.30	1.99	1.67	1.34	.96	.43
5	4.70	3.87	3.43	2.90	2.53	2.21	1.91	1.61	1.29	.92	.42
6	4.55	3.75	3.33	2.82	2.46	2.15	1.86	1.57	1.26	.90	.41
7	4.45	3.67	3.26	2.77	2.41	2.11	1.82	1.54	1.24	.89	.40
8	4.38	3.62	3.21	2.73	2.38	2.08	1.80	1.52	1.22	.88	.40
9	4.32	3.58	3.18	2.70	2.35	2.06	1.78	1.51	1.21	.87	.39
10	4.28	3.54	3.15	2.67	2.33	2.04	1.77	1.49	1.20	.86	.39
11	4.25	3.52	3.13	2.65	2.31	2.02	1.75	1.48	1.19	.86	.39
12	4.22	3.50	3.11	2.64	2.30	2.01	1.74	1.47	1.19	.85	.38
13	4.20	3.48	3.09	2.63	2.29	2.00	1.74	1.47	1.18	.85	.38
14	4.18	3.46	3.08	2.62	2.28	2.00	1.73	1.46	1.18	.84	.38
15	4.17	3.45	3.07	2.61	2.27	1.99	1.72	1.46	1.17	.84	.38
16	4.16	3.44	3.06	2.60	2.27	1.98	1.72	1.45	1.17	.84	.38
17	4.14	3.43	3.05	2.59	2.26	1.98	1.71	1.45	1.17	.84	.38
18	4.13	3.42	3.04	2.59	2.26	1.97	1.71	1.45	1.16	.83	.38
19	4.12	3.41	3.04	2.58	2.25	1.97	1.71	1.44	1.16	.83	.38
20	4.12	3.41	3.03	2.58	2.25	1.97	1.70	1.44	1.16	.83	.38
21	4.11	3.40	3.03	2.57	2.24	1.96	1.70	1.44	1.16	.83	.38
22	4.10	3.40	3.02	2.57	2.24	1.96	1.70	1.44	1.16	.83	.37
23	4.10	3.39	3.02	2.56	2.24	1.96	1.69	1.43	1.15	.83	.37
24	4.09	3.39	3.01	2.56	2.23	1.95	1.69	1.43	1.15	.83	.37
25	4.09	3.38	3.01	2.56	2.23	1.95	1.69	1.43	1.15	.83	.37
26	4.08	3.38	3.01	2.55	2.23	1.95	1.69	1.43	1.15	.82	.37
27	4.08	3.38	3.00	2.55	2.23	1.95	1.69	1.43	1.15	.82	.37
28	4.07	3.37	3.00	2.55	2.22	1.95	1.69	1.43	1.15	.82	.37
29	4.07	3.37	3.00	2.55	2.22	1.94	1.68	1.42	1.15	.82	.37
30	4.07	3.37	3.00	2.54	2.22	1.94	1.68	1.42	1.15	.82	.37
40	4.04	3.35	2.98	2.53	2.21	1.93	1.67	1.42	1.14	.82	.37
60	4.02	3.33	2.96	2.52	2.19	1.92	1.66	1.41	1.13	.81	.37
100	4.00	3.31	2.95	2.50	2.18	1.91	1.66	1.40	1.13	.81	.37
∞	3.97	3.29	2.93	2.49	2.17	1.90	1.64	1.39	1.12	.80	.36

$\Pr[\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)] = 1 - \beta.$

Table D7. (continued) **Values of the noncentrality parameter for a one-sided test with $\alpha = 0.025$.** (Source: Owen, 1965)

f	β = type 2 error										
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
1	32.83	24.98	20.96	16.33	13.21	10.73	8.60	6.68	4.91	3.22	1.58
2	9.67	7.77	6.80	5.65	4.86	4.21	3.63	3.07	2.50	1.88	1.09
3	6.88	5.65	5.01	4.26	3.72	3.28	2.87	2.47	2.05	1.57	.94
4	5.94	4.93	4.40	3.76	3.31	2.93	2.58	2.23	1.86	1.44	.87
5	5.49	4.57	4.09	3.51	3.10	2.75	2.43	2.11	1.76	1.37	.82
6	5.22	4.37	3.91	3.37	2.98	2.64	2.34	2.03	1.70	1.32	.80
7	5.06	4.23	3.80	3.27	2.89	2.57	2.27	1.98	1.66	1.29	.78
8	4.94	4.14	3.71	3.20	2.83	2.52	2.23	1.94	1.63	1.27	.77
9	4.85	4.07	3.65	3.15	2.79	2.48	2.20	1.91	1.60	1.25	.76
10	4.78	4.01	3.60	3.11	2.75	2.45	2.17	1.89	1.59	1.23	.75
11	4.73	3.97	3.57	3.08	2.73	2.43	2.15	1.87	1.57	1.22	.74
12	4.69	3.93	3.54	3.05	2.70	2.41	2.13	1.85	1.56	1.21	.74
13	4.65	3.91	3.51	3.03	2.69	2.39	2.12	1.84	1.55	1.21	.73
14	4.62	3.88	3.49	3.01	2.67	2.38	2.11	1.83	1.54	1.20	.73
15	4.60	3.86	3.47	3.00	2.66	2.37	2.09	1.82	1.53	1.19	.72
16	4.58	3.84	3.46	2.98	2.65	2.36	2.09	1.81	1.53	1.19	.72
17	4.56	3.83	3.44	2.97	2.64	2.35	2.08	1.81	1.52	1.18	.72
18	4.54	3.82	3.43	2.96	2.63	2.34	2.07	1.80	1.52	1.18	.72
19	4.52	3.80	3.42	2.95	2.61	2.33	2.06	1.80	1.51	1.17	.71
20	4.51	3.79	3.41	2.95	2.61	2.33	2.06	1.79	1.51	1.17	.71
21	4.50	3.78	3.40	2.93	2.60	2.32	2.05	1.79	1.50	1.17	.71
22	4.49	3.77	3.39	2.93	2.60	2.32	2.05	1.78	1.50	1.17	.71
23	4.48	3.77	3.39	2.93	2.59	2.31	2.05	1.78	1.50	1.17	.71
24	4.47	3.76	3.38	2.92	2.59	2.31	2.04	1.78	1.50	1.16	.71
25	4.46	3.75	3.37	2.92	2.58	2.30	2.04	1.77	1.49	1.16	.71
26	4.46	3.75	3.37	2.92	2.58	2.30	2.04	1.77	1.49	1.16	.70
27	4.45	3.74	3.36	2.91	2.58	2.30	2.03	1.77	1.49	1.16	.70
28	4.44	3.73	3.36	2.90	2.57	2.29	2.03	1.77	1.49	1.16	.70
29	4.44	3.73	3.35	2.90	2.57	2.29	2.03	1.77	1.48	1.16	.70
30	4.43	3.73	3.35	2.90	2.57	2.29	2.02	1.76	1.48	1.16	.70
40	4.39	3.69	3.32	2.87	2.55	2.27	2.01	1.75	1.47	1.15	.69
60	4.36	3.66	3.29	2.85	2.53	2.25	1.99	1.73	1.46	1.14	.69
100	4.33	3.64	3.27	2.83	2.51	2.23	1.98	1.73	1.45	1.12	.68
∞	4.29	3.60	3.24	2.80	2.48	2.21	1.96	1.71	1.44	1.12	.68

$\Pr[\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)] = 1 - \beta.$

Table D7. (continued) **Values of the noncentrality parameter for a one-sided test with $\alpha = 0.010$.** (Source: Owen, 1965)

f	β = type 2 error										
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
1	82.00	62.40	52.37	40.80	33.00	26.79	21.47	16.69	12.27	8.07	4.00
2	15.22	12.26	10.74	8.96	7.73	6.73	5.83	4.98	4.12	3.20	2.08
3	9.34	7.71	6.86	5.87	5.17	4.59	4.07	3.56	3.03	2.44	1.66
4	7.52	6.28	5.64	4.88	4.34	3.88	3.47	3.06	2.63	2.14	1.48
5	6.68	5.62	5.07	4.40	3.93	3.54	3.17	2.81	2.42	1.98	1.38
6	6.21	5.25	4.74	4.13	3.70	3.33	2.99	2.66	2.30	1.88	1.32
7	5.91	5.01	4.53	3.96	3.55	3.20	2.88	2.56	2.22	1.82	1.27
8	5.71	4.85	4.39	3.84	3.44	3.11	2.80	2.49	2.16	1.77	1.24
9	5.56	4.72	4.28	3.75	3.37	3.04	2.74	2.43	2.11	1.74	1.22
10	5.45	4.63	4.20	3.68	3.31	2.99	2.69	2.39	2.08	1.71	1.20
11	5.36	4.56	4.14	3.63	3.26	2.94	2.65	2.36	2.05	1.69	1.18
12	5.29	4.50	4.09	3.58	3.22	2.91	2.62	2.33	2.03	1.67	1.17
13	5.23	4.46	4.04	3.55	3.19	2.88	2.60	2.31	2.01	1.65	1.16
14	5.18	4.42	4.01	3.51	3.16	2.86	2.57	2.29	1.99	1.64	1.15
15	5.14	4.38	3.98	3.49	3.14	2.84	2.56	2.28	1.98	1.63	1.14
16	5.11	4.35	3.95	3.47	3.12	2.82	2.54	2.26	1.97	1.62	1.14
17	5.08	4.33	3.93	3.45	3.10	2.80	2.53	2.25	1.96	1.61	1.13
18	5.05	4.31	3.91	3.43	3.09	2.79	2.52	2.24	1.95	1.60	1.13
19	5.03	4.29	3.89	3.42	3.07	2.78	2.50	2.23	1.94	1.60	1.12
20	5.01	4.27	3.88	3.40	3.06	2.77	2.50	2.22	1.93	1.59	1.12
21	4.99	4.25	3.86	3.39	3.05	2.76	2.49	2.22	1.92	1.59	1.11
22	4.97	4.24	3.85	3.38	3.04	2.75	2.48	2.21	1.92	1.58	1.11
23	4.96	4.23	3.84	3.37	3.03	2.74	2.47	2.20	1.91	1.58	1.11
24	4.94	4.22	3.83	3.36	3.02	2.73	2.47	2.20	1.91	1.57	1.11
25	4.93	4.20	3.82	3.35	3.02	2.73	2.46	2.19	1.90	1.57	1.10
26	4.92	4.19	3.81	3.34	3.01	2.72	2.45	2.19	1.90	1.57	1.10
27	4.91	4.19	3.80	3.34	3.00	2.72	2.45	2.18	1.90	1.56	1.10
28	4.90	4.18	3.79	3.33	3.00	2.71	2.44	2.18	1.89	1.56	1.10
29	4.89	4.17	3.79	3.32	2.99	2.71	2.44	2.17	1.89	1.56	1.10
30	4.88	4.16	3.78	3.32	2.99	2.70	2.44	2.17	1.89	1.55	1.09
40	4.82	4.11	3.74	3.28	2.95	2.67	2.41	2.15	1.86	1.54	1.08
60	4.76	4.06	3.69	3.24	2.92	2.64	2.38	2.12	1.84	1.52	1.07
100	4.72	4.03	3.66	3.21	2.89	2.62	2.36	2.10	1.83	1.51	1.06
∞	4.65	3.97	3.61	3.17	2.85	2.58	2.33	2.07	1.80	1.48	1.04

$\Pr[\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)] = 1 - \beta.$

Table D8. Binomial distribution with y successes out of n trials.

		$p(Y<y)$									
n	y	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$								
n	y	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
1	0	0.4500	0.4000	0.3500	0.3000	0.2500	0.2000	0.1500	0.1000	0.0500
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.2025	0.1600	0.1225	0.0900	0.0625	0.0400	0.0225	0.0100	0.0025
	1	0.6975	0.6400	0.5775	0.5100	0.4375	0.3600	0.2775	0.1900	0.0975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.0911	0.0640	0.0429	0.0270	0.0156	0.0080	0.0034	0.0010	0.0001
	1	0.4253	0.3520	0.2818	0.2160	0.1563	0.1040	0.0607	0.0280	0.0072
	2	0.8336	0.7840	0.7254	0.6570	0.5781	0.4880	0.3859	0.2710	0.1426
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.0410	0.0256	0.0150	0.0081	0.0039	0.0016	0.0005	0.0001	0.0000
	1	0.2415	0.1792	0.1265	0.0837	0.0508	0.0272	0.0120	0.0037	0.0005
	2	0.6090	0.5248	0.4370	0.3483	0.2617	0.1808	0.1095	0.0523	0.0140
	3	0.9085	0.8704	0.8215	0.7599	0.6836	0.5904	0.4780	0.3439	0.1855
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.0185	0.0102	0.0053	0.0024	0.0010	0.0003	0.0001	0.0000	0.0000
	1	0.1312	0.0870	0.0540	0.0308	0.0156	0.0067	0.0022	0.0005	0.0000
	2	0.4069	0.3174	0.2352	0.1631	0.1035	0.0579	0.0266	0.0086	0.0012
	3	0.7438	0.6630	0.5716	0.4718	0.3672	0.2627	0.1648	0.0815	0.0226
	4	0.9497	0.9222	0.8840	0.8319	0.7627	0.6723	0.5563	0.4095	0.2262
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.0083	0.0041	0.0018	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
	1	0.0692	0.0410	0.0223	0.0109	0.0046	0.0016	0.0004	0.0001	0.0000
	2	0.2553	0.1792	0.1174	0.0705	0.0376	0.0170	0.0059	0.0013	0.0001
	3	0.5585	0.4557	0.3529	0.2557	0.1694	0.0989	0.0473	0.0158	0.0022
	4	0.8364	0.7667	0.6809	0.5798	0.4661	0.3446	0.2235	0.1143	0.0328
	5	0.9723	0.9533	0.9246	0.8824	0.8220	0.7379	0.6229	0.4686	0.2649
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.0037	0.0016	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.0357	0.0188	0.0090	0.0038	0.0013	0.0004	0.0001	0.0000	0.0000
	2	0.1529	0.0963	0.0556	0.0288	0.0129	0.0047	0.0012	0.0002	0.0000
	3	0.3917	0.2898	0.1998	0.1260	0.0706	0.0333	0.0121	0.0027	0.0002
	4	0.6836	0.5801	0.4677	0.3529	0.2436	0.1480	0.0738	0.0257	0.0038
	5	0.8976	0.8414	0.7662	0.6706	0.5551	0.4233	0.2834	0.1497	0.0444
	6	0.9848	0.9720	0.9510	0.9176	0.8665	0.7903	0.6794	0.5217	0.3017
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0181	0.0085	0.0036	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000
	2	0.0885	0.0498	0.0253	0.0113	0.0042	0.0012	0.0002	0.0000	0.0000
	3	0.2604	0.1737	0.1061	0.0580	0.0273	0.0104	0.0029	0.0004	0.0000
	4	0.5230	0.4059	0.2936	0.1941	0.1138	0.0563	0.0214	0.0050	0.0004
	5	0.7799	0.6846	0.5722	0.4482	0.3215	0.2031	0.1052	0.0381	0.0058
	6	0.9368	0.8936	0.8309	0.7447	0.6329	0.4967	0.3428	0.1869	0.0572
	7	0.9916	0.9832	0.9681	0.9424	0.8999	0.8322	0.7275	0.5695	0.3366
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$									
n	y	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$								
n	y	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
9	0	0.0008	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0091	0.0038	0.0014	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.0498	0.0250	0.0112	0.0043	0.0013	0.0003	0.0000	0.0000	0.0000
	3	0.1658	0.0994	0.0536	0.0253	0.0100	0.0031	0.0006	0.0001	0.0000
	4	0.3786	0.2666	0.1717	0.0988	0.0489	0.0196	0.0056	0.0009	0.0000
	5	0.6386	0.5174	0.3911	0.2703	0.1657	0.0856	0.0339	0.0083	0.0006
	6	0.8505	0.7682	0.6627	0.5372	0.3993	0.2618	0.1409	0.0530	0.0084
	7	0.9615	0.9295	0.8789	0.8040	0.6997	0.5638	0.4005	0.2252	0.0712
	8	0.9954	0.9899	0.9793	0.9596	0.9249	0.8658	0.7684	0.6126	0.3698
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0045	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0274	0.0123	0.0048	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
	3	0.1020	0.0548	0.0260	0.0106	0.0035	0.0009	0.0001	0.0000	0.0000
	4	0.2616	0.1662	0.0949	0.0473	0.0197	0.0064	0.0014	0.0001	0.0000
	5	0.4956	0.3669	0.2485	0.1503	0.0781	0.0328	0.0099	0.0016	0.0001
	6	0.7340	0.6177	0.4862	0.3504	0.2241	0.1209	0.0500	0.0128	0.0010
	7	0.9004	0.8327	0.7384	0.6172	0.4744	0.3222	0.1798	0.0702	0.0115
	8	0.9767	0.9536	0.9140	0.8507	0.7560	0.6242	0.4557	0.2639	0.0861
	9	0.9975	0.9940	0.9865	0.9718	0.9437	0.8926	0.8031	0.6513	0.4013
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0022	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0148	0.0059	0.0020	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.0610	0.0293	0.0122	0.0043	0.0012	0.0002	0.0000	0.0000	0.0000
	4	0.1738	0.0994	0.0501	0.0216	0.0076	0.0020	0.0003	0.0000	0.0000
	5	0.3669	0.2465	0.1487	0.0782	0.0343	0.0117	0.0027	0.0003	0.0000
	6	0.6029	0.4672	0.3317	0.2103	0.1146	0.0504	0.0159	0.0028	0.0001
	7	0.8089	0.7037	0.5744	0.4304	0.2867	0.1611	0.0694	0.0185	0.0016
	8	0.9348	0.8811	0.7999	0.6873	0.5448	0.3826	0.2212	0.0896	0.0152
	9	0.9861	0.9698	0.9394	0.8870	0.8029	0.6779	0.5078	0.3026	0.1019
	10	0.9986	0.9964	0.9912	0.9802	0.9578	0.9141	0.8327	0.6862	0.4312
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0079	0.0028	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0356	0.0153	0.0056	0.0017	0.0004	0.0001	0.0000	0.0000	0.0000
	4	0.1117	0.0573	0.0255	0.0095	0.0028	0.0006	0.0001	0.0000	0.0000
	5	0.2607	0.1582	0.0846	0.0386	0.0143	0.0039	0.0007	0.0001	0.0000
	6	0.4731	0.3348	0.2127	0.1178	0.0544	0.0194	0.0046	0.0005	0.0000
	7	0.6956	0.5618	0.4167	0.2763	0.1576	0.0726	0.0239	0.0043	0.0002
	8	0.8655	0.7747	0.6533	0.5075	0.3512	0.2054	0.0922	0.0256	0.0022
	9	0.9579	0.9166	0.8487	0.7472	0.6093	0.4417	0.2642	0.1109	0.0196
	10	0.9917	0.9804	0.9576	0.9150	0.8416	0.7251	0.5565	0.3410	0.1184
	11	0.9992	0.9978	0.9943	0.9862	0.9683	0.9313	0.8578	0.7176	0.4596
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$									
n	y	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$								
n	y	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
13	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0041	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0203	0.0078	0.0025	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	4	0.0698	0.0321	0.0126	0.0040	0.0010	0.0002	0.0000	0.0000	0.0000
	5	0.1788	0.0977	0.0462	0.0182	0.0056	0.0012	0.0002	0.0000	0.0000
	6	0.3563	0.2288	0.1295	0.0624	0.0243	0.0070	0.0013	0.0001	0.0000
	7	0.5732	0.4256	0.2841	0.1654	0.0802	0.0300	0.0075	0.0009	0.0000
	8	0.7721	0.6470	0.4995	0.3457	0.2060	0.0991	0.0342	0.0065	0.0003
	9	0.9071	0.8314	0.7217	0.5794	0.4157	0.2527	0.1180	0.0342	0.0031
	10	0.9731	0.9421	0.8868	0.7975	0.6674	0.4983	0.3080	0.1339	0.0245
	11	0.9951	0.9874	0.9704	0.9363	0.8733	0.7664	0.6017	0.3787	0.1354
	12	0.9996	0.9987	0.9963	0.9903	0.9762	0.9450	0.8791	0.7458	0.4867
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0022	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0114	0.0039	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0426	0.0175	0.0060	0.0017	0.0003	0.0000	0.0000	0.0000	0.0000
	5	0.1189	0.0583	0.0243	0.0083	0.0022	0.0004	0.0000	0.0000	0.0000
	6	0.2586	0.1501	0.0753	0.0315	0.0103	0.0024	0.0003	0.0000	0.0000
	7	0.4539	0.3075	0.1836	0.0933	0.0383	0.0116	0.0022	0.0002	0.0000
	8	0.6627	0.5141	0.3595	0.2195	0.1117	0.0439	0.0115	0.0015	0.0000
	9	0.8328	0.7207	0.5773	0.4158	0.2585	0.1298	0.0467	0.0092	0.0004
	10	0.9368	0.8757	0.7795	0.6448	0.4787	0.3018	0.1465	0.0441	0.0042
	11	0.9830	0.9602	0.9161	0.8392	0.7189	0.5519	0.3521	0.1584	0.0301
	12	0.9971	0.9919	0.9795	0.9525	0.8990	0.8021	0.6433	0.4154	0.1530
	13	0.9998	0.9992	0.9976	0.9932	0.9822	0.9560	0.8972	0.7712	0.5123
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0063	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0255	0.0093	0.0028	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	5	0.0769	0.0338	0.0124	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000
	6	0.1818	0.0950	0.0422	0.0152	0.0042	0.0008	0.0001	0.0000	0.0000
	7	0.3465	0.2131	0.1132	0.0500	0.0173	0.0042	0.0006	0.0000	0.0000
	8	0.5478	0.3902	0.2452	0.1311	0.0566	0.0181	0.0036	0.0003	0.0000
	9	0.7392	0.5968	0.4357	0.2784	0.1484	0.0611	0.0168	0.0022	0.0001
	10	0.8796	0.7827	0.6481	0.4845	0.3135	0.1642	0.0617	0.0127	0.0006
	11	0.9576	0.9095	0.8273	0.7031	0.5387	0.3518	0.1773	0.0556	0.0055
	12	0.9893	0.9729	0.9383	0.8732	0.7639	0.6020	0.3958	0.1841	0.0362
	13	0.9983	0.9948	0.9858	0.9647	0.9198	0.8329	0.6814	0.4510	0.1710
	14	0.9999	0.9995	0.9984	0.9953	0.9866	0.9648	0.9126	0.7941	0.5367
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$									
n	y	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$								
n	y	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
16	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0035	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0149	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0486	0.0191	0.0062	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000
	6	0.1241	0.0583	0.0229	0.0071	0.0016	0.0002	0.0000	0.0000	0.0000
	7	0.2559	0.1423	0.0671	0.0257	0.0075	0.0015	0.0002	0.0000	0.0000
	8	0.4371	0.2839	0.1594	0.0744	0.0271	0.0070	0.0011	0.0001	0.0000
	9	0.6340	0.4728	0.3119	0.1753	0.0796	0.0267	0.0056	0.0005	0.0000
	10	0.8024	0.6712	0.5100	0.3402	0.1897	0.0817	0.0235	0.0033	0.0001
	11	0.9147	0.8334	0.7108	0.5501	0.3698	0.2018	0.0791	0.0170	0.0009
	12	0.9719	0.9349	0.8661	0.7541	0.5950	0.4019	0.2101	0.0684	0.0070
	13	0.9934	0.9817	0.9549	0.9006	0.8029	0.6482	0.4386	0.2108	0.0429
	14	0.9990	0.9967	0.9902	0.9739	0.9365	0.8593	0.7161	0.4853	0.1892
	15	0.9999	0.9997	0.9990	0.9967	0.9900	0.9719	0.9257	0.8147	0.5599
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0086	0.0025	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0301	0.0106	0.0030	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	6	0.0826	0.0348	0.0120	0.0032	0.0006	0.0001	0.0000	0.0000	0.0000
	7	0.1834	0.0919	0.0383	0.0127	0.0031	0.0005	0.0000	0.0000	0.0000
	8	0.3374	0.1989	0.0994	0.0403	0.0124	0.0026	0.0003	0.0000	0.0000
	9	0.5257	0.3595	0.2128	0.1046	0.0402	0.0109	0.0017	0.0001	0.0000
	10	0.7098	0.5522	0.3812	0.2248	0.1071	0.0377	0.0083	0.0008	0.0000
	11	0.8529	0.7361	0.5803	0.4032	0.2347	0.1057	0.0319	0.0047	0.0001
	12	0.9404	0.8740	0.7652	0.6113	0.4261	0.2418	0.0987	0.0221	0.0012
	13	0.9816	0.9536	0.8972	0.7981	0.6470	0.4511	0.2444	0.0826	0.0088
	14	0.9959	0.9877	0.9673	0.9226	0.8363	0.6904	0.4802	0.2382	0.0503
	15	0.9994	0.9979	0.9933	0.9807	0.9499	0.8818	0.7475	0.5182	0.2078
	16	1.0000	0.9998	0.9993	0.9977	0.9925	0.9775	0.9369	0.8332	0.5819
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0010	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0183	0.0058	0.0014	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0537	0.0203	0.0062	0.0014	0.0002	0.0000	0.0000	0.0000	0.0000
	7	0.1280	0.0576	0.0212	0.0061	0.0012	0.0002	0.0000	0.0000	0.0000
	8	0.2527	0.1347	0.0597	0.0210	0.0054	0.0009	0.0001	0.0000	0.0000
	9	0.4222	0.2632	0.1391	0.0596	0.0193	0.0043	0.0005	0.0000	0.0000
	10	0.6085	0.4366	0.2717	0.1407	0.0569	0.0163	0.0027	0.0002	0.0000
	11	0.7742	0.6257	0.4509	0.2783	0.1390	0.0513	0.0118	0.0012	0.0000
	12	0.8923	0.7912	0.6450	0.4656	0.2825	0.1329	0.0419	0.0064	0.0002
	13	0.9589	0.9058	0.8114	0.6673	0.4813	0.2836	0.1206	0.0282	0.0015
	14	0.9880	0.9672	0.9217	0.8354	0.6943	0.4990	0.2798	0.0982	0.0109
	15	0.9975	0.9918	0.9764	0.9400	0.8647	0.7287	0.5203	0.2662	0.0581
	16	0.9997	0.9987	0.9954	0.9858	0.9605	0.9009	0.7759	0.5497	0.2265
	17	1.0000	0.9999	0.9996	0.9984	0.9944	0.9820	0.9464	0.8499	0.6028
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$									
n	y	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	0.9868	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022
	4	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D8. Binomial distribution with y successes out of n trials (continued).

		$p(Y<y)$								
n	y	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
19	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0028	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0109	0.0031	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0342	0.0116	0.0031	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	7	0.0871	0.0352	0.0114	0.0028	0.0005	0.0000	0.0000	0.0000	0.0000
	8	0.1841	0.0885	0.0347	0.0105	0.0023	0.0003	0.0000	0.0000	0.0000
	9	0.3290	0.1861	0.0875	0.0326	0.0089	0.0016	0.0001	0.0000	0.0000
	10	0.5060	0.3325	0.1855	0.0839	0.0287	0.0067	0.0008	0.0000	0.0000
	11	0.6831	0.5122	0.3344	0.1820	0.0775	0.0233	0.0041	0.0003	0.0000
	12	0.8273	0.6919	0.5188	0.3345	0.1749	0.0676	0.0163	0.0017	0.0000
	13	0.9223	0.8371	0.7032	0.5261	0.3322	0.1631	0.0537	0.0086	0.0002
	14	0.9720	0.9304	0.8500	0.7178	0.5346	0.3267	0.1444	0.0352	0.0020
	15	0.9923	0.9770	0.9409	0.8668	0.7369	0.5449	0.3159	0.1150	0.0132
	16	0.9985	0.9945	0.9830	0.9538	0.8887	0.7631	0.5587	0.2946	0.0665
	17	0.9998	0.9992	0.9969	0.9896	0.9690	0.9171	0.8015	0.5797	0.2453
	18	1.0000	0.9999	0.9997	0.9989	0.9958	0.9856	0.9544	0.8649	0.6226
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0064	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0214	0.0065	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0580	0.0210	0.0060	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
	8	0.1308	0.0565	0.0196	0.0051	0.0009	0.0001	0.0000	0.0000	0.0000
	9	0.2493	0.1275	0.0532	0.0171	0.0039	0.0006	0.0000	0.0000	0.0000
	10	0.4086	0.2447	0.1218	0.0480	0.0139	0.0026	0.0002	0.0000	0.0000
	11	0.5857	0.4044	0.2376	0.1133	0.0409	0.0100	0.0013	0.0001	0.0000
	12	0.7480	0.5841	0.3990	0.2277	0.1018	0.0321	0.0059	0.0004	0.0000
	13	0.8701	0.7500	0.5834	0.3920	0.2142	0.0867	0.0219	0.0024	0.0000
	14	0.9447	0.8744	0.7546	0.5836	0.3828	0.1958	0.0673	0.0113	0.0003
	15	0.9811	0.9490	0.8818	0.7625	0.5852	0.3704	0.1702	0.0432	0.0026
	16	0.9951	0.9840	0.9556	0.8929	0.7748	0.5886	0.3523	0.1330	0.0159
	17	0.9991	0.9964	0.9879	0.9645	0.9087	0.7939	0.5951	0.3231	0.0755
	18	0.9999	0.9995	0.9979	0.9924	0.9757	0.9308	0.8244	0.6083	0.2642
	19	1.0000	1.0000	0.9998	0.9992	0.9968	0.9885	0.9612	0.8784	0.6415
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table D9. Population correlation coefficients. (Source: Remington and Schork, 1970)

CHARTS FOR COMPUTING CONFIDENCE LIMITS FOR THE
POPULATION CORRELATION COEFFICIENT ρ
CONFIDENCE LEVEL 95%

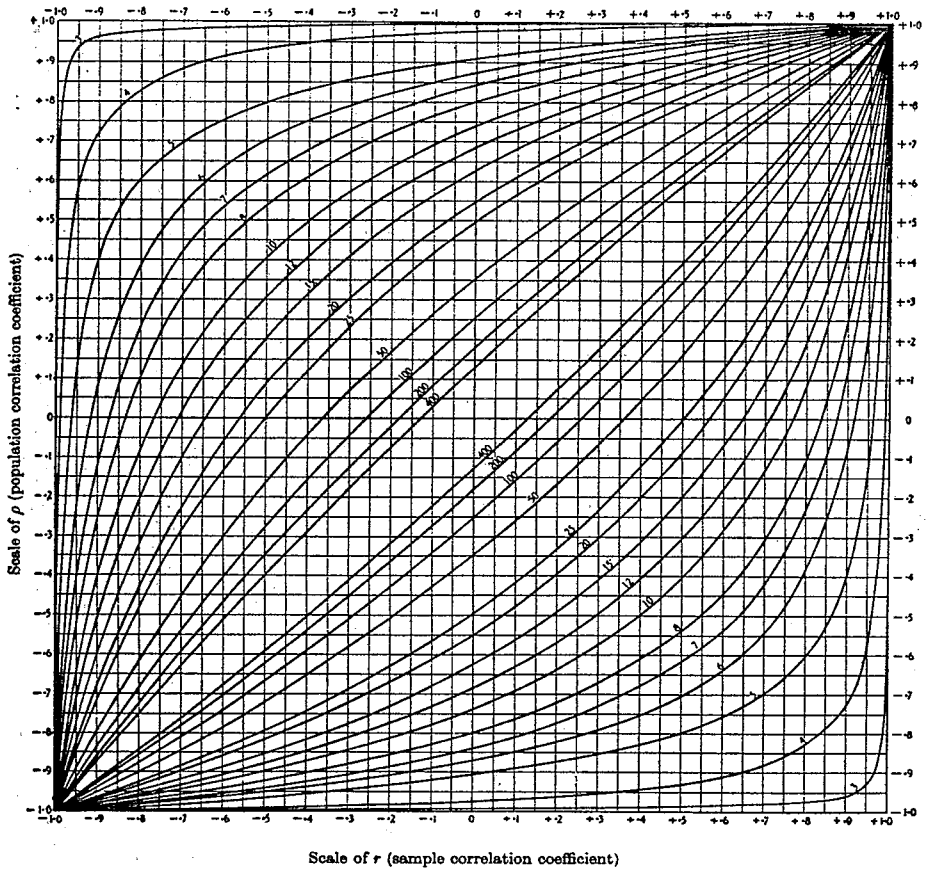


Table D9. (continued)

CHARTS FOR COMPUTING CONFIDENCE LIMITS FOR THE
POPULATION CORRELATION COEFFICIENT ρ
CONFIDENCE LEVEL 99%

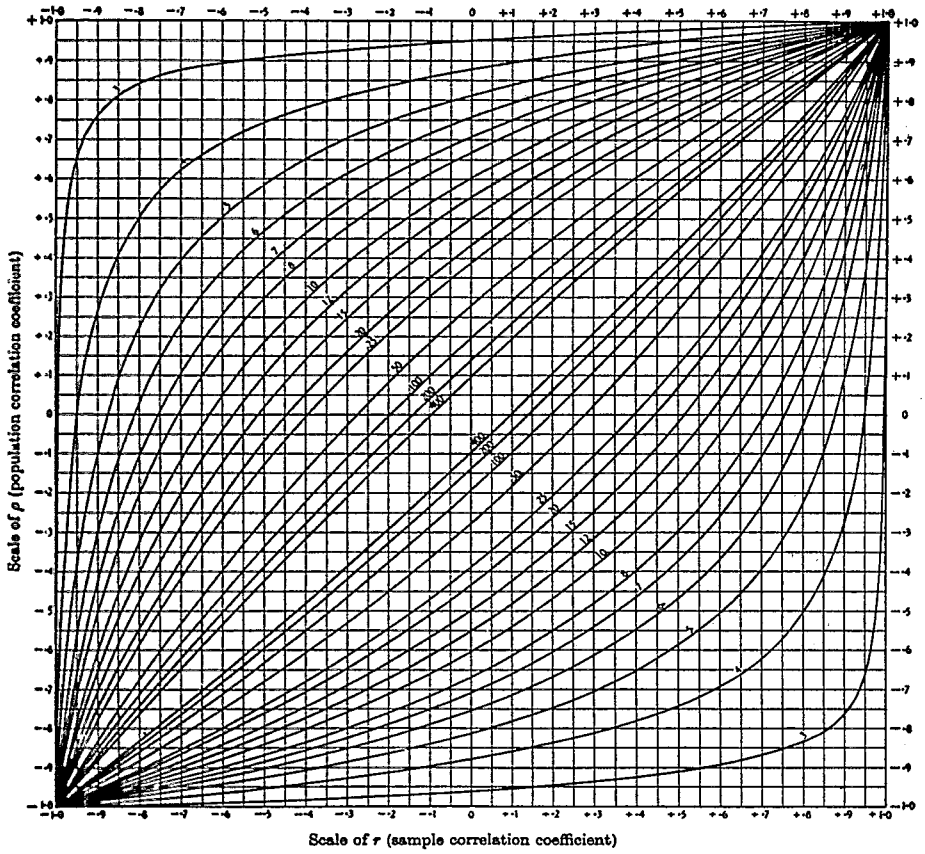


Table D10. Quantiles of the Spearman test statistic. (Source: Conover, 1980)

n	$p = .900$	.950	.975	.990	.995	.999
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

For n greater than 30 the approximate quantiles of ρ may be obtained from

$$w_p \cong \frac{x_p}{\sqrt{n-1}}$$

where x_p is the p quantile of a standard normal random variable obtained from Table 1.

[a] The entries in this table are selected quantiles w_p of the Spearman rank correlation coefficient ρ when used as a test statistic. The lower quantiles may be obtained from the equation

$$w_p = -w_{1-p}$$

The critical region corresponds to values of ρ smaller than (or greater than) but not including the appropriate quantile. Note that the median of ρ is 0.

Table D11. Sample sizes for one-sided nonparametric tolerance limits[a] (Source: *Practical Nonparametric Statistics*, W.J. Conover, Copyright © (1980 John Wiley & Sons, Inc.). Reprinted by permission of John Wiley & Sons, Inc.)

$1-\alpha$	$q = .500$	.700	.750	.800	.850	.900	.950	.975	.980	.990
.500	1	2	3	4	5	7	14	28	35	69
.700	2	4	5	6	8	12	24	48	60	120
.750	2	4	5	7	9	14	28	55	69	138
.800	3	5	6	8	10	16	32	64	80	161
.850	3	6	7	9	12	19	37	75	94	189
.900	4	7	9	11	15	22	45	91	144	230
.950	5	9	11	14	19	29	59	119	149	299
.975	6	11	13	17	23	36	72	146	183	368
.980	6	11	14	18	25	38	77	155	194	390
.990	7	13	17	21	29	44	90	182	228	459
.995	8	15	19	24	33	51	104	210	263	528
.999	10	20	25	31	43	66	135	273	342	688

[a] The quantity tabled is the sample size n such that $q^n \leq \alpha$, for use in finding the tolerance limits

$$P(X^{(1)} \leq p \text{ of the population}) \geq 1-a$$

or

$$P(q \text{ of the population} \leq X^{(n)}) \geq 1-a$$

*U.S. GOVERNMENT PRINTING OFFICE: 1997-521-701/90285