United States Environmental Protection Agency
Office Of Water (4503F) Washington, DC
September 1995 FINAL REVIEW DRAFT

EPA

NONPOINT SOURCE MONITORING AND EVALUATION GUIDE

823D95001

NONPOINT SOURCE MONITORING AND EVALUATION GUIDE

FINAL REVIEW DRAFT

September, 1995

Nonpoint Source Control Branch
U.S.E.P.A. Headquarters
401 M Street, S.W.
Washington, DC 20460

PURPOSE

This nonpoint source (NPS) monitoring and evaluation guide is written for use by both those who fund and approve monitoring and evaluation (M&E) plans and those who perform the monitoring and evaluation. For example, federal, state, and regional agencies that support M&E activities may use the guide to assess the technical merit of proposed M&E plans. Those agencies, private groups, and university personnel that perform M&E may use this guide to formulate their M&E plans. This guide is in no way intended to supersede proven nonpoint source M&E plans currently in use, but it is intended as both a check against existing M&E plans and an outline for developing new NPS M&E plans.

TABLE OF CONTENTS

Page

LIST OF TABLES iii

LIST OF FIGURES vii

I. OVERVIEW OF NPS PROBLEM I-1
A. Definition of a Nonpoint Source I-1
B. Nature and Scope of the Nonpoint Source Problem I-1

II. GOALS AND OBJECTIVES OF NPS MONITORING PROGRAMS II-1
A. Monitoring Goals II-1
B. Monitoring Objectives II-3
C. Monitoring Level II-7
D. Monitoring Scale II-9

III. REVIEW OF SELECT AVAILABLE MONITORING GUIDANCES III-1
A. Introduction III-1
B. Categorical Listing of Monitoring Guidances Reviewed . III-2
C. Guidance Reviews III-5

IV. WATER RESOURCE CONSIDERATIONS IV-1
A. Variability IV-1
B. Streams IV-7
C. Lakes IV-7
D. Estuaries IV-17
E. Coastal Waters IV-21
F. Ground Water IV-21

V. DATA NEEDS AND DATA SOURCES V-1
A. Data Needs V-1
B. Data Types V-1
C. Data Sources V-3

VI. MONITORING RECOMMENDATIONS VI-1
A. Samples and Sampling VI-1
B. Monitoring Approaches VI-3
C. Ground Water Monitoring VI-30
D. Example Monitoring Programs VI-31

VII. BIOLOGICAL MONITORING AND ASSESSMENT VII-1
A. Introduction VII-1
B. Habitat Assessment VII-4
C. Overview of Biological Assessment Approaches VII-6
D. Reference Sites and Conditions VII-10
E. Rapid Bioassessment Protocols VII-15
F. The Multimetric Approach for Biological Assessment . VII-17
G. Sampling Considerations VII-20
H. Biomonitoring Program Design VII-30
I. Monitoring Trends in Biological Conditions VII-37

J. Overview of Some State Programs VII-43

VIII. DATA ANALYSIS VIII-1
A. Introduction VIII-1
B. Exploratory Analysis VIII-10
C. Summary (Descriptive) Statistics VIII-49
D. Trend Testing VIII-63
E. Extreme Analysis VIII-144
F. Multivariate Analyses VIII-158
G. Special Treatment of Environmental Data VIII-163

IX. QUALITY ASSURANCE IX-1
A. Introduction IX-1
B. Data Quality Objectives IX-4
C. Elements of a Quality Assurance Project Plan IX-10
D. Field Operations Program IX-25
E. Laboratory Operations IX-30
F. Data and Reports IX-34

APPENDICES
Appendix A. Chain of Custody Procedures A-1
Appendix B. Normal Distribution B-1
Appendix C. t-Distribution C-1
Appendix D. Chi-square Distribution D-1
Appendix E. F Distribution E-1
Appendix F. Skewness Test F-1
Appendix G. Kurtosis Test G-1
Appendix H. Coefficients for W Test for Normality H-1
Appendix I. W Test Percentage Points I-1
Appendix J. Noncentral t-Distribution J-1
Appendix K. Population Correlation Coefficient K-1
Appendix L. Transformation of Linearization L-1
Appendix M. Multiple Regression Example - Hand Calculated . . M-1
Appendix N. Multiple Regression Example - Computerized . . . N-1
Appendix O. Example Use of R-Square O-1
Appendix P. Confidence Limits for the Median of Any Continuous Distribution P-1
Appendix Q. Quantities of the Spearman Test Statistic Q-1
Appendix R. Probabilities for the Mann-Kendall Nonparametric Test for Trend R-1
Appendix S. Quantiles of the Spearman Test Statistic S-1
Appendix T. Quantiles of the Wilconxon Signed Rank Test Statistic T-1
Appendix U. Nonparametric 95 and 90 percentile confidence intervals on a proportions U-1
Appendix V. Sample sizes for one-sided nonparametric tolerance limits V-1

REFERENCES REF-1

LIST OF TABLES

Table		Page
I-1	National Summary of Nonpoint Source Problems	I-4
II-1	Relationship of monitoring objectives and scales	II-11
III-1	Guidance Review Chart	III-36
IV-1	Comparison of means by site for CAC (ug/l) at Highland Silver Lake, May 1981 - March 1984.	IV-17
VI-1	Applications of six probability sampling designs to estimate means and totals.	VI-14
VI-2	Long-term monitored parameters for each level in the St. Albans Bay watershed.	VI-35
VI-3	Related long-term monitored parameters in the St. Albans Bay watershed.	VI-36
VI-4	Monitored parameters for the short-term intensive studies.	VI-36
VI-5	Water quality sampling schedule for the St. Albans watershed.	VI-37
VI-6	Years of sampling for each monitored level in the St. Albans watershed.	VI-38
VI-7	Methods of water quality analysis.	VI-39
VI-8	Specific problems to be addressed for each monitoring component.	VI-49
VI-9	Monitoring plan for regional network.	VI-50
VI-10	Monitoring plan for small watershed site.	VI-52
VI-11	Monitoring plan for field site 1.	VI-54
VI-12	Monitoring plan for field site 2.	VI-56
VII-1	General strenths and limitations of biological monitoring and assessment approaches.	VII-5

VII-2	Five tiers of the rapid bioassessment protocols.VII-17	
VII-3	Fish IBI metrics used in various regions of North America.	VII-22
VII-4	Examples of metric suites used for analysis of macroinvertebrate assemblages	VII-24
VII-5	Scoring criteria for the core metrics as determined by the 25th percentile of the metric values from the Middle Rockies-Central Ecoregion.	VII-25
VII-6	Scoring criteria for the metrics as determined by the 25th percentile of the metric values for the aggregated subecoregions for Florida streams.	VII-26
VII-7	Comparison of probabalistic and targeted monitoring designs.	VII-32
VII-8	Waterbody stratification hierarchy.	VII-34
VII-9	Summary of the primary technical issues related to biological monitoring for nonpoint source evaluations.	VII-45
VII-10	Selected biomonitoring program components.	VII-46
VIII-1	Methods for characterizing data.	VIII-4
VIII-2	Methods for routing data analysis.	VIII-7
VIII-3	Raw data by time period	VIII-22
VIII-4	Normalized data.	VIII-22
VIII-5	Total nitrogen runoff concentrations for a single storm event in Florida.	VIII-23
VIII-6	Total nitrogen runoff concentrations for a single storm evnet in Florida and example calculations for the EMC.	VIII-25
VIII-7	Selected summary statistics for mean abundance per grab from 1991 through 1992 for the Louisiana Province of EMAP from large rivers and small rivers.	VIII-27

VIII-8	Example analysis of the Shapiro-Wilk W test using mean abundance per grab data for the Louisiana Province of EMAP for small rivers.	VIII-30
VIII-9	Example application of Spearman's rho test for evaluating independence.	VIII-36
VIII-10	Sample sizes required for estimating the true mean.	VIII-63
VIII-11	Errors in hypothesis testing.	VIII-45
VIII-12	Highland Silver Lake TSS data for Site 1.	VIII-68
VIII-13	Nonparametric evaluation of post-implementation data using the Wilcoxon Signed ranks test.	VIII-76
VIII-14	Summary of parametric tests used to evaluate difference between means.	VIII-79
VIII-15	Nonparametric evaluation of post-implementation data using the Mann-Whitney test.	VIII-82
VIII-16	Sign test for comparing daily BOD5 concentrations at two locations.	VIII-86
VIII-17	Spearman's test for total nitrogen and total phosphorus correlation.	VIII-88
VIII-18	Stream trout population.	VIII-93
VIII-19	One-way ANOVA.	VIII-93
VIII-20	Waller-Duncan k-ratio and Bonferroni t-test results.	VIII-95
VIII-21	Two-way ANOVA table.	VIII-98
VIII-22	Two-way ANOVA.	VIII-98
VIII-23	Waller-Duncan k-ratio and Bonferroni t-test results.	VIII-99
VIII-24	Waller-Duncan k-ratio and Bonferroni t-test results.	VIII-100
VIII-25	Two-way ANOVA with interaction.	VIII-102
VIII-26	Two-way ANOVA with interaction.	VIII-102

VIII-27	Annual total rainfall for 21 years.	VIII-116
VIII-28	Sign calculation for Cox and Stuart test.	VIII-117
VIII-29	Data difference calculation for Mann-Kendall test.	VIII-118
VIII-30	Runoff sampler calibration data.	VIII-123
VIII-31	Linear regression output.	VIII-125
VIII-32	Theoretical log-probability frequency factors.	VIII-145
VIII-33	Linearized rainfall frequency variate for equation 8.142.	VIII-145
VIII-34	Linearized rainfall duration variate for equation 8.142.	VIII-150
IX-1	Common QA and QC activities.	IX-2
IX-2	Elements required in an EPA Quality Assurance Project Plan.	IX-11
IX-3	Checklist of items that should be considered in the field operations section of a QA/QC program.	IX-26
IX-4	Checklist of items that should be considered in the laboratory operations section of a QA/QC program.	IX-33

LIST OF FIGURES

Figure		Page
II-1	Monitoring in perspective	II-2
IV-1	Flow and pollutant concentrations for a single storm event in the Honey Creek watershed, Ohio	IV-2
IV-2	Time course of phosphorus loading from the West Branch Delaware River and output from Cannonsville Reservoir	IV-3
IV-3	Mean total suspended solids concentrations (mg/L) in Highland Silver Lake, Illinois, May 1981-March 1984	VI-4
IV-4	Mixing of salt water and fresh water in an estuary	IV-5
IV-5	Contour map of nitrate concentrations recorded from Galena aquifer inventory water samples (Iowa)	IV-6
IV-6	Seasonal adjustment to discharges of sediment, in clay and silt sizes, White River near Kadoka, SD	IV-8
IV-7	Seasonal relationship between sediment concentration, sediment load, discharge, and precipitation for Fern Ridge Dam, Oregon	IV-9
IV-8	Schematic diagram of stream vertical showing relative position of sediment load terms	IV-10
IV-9	Vertical sediment concentration and flow velocity distribution in a typical stream cross section	IV-11
IV-10	Phytoplankton chlorophyll a concentration in Chautauqua Lake's northern basin and southern basin, 1977	IV-13
IV-11	Typical thermal stratification of a lake into the epilimnetic, metalimnetic, and hypolimnetic water strata	IV-14
IV-12	Temperature and dissolved oxygen profiles for Barren River Lake in 1979	IV-15

IV-13 Factors contributing to lateral differences in lake quality IV-16

IV-14 Chesapeake Bay salinity levels over time and space IV-19

IV-15 Physical features affecting lateral salinity gradients in an estuary IV-20

IV-16 Permeable sand layer underlain by a clay layer IV-23

IV-17 Two-aquifer system with opposite flow directions IV-24

IV-18 Effect of permeability change on shape of pollution plume IV-25

VI-1 Map of the Rock Creek Rural Clean Water Program study area, Twin Falls County, Idaho VI-21

VI-2 St. Albans Bay watershed, Franklin County, Vermont sampling locations VI-23

VI-3 State locus of Bellevue NURP VI-41

VI-4 Bellevue stream systems VI-42

VI-5 Bellevue sampling sites VI-44

VI-6 Location of Conestoga River headwaters area and four monitoring components VI-48

VI-7 Location of monitoring facilities for regional network VI-51

VI-8 Location of monitoring facilities for small watershed sites VI-53

VI-9 Location of monitoring facilities for field site 1 . . . . . VI-55

VI-10 Field site 2, Lancaster Co., PA VI-57

VII-1 Approaches to establishing reference conditions VII-16

VII-2 Selection and application of the different tiers of RBP depend on monitoring objectives VII-18

VII-3 Organizational structure of attributes that can serve as metrics VII-20

VII-4 Regions in which various fish IBI metrics have been used. VII-21

VII-5 Some trends that might be observed during the course of a biological monitoring program VII-39

VII-6 Trend in biological condition (based on the multimetric ICI) of the Cuyohaga River at Independence, Ohio. VII-44

VIII-1 Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware using a time series plot. VIII-12

VIII-2 Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware using a histogram VIII-13

VIII-3 Stem and leaf plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, DelawareVIII-13

VIII-4 Box and whisker plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, DelawareVIII-15

VIII-5 Box and whisker plot of mean abundance per grab from 1991 through 1992 for the Louisiana Province of the Environmental Monitoring and Assessment Program VIII-17

VIII-6 Bivariate scatter plot of total suspended solids and flow at 36th Street storm sewer in Denver, Colorado VIII-18

VIII-7 Time series plot of dissolved orthophosphate from 1989 through 1994 for portions of the Delaware River VIII-19

VIII-8 Precipitation, runoff, total nitrogen, and total phosphorus from a single storm event in Florida VIII-24

VIII-9 Examples of several frequency distributions and their respective cumulative frequency distributions VIII-32

VIII-10 Autocorrelation (correlogram) for dissolved oxygen data from 1980 through 1989 for the Delaware River at Reedy Island, Delaware . VIII-38

VIII-11 A normal distribution VIII-41

VIII-12 The effect of changing μ in the normal distribution while keeping σ constant . . VIII-42

VIII-13 The effect of changing σ in the normal distribution while keeping μ constant . . VIII-43

VIII-14 Change of scale VIII-45

VIII-15 Standard normal distribution and example of t-distribution VIII-47

VIII-16 Chi-square distribution VIII-48

VIII-17 F distribution VIII-50

VIII-18 Some examples of symmetric distributions . VIII-55

VIII-19 Examples of skewed distributions VIII-56

VIII-20 Symmetric distributions of varying kurtosis VIII-57

VIII-21 Normal distribution VIII-59

VIII-22. Relationship between α and β. VIII-66

VIII-23. Pre-implementation data set. VIII-69

VIII-24. Post-implementation data set. VIII-70

VIII-25. Log-transformed pre-implementation data set VIII-72

VIII-26. Log-transformed post-implementation data set VIII-73

VIII-27. One-sided t-test. VIII-75

VIII-28. Parameter Y vs. X for three time periods. VIII-107

VIII-29. General representation of regression between paired observations of parameters, Y and X, each unit of time has its unique slope. VIII-110

VIII-30. Types of time series VIII-113

VIII-31 Split versus flowrate VIII-124

VIII-32. Plot of residuals versus predicted values. VIII-127

VIII-33. Plot of residuals versus predicted values. VIII-128

VIII-34. Plot of residuals versus time. VIII-129

VIII-35. TIme sequence plot of residuals. VIII-130

VIII-36. Regression line with confidence intervals. VIII-138

VIII-37. Regression line with confidence intervals. VIII-138

VIII-38. Plot of mean and individual value confidence limits. VIII-139

VIII-39. One-hour rainfall to be expected at a return period of 2 years VIII-146

VIII-40. 24-hour rainfall to be expected at a return period of 2 years VIII-147

VIII-41. One-hour rainfall to be expected at a return period of 100 years VIII-148

VIII-42. 24-hour rainfall to be expected at a return period of 100 years VIII-149

IX-1 Sample organization chart for a quality assurance project plan IX-13

IX-2 Sample qualtiy assurance objectives IX-15

IX-3 Sample custody chart IX-19

CHAPTER 1

OVERVIEW OF THE NONPOINT SOURCE PROBLEM

I. OVERVIEW OF THE NONPOINT SOURCE PROBLEM

A. Definition of a Nonpoint Source

Nonpoint sources of water pollution are both diffuse in nature and difficult to define. Technically, 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. That definition states:

> 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 generally 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 permit requirements of the Clean Water Act. In contrast, nonpoint sources are not subject to federal permit requirements.

Nonpoint source pollution is the pollution of our nation's waters caused by rainfall or snowmelt moving over and through the ground. As the runoff moves, 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. In addition, hydrologic modification is a form of nonpoint source pollution that often adversely affects the biological and physical integrity of surface waters.

B. Nature and Scope of the Nonpoint Source Problem

Nonpoint sources can generate both conventional (e.g., bacteria, oxygen-demanding substances) and toxic pollutants, just as point sources do. Although nonpoint sources can contribute many of the same kinds of pollutants, these pollutants are generated in different volumes, combinations, and concentrations. Pollutants from nonpoint sources are mobilized primarily during storm or snowmelt events. Consequently, NPS pollution episodes are generally less frequent and shorter in duration than continuous point source discharges. However, it should be noted that NPS impacts on a downstream or down gradient water resource can be continuous due to, for example, (1) upstream/upgradient mixing, (2) multiple sources, (3) flow dynamics, and (4) chemical, physical, or biological processes.

Over the last decade, significant achievements have been made nationally in the protection and enhancement of water quality. Much of this progress, however, has been accomplished by controlling point sources of pollution. While some state and local NPS management programs have been developed and 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 from the *National Water Quality Inventory 1992 Report to Congress* (USEPA, 1992) indicate that nonpoint sources negatively affect rivers and streams in 46 of 55 states and territories; lakes, reservoirs, and ponds in 41 of 55 states and territories (including the District of Columbia); estuaries and coastal waters in 16 of 29 coastal states and territories; and the Great Lakes in 5 of 8 Great Lakes states. The categories of nonpoint source pollution affecting these waterbodies included agriculture, resource extraction, silviculture, hydrologic and habitat modification, construction, atmospheric deposition, flow modification, land disposal, and onsite disposal systems.

According to the report, agriculture is the primary nonpoint pollution source affecting rivers and streams. Thirty-one states and territories list it as a major source of water quality impairment. Seventy-two percent of the impaired river and stream miles were reported to be affected by agricultural sources of pollution. Other nonpoint sources that affect rivers and streams include resource extraction, silviculture, and hydrologic and habitat modification, which are listed as major sources of water quality impairment by 21, 13, and 22 states, respectively.

Agriculture was also considered the most significant nonpoint source affecting lakes, reservoirs, and ponds. Fifty-six percent of impaired lake, reservoir, and pond surface acres are negatively impacted by agriculture, followed by hydrologic and habitat modification (23%), onsite disposal systems (OSDS, 16%), flow modification (13%), and resource extraction (6%). Thirty states listed agriculture as a major source of impairment to the water quality of lakes, reservoirs, and ponds; 15 states listed hydrologic and habitat modification; 13 listed resource extraction, 6 listed OSDS; and 5 listed flow modification.

The water quality of estuaries and coastal waters is also more impaired by agriculture than by any other nonpoint source. Agriculture negatively impacts 43 percent of the impaired estuarine and coastal water area and is listed as a major source of impairment by seven coastal states (not including the Great Lakes states). Resource extraction is listed as a major source of impairment to estuarine and coastal waters by three coastal states, construction is listed as major by one coastal state, atmospheric deposition is listed as major by one territory, and hydrologic and habitat modification is listed as major by one coastal state.

Five of the eight Great Lakes states list nonpoint source pollution as affecting the quality of their Great Lakes shoreline

miles. Each of the nonpoint sources atmospheric deposition, land disposal, and construction is listed by one Great Lakes state as a major source of impairment to Great Lakes shoreline water quality. Atmospheric deposition is reported to be a major source of impairment to 50 percent of the impaired Great Lakes shoreline miles, followed by land disposal, which affects 31 percent, and construction, which affects 2 percent.

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

1. Effects of Nonpoint Source Pollutants

Sediment, nutrients, pathogens, habitat alteration, salts, toxic substances, petroleum products, pesticides and herbicides, and hydrologic modifications are the pollutants contributed to surface and ground waters by various nonpoint sources. Each of these pollutants can have negative effects on aquatic systems and in some cases on human health. The nutrients nitrogen and phosphorus are contained in commercial fertilizers and manure. Addition of excessive amounts of nutrients to marine and freshwater systems, where nitrogen and phosphorus, respectively, are generally limiting to plant growth, can lead to eutrophication. Animal wastes from agriculture and pets also contain oxygen-demanding substances that deplete dissolved oxygen when carried into aquatic systems. Suspended sediments in runoff from nonirrigated cropland, overgrazing on rangeland and pastureland, livestock grazing along stream banks, urban runoff, and construction reduce sunlight to aquatic plants, smother fish spawning areas, and clog filter feeders and fish gills. Turbidity also reduces the aesthetic attractiveness of recreational surface waters. Salts from irrigation water concentrate at the soil surface through the evapotranspiration process. Salts used on roads accumulate along the edges of roads or are carried via storm sewer systems to surface waters. Salts cause soil structure to break down, decrease water infiltration, and in high concentrations can be toxic to plants or decrease productivity. Some pesticides are persistent in aquatic systems and biomagnify in animal tissue (primarily fish) as they are passed up through the food chain. This has detrimental physiological effects 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, habitat impacts from soil trampling, stream bank erosion after excessive grazing, conversion of natural habitats to agricultural systems, urbanization, and silvicultural practices diminish riparian vegetation, increase erosion, and alter plant species composition in riparian areas (Khaleel et al., 1980).

Table I-1. Sources of nonpoint source pollution and their contribution to the impairment of the Nation's water quality.

		RIVERS & STREAMS	LAKES, PONDS & RESERVOIRS	ESTUARIES & COASTAL	GREAT LAKES
AGRICULTURE	Major*	31	30	7	
	Percent*	159,353 miles (72%)	3,091,585 acres (56%)	3,539 square miles (43%)	
RESOURCE EXTRACTION	Major*	21	13	3	
	Percent*	23,697 miles (11%)	343,915 acres (6%)	998 square miles (12%)	
HYDROLOGIC & HABITAT MODIFICATION	Major*	22	15	1	
	Percent*	15,119 miles (7%)	1,275,583 acres (23%)	457 square miles (6%)	
SILVICULTURE	Major*	13			
	Percent*	16,236 miles (7%)			
ONSITE DISPOSAL SYSTEMS	Major*		6		
	Percent*		874,265 acres (16%)		
FLOW MODIFICATION	Major*		5		
	Percent*		716,189 acres (13%)		
CONSTRUCTION	Major*			1	1
	Percent*			664 square miles (8%)	41 shoreline miles (2%)
ATMOSPHERIC DEPOSITION	Major*			1	1
	Percent*			466 square miles (6%)	946 shoreline miles (50%)
LAND DISPOSAL	Major*				1
	Percent*				576 shoreline miles (31%)
TOTAL IMPAIRED		221,877 impaired miles	8,071,260 impaired acres	8,303 impaired square miles	5,171 impaired shoreline miles

(Source: USEPA 1992)

*Listed as major factor contributing to water quality impairment

*Percent of total impaired length or area impaired by the particular source

2. Agricultural Nonpoint Source Pollution

Nonpoint source pollution from agriculture comes from several individual sources, including nonirrigated cropland (both row and field), irrigated cropland, rangeland and pastureland, and livestock facilities. The primary pollutants associated with agriculture are sediment, nutrients, animal wastes, salts, and pesticides, which are contributed to surface and ground waters in varying degrees by individual agricultural sources (Bailey and Waddell, 1979; Dressing et al., 1982; Kreglow et al., 1982; Maas et al., 1982; Maas et al., 1984). Agriculture also has direct impacts on the habitats of aquatic species.

3. Urban Nonpoint Source Pollution

Urbanization adversely affects surface waters through the addition of many of the pollutants discussed above and through dramatic changes to stream hydrology in urbanized areas. From the perspective of nonpoint source pollution impacts, urbanization is best characterized as a process by which pervious, naturally vegetated land is converted to impervious surfaces through which storm water cannot percolate to reach the soil beneath and upon which numerous pollutants are deposited and then carried by storm water quickly and directly to surface waters. These changes result in dramatic hydrologic, water quality, and aquatic habitat impacts (Field et al., 1977).

Storm water that runs off roofs, lawns, streets, industrial sites, and other pervious and impervious areas washes a number of pollutants into urban lakes and streams (Hawkins et al., 1974). Runoff that percolates through the soil to ground water supplies carries these pollutants with it. The hydrologic changes resulting from urbanization include increased peak discharges of storm water runoff, increased runoff volume, increased runoff velocity, decreased time before runoff reaches surface waters, and reduced stream flow during low-flow periods.

A large percentage of the pollutant load in urban runoff is composed of sediment and debris from decaying pavements and buildings that can clog sewers and waterways, reduce hydraulic capacity, increase the chance of flooding, and degrade aquatic habitats. Other urban pollutants include heavy metals and inorganic chemicals (including copper, lead, zinc, and cyanides) from brake linings on automobiles, industrial activities, and building materials (USEPA, 1983b); oils, greases, antifreeze, and other automotive fluids that drip or are disposed of onto pavements; nutrients in fertilizers that are applied to residential lawns, parks, golf courses, and along transportation rights-of-way; fecal bacteria that are contained in pet wastes and leaking septic tanks (which also contain excess nutrients); and road salts, which are a problem primarily in northern urban areas.

Urbanization often entails modifications to stream courses in the urban area to accommodate roadways, buildings, waterfronts, residential developments, and the like. Stream course modifications are also undertaken to protect urban areas from flood damage, but the modifications cause their own set of NPS impacts. Levees along streams serve to protect areas from flooding, but also decrease the natural supply of sediment reaching floodplain areas. Rather than being dispersed over a floodplain, the sediment contained in flowing water during high-flow periods is carried to stream and river deltas, where it accumulates due to the natural loss of flow velocity at deltas. This sediment can impede navigation and smother biological communities in these areas. The sediment that is redirected downstream is lost to the floodplain. A periodic resupply of nutrient-rich sediment is important to wetland and floodplain ecosystem function, and its loss has negative ecological consequences. In coastal marshes, this loss of a natural resupply of sediment and fresh water leads to subsidence of marshlands and infiltration of saline waters. The biological communities in these areas are unable to survive the resultant increase in salinity, and there is a shift toward marine ecosystems. Coastal marshes are biologically diverse and productive, and their loss can result in severe economic loss.

4. Mining Nonpoint Source Pollution

Although mining activities are not as widespread as agriculture, the water quality effects resulting from mining are usually much more harmful. Sedimentation rates from mining can be extraordinarily high (Mitsch et al., 1983). Entire streams might become biologically dead as a result of acidic mine drainage (water with low pH released from mining areas). Other pollutants in mining runoff with potentially serious effects include heavy metals and radioactive materials. Active mine sites can pose a number of potential water quality problems. Most of these, however, are considered to be point source problems and are regulated under state and federal National Pollutant Discharge Elimination System (NPDES) permits (USEPA, 1978a). The primary NPS pollution problems associated with mining are sediment-laden runoff from haul roads and runoff polluted with acid, sediment, salts, and metals from the spoil and tailings piles. Both of these problems arise from both active and inactive mines, and both ground and surface waters are adversely affected. In the western United States, water quality effects from metal and uranium mining are more serious than those from other types of mining. Acid drainage from coal mines is a greater problem in the eastern and midwestern states, where the most damage occurs in the heavily mined areas of the mid-Atlantic and Appalachian regions. Stream quality in these regions is severely degraded by drainage from abandoned coal mines.

5. Silvicultural Nonpoint Source Pollution

The smaller areal extent of forest management activities, less intensive site preparation, infrequent harvest, and lower frequency

of pesticide and nutrient applications in a given year all result in silviculture generating a smaller volume of total NPS pollutants than agriculture nationwide (Bailey and Waddell, 1979). Silvicultural activities can nonetheless cause significant localized NPS pollution problems (USEPA, 1978a). The impact of this localized NPS pollution is often disproportionately large compared to NPS impacts in other areas because forested watersheds often contain the Nation's highest quality waters, which are the source of many municipal water supplies and are prized for cold-water fisheries, high-economic-value recreation, and aesthetics.

Silvicultural activities that generate NPS pollution include road building, pesticide and herbicide application, harvesting and logging operations, site preparation to promote revegetation, and timber stand improvement (USDA, 1981). Sediment, which results primarily from road construction and harvesting, is the major pollutant by volume from silvicultural activities (USEPA, 1980a). Soil type, slope, climate, and the best management practices used on a site have a large effect on the delivery of sediment to forested water courses. The Illinois Environmental Protection Agency found that road building and road use are the major causes of erosion and soil loss from silvicultural activities (IEPA, 1979).

Fertilizers and pesticides are used increasingly in silviculture, but unlike annual agricultural applications they are typically applied only once or twice during a 20- to 35-year period on any given harvestable forested tract of land. The contribution of chemicals to lakes and streams is less frequently a problem for silviculture than for agriculture, though localized problems can result if practices such as aerial spraying are employed near a water course. Debris from silvicultural operations that is left in or is carried to streams contributes excessive organic matter to waterbodies, thereby decreasing dissolved oxygen levels and causing changes in flow patterns. Water temperature increases occur when streamside vegetation is removed during harvests, and this can severely impact cold-water aquatic ecosystems that are adapted to narrow temperature ranges.

6. Construction Nonpoint Source Pollution

On a national basis, the water quality degradation caused by NPS pollution from construction activities is not nearly as great as the amount caused by other major nonpoint sources (Bailey and Waddell, 1979). This is due, in part, to improved construction erosion control by local jurisdictions. In addition, construction sites are often dispersed and, as a result, usually create only localized problems. States reported construction as a source of degradation to only estuaries, coastal waters, and wetlands, and estimated that it degrades only 8 percent of impacted estuarine square miles and 7 percent of impacted coastal shoreline miles (USEPA, 1992). Eight states reported road construction as a source of degradation to wetlands.

Where construction activities are intensive, localized impacts on water quality can be severe because excessive loads are discharged in short time spans. Erosion rates from construction sites typically are 10 to 20 times those of agricultural lands (U.S. House of Representatives, 1980). Sediment is the main construction site pollutant, though sediment from construction sites accounts for only about 4 to 5 percent of nationwide sediment loads in receiving waters (USEPA, 1980b). Other pollutants generated by construction activities are nutrients (primarily phosphorus and nitrogen) from fertilizers, which bind to sediment particles or dissolve in solution (USEPA, 1978a); chemicals in pesticides that are used to control weeds and insects at construction sites; petroleum products and construction chemicals, such as cleaning solvents, paints, asphalt, acids, and salts; and solid wastes, ranging from coffee cups to trees and other debris left at construction sites (USEPA, 1984e). Some of these materials are toxic to aquatic organisms (including shellfish and fish) and make them unfit for human consumption. These pollutants also make water unfit for use as drinking water and for contact recreation.

CHAPTER 2

GOALS AND OBJECTIVES OF NPS MONITORING PROGRAMS

II. GOALS AND OBJECTIVES OF NPS MONITORING PROGRAMS

A. Monitoring Goals

Overall goals established in the EPA Deputy Administrator's 1983 monitoring policy memo (USEPA, 1985c) are to:

- Meet the full range of current and future Agency needs for environmental data.
- Ensure monitoring is technically and scientifically sound.
- Ensure environmental monitoring data are managed to facilitate both access and appropriate use in Agency decision making.
- Ensure effective and coordinated Agency-wide processes for planning and execution of monitoring activities.
- Ensure that roles and responsibilities are clear in regard to monitoring management and implementation by EPA and state officials.

An attempt to put monitoring into perspective is presented in Figure II-1, taken from the National Water Monitoring Panel (USEPA, 1975). It is obvious that if each functional purpose is to be productive, the proper information must be provided by the monitoring program. It also should be clear that persons responsible for monitoring must maintain a frequent and substantive contact with those programs requiring information. Finally, there is need for coordination among all aspects of a monitoring program.

The following, taken from a monitoring and evaluation workshop manual (Simons et al., 1981), discusses the conceptual design of a monitoring system:

> In order for any monitoring network to be designed and implemented, the overall objective as well as the monitoring objective should be analyzed to assess what type of data [are] required to fulfill these objectives. An overall objective is of broad scope; for example, a water supply system or a sewage treatment system. A much smaller objective which is formulated under the overall objective is the monitoring objective. Examples are: stream flow measurement; water quality measurements; precipitation measurements; etc. It must be kept in mind that the overall objective determines the nature of the monitoring objective. For example, if the overall objective was enforcement of water quality laws, the monitoring objective would be to monitor the water quality variables in the jurisdiction area; but how is the monitoring to be done? In this case, sampling every night may miss industrial pollution loads, and monitoring upstream of industrialized areas would do the same. Here the overall objective conceptually defines the sampling frequency and spatial location of the monitoring network. The exact number, location and sampling frequency for monitoring stations must bow to social, political, legal, and financial considerations.

(TO BE INSERTED)

Figure II-1. Monitoring in perspective. (SOURCE: USEPA, 1975).

B. Monitoring Objectives

Identifying and concisely stating the monitoring objectives are critical steps in the development of a monitoring program. Numerous authors attest to the importance of clearly specifying monitoring program objectives prior to designing the program or collecting any data. Plafkin et al. (1989, p. 3-1) state "...selection of the appropriate bioassessment approach depends on the objectives of the study." A monitoring guidance published by the U.S. Department of Agriculture (USDA-SCS, 1989, p. iv (draft)) states "Setting objectives for monitoring clarifies the purposes of the study and keeps it on track." Monitoring program objectives must be detailed enough to define precisely what data and information will have to be gathered and how that data and information will help address the NPS problem at hand.

Monitoring program design includes selection of sampling locations, sampling strategy, sampling variables, data analysis techniques, length of monitoring program, and level of effort and resources. Vague or inaccurate statements of objectives lead to program designs that only partially fulfill management needs or provide redundant data that would not help address the NPS problem. If the monitoring objective is to determine presence or absence of an NPS impact, a simple sampling survey might be conducted in a few select locations. If the objective is to calibrate or verify a model, more extensive sampling over a well-defined area might be necessary. Depending on the type of predictive modeling, a strategic sampling program might be performed within a short time frame to allow calibration and verification of the model. In this way monitoring program design and scale are dependent on the stated monitoring objectives.

Monitoring programs can be implemented for one or many reasons. The more common objectives of monitoring programs are discussed briefly below.

1. Monitoring Objective Category: Problem Identification

    1.1. Objective A: 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. Biological monitoring is particularly useful when this is the objective.

    1.2. Objective B: Determine whether further investigation is necessary

It might be necessary to determine whether waterbody impairment is severe enough to warrant management attention. Preliminary monitoring might reveal that a problem is more complicated or more serious than originally thought and that more intensive monitoring studies are called for.

2. Monitoring Objective Category: Regulatory Compliance

2.1. Objective C: Assess compliance with water quality standards

Measurements of individual pollutants in waterbodies are taken to determine whether violations of water quality standards are occurring. Monitoring for this purpose is usually conducted on a small (plot) scale. Caution is necessary to be certain that measured water quality is causally linked to the source being assessed for compliance.

3. Monitoring Objective Category: Management Benefit Evaluation

3.1. Objective D: Measure the effectiveness of individual conservation practices

Individual BMPs are monitored to determine the extent of pollution control attributable to them. Monitoring can typically be conducted at a plot or field scale, and measurements should be taken as close as possible to where the BMP is implemented. Studies of some individual practices can often be conducted in a relatively short time (< 5 yr), but this depends on the BMP being investigated.

3.2. Objective E: Measure BMP program effectiveness

This monitoring usually must be conducted on a watershed scale because the combined effect of a few or several BMPs is being investigated. It must be conducted over a long term (< 5 yr) because BMP implementation can take years to affect water quality. This type of monitoring is difficult because of a lack of control conditions, the presence of reserves of pollutants in soil and sediments, the effect of there being many land uses within a study area, and the necessity to account for the effects of many landowners with differing means of implementing the same BMPs.

4. Monitoring Objective Category: Cause-and-Effect Characterization

4.1. Objective F: Define a water quality problem

Monitoring might be required to determine the cause of an environmental problem, such as degraded fish habitat or an algal bloom. 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.

4.2. Objective G: Determine the geographical extent of the problem

Even if a problem is known to exist, the areal extent of the problem might not be known. Does it affect only a stream reach, or are problems in the downstream lake attributable to the same source

that affects the stream reach? Monitoring might be necessary to determine this.

4.3. Objective H: Determine the temporal variability of the problem and its time of occurrence

Some pollution sources are emitted only during certain parts of the year or in association with certain events, such as storms. Similarly, a constant source of pollution might be a problem only during a particular time of the year, such as during fish spawning. Determining the temporal aspect of a pollution problem will help focus management on times when it will have the most impact.

4.4. Objective I: Identify the sources of the problem

Determining that a problem exists is often far easier than determining its source because there are often many potential sources. Monitoring will help identify the sources of a problem, and further monitoring (under other objectives) can determine their individual significance.

4.5. Objective J: Quantify the relative contribution of nonpoint sources

Point and nonpoint sources often affect the same waterbody, and monitoring might be required to determine the contribution of each to water quality impairment. Quantitative determination of the contribution of each source of a particular pollutant is important. This can often be accomplished with several coordinated plot studies and within a short time frame.

4.6. Objective K: Determine the relative impact of nonpoint sources

This objective requires more intensive monitoring because it is focused on the relative impact or importance of each source of a pollutant to the impairment of a waterbody. Factors such as the timing of pollutant contributions relative to the hydrological cycle of the waterbody and the ecology of the biological communities must be factored into the analysis. The distance of pollutant sources to receiving waters, the fate and transport of pollutants from different sources, the magnitude of pollutant contribution 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) are all important factors to consider when determining the relative impacts of different sources.

5. Monitoring Objective Category: Management Alternatives Identification

5.1. Objective L: Determine fate and transport

Determination of where a pollutant goes after its introduction into the environment, how it changes chemically, where it ends up, and

by what means it travels will indicate what sources need to be controlled. This does not usually require a long monitoring period; it can be done over a short time (1-2 yr or more) and in a small area with frequent sampling.

5.2. Objective M: Locate critical areas

A determination of which areas within a watershed are most critical in causing waterbody impairment, and therefore should be the focus of pollution control efforts, is important to the development of a management strategy. 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. Monitoring for this purpose is usually conducted over a short time period, so it is critical that climatic factors be accounted for. Monitoring to locate critical areas should be basinwide to target subwatersheds for BMP implementation.

5.3. Objective N: Analyze trends

The objective here is to answer the question, "How is water quality changing over time?" Baseline monitoring is part of trend analysis because establishing a baseline is crucial 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.

5.4. Objective O: Determine the level of control needed

Monitoring for this objective involves not only an assessment of the degree of impairment and the degree of pollutant reduction needed to achieve a specified level of environmental quality, but also an analysis of the pollutant reduction capabilities of individual BMPs. The latter can be seen as a separate step, i.e., the first step being to determine the amount of pollution reduction required, and the next a determination of what BMPs can be used to achieve the reduction. Because BMPs are the pollutant control mechanisms that must be used to reduce pollutant inputs, however, the two exercises are really two aspects of a single management goal.

5.5. Objective P: Make wasteload allocations

This objective is to determine the land uses or intensity of existing land uses that can be practiced while maintaining the ecological integrity of receiving waterbodies. This monitoring objective is often associated with point sources, but it is a valuable tool for NPS pollutant control as well. Receiving waterbodies must be monitored to achieve this objective, and a good knowledge of the actual pollutant contributions of individual sources is required. For NPS pollution waste load allocation, extensive monitoring is often required.

6. Monitoring Objective Category: Model Development

6.1. Objective Q: Calibrate models

Monitoring is necessary to determine values for uncertain or unmeasurable parameters in a model before it can be used to predict the effect of management techniques or activities.

6.2. Objective R: Verify models

Model verification involves the simulation of cases covering the full range of model uses and comparison to observed data to establish confidence in the model. If premodeling monitoring data are not available, observed data will have to be collected through a monitoring program for comparison to model runs. Calibration might be an aspect of model verification. Once a model has been verified, it can be used to assist managers with management decisions.

6.3. Objective S: Test models

Monitoring is needed to fit or calibrate a model to local conditions and to verify a model's applicability to a particular situation. One problem encountered in model calibration and verification is that many models are meant to simulate long-term average conditions, but monitoring data are often restricted to a short time frame.

7. Monitoring Objective Category: Research

7.1. Objective T: Conduct research

Research monitoring is done to address specific research questions. Research monitoring is generally 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 of relevant and well-conducted research projects in which monitoring is a key element can provide excellent guidance for the design of a monitoring program.

C. Monitoring Level

Monitoring level refers to the level of effort required for data collection, the amount and type of data that must be collected, and the equipment and personnel needed to conduct a monitoring program. A low monitoring level might entail no more than a site visit to qualitatively describe riparian vegetation condition. A high monitoring level might involve extensive benthic macroinvertebrate sampling and taxonomic analysis coupled with physical and chemical water quality sampling conducted over a period of years.

Monitoring level, like monitoring program design, is also largely determined by a monitoring program's objectives, though

there is some leeway in choice of monitoring level for achieving most monitoring objectives. A high monitoring level is required to achieve some monitoring objectives, while other objectives can be achieved with a lower monitoring level. However, when a low monitoring level can be used, the use of a higher monitoring level is not precluded. In addition to monitoring objectives, the monitoring level is chosen based on availability of staff, time, money, and equipment; physical location of the resource being monitored; season; and complexity of the problem being monitored.

Three levels of monitoring are described here. Monitoring programs have been divided elsewhere into fewer (two) or more (four) monitoring levels, but a division into three levels allows for adequate distinction between the levels without drawing unnecessarily fine distinctions that would be applicable only to limited circumstances.

1. Level I: Preliminary Assessment

Level I monitoring relies largely on empirical observations and best professional judgment (BPJ). The use of data is limited to those readily available, such as historical data and information available from sources such as those listed in Chapter 5. An assessment of this historical data is part of a preliminary assessment. Limited field work is involved, such as a site survey to verify land use or vegetation type. Level I monitoring can be used for a variety of purposes, including:

- determine whether a water quality problem exists
- determine whether water quality goals are being met
- assess current conditions
- document water quality problems
- document water quality standards violations
- determine general water quality.

2. Level II: Problem Characterization

Level II monitoring uses qualitative and quantitative data in addition to level I monitoring information. Example uses of level II monitoring include:

- locate critical areas
- identify sources and evaluate their relative magnitudes
- characterize critical conditions
- characterize pollutant fate and transport.

Data collection is necessary for level II monitoring but is limited to parameters that can be measured with limited equipment, time, and staff. Biological monitoring parameters for level II monitoring might include coliform bacteria count, phytoplankton and zooplankton density, areal extent and species of macrophytes, benthic macroinvertebrate families and distribution, and fish species. Habitat quality monitoring parameters might include stream substrate type, flow, channel morphology, riparian cover density, and shoreline habitat condition. Physical and chemical monitoring variables for level II monitoring include water

temperature, sedimentation rate, water transparency, turbidity, phosphorus concentration, and nitrogen concentration.

3. Level III: Long-term Monitoring (>5 years)

Level III monitoring relies primarily on quantitative data to provide detailed analyses of water quality problems and to build time series data for assessing trends, determining management program effectiveness, and regulatory compliance. Some uses of level III monitoring include:

- perform trend analysis
- determine causality
- demonstrate BMP program effectiveness
- conduct research
- define biological criteria for water quality standards
- define a desired future condition
- determine compliance with state water quality standards
- determine whether it is necessary to modify a BMP system
- make wasteload allocations.

Biological parameters for level III monitoring include those from levels I and II plus pathogens, viruses, intestinal parasites, periphyton, and a more detailed analysis of macrophytes, such as a calculation of biomass from in-water plot collections. Level III monitoring habitat parameters might include flow at high and low flow events, water velocity, water depth, coarse particulate organic matter (CPOM), and interstitial water. A substratum habitat description, including parent material and presence of trash, organic matter, odors, oils, and sewage might be necessary, as well as a determination of anaerobic conditions.

D. Monitoring Scale

The scale of a monitoring program has two components: a time or temporal scale and an areal or geographic scale. The temporal scale is the amount of time required to accomplish the program objectives. It can vary from an afternoon to many (>5) 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 watershed. 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. However, as with monitoring level, there is often some leeway in selecting both temporal and geographic scales to achieve specific objectives. Other factors that should be considered to determine appropriate temporal and geographic scales include: (1) available resources (staff and money); (2) time frame within which managers require specific information; (3) type of water resource being monitored; and (4) complexity of the NPS problem.

As mentioned above, both scales are dependent on the objectives of the program. If the main objective is to determine

the present biological condition of a stream, then sampling at a few stations in a stream reach over 1 or 2 days might suffice. If the objective is to determine the effectiveness of a nutrient management program for reducing nutrient inputs to a downstream lake, then monitoring a subwatershed for 5 years or longer 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. 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.

A monitoring program to be 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 of these factors is important because each has an influence on stream characteristics, while no direct relationship exists between watershed area and stream characteristics. Watersheds larger than 50,000 acres might contain enough heterogeneity to preclude the use of a single monitoring program.

The temporal and geographic scales of a monitoring program will usually be similar to the level of the monitoring program (I, II or III). That is, level I programs are more likely to be of short duration and limited geographic extent, while level III programs are more likely to be of longer duration and larger areal extent. Studies to assess the present condition of a waterbody can often be completed in a season, or less than a year, and by sampling in select places. Problem characterization studies or trend analyses take longer and are more involved in terms of the parameters to be monitored, the area that must be monitored, and the amount of data that must be collected. Model validation monitoring includes both a first phase of model calibration and a second phase of model verification, and overall can take anywhere from a few months to several years.

Table II-1 presents approximate monitoring scales for the monitoring objectives listed above.

Table II-1. Relationship of monitoring objectives and scales.

		GEOGRAPHICAL		
		Stream Reach	Subwatershed	Watershed
TEMPORAL	Days	A B C		
	Weeks		Q S	G M
	Months	D T	F O P	H I
	Years			E J K L N R

KEY
A: Determine whether an impairment exists
B: Determine whether further investigation is necessary
C: Assess water quality standards compliance
D: Measure individual BMP effectiveness
E: Measure BMP program effectiveness
F: Define a water quality problem
G: Determine the geographical extent of a problem
H: Determine the temporal variability and time of occurrence of a problem
I: Identify sources of a problem
J: Quantify the relative contribution of nonpoint sources
K: Determine the relative impact of nonpoint sources
L: Determine fate and transport
M: Locate critical areas
N: Analyze trends
O: Determine the level of control needed
P: Make watershed allocations
Q: Calibrate models
R: Verify models
S: Test models
T: Conduct research

CHAPTER 3

REVIEW OF SELECTED AVAILABLE MONITORING GUIDANCES

III. REVIEW OF SELECT AVAILABLE MONITORING GUIDANCES

A. Introduction

Numerous guidance documents have been developed, or are in development, to assist resource managers in developing and implementing monitoring programs for both point and nonpoint sources of pollution. This chapter presents a review of numerous monitoring guidances for both point and nonpoint pollution. Guidances specific to point source control program monitoring are reviewed because very often receiving waterbodies contain contaminants from both point and nonpoint sources of pollution and resource managers must take both point and nonpoint sources into account when devising a monitoring program. Since point source monitoring programs are not specifically the topic of this guidance document, one or more of the guidances reviewed here for designing and implementing point source monitoring programs should be consulted if point source monitoring is required as part of a monitoring program.

Despite their different focuses, point and nonpoint source pollution monitoring programs often have similar or identical designs, objectives, parameters to be measured, and statistical analyses. Therefore, the guidance provided in this document is generally applicable to point source monitoring programs, and guidance for point source monitoring is generally applicable to nonpoint source monitoring programs. A possible exception is monitoring programs for a specific mandated program, such as the 301(h) or 403(c) program. These programs often have specific legal monitoring requirements, and the appropriate monitoring guidance should be consulted when designing and implementing a monitoring program for one of the programs.

The guidances reviewed in this chapter discuss virtually every aspect of NPS pollution monitoring, including monitoring program design and objectives, sample types and sampling methods, chemical and physical water quality parameters, biological monitoring, data analysis and management, and quality assurance and quality control. However, all of these points are not discussed in each guidance reviewed. Therefore, to help the reader find the guidance(s) most appropriate to specific needs, each review highlights those aspects of monitoring which are stressed in the particular guidance.

Grouping the guidances according to some seemingly logical system is both easy and difficult: easy because they can be grouped in any of a hundred ways with no one grouping being superior and difficult for the same reason—they follow no natural organization. Therefore, for lack of a better system, they 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

There is some reasoning behind this grouping. The reviews have been organized primarily according to the type of waterbody or waterbody system to which the guidance is most applicable, i.e., 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 NPDES program or the National Estuary Program, are listed separately.

A complete listing of the guidances reviewed in this chapter is presented first, followed by the individual guidance reviews. A table at the end of the chapter 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.

B. Categorical Listing of Monitoring Guidances Reviewed

1. General Monitoring Guidances and References

1. *Water Quality Monitoring*. USDA-SCS Draft.

2. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring*. Coffey, Spooner, and Smolen 1993.

3. *Designing Effective Nonpoint Source Water Quality Monitoring Programs*. Maas 1989.

4. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document*. Minnesota Pollution Control Agency 1989.

5. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards*. Dissmeyer 1994.

6. *Volunteer Water Monitoring: A Guide for State Managers*. USEPA 1990.

7. *Volunteer Estuary Monitoring: A Methods Manual*. USEPA 1993.

8. *The Volunteer Monitor*. The National Newsletter of Volunteer Water Quality Monitoring.

9. *Water-Quality Monitoring in the United States*. Intergovernmental Task Force on Monitoring Water Quality 1994.

10. *Water-Quality Monitoring in the United States. Technical Appendixes.* Intergovernmental Task Force on Monitoring Water Quality 1994.

11. *Guidance for State Water Monitoring and Wasteload Allocation Programs.* USEPA 1985.

12. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects.* USEPA 1981.

2. Guidances for Stream and River Monitoring

13. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* MacDonald, Smart and Wissmar 1991.

14. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams.* USEPA Draft.

15. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* USEPA 1989.

16. *An Improved Biotic Index of Organic Stream Pollution.* Hilsenhoff 1987.

17. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin.* Lyons 1992.

18. *Handbook: Stream Sampling for Waste Load Allocation Applications.* USEPA 1986.

19. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects.* Simonson and Lyons 1992.

20. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects.* Simonson and Lyons 1992.

21. *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.

3. Guidances for Lake and Reservoir Monitoring

22. *Monitoring Lake and Reservoir Restoration.* USEPA 1990.

23. *Volunteer Lake Monitoring: A Methods Manual.* USEPA 1991.

24. *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses.* Volumes I-III. USEPA 1983-1984.

4. Guidances for Watershed Monitoring

25. *Monitoring Primer for Range Watersheds*. Bedill and Buckhouse Draft.

26. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 years of controlling agricultural nonpoint source pollution: The RCWP experience*. USEPA 1992.

5. Guidances for Ground Water Monitoring

27. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water*. USEPA 1991.

6. Guidances for Biological Monitoring

28. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data*. Volume I. USEPA 1987.

29. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. USEPA 1993.

30. *Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents*. USEPA 1973.

31. *Methods for Sampling Fish Communities as a Part of the National Water-quality Assessment Program*. Meador, Cuffney and Gurtz 1993.

7. Program-specific Monitoring Guidances

32. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects*. USEPA 1991.

33. *Monitoring Guidance for the National Estuary Program*. USEPA 1991.

34. *QC-QA Plan Rural Nonpoint*. USGS 1992.

35. *NPDES Storm Water Sampling Guidance Document*. USEPA 1992.

36. *Water Quality Standards Handbook*. USEPA 1983.

37. *Environmental Monitoring and Assessment Program: Ecological Indicators*. USEPA 1990.

38. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document*. USEPA 1986.

39. *Summary of U.S. EPA-Approved Methods, Standard Methods, and Other Guidance for 301(h) Monitoring Variables*. USEPA 1985.

40. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document*. USEPA 1989.

41. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance.* USEPA 1989.

42. *CWA Section 403: Procedural and Monitoring Guidance.* USEPA 1994.

C. Guidance Reviews

1. General Monitoring Guidances and References

1. *Water Quality Monitoring.* USDA-SCS, Draft.

REFERENCE: USDA-SCS. 1993. *Water Quality Monitoring.* Draft. June.

CONTENTS: Defining water quality problem; Objectives; Scale of study; Variable selection; Sample type; Sampling location; Sampling frequency and duration; Monitoring study designs; 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 is divided into two parts. Part 1 contains 12 chapters, each of which covers a specific step of monitoring study design: (1) define water quality problem, (2) define monitoring objectives, (3) determine study scale, (4) select water quality variables, (5) select sample type, (6) determine sampling locations, (7) determine sampling frequency and duration, (8) describe statistical design, (9) design station type, (10) describe sample collection and analysis methods, (11) describe land use monitoring, and (12) describe data management. Part 2 discusses statistical analysis of monitoring results.

2. *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, 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.

3. *Designing Effective Nonpoint Source Water Quality Monitoring Programs*. Maas 1989.

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

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.

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.

4. *Water Quality Monitoring for the Clean Water Partnership. A Guidance Document.* Minnesota Pollution Control Agency 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.

5. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards.* Dissmeyer 1994.

> **REFERENCE:** USDA-FS. 1994. *Evaluating the Effectiveness of Forestry Best Management Practices in Meeting Water Quality Goals or Standards.* By G.E. Dissmeyer, 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 BMP effectiveness longevity. 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.

6. *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

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.

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.

7. *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. 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* (in progress as of 12/93). 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 nations'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.

8. *The Volunteer Monitor*. The National Newsletter of Volunteer Water Quality Monitoring.

REFERENCE: *The Volunteer Monitor*. Published by the Coastal Resources Center, The University of Rhode Island, Rhode Island Sea Grant Program, Narragansett, Rhode Island, 02882.

CONTENTS: Vary from issue to issue. Example issue (vol. 3, no. 1, Fall 1991): 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 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.

9. *Water-Quality Monitoring in the United States*. Intergovernmental Task Force on Monitoring Water Quality 1994.

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 second annual report of the ITFM. It describes the progress of the ITFM and its task groups in developing concepts, guidelines, and procedures for use in a nationwide monitoring strategy. The recommendations of the ITFM will be refined in 1994, its third and last year, and published in its final report. The information in

REFERENCE: ITFM. 1994. *Water-quality monitoring in the United States.* Intergovernmental Task Force on Monitoring Water Quality, Interagency Advisory Committee on Water Data, Water Information Coordination Program. Washington, DC. January.

CONTENTS: Executive summary; Second year's progress—answering the questions; Third year plan; Conclusions and recommendations.

MAIN FOCUS: In-depth discussions of monitoring program problems that program managers need to consider.

this document specifically addresses the needs and direction of the nationwide monitoring effort, but members of the ITFM have given careful thought to problems that plague many monitoring programs and the monitoring program concepts and issues discussed in this report should be considered seriously by anyone undertaking or modifying a monitoring program.

10. *Water-Quality Monitoring in the United States. Technical Appendixes.* Intergovernmental Task Force on Monitoring Water Quality 1994.

REFERENCE: ITFM. 1994. *Water-quality monitoring in the United States. Technical appendixes.* Intergovernmental Task Force on Monitoring Water Quality, Interagency Advisory Committee on Water Data, Water Information Coordination Program. Washington, DC. January.

CONTENTS: Glossary; Framework for water-quality monitoring; Charter for the National Water-quality Monitoring Council; Indicators for meeting management objectives—summary matrix; Indicators for meeting management objectives—rationale matrix; Indicator selection criteria; Ecoregions, reference conditions, and index calibration; Multimetric approach for describing ecological conditions; Charter for Methods and Data Comparability Council; 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; Ground-water indicators.

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.

This companion document to the ITFM 1993 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 document in a single place.

11. *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 (TMDL) and waste load allocations for point sources (wasteload allocations, WLA) and nonpoint sources (load allocations, LA) 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 type 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.

12. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects*. USEPA 1981.

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

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 that should be monitored of each receiving water type are discussed.

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 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 to 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 a NPS pollution control project. Two levels of socioeconomic evaluations are discussed. A level I evaluation considers the effects of a 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.

2. Guidances for Stream and River Monitoring

13. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska*. MacDonald, Smart, and Wissmar 1991.

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.

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.

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.

14. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams.* USEPA Draft.

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

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.

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 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 are then thoroughly described, including data collection and analysis procedures. 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.

15. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish*. USEPA 1989.

This is a comprehensive technical reference for bioassessment procedures for streams and rivers. Three macroinvertebrate protocols and two fish protocols are presented. The

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.

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 various 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.

16. *An Improved Biotic Index of Organic Stream Pollution.* Hilsenhoff 1987.

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 1000 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.

REFERENCE: Hilsenhoff, W.L. 1987. *An Improved Biotic Index of Organic Stream Pollution.* The Great Lakes Entomologist ?: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.

17. *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.

18. *Handbook: Stream Sampling for Waste Load Allocation Applications.* USEPA 1986.

This report discusses sampling requirements in support of waste load allocation studies in rivers and streams. Two approaches to waste load allocation are addressed: the chemical-specific

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 waste load 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.

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 waste load applications. The data collection process required to calibrate, verify, and apply models used for waste load allocations to critical design conditions is described.

19. *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 NPSP. 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.

20. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects.* Simonson and Lyons 1992.

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.

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.

21. *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 abstract)

3. Guidances for Lake and Reservoir Monitoring

22. *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) and 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.

23. *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.

24. *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

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.

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.

4. Guidances for Watershed Monitoring

25. *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. While 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.

26. *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.

5. Guidances for Ground Water Monitoring

27. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* USEPA 1991.

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

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.

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.

6. Guidances for Biological Monitoring

28. *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 recent (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.

29. *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.

30. *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.* U.S. Environmental Protection Agency, National Environmental Research Center, Office of Research and Development, Cincinnati, Ohio. Program Element 1BA027. C.I. Weber, ed. 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, and 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.

31. *Methods for Sampling Fish Communities as a Part of the National Water-quality Assessment Program.* Meador, Cuffney and Gurtz 1993.

This document provides detailed procedures for evaluations of fish community structure. The procedures are intended to facilitate standardization of collection methods and fish community descriptions to facilitate unbiased evaluations of relations among physical, chemical, and biological components of water-quality conditions. The methods described are standard for use in streams ranging from headwaters to large rivers. Electrofishing and seining are the methods stressed in the document, but others are considered. Sampling program design and sample processing are discussed in detail, and forms for recording data are provided. (from introduction)

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.* U.S. Geological Survey Open-file Report 93-104.

CONTENTS: 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; Summary.

MAIN FOCUS: This manual documents standard procedures for evaluating stream fish communities. It focuses on electrofishing and seining. Taxonomic identification, measurements, and external examination of fish specimens are discussed and forms are provided for data entry.

7. Program-specific Monitoring Guidances

32. *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 § 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.

33. *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. 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 citizen's 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 mangers, 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, etc. 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.

34. *QC-QA Plan Rural Nonpoint*. USGS 1992.

REFERENCE: Not available.

CONTENTS: QC Procedures; Field Procedures; Record Keeping; Sample Processing; Quality Assurance; Blank Sample Processing.

MAIN FOCUS: Provides a detailed description of field sampling and laboratory sample analysis techniques. Analysis of blank samples for quality assurance is discussed in detail as well.

This short (5-page) document describes the quality control and quality assurance procedures that apply to all USGS rural nonpoint data collection activities. Field procedures for sample collection are described in detail. Sampling frequency, equipment specifications, and general sampling procedures are discussed to ensure consistent sample collection techniques at all USGS rural nonpoint collection points. Laboratory procedures for sample analysis are discussed in similar detail. The document notes what records are to be kept. Quality assurance consists of regular analysis of blank samples and equipment maintenance, both of which are discussed.

35. *NPDES Storm Water Sampling Guidance Document*. USEPA 1992.

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

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.

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.

36. *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 water body 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 a waterbody 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.

37. *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) was initiated by EPA's Office of Research and Development to provide data for improved assessment of the Nation's ecological resources. Under the program, several integrated monitoring networks will be

implemented by the mid-1990s to (1) estimate current status and trends in indicators of ecological condition and (2) monitor indicators of pollutant exposure and habitat condition and identify cause-and-effect associations between indicators and adverse conditions. 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 near-coastal waters, inland surface waters, wetlands, forests, arid lands, and agroecosystems.

EMAP is a large-scale national monitoring program whose scope is beyond the reach of most state, local, or regional monitoring efforts. Therefore, the objectives and design of the program will not be similar to many smaller monitoring programs. However, the discussions of monitoring indicators for the six ecosystem types mentioned above provide valuable information on types of indicators to choose for different monitoring goals. Indicators that were considered for EMAP but were not chosen for it are also discussed, which provides valuable information as well in that it points out for what purposes these monitoring parameters are 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.

38. *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. While 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.

39. *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 is a short (16-page) document that 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.

40. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* USEPA 1989.

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

REFERENCE: USEPA. 1989. *Ecological Assessments of Hazardous Waste Sites: A Field and Laboratory Reference Document.* U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. EPA/600/3-89/013. March. W. Warren-Hicks, B.R. Parkhurst, and S.S. Baker, Jr., eds.

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.

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.

41. *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.

42. *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.

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																		
	1	2	3	4	5	6	7	9	10	11	12	13	14	15	17	18	19	20	21	22
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	9	10	11	12	13	14	15	17	18	19	20	21	22
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. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring.* Coffey, Spooner and Smolen 1993.

3. *Designing effective nonpoint source water quality monitoring programs.* Maas 1989.

4. *Water quality monitoring for the Clean Water Partnership. A guidance document.* Minnesota Pollution Control Agency 1989.

5. *Evaluating the effectiveness of forestry best management practices in meeting water quality goals or standards.* Dissmeyer 1994.

6. *Volunteer Water Monitoring: A Guide for State Managers.* USEPA 1990.

7. *Volunteer estuary monitoring: A methods manual.* USEPA 1993.

9. *Water-quality monitoring in the United States.* Intergovernmental Task Force on Monitoring Water Quality 1994.

10. *Water-quality monitoring in the United States. Technical appendixes.* Intergovernmental Task Force on Monitoring Water Quality 1994.

11. *Guidance for State Water Monitoring and Wasteload Allocation Programs.* USEPA 1985.

12. *Guidelines for Evaluation of Agricultural Nonpoint Source Water Quality Projects.* USEPA 1981.

13. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska.* MacDonald, Smart and Wissmar 1991.

14. *Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management of Western Rangeland Streams.* USEPA Draft.

15. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish.* USEPA 1989.

17. *Using the Index of Biotic Integrity (IBI) to Measure Environmental Quality in Warmwater Streams of Wisconsin.* Lyons 1992.

18. *Handbook: Stream Sampling for Waste Load Allocation Applications.* USEPA 1986.

19. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects.* Simonson and Lyons 1992.

20. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects.* Simonson and Lyons 1992.

21. *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.

22. *Monitoring Lake and Reservoir Restoration.* USEPA 1990.

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																		
	23	24	25	26	27	28	29	30	31	32	33	35	36	37	38	39	40	41	42
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																			

SUBJECTS DISCUSSED	GUIDANCE REFERENCE NUMBER																			
	23	24	25	26	27	28	29	30	31	32	33	35	36	37	38	39	40	41	42	
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

23. *Volunteer Lake Monitoring: A Methods Manual.* USEPA 1991.

24. *Technical support manual: Waterbody surveys and assessments for conducting use attainability analyses.* Volumes I-III. USEPA 1983-1984.

25. *Monitoring Primer for Range Watersheds.* Bedill and Buckhouse Draft.

26. *Seminar Publication: The National Rural Clean Water Program Symposium. 10 years of controlling agricultural nonpoint source pollution: The RCWP experience.* USEPA 1992.

27. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* USEPA 1991.

28. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* Volume I. USEPA 1987.

29. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters.* USEPA 1993.

30. *Biological field and laboratory methods for measuring the quality of surface waters and effluents.* USEPA 1973.

31. *Methods for sampling fish communities as a part of the National Water-quality Assessment Program.* Meador, Cuffney and Gurtz 1993.

32. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects.* USEPA 1991.

33. *Monitoring Guidance for the National Estuary Program.* USEPA 1991.

35. *NPDES Storm Water Sampling Guidance Document.* USEPA 1992.

36. *Water Quality Standards Handbook.* USEPA 1983.

37. *Environmental Monitoring and Assessment Program: Ecological Indicators.* USEPA 1990.

38. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document.* USEPA 1986.

39. *Summary of U.S. EPA-approved methods, standard methods, and other guidance for 301(h) monitoring variables.* USEPA 1985.

40. *Ecological assessments of hazardous waste sites: A field and laboratory reference document.* USEPA 1989.

41. *Statistical analysis of ground-water monitoring data at RCRA facilities - Interim final guidance.* USEPA 1989.

42. *CWA section 403: Procedural and Monitoring guidance.* USEPA 1994.

CHAPTER 4

WATER RESOURCES CONSIDERATIONS

IV. WATER RESOURCE CONSIDERATIONS

The water resource to be monitored must be understood before a monitoring plan that provides sufficient information for meeting the monitoring objectives can be developed. Options for tracking water quality vary with the type of water resource. For example, a monitoring program for ephemeral streams can be different from that for perennial streams or large rivers. Lakes, wetlands, riparian zones, estuaries, and near-shore coastal waters all present different monitoring considerations. Although upstream-downstream designs work on rivers and streams, they are generally less effective on natural lakes where linear flow is not prevalent. Similarly, estuaries present difficulties in monitoring loads because of the shifting flows and changing salinity caused by the tides. A successful monitoring program must take into consideration the unique features of the water resources involved and must be structured to either adapt to those features or avoid them.

A. Variability

Monitoring programs must address both temporal and spatial variability because these variabilities are key factors that determine the impact that NPS pollution has on a water resource. Temporal variability can be of short or long duration. As shown in Figure IV-1, water quality can vary tremendously over the course of a single storm event. Many water resources also undergo seasonal changes or changes that occur over the course of several years, decades, or even centuries (Figure IV-2). A water resource can vary in the short or long term, or both. There is often insufficient information to describe much of the very long-term variability (years or decades) associated with a given water resource. The temporal variability referred to in this document is that which generally occurs on the minutes-to-a year time scale.

Spatial variability is the other major component of NPS water quality variability. Within a given stream, lake, estuary, or aquifer there is usually some spatial differential in water quality. For example, in Figure IV-3 the total suspended solids concentrations vary across the reservoir. Inlets to the reservoir (points of concentration) are closer to higher concentrations of suspended materials, and thus they show higher total suspended solids levels than areas near the dam outlet where substantial settling of larger particles, dilution, and dispersion have already occurred. As another example, salinity levels in an estuary vary with depth (spatial variability must be considered in three dimensions) at the salt wedge as the more saline ocean water flows under the fresh water from the stream (Figure IV-4). Water quality can also vary considerably in an aquifer underlying agricultural lands. The example in Figure IV-5 shows how nitrate levels can differ substantially from location to location.

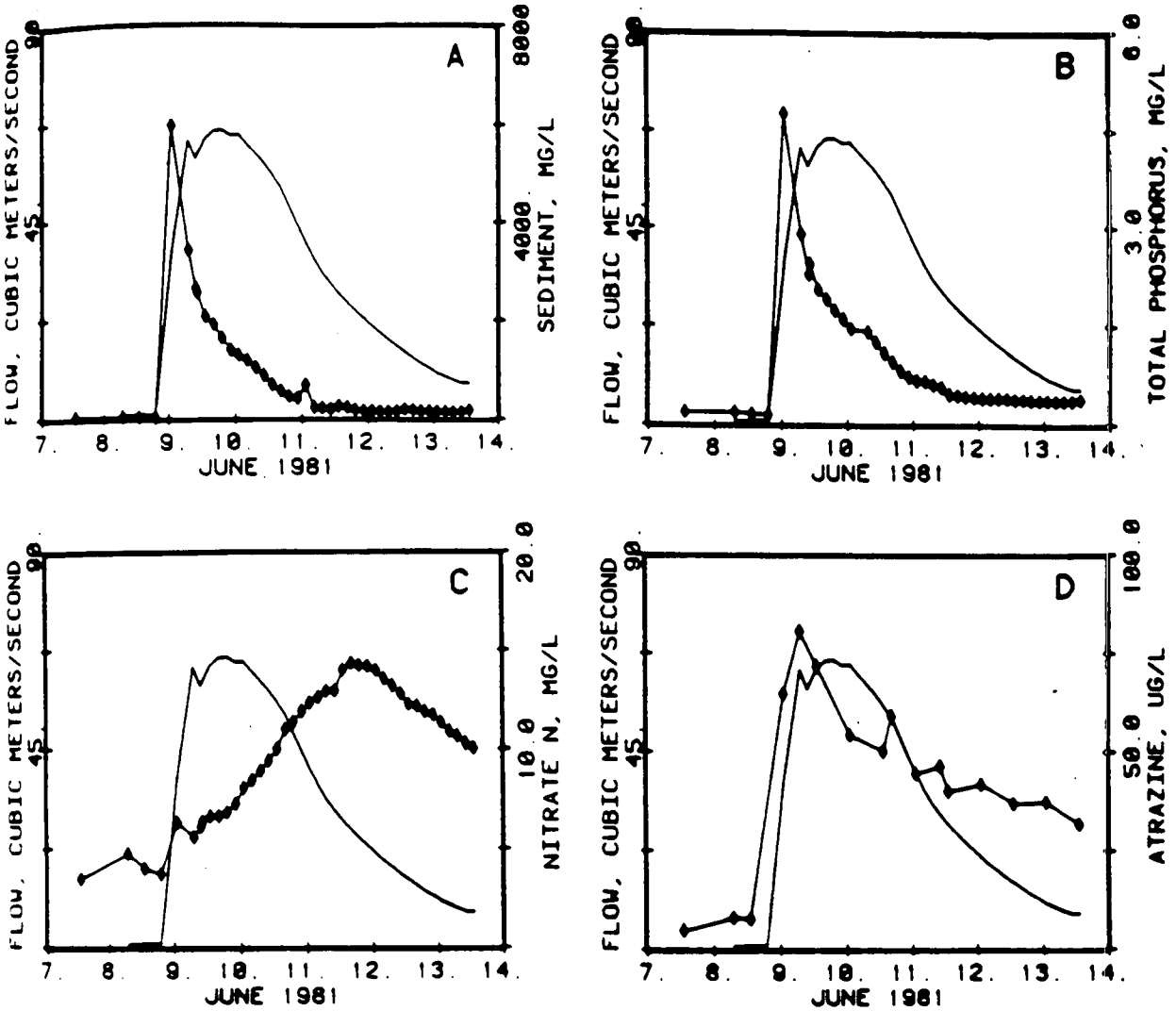


Figure IV-1. Flow and pollutant concentrations for a single storm event in the Honey Creek watershed, Ohio. (SOURCE: Baker et al., 1985)

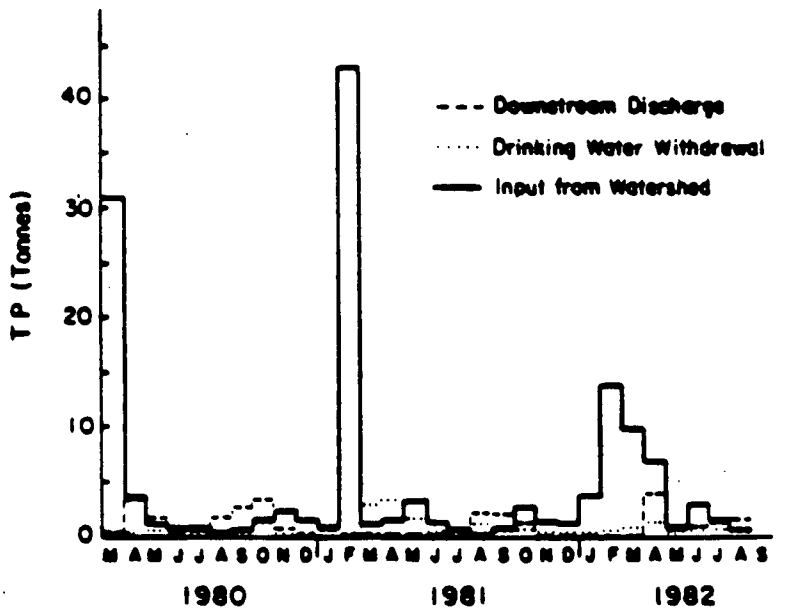


Figure IV-2. Time course of phosphorus loading from the West Branch Delaware River and output from Cannonsville Reservoir. SOURCE: Brown et al., 1986)

(TO BE INSERTED)

Figure IV-3. Mean total suspended solids concentrations (mg/L) in Highland Silver Lake, Illinois, May 1981-March 1984. (SOURCE: Davenport and Kelly, 1984)

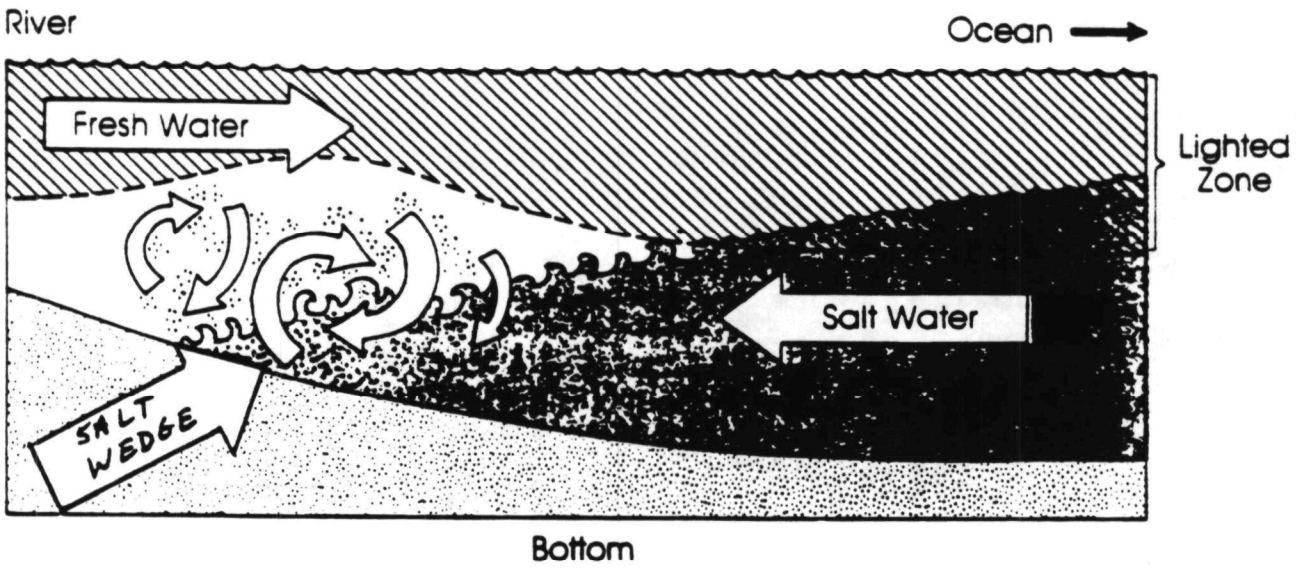


Figure IV-4. Mixing of salt water and fresh water in an estuary. SOURCE: USEPA, 1983a)

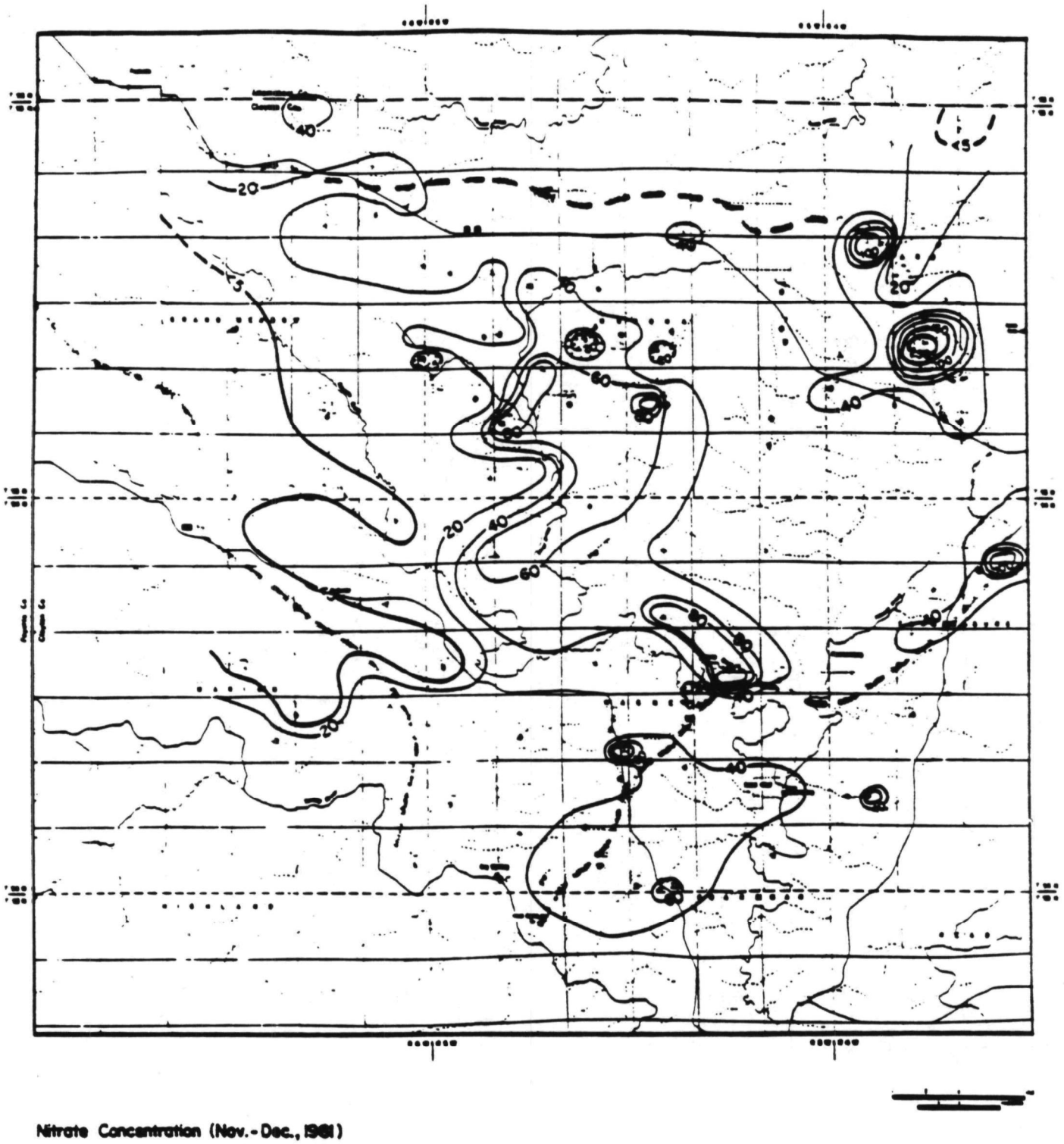


Figure IV-5. Contour map of nitrate concentrations recorded from Galena aquifer inventory water samples (Iowa). (SOURCE: Hallberg et al., 1983)

B. Streams

Freshwater streams can be classified on the basis of flow attributes as intermittent or perennial streams. Intermittent streams do not flow at all times and serve as conveyance systems for runoff. Perennial streams always flow and usually have significant inputs from ground water or interflow.

For intermittent streams, seasonal variability is a very significant factor in determining pollutant loads and water quality. During some periods sampling might be impossible due to the absence of flow. Seasonal flow variability in perennial streams can be caused by seasonal patterns in precipitation or snowmelt, reservoir discharges, or irrigation practices.

For many streams the greatest concentrations of suspended sediment and other pollutants occur during spring runoff or snowmelt periods. Figure IV-6 shows the monthly variations in the sediment-discharge relationship for the White River, South Dakota. Figure IV-7 shows the seasonal relationship between sediment concentrations and discharge for the Fern Ridge Dam, Oregon. Concentrations of both particulate and soluble chemical parameters have been shown to vary throughout the course of a rainfall event in many studies across the Nation. This short-term variability should be considered in developing monitoring programs for flowing (lotic) waterbodies.

Spatial variability is largely lateral for both intermittent and perennial streams. Vertical variability does exist, however, and can be very important in both stream types (e.g., during runoff events, in tidal waters, and in deep, slow-moving streams). Intake depth is often a key factor in stream sampling. For example, slow-moving, larger streams can show considerable water quality variability with depth, particularly for parameters such as suspended solids, dissolved oxygen, and algal productivity. Figure IV-8 illustrates the relative location of suspended load and bed load in a stream vertically. Suspended sediment samples must be taken with an understanding of the vertical distribution of both sediment concentration and flow velocity (Brakensiek et al., 1979), as illustrated in Figure IV-9. When sampling bed sediment or monitoring biological parameters, it is important to recognize the potential for significant lateral and vertical variation in the toxicity and contaminant levels of bed sediments (USEPA, 1987).

C. Lakes

Lakes can be categorized in several ways, but a useful grouping for monitoring guidance is related to the extent of vertical and lateral mixing of the waterbody. Therefore, lakes are considered to be either mixed or stratified for the purpose of this guidance. Mixed lakes are those lakes in which water quality (as determined by measurement of the parameters and attributes of interest) is homogenous throughout, and stratified lakes are

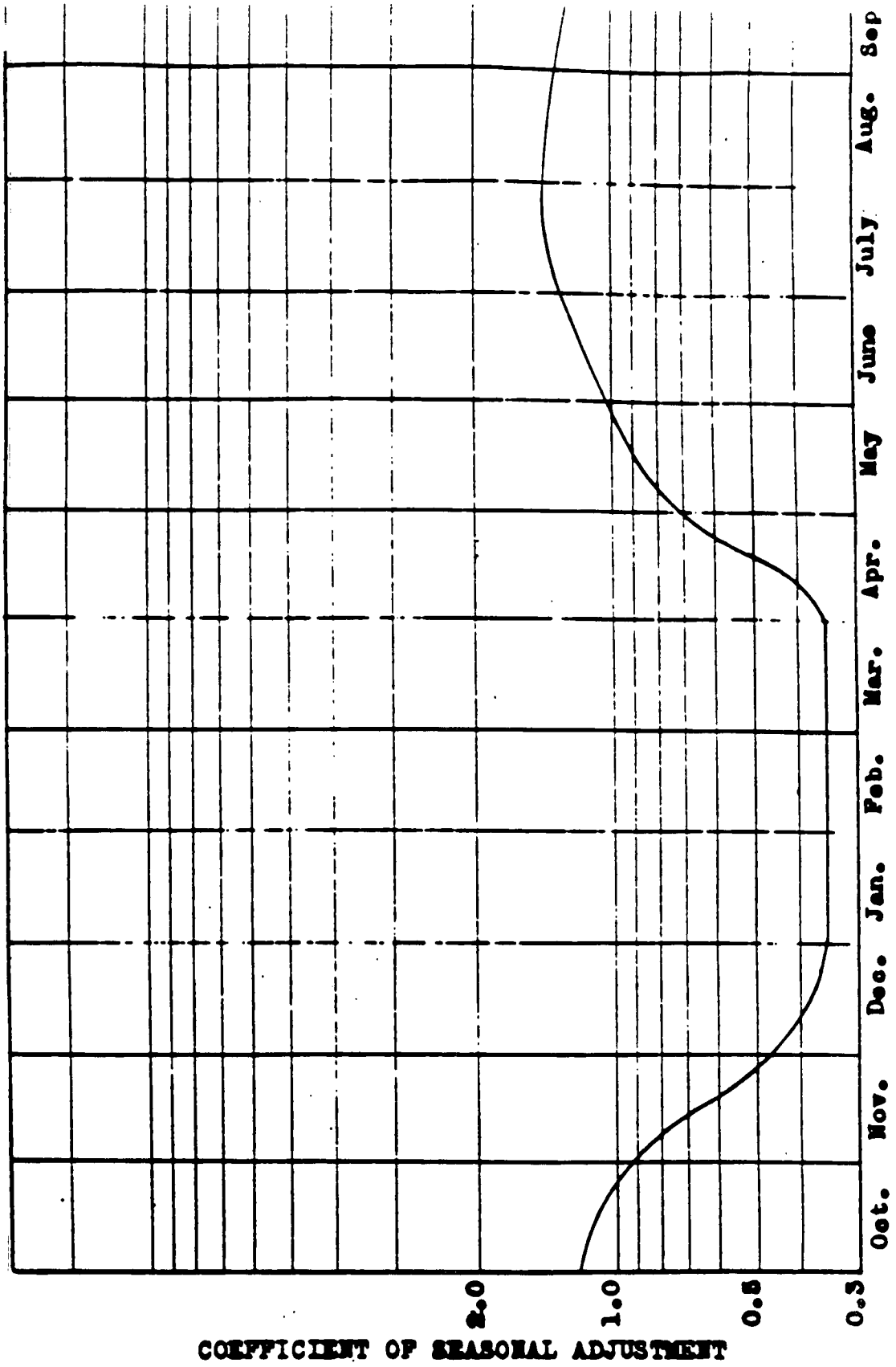


Figure IV-6. Seasonal adjustment to discharges of sediment, in clay and silt sizes, White River near Kadoka, South Dakota. (SOURCE: Colby, 1956)

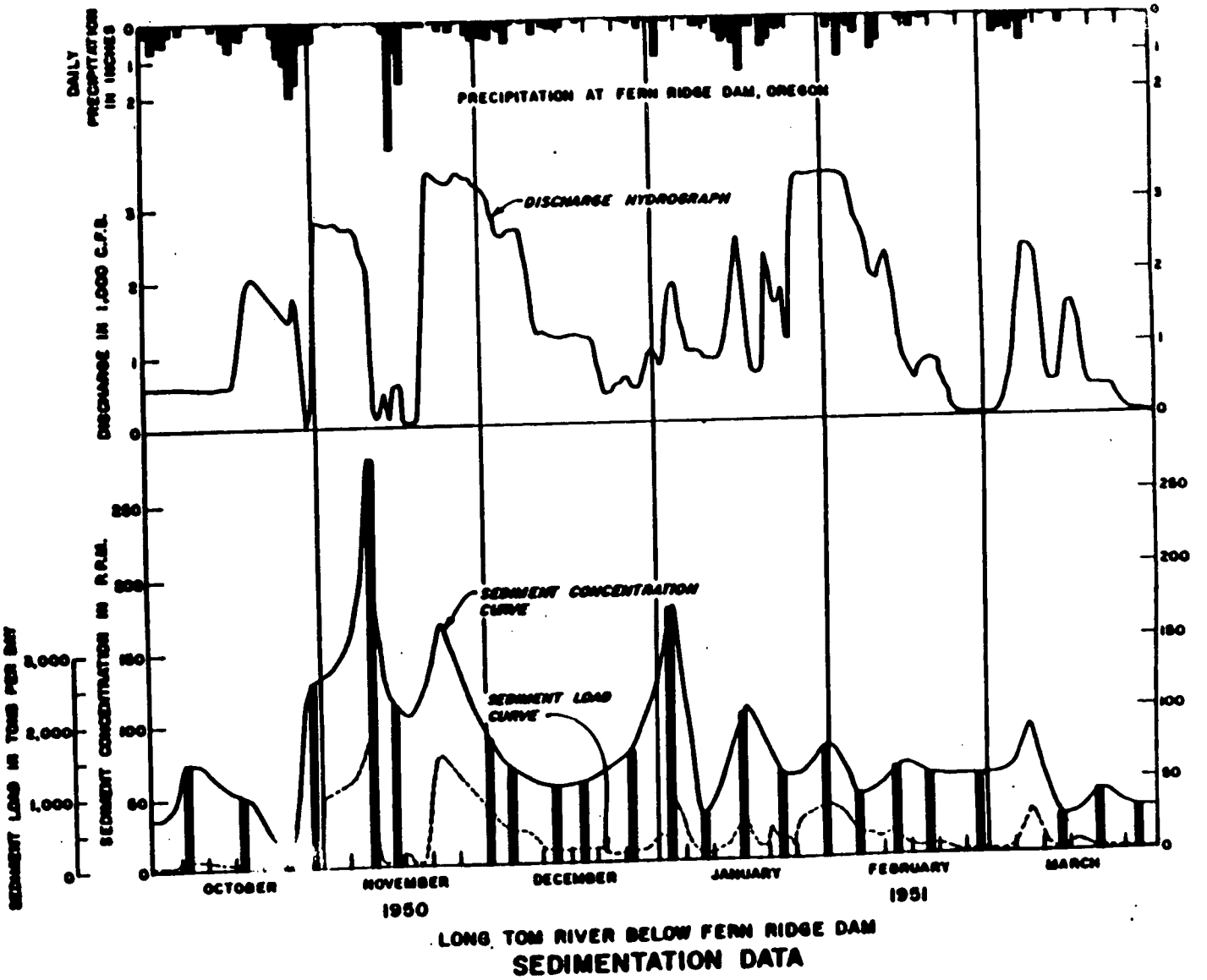


Figure IV-7. Seasonal relationship between sediment concentration, sediment load, discharge, and precipitation for Fern Ridge Dam, Oregon. (SOURCE: Thomas, 1970)

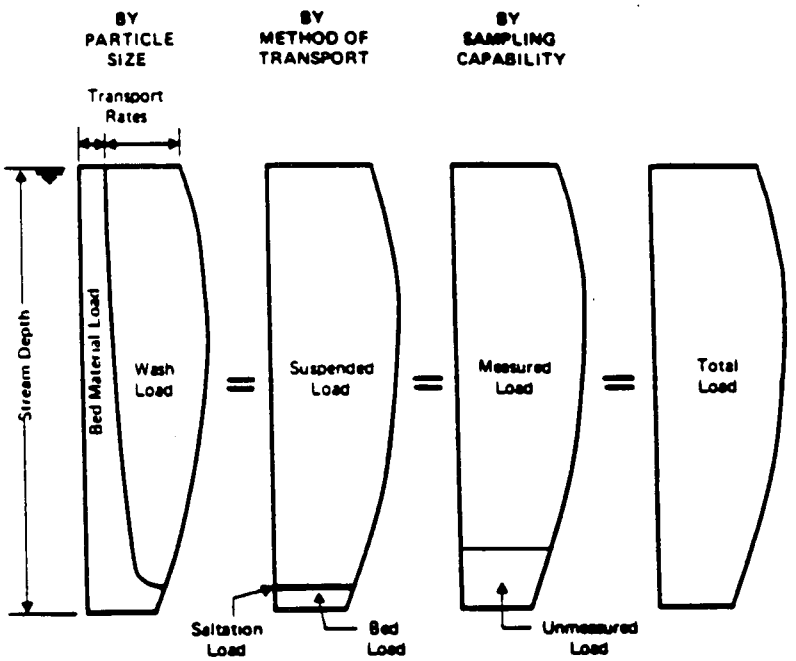
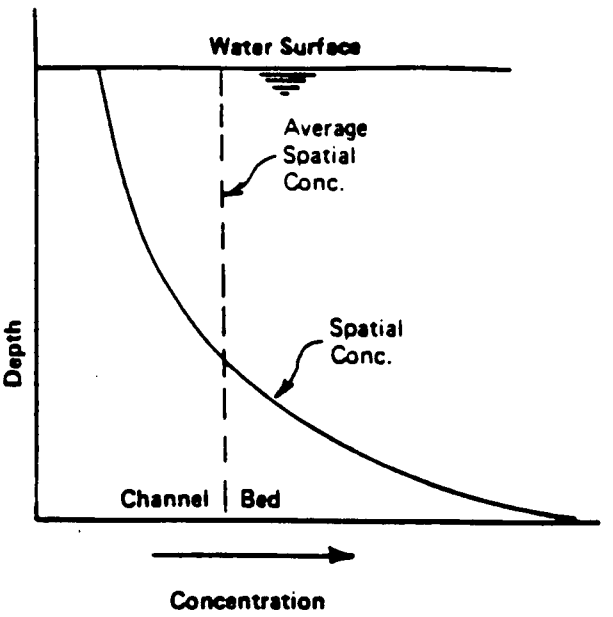


Figure IV-8. Schematic diagram of stream vertical showing relative position of sediment load terms. (SOURCE: Brakensiek et al., 1979)



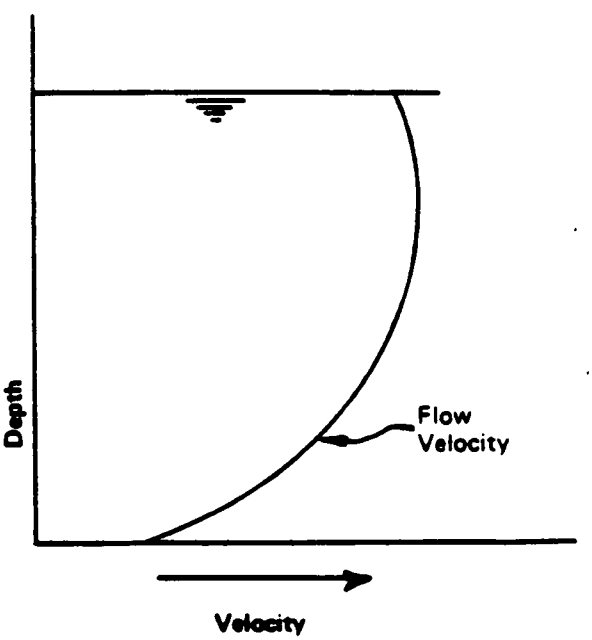


Figure IV-9. Vertical sediment concentration and flow velocity distribution in a typical stream cross section. (SOURCE: Brakensiek et al., 1979)

considered to be those lakes which have lateral or vertical water quality differentials in the lake parameters and attributes ofinterest. Totally mixed lakes, if they exist, are certainly few in number, but it might be useful to perform monitoring in selected homogenous portions of stratified lakes to simplify data interpretation. Similarly, for lakes that exhibit significant seasonal mixing, it might be beneficial to monitor during a time period in which they are mixed. For some monitoring objectives, however, it might be best to monitor during periods of peak stratification.

Temporal variability concerns are similar for mixed and stratified lakes. Seasonal changes are often obvious, but should not be assumed to be similar for all lakes or even the same for different parts of any individual lake (Figure IV-10). Due to the importance of factors such as precipitation characteristics, climate, lake basin morphology, and hydraulic retention characteristics, seasonal variability should be at least qualitatively assessed before any lake monitoring program is initiated.

Short-term variability is also an inherent characteristic of most still (lentic) waterbodies. Parameters such as pH, dissolved oxygen, and temperature can vary considerably over the course of a day. Monitoring programs targeted toward biological parameters should be structured to account for this short-term variability. It is often the case that small lakes and reservoirs respond rapidly to runoff events. This factor can be very important in cases where lake water quality will be correlated to land treatment activities or stream water quality.

In stratified lakes spatial variability can be lateral or vertical. The classic stratified lake is one in which there is an epilimnion and a hypolimnion (Figure IV-11). Water quality can vary considerably between the two strata, so sampling depth is an important consideration when monitoring vertically stratified lakes. Temperature and dissolved oxygen profiles of a flood-control reservoir in Kentucky illustrate vertically stratified and mixed conditions for these two parameters (Figure IV-12).

Lateral variability is probably as common as vertical variability, particularly in lakes and ponds receiving inflow of varying quality. Figure IV-13 illustrates the types of factors that contribute to lateral variability in lake water quality. In reservoir systems, storm plumes can cause significant lateral variability.

Davenport and Kelly (1984) explained the lateral variability in chlorophyll *a* concentrations in an Illinois reservoir based on water depth and the time period that phytoplankters spend in the photic zone. Refer to Figure IV-3 for monitoring site locations, and note from Table IV-1 that sites 3 and 9 had significantly higher annual mean chlorophyll *a* concentrations than all other

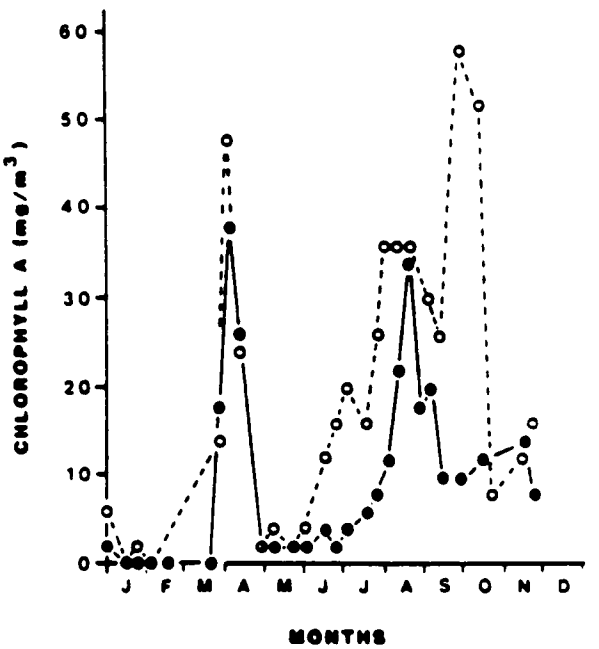


Figure IV-10. Phytoplankton chlorophyll a concentration in Chautauqua Lake's northern basin (●-●) and southern basin (o-o), 1977. (SOURCE: Storch, 1986)

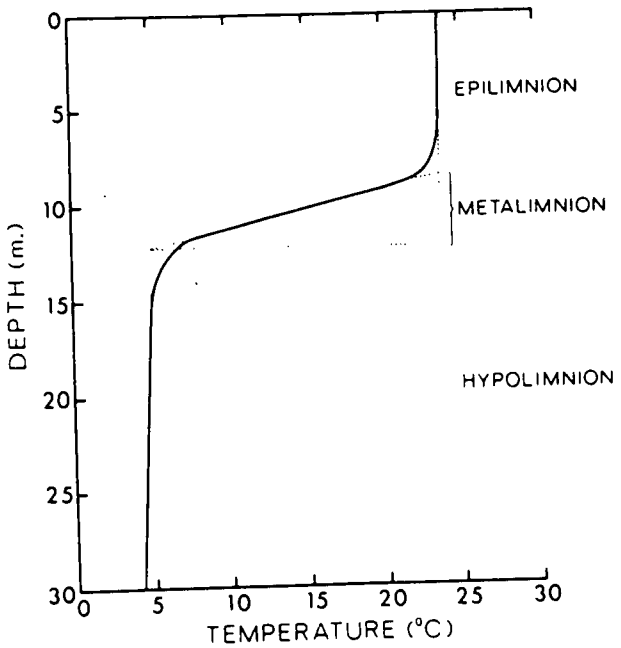


Figure IV-11. Typical thermal stratification of a lake into the epilimnetic, metalimnetic, and hypolimnetic water strata. (SOURCE: Wetzel, 1975)

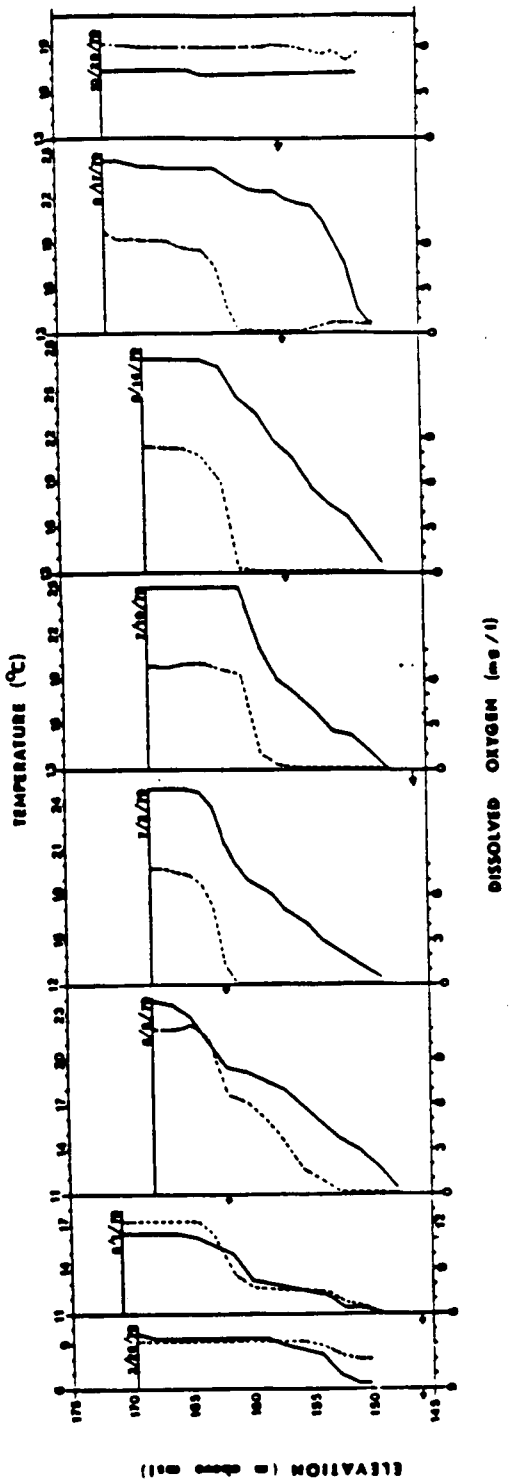


Figure IV-12. Temperature (solid line) and dissolved oxygen (dashed line) profiles for Barren River Lake in 1979. Arrows indicate levels of water release from the reservoir on sample data. (SOURCE: Walburg et al., 1983)

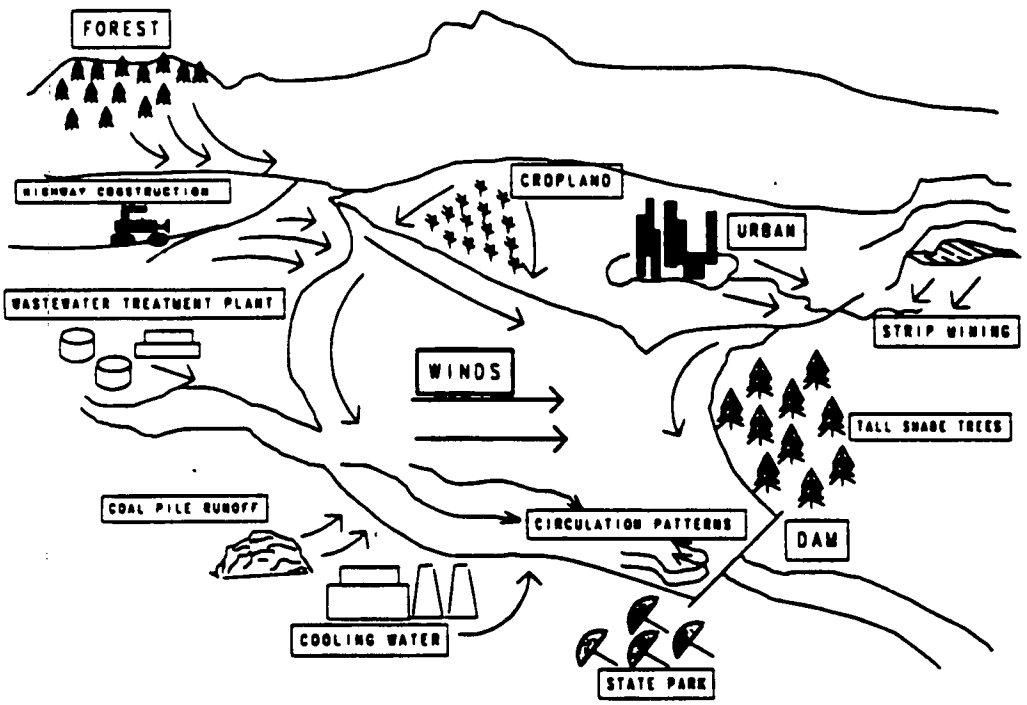


Figure IV-13. Factors contributing to lateral differences in lake quality.

Table IV-1. Comparison of means by site for chlorophyll *a* concentrations (CAC) (ug/l) at Highland Silver Lake, May 1981 March 1984. Duncan's Multiple Range Test was used to compare means for log transformed data. Means with same letter are not significantly different.

Site	Mean (n)	Mean of Log (CAC+1)	Min.	Max.	Std. Dev.	Grouping
3	31.6(20)	1.172	0.00	143.79	39.04	A
9	27.5(17)	1.224	0.00	79.41	23.45	A
10	14.3(21)	0.988	0.65	57.85	16.04	B
7	12.8(20)	0.976	1.25	45.00	12.55	B
6	12.3(21)	0.949	0.94	47.74	12.96	B
2	10.5(20)	0.855	0.00	44.55	11.91	B
5	9.6(21)	0.841	1.10	41.54	10.28	B
4	8.9(21)	0.828	1.25	39.81	10.24	B
1	6.7(21)	0.680	0.00	30.18	8.98	B

(SOURCE: Davenport and Kelly, 1984)

sites in the reservoir. A horizontal gradient of sediment, nutrient, and chlorophyll *a* concentrations in St. Albans Bay, Vermont, was related to mixing between Lake Champlain and the Bay (Clausen, 1985). It is important to note that there is frequently significant lateral and vertical variation in the toxicity and contaminant levels of bed sediments (USEPA, 1987).

Despite the distinction made between mixed and stratified lakes, there is considerable gray area between these groups. For example, thermally stratified lakes might be assumed to be mixed during periods of overturn, and laterally stratified lakes can sometimes be treated as if the different lateral segments are sublakes. In any case, it is important that the monitoring team knows what parcel of water is being sampled when the program is implemented. It would be inappropriate, for example, to assign the attributes of a surface sample to the hypolimnion of a stratified lake due to the differences in temperature and other parameters between the upper and lower waters.

D. Estuaries

Estuaries can be very complex systems, particularly large ones such as the Chesapeake Bay. Like lakes and streams, estuaries exhibit temporal and spatial variability. Physically, the major differences between estuaries and fresh waterbodies are related to the mixing of fresh water with salt water and the influence of

tides. These factors increase the complexity of spatial and temporal variability within an estuary.

Short-term variability in estuaries is related directly to the tidal cycles, which can have an effect on both the mixing of the fresh and saline waters and the position of the freshwater-saltwater interface (USEPA, 1982a). The same considerations made for lakes regarding short-term variability of parameters such as temperature, dissolved oxygen, and pH should also be made for estuaries.

Temperature profiles such as those found in stratified lakes can also change seasonally in estuaries. The resulting circulation dynamics must be considered when developing monitoring programs. The effects of season on the quantity of freshwater runoff to an estuary can be profound. In the Chesapeake Bay, for example, salinity is generally lower in the spring and higher in the fall due to the changes in freshwater runoff from such sources as snowmelt runoff and rainfall (USEPA, 1982a). Figure IV-14, developed from Chesapeake Bay investigations, illustrates the short-term temporal variability of salinity levels in various segments of the Bay.

Spatial variability in estuaries has both significant vertical and lateral components. Typical estuarine circulation exhibits a spatial pattern in which incoming tides push salt water into the estuary while freshwater flow from rivers moves outward. Since fresh water is lighter than salt water, it flows over the salt water and displaces the lower saltwater flow shoreward. On outgoing tides the saltwater mass is pulled outward and the freshwater flow drops vertically and moves outward with the saltwater flow (Cohen, 1990). Both small and large estuaries (e.g., Chesapeake Bay, San Francisco Bay, Puget Sound) exhibit this pattern. Stratification during the summer creates vertical variability during part of the year as well (USEPA, 1982a). Freshwater and saltwater flows might also be diverted to opposite shores of an estuary by a combination of tributary location, the earth's rotation, and barometric pressure (Figure IV-15). In addition, lateral variability in salinity can be caused by different levels of mixing between saltwater and freshwater inputs. For instance, in Puget Sound the numerous sills, inlets, channels, and islands disrupt the natural estuarine circulation. Strong mixing occurs around these features, and as a result the outward freshwater flow—as much as two-thirds of it in places—is mixed with the deeper salt water and flows back into the Sound. Pollutants contained in this water or bound to sediments carried in the fresh water return to the Sound. The San Francisco estuary is complicated in another way. The northern reach of the estuary exhibits fairly typical estuarine flow, but the southern reach receives only about one-tenth the amount of freshwater flow as the northern reach. This quantity is too small to establish estuarine flow or create a salinity gradient. Also, water stays in the southern reach up to four times as long as in the northern reach.

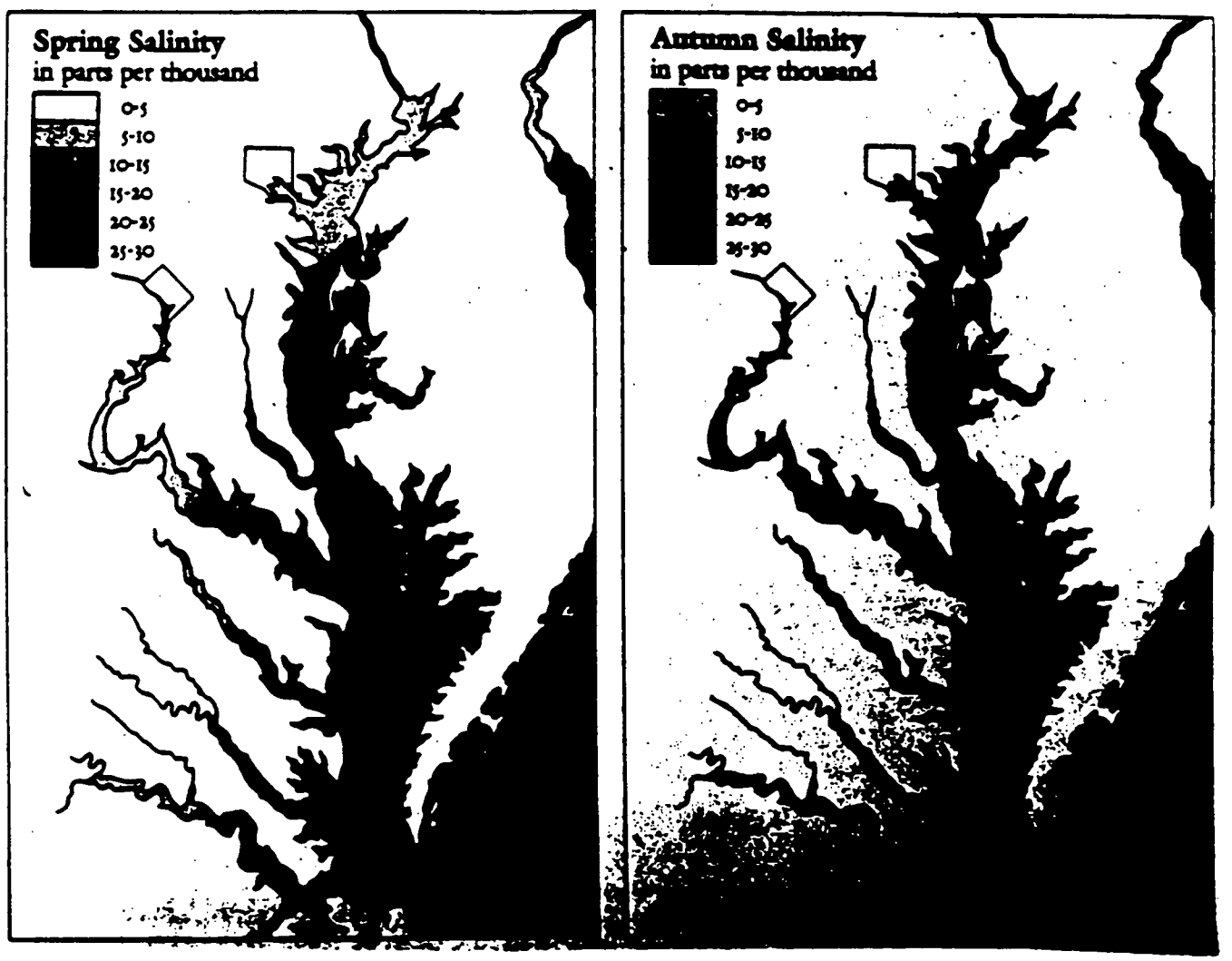


Figure IV-14. Chesapeake Bay salinity levels over time and space. (SOURCE: USEPA, 1982a)

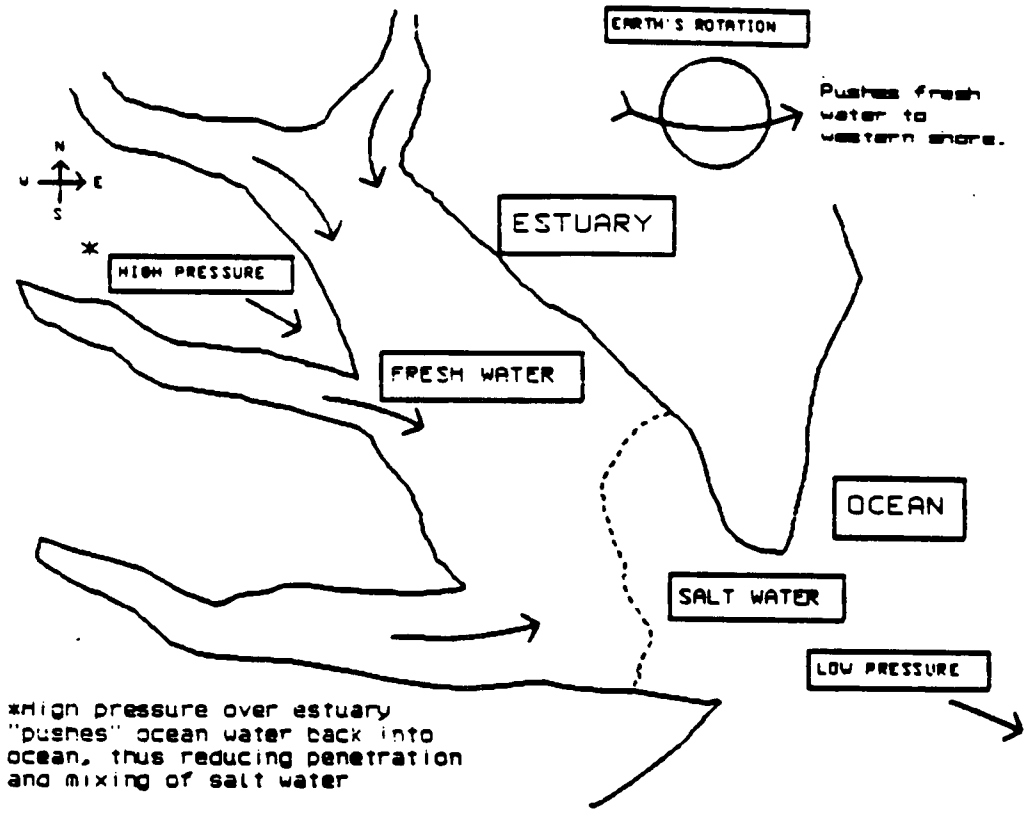


Figure IV-15. Physical features affecting lateral salinity gradients in an estuary.

This has consequences for water quality because the southern reach receives three times as much sewage effluent as the northern reach.

As noted for streams and lakes, the lateral and vertical variation in the toxicity and contaminant levels of bed sediments should be considered (USEPA, 1987).

E. Coastal Waters

Researchers and government agencies are collectively devoid of significant experience in evaluating the effectiveness of nonpoint source pollution control efforts through the monitoring of near-shore and off-shore coastal waters. Our understanding of the factors to consider when performing such monitoring is therefore very limited.

As for other waterbody types, it is important to understand the hydrology, chemistry, and biology of the coastal system in order to develop an effective monitoring program. Of particular importance is the ability to identify discrete populations from which to sample. For trend analysis it is essential that the researcher be able to track over time the conditions of a clearly identifiable segment or unit of coastal water. This may be accomplished by monitoring a semienclosed near-shore embayment or similar system. Knowledge of salinity and circulation patterns should be useful in identifying such areas. Secondly, monitoring should be focused on those segments or units of coastal water for which there is a reasonable likelihood that changes in water quality will result from the implementation of management measures. Segment size, circulation patterns, and freshwater inflows should be considered when estimating the chances for such water quality improvements.

Near-shore coastal waters can exhibit salinity gradients similar to those of estuaries due to the mixing of fresh water with salt water. Currents and circulation patterns can create temperature gradients as well. Farther from shore, salinity gradients are less likely, but gradients in temperature can occur. In addition, vertical gradients in temperature and light might be significant. These and other biological, chemical, and physical factors should be considered in the development of monitoring programs for coastal waters.

F. Ground Water

Aquifers are probably the least frequently monitored water resources because of both the expense associated with drilling wells and the "out of sight out of mind" approach to problem finding. Today, however, ground water monitoring has become very important because contamination of ground water drinking supplies has been discovered in many areas across the United States. For example, in the Big Spring area of Iowa, ground water is contaminated with high nitrate levels and pesticides such as

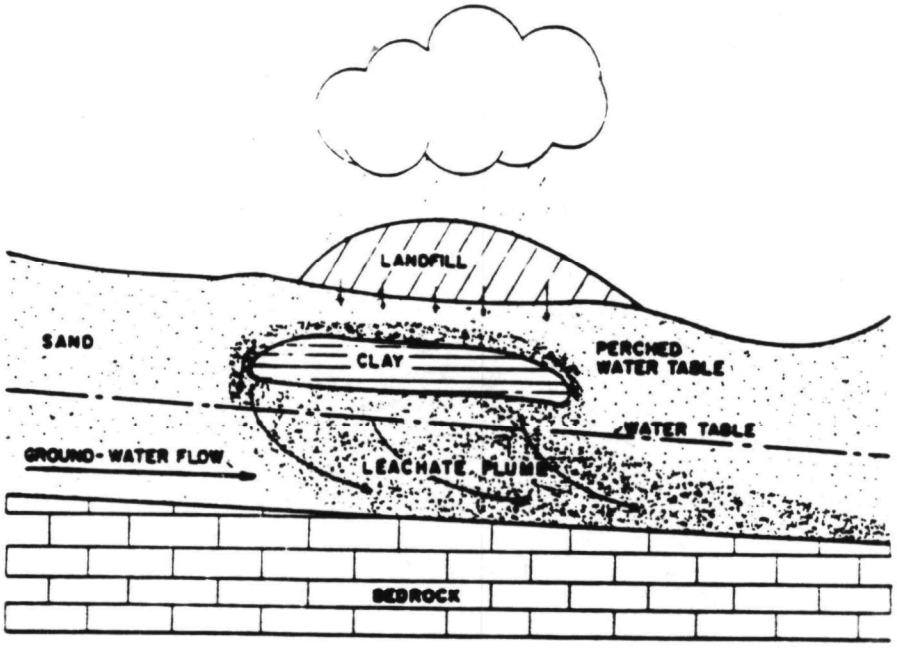
atrazine, alachlor, cyanazine, and metolachlor (Hallberg et al., 1984). In a nation where ground water withdrawals account for 35 percent of all public water supply use (more than 50 percent in 11 states), it is increasingly important to monitor the quality of this resource (USGS, 1985, p. 120).

Ground water monitoring for NPS pollution studies serves nunerous purposes, such as to determine the ground water component of a hydrologic/chemical budget for a surface waterbody, to document the impact of a pollution activity, to identify background water quality, to identify trends and variations in water quality, or to determine the effectiveness of land use or best management practices.

Spatial and temporal variability in aquifers cannot be generalized. Some aquifers respond to precipitation events very quickly while others respond quite slowly. If the response rate of an aquifer to outside influences, such as fertilizer applications on agricultural fields, is not known, sampling at least weekly is recommended. Otherwise, significant fluctuations in ground water concentrations might be missed (Clausen et al., 1993).

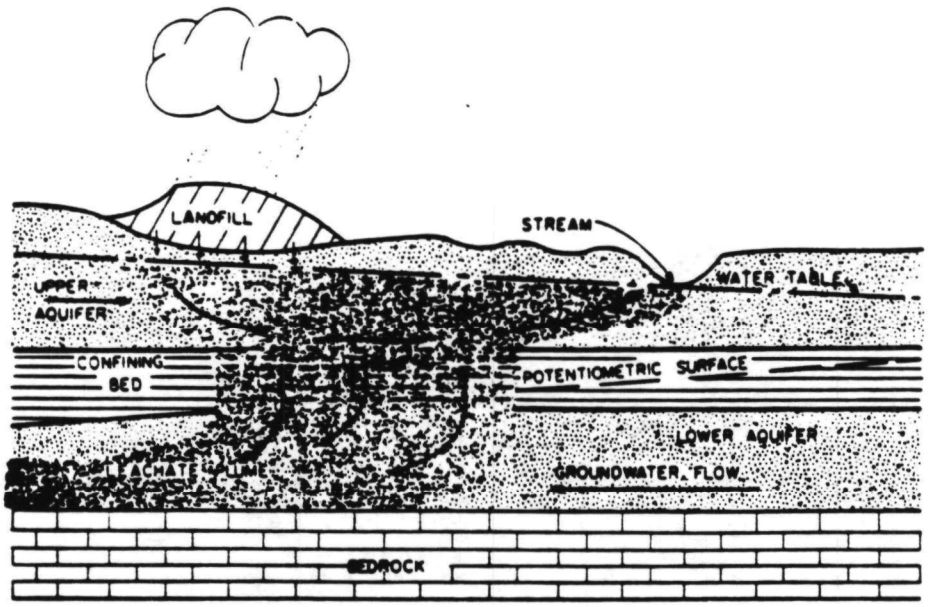
For the purposes of this brief discussion, aquifers are divided into two major types: (1) those with intergranular porosity such as sands, gravels, sandstones, and silts, and (2) those with solution porosity or fractured rock (Scalf et al., 1981). Both types of aquifers exhibit short- and long-term temporal variability and vertical and lateral spatial variability. The differences between the two types of aquifers largely relate to flow patterns. In general, aquifers with intergranular porosity have more predictable flow patterns than do fractured rock aquifers (Scalf et al., 1981).

The selection of well locations is highly dependent on the spatial variability of aquifer water quality. Factors that can complicate well monitoring efforts include the presence of confining beds (Figure IV-16), the possibility of multiple aquifer systems (Figure IV-17), the effects of pollutant density on pollutant transport, and changes in permeability (Figure IV-18). Because of these complicating factors, hydrologic information specific to the site to be monitored is needed to design a ground water monitoring program. Only site-specific information can create an accurate picture of the surrounding geology, which is essential to be certain that sampling locations are properly located.



The water table is deep. Leachate percolates downward under the landfill, forming a perched water table before finally reaching the actual water table.

Figure IV-16. Permeable sand layer underlain by a clay layer. (SOURCE: Scalf et al., 1981)



Leachate first moves into and flows with the ground water in the upper aquifer. Some of the leachate eventually moves through the confining bed into the lower aquifer where if flows back beneath the landfill and away in the other direction.

Figure IV-17. Two-aquifer system with opposite flow directions. (SOURCE: Scalf et al., 1981)

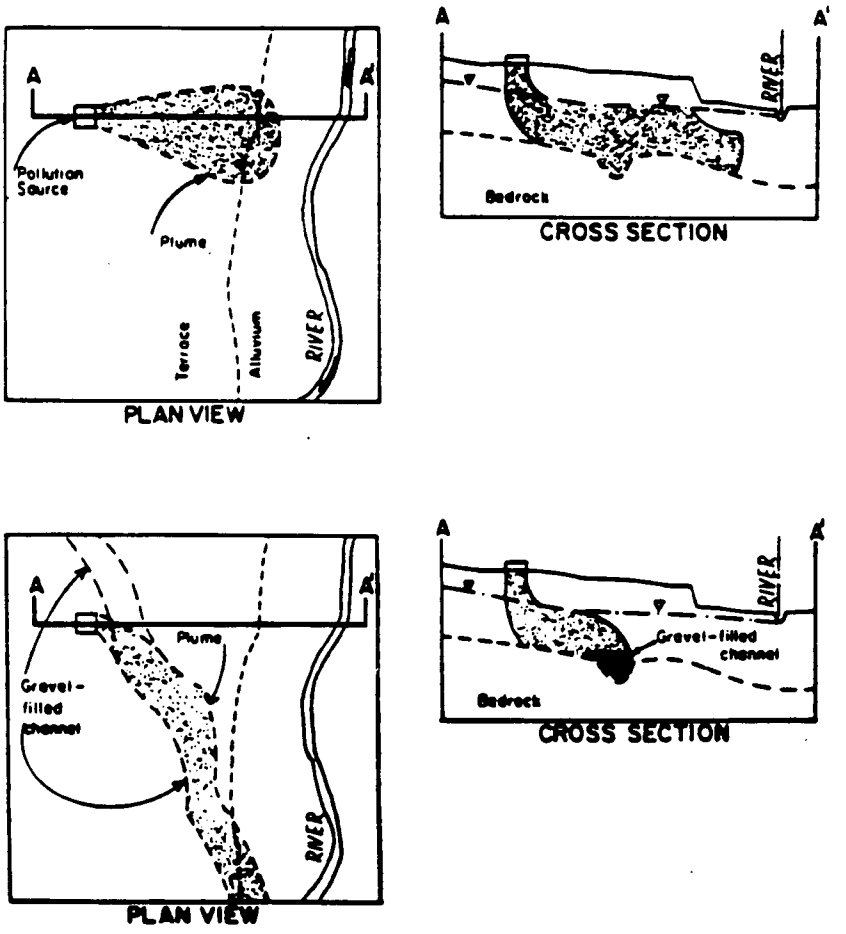


Figure IV-18. Effect of permeability change on shape of pollution plume. (SOURCE: Scalf et al., 1981)

CHAPTER 5

DATA NEEDS AND DATA SOURCES

V. DATA NEEDS AND DATA SOURCES

A. Data Needs

Data needs are a direct function of monitoring goals and objectives. Thus, data needs cannot be established until specific goals and objectives are defined. Furthermore, data analyses should be planned before data types and data collection protocols are agreed upon. In short, the experimental approach should be followed in establishing all monitoring programs. As defined by Random House, an experiment is "a test, trial, or tentative procedure; and act or operation for the purpose of discovering something unknown or of testing a principle, supposition, etc." (Stein, 1980). Rarely will a monitoring program have a purpose that does not fall within the above definition of an experiment.

The following discussion touches upon the types of information that should be considered in nonpoint source monitoring programs. It is impossible to recommend in advance a complete suite of parameters to be monitored for all possible program objectives. The monitoring group should select parameters based upon their monitoring objectives and their particular situation.

B. Data Types

Many types of data may be required for nonpoint source monitoring and evaluation depending upon the program and the monitoring and evaluation goals and objectives. Adequate water quality data are obviously required, but the form and scope of needed water quality data should be determined as a function of quantitative monitoring and evaluation objectives and planned data analysis.

In some cases, it may be more beneficial to use surrogate measures instead of the parameters mentioned in the monitoring and evaluation goals and objectives. In these cases, objectives for the surrogates should be established which are consistent with the overall monitoring and evaluation goals. For example, a program goal may be to improve the trophic state of a lake from eutrophic to mesotrophic, with a monitoring and evaluation goal to document this change. Surrogate measures for trophic state may be phosphorus and chlorophyll-*a*. The corresponding surrogate program goal may be to reduce phosphorus and chlorophyll-*a* concentrations to levels deemed sufficient to change the trophic state to mesotrophic. The monitoring and evaluation goal would then be to document these changes in phosphorus and chlorophyll-*a* levels.

The key to using surrogate measures is that a reliable relationship must exist between the true measure and the surrogate measure. An example of poor surrogate selection is the unqualified use of erosion rates as estimators of sediment delivery to water resources. Without the existence of a reliable relationship between these two measures (i.e., sediment delivery ratio), the surrogate will produce misleading results.

Sherwani and Moreau (1975, p. 34) illustrate how interparameter correlations can be used to optimize parameter selection. In essence, one parameter value is used to estimate the value for a correlated parameter. The authors noted that higher order streams showed better interparameter correlations than lower order streams.

1. Hydrologic Data

Hydrologic data are critical to many nonpoint source monitoring and evaluation efforts. Due to the relationships between flow parameters and pollutant characteristics, hydrologic data are almost always required in some form at some state of nonpoint source monitoring and evaluation. It may be necessary to establish watershed water budgets in order to determine the location and magnitude of nonpoint source or background sources. In other cases, the extent of the floodplain may prove critical to assessments of BMP control needs. For example, in the New York MIP project it was discovered that flooded cropland was a much lager source of phosphorus to Cannonsville Reservoir than was expected (Brown, et al. 1986). As a result, nonpoint source control efforts would have to be redirected to cover manure-spread cropland as well as barnyard runoff which was the original focus of BMP implementation.

2. Biological Data

Biological data are very useful in evaluating impairment to the water resource from nonpoint source impacts, because aquatic communities integrate the exposure to various nonpoint sources over time. An assessment of biological condition is crucial to the judgment of a waterbody to maintain and protect a healthy aquatic ecosystem. Monitoring of changes in the aquatic community over time provides an understanding of the improvement due to BMPs and allows a judgment of the effectiveness of various mitigative measures. Biological survey approaches will differ depending upon the waterbody—i.e., streams, rivers, lakes, wetlands, or estuaries.

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 NURP, MIP, and RCWP studies. The Universal Soil Loss Equation (USLE) is based on research showing that there is a direct linear relationship between average annual soil loss and a rainfall runoff factor (R) (Wischmeier and Smith, 1978). The R factor accounts for the cumulative energies and intensities of a year's storms, plus the erosive forces of runoff from thaw, snowmelt, and irrigation. The erosion index found in the USLE is also used in several nonpoint source models, including the Agricultural Nonpoint Source Pollution Model (AGNPS) (Young, et al., 1985). A procedure derived from the NURP program uses storm frequency and other factors to determine recurrence intervals for in-stream pollutant

concentrations resulting from urban nonpoint source pollution (USEPA, 1984a).

4. Land Use Data

Land use and land treatment data are important in monitoring and evaluation efforts designed to test the effectiveness of implementation programs. In RCWP, for example, it has been discovered that traditional USDA reporting schemes may not provide sufficient land treatment data for use in analyses relating BMP implementation to water quality changes. New approaches are needed to report land treatment and land use in ways that are quantitative and transferable. This problem was addressed and recommendations were made in an RCWP workshop, but the issue was not resolved.

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).

6. Soil Characteristics Data

Other data such as soil chemistry and soil physical characteristics may 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 (Dean, et al, 1984; Aller, et al. 1985).

C. Data Sources

Many of the data requirements for nonpoint source monitoring and evaluation efforts can be met using nationally available data sources. This section describes some of these data sources and includes contacts for those interested in accessing the data. There are likely very many other data sources available to nonpoint source professionals—this guidance focuses only on some of the major data sources made available to EPA or known to reviewers of this document.

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 is the main component of STORET and 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: 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 System: DFS contains daily observations of stream flow and miscellaneous water quality parameters collected at gaging stations belonging to the 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 data bases.

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 submit information to STORET, including federal, state, interstate, and international agencies. 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 has been checked for invalid data ranges or missing mandatory fields before being 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: Bob King
Office of Wetland, Oceans, and Watersheds
U.S. Environmental Protection Agency
The Fairchild Building
499 South Capitol Street, SW, 4503F
Washington, DC 20460

Phone: (202) 260-7028

2. Water Quality Analysis System (WQAS) USEPA/OWOW

The Water Quality Analysis Branch has developed several data bases and procedures that operate under the umbrella of the Water Quality Analysis System (WQAS). WQAS data bases 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 data bases; the following data bases are a sample of these that 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: Bob King
Office of Wetland, Oceans, and Watersheds
U.S. Environmental Protection Agency
The Fairchild Building
499 South Capitol Street, SW, 4503F
Washington, DC 20460
Phone: (202) 260-7028

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): This system 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 also 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, locations, 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 may 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 plus expanded capability for processing and managing additional chemical constituent, sediment, biological, and spatial data.

Contact: Tom Yorke, Chief
National Water Information System
U.S. Geological Survey
437 National Center
Reston, VA 22092
703-648-5659

4. National Stream Quality Accounting Network (NASQAN)/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 the stations which are generally located on major rivers at the downstream end of the accounting unit.

Physical constituents monitored in freshwater and analyzed for trends are pH, alkalinity, sulfate, nitrate, phosphorus, calcium, magnesium, sodium, potassium, chloride, suspended sediment, fecal coliform bacteria, fecal streptococcal bacteria, dissolved oxygen, and dissolved oxygen deficit. The following radionuclides are also monitored at 46 sites, but have not been analyzed for trends: gross alpha, gross beta, radium-226, and uranium. Data collection stations are maintained at selected locations to provide standardized records on surface and ground water conditions. A variety of automated instruments are used to measure and record

water conditions. Data are collected at 58 percent of sites and quarterly at 42 percent of sites. Trace element collection is quarterly and radionuclides are collected semiannually. Yearly data summaries are available for each state.

Contact: Richard A. Smith, Hydrologist
Water Resources Division
U.S. Geological Survey
410 National Center
Reston, VA 22092
Phone: 703-648-6870

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 know 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 may, 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 low acid-neutralizing capacity waters.

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 with 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: Dixon Landers
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 USEPA/OWRS

The National Coastal Pollutant Discharge Inventory (NCPDI) Program is a series of data base development and analytical

activities within the Strategic Assessment Program for coastal and estuarine areas of NOAA. The cornerstone of the program is a comprehensive data base and computational framework that has been developed over the last nine years. The data base 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 the following base years for each coastal component:

- East Coast - 1982,
- West Coast - 1984, and
- Gulf Coast - 1987.

The estimates can be considered to approximate pollutant discharge conditions for a five-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 spill, and
- 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), and
- 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 1987for the coastal areas of the Gulf of Mexico and for 1989 for the East Coast are being prepared.

Contact: Daniel R. Farrow, Chief
Pollutant Sources Characterization Branch
National Oceanic and Atmospheric Administration
6001 Executive Boulevard, Room 220

Rockville, M 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: Bob King
Office of Wetland, Oceans, and Watersheds
U.S. Environmental Protection Agency
The Fairchild Building
499 South Capitol Street, SW, 4503F
Washington, DC 20460
Phone: (202) 260-7028

8. Waterbody System (WBS) USEPA/OWOW

The Waterbody System (WBS) is an automated data base 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 water body. It may 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 loads 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 sued as the basis for the 305(b) report to Congress every two 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.

WBS is a voluntary program currently used by approximately 40 states, territories, and river basin commissions. the data base consists of assessments rather than monitoring data and includes many optional fields. Consistency is good within a state. Those wishing to aggregate to a regional or national level should discuss data characteristics with the WBS coordinator.

Contact: Jack Clifford
Office of Wetland, Oceans, and Watersheds
U.S. Environmental Protection Agency
The Fairchild Building
499 South Capitol Street, SW, 4503F
Washington, DC 20460
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 utilized. PIN also provides federal, state, and local agencies with a means to 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: Constance A. Hoheisel
Office of Pesticide Programs
U.S. Environmental Protection Agency
401 M Street, SW, H7507C
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 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 may threaten fish and wildlife. NCBP data is maintained by the ECDMS. The NCBP is the USFWS segment of the National Pesticide Monitoring Program, a multi-agency 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 Untied States are collected to assess body burden of organochlorine compounds in migratory birds. This monitoring program has continued at two- to four-year intervals since 1965. The USFWS is in the process of reviewing the NCBP and an agency initiative was approved for fiscal year 1992.

Contact: James K. Andreasen
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,, conditional approved, restricted);
- size; and
- pollution sources (identified for all non-approved 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 that were reclassified as a result or 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; and
- other factors that may influence classification.

Data are collected by questionnaire and followed by interviews. Classifications are noted on 265 nautical charts. Data were compiled in 1966, 1971, 1974, 1980, 1985, and 1990. The next survey is scheduled for 1995.

Contact: Sharon Adamany, Environmental Analyst
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 trends data are available from the Mussel Watch and Benthic Surveillance for four major elements, twelve 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 fifty Benthic Surveillance sites and since 1986 at about 150 Mussel Watch sites.

Sediment samples are collected at all sites. At Benthic Surveillance sites, benthic fishes 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: Thomas P. O'Connor
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 comprised 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 their data on CD-ROM's so that users could look at a large amount of climatic data at one time. The NCDC plans to archive new data sets using the latest technical advances available. Observations are taken at varying intervals, from every fifteen minutes to once per month. Collections are daily or monthly depending on type and source of information.

Contact: National Climatic Data Center
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), 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: Dennis Athayde
SYNOP
U.S. Environmental Protection Agency
401 M Street, SW, H7507C
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, United States 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 six 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, and
- chloride, magnesium, sodium, and potassium ions.

The NADP/NTN monitoring program has developed criteria and protocols which 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; and
- ion chromatographic detection of sulfate, nitrate, and chloride.

Samples are collected weekly. Data from some sites are available from 1979. The data is maintained on the Acid Deposition System (ADS).

Contact: Ranard J. Pickering
U.S. Geological Survey
416 National Center
Reston, VA 22092
Phone: 703-648-6875

16. Major Land Uses in the United States (MLU) USDA/ERS

For more than fifty years, the 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; and
- unclassified use.

Each major class is further classified by specified uses and some 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; and
- miscellaneous agricultural uses (farmsteads, farm roads, and farm lands).

Special uses include:

- forest land not grazed;
- intensive uses (highways and roads, railroads, and airports); and

extensive uses (national parks, state parks, wilderness areas, federal wildlife areas, state wildlife areas, national defense areas, and federal industrial facilities).

Unclassified other land uses include urban and other 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 five 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: Arthur B. Daugherty, Agricultural Economist
Economic Research Service
U.S. Department of Agriculture
1301 New York Avenue, N.W., Room 408
Washington, DC 20005-4788
Phone: 202-219-0424

17. Agricultural Census (AgCensus) USDOC/BOC

The Agricultural Census (AgCensus) data base 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 ten thousand dollars.

AgCensus does ont 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 data base.

Contact: Charles P. Pautier, Jr., Chief
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, and
- 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 three to five years.

Contact: Robert M. Williamson, Director
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 and, by the 1960s, inventories were completed for all of the 48 conterminous states and many of the important forested states had been re-inventoried. 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, and
- 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 are estimated. Estimates also are made of areas harvested or otherwise disturbed, regenerated to forest, or cleared for other use. Additional estimates of recreation use, wildlife values, site productivity, physiographic characteristics, and other items are made. The data is maintained in three possible data bases:

- 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 one acre. Depending upon 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 fifty 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 fifty million acres are inventoried in the conterminous Untied States. Currently this rate of coverage translates into an inventory cycle of twelve years for the nation.

Contact: James T. Bones
U.S. Forest Service
U.S. Department of Agriculture
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, and
- 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: Richard L. Kleckner
Office of Geographic and Cartographic Research
U.S. Geological Survey
590 National Center
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 pervious 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: Philip N. Fulton, Assistant Division Chief for Census Programs
Population Division
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233
Phone: 301-763-7890

22. National Resources Inventory (NRI) USDA/SCS

For fifty years, 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 of 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); and
- potential for conversion to cropland.

The NRI is a multi-resource 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 five-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 has a high degree of reliability at the state level and the 1982 NRI provides a high degree of reliability at the multi-county level. Data estimates can be made by Major Land Resource Areas (MLRA), SCS Administrative Areas, Water Resources Council Aggregated Subareas, and other multi-county geographic subdivisions.

Contact: Jeff Goebel
Resources Inventory and Geographic Information Systems Division
Soil Conservation Service
U.S. Department of Agriculture
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 may 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 ARC-INFO format for use by individual users.

Contact: David Larsen
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 (USFWS) established the National Wetlands Inventory (NWI) to develop technically sound and comprehensive information 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 with updates on a ten-year cycle. 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: Thomas E. Dahl
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

CHAPTER 6

MONITORING RECOMMENDATIONS

VI. MONITORING RECOMMENDATIONS

The following, taken from Sherwani and Moreau (1975, p. 3), describes the need for water quality data:

All phases of planning, development, and operation of a water resource management system, including water quality control, require the acquisition and processing of data to quantify existing states of the system, to forecast future changes and trends, and to predict responses of the system to interventions for development and control. The kind of data required and their spatial and temporal resolution depend upon [the] nature of the problems and [the] kinds of interventions under consideration. The word monitoring has been defined as this activity of making "systematic observations of parameters related to a specific problem, designed to provide information on the characteristics of the problem and their changes with time."

Before getting into specific monitoring guidance it is important to note that these recommendations are just that, recommendations. As stated by Sherwani and Moreau (1975, p. ix), "there is no single specification of a [monitoring] network which is optimal for all purposes." The researcher should tailor these general guidelines to adjust for the specific objectives, water resource characteristics, climate, financial resources, and other factors unique to the area being monitored.

A. Samples and Sampling

1. Samples

Random House (Stein, 1980, p. 1165) defines a sample as "a small part of anything or one of a number, intended to show the quality, style, or nature of the whole." We collect environmental samples for both economic and practical reasons; that is, we cannot afford to inspect the "whole" and we usually have neither the time and resources nor the capability to even try to inspect the "whole." Besides, we often find that a sample or collection of samples will provide sufficient information about the "whole" to allow us to make decisions regarding actions which should or should not be taken.

In a statistical sampling program, the "whole" is called the "population" or "target population," and it consists of the set of "population units" about which inferences will be made (Gilbert, 1987, p. 7). As an example, population units could be defined as macroinvertebrate populations on square-meter sections of river bottom, phosphorus concentrations in 1-liter grab samples, or hourly-mean flow values at a specific gaging station. Gilbert (1987, p. 7) refers to the "sampled population" as the set of population units directly available for measurement.

2. Sampling Objectives

Gaugush (1986, p. 189) states that "the major objective in sampling program design is to obtain as accurate or unbiased an estimate as possible, and at the same time reduce or explain as much of the variability as possible in order to improve the precision of the estimates." Accuracy is a measure of how close the sample value is to the true population value, whereas precision refers to the repeatability of sample values. Biased estimates are those which are consistently higher or lower than the true population value. According to Cochran (1977, p. 11), an estimator is unbiased if its mean value, taken over all possible samples, is equal to the population statistic which it estimates.

In the real world it is necessary to design sampling programs which meet accuracy and precision requirements, while not placing unreasonable burdens on sampling personnel or sampling budgets. As stated by Gaugush (1986, p. 190), budget constraints may force the issue of whether sampling results will produce information sufficient to meet the study objectives.

Gaugush (1986, p. 190-194) describes in some detail specific points to consider in defining study objectives. He notes that "sampling is facilitated by specifying the narrowest possible set of objectives which will provide the desired information." First, he recommends that the target population be defined as a key step in limiting the variability encountered in the sampling program. 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. This specificity focuses the sampling program better, eliminating the need to monitor at upstream and in-lake sites, and during baseflow conditions. The definition also specifies the water quality parameter of interest. Note that both spatial and temporal limits should be established in defining the target population. As a further refinement, the researcher may define the population units as the dissolved phosphorus concentrations in half-hour, composite samples taken during all storms ("storms" as defined by the researcher).

The next step, according to Gaugush (1986, p. 192), is to decide whether parameter estimation or hypothesis testing is the primary analytic goal. This choice will have an impact on the sampling design. As an example, Gaugush points out that balanced designs are desirable for hypothesis testing, whereas parameter estimation may require unbalanced sample allocations to account for the spatial variability of variable levels. Hypothesis testing is likely to be used in 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.

Finally, Gaugush (1986, p. 193) recommends that exogenous variables and sampling strata be defined. Exogenous variables are used to explain some of the variability in the measured parameter

of interest. As an example, total suspended solids (TSS) concentration is typically a covariate of total phosphorus (TP) concentration in watersheds impacted by agricultural runoff. Measurement of TSS may help increase the precision of TP estimates.

3. Sampling Error and Variability

Gilbert (1987, p. 10-13) discusses five general sources of variability and error in environmental studies:

- Environmental variability
- Bias, precision, and accuracy
- Statistical bias
- Random sampling errors
- Gross errors and mistakes

The author describes environmental variability as "the variation in true pollution levels from one population unit to the next." This can be 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; and temporal and spatial variation in background levels. Gilbert (1987, p. 11) notes that existing information on environmental variability can be used to "design a plan that will estimate population parameters with greater accuracy and less cost than can otherwise be achieved."

Accuracy, precision, and measurement bias are described above (Section III.C.1.b.). Gilbert (1987, p. 11) uses an illustration to point out that high accuracy can only be obtained when measurement bias is low and precision is high. Cochran (1977, p. 12-15) provides a simple, yet convincing, demonstration of the impact of bias on the probability of estimation errors.

Random sampling errors (e.g., variability in sample means for different random samples from the same population) arise from the environmental variability of population units (Gilbert, 1987, p. 12). By definition, random sampling error is zero if all (N) population units are measured, but will never be zero if N is infinitely large (Gilbert, 1987, p. 12).

Gross mistakes can occur at any point in the process beginning with sample collection and ending with the reporting of study results (Gilbert, 1987, p. 12). Adherence to accepted sampling and laboratory protocol, combined with thorough quality control and data screening procedures, will minimize the chances for gross errors.

B. Monitoring Approaches

1. General Approaches

Reviewers of three general types of monitoring designs for documenting water quality changes related to best management

practice (BMP) implementation have concluded that (Spooner et al., 1985):

- Monitoring above and below an implementation site is generally more useful for documenting the severity of a nonpoint source than for documenting BMP effectiveness.

- Time trends may be helpful, but variables associated with land treatment, hydrology, and meteorology should be accounted for to increase the likelihood of successful documentation of water quality-BMP relationships.

- Paired watershed designs show the greatest potential for documenting improvements from BMP implementation due to the ability to control for meteorologic and hydrologic variability.

The paired watershed approach is currently being used in the St. Albans Bay, Vermont RCWP project (Clausen, 1985). In this study, two small watersheds received proper manure management during a two-year calibration period, followed by a period in which one watershed received winter-spread manure. This is an interesting approach in that BMPs were removed from instead of applied to a watershed after the calibration period. The anticipated analytic result in using the paired watershed design is a demonstrated water quality change attributable to management differences because all other key factors are very similar.

Data from the Rock Creek, Idaho RCWP were collected in upstream-downstream pairs. These paired data were utilized in regressions of water quality against time. The downstream concentrations (below NPS) were adjusted for upstream concentrations (above NPS), transformed, and then regressed against time as a continuous variable (Spooner et al., 1986). Results of this approach were used to indicate decreasing pollutant contributions from NPS. This example illustrates uses of both upstream-downstream monitoring and time trends.

Erlebach (1979, p. 7) has described a systematic approach to monitoring surface water quality trends associated with urban, industrial, and agricultural developments in a river basin. The steps in this approach are:

- identify developments which may impair or change water quality

- assign priorities to areas of concern and select study areas

- examine background data and literature to assess gaps in required information and develop a conceptual model

- perform a basin reconnaissance to confirm background assessment, to determine logistic requirements, and to select sites and parameters

- propose and implement coordinated pilot studies to develop a monitoring approach, to determine spatial, short term, and seasonal variations in relevant parameters of water quality and quantity and to investigate relationships between parameters

- design and implement a trend monitoring program to investigate future needs for remedial action or to show that remedies are effective.

Richards (1986) has used Monte Carlo models to determine the best of seven sampling strategies for estimating tributary loads. Applying the bias and precision statistics from the Monte Carlo simulation to datasets covering five parameters for three rivers, the researcher found that bias and precision of loading estimates are affected by the frequency and pattern of sampling, the calculation approach used, watershed size, and the parameter being monitored. Richards concludes that flow-stratified sampling and calculations using the Beale Ratio Estimator provided the most precise results in the study. A further point made is that initial sampling should be intense such that sampling theory calculations can then be used to adjust the program.

Meals (1987, personal communication) notes that flow-proportional sampling is not possible unless the stage-discharge relationship is known. His experiences in nonpoint source monitoring of agricultural watersheds in Vermont has led him to conclude that researchers should use time-based sampling, even after a stage-discharge relationship is established. In addition, Meals recommends a greater emphasis on runoff events versus baseflow conditions, particularly where budget limitations are severe.

2. Parameters

In these days of increasing M&E needs and relatively small M&E budgets it is extremely important for researchers to design efficient M&E programs. The menu of parameters selected for a monitoring program should be tied directly to the monitoring objectives. It is often the case that parameters in addition to those of prime interest are monitored because these other parameters are cheap to monitor and may provide some useful information for purposes not yet outlined. This is generally reasonable, but the researcher should:

- Anticipate these undefined purposes such that the extra parameters are monitored in a manner that truly does yield useful information (e.g., support statistical analyses).

- Make sure that the extra cost associated with additional parameters does not preclude necessary expansions and/or extensions of the monitoring and evaluation program for the parameters of prime interest.

Sherwani and Moreau (1975, p. 24) indicate that parameter selection can be governed by (1) water use requirements, (2) water quality standards, (3) type of pollutant source, and (4) the other constituents normally present in the waterbody. For enforcement purposes, the parameters measured should be those specified in water quality standards. For use support analyses, additional parameters may be required since a broader definition of water quality is applied. Both Sherwani and Moreau (1975, p. 27) and an EPA Interagency Task Force (USEPA, 1981) recommend key parameters to be measured for various designated uses.

Parameter selection should reflect the NPS M&E objectives and data analysis plans. For example, if the objective is to monitor the condition of salmon spawning areas, then parameters such as streambank undercut, embeddedness, and vegetation overhang may be important (Platts et al., 1983). If the goal is to assess the impact of NPS controls in terms of standards violations, then the parameters selected should be only those required for the analysis of standards violations. If the M&E plan involves data normalization or grouping prior to data analysis then the parameter list should include those parameters used in normalization and/or grouping. Some analyses may require discrete variables, whereas others may use continuous variables; these issues should be addressed in choosing parameters to monitor.

Both surface and ground water monitoring require that sampling sites are characterized sufficiently for meaningful data interpretation. 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, p. 2-1). 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, lake depth, etc.

Water level measurements should be included in all ground water studies. These measurements are used, among other things, to: (1) indicate the directions of ground water flow and areas of recharge and discharge, (2) evaluate the effects of manmade and natural stresses on the ground water system, (3) define the hydraulic characteristics of aquifers, and (4) evaluate stream-aquifer relations (USGS, 1977, p. 2-1).

The reader is referred to Section III.B. for additional discussion of data types or parameters.

Sherwani and Moreau (1975, p. 34) illustrate how inter-parameter correlations can be used to optimize parameter selection. In essence, one parameter value is used to estimate the value for a correlated parameter. The authors noted that higher order streams in North Carolina showed better inter-parameter correlations than lower order streams. In addition to inter-parameter correlations, the authors recommend principal components and factor analysis as a means for identifying the most important

parameters and clusters of related parameters (Sherwani and Moreau, 1975, p. 37).

3. Sampling Methodology

In the context of this NPS M&E Guidance, the purpose of sampling is to obtain information sufficient to describe selected characteristics of a system over pre-defined temporal and spatial limits. The scope and quality required of these descriptions are a direct function of the M&E objectives. For example, objectives which incorporate hypothesis testing will usually require more extensive data bases than objectives which entail only general assessments of water quality. Hence, the sampling methodologies required to meet different M&E objectives for the same waterbody, and especially for different waterbodies, may differ considerably. This may seem to be a statement of the obvious, but all too often sampling programs are established with insufficient consideration of the match between data needs and the data to be collected.

As stated earlier, the target population is the set of population units about which inferences are to be made. Following the steps outlined by Gaugush (1986, p. 190-194), researchers should define the target population and population units before establishing the sampling program. While this step in the process ensures that the "right stuff" is being sampled, it does not address other sampling issues such as "when," "where," and "how much" to sample. Researchers should design their sampling programs to develop data bases describing the "right stuff" in such a way that statistical and other planned analyses can be performed in support of the M&E objectives. That is, the sampling program should provide "representative" and "sufficient" data.

The question of whether the data are sufficient is addressed in subsequent sections of this guidance. The representativeness of collected data is discussed in considerable detail in section III.A., but warrants some additional comments as follow.

Site location and sampling frequency are often considered sufficient to describe the "where" and "when" of sampling programs. While this is certainly true to a large extent, these two factors, taken alone, do not describe fully where and when samples are collected. Additional considerations include the depth of sampling and the surface or ground water stratum in which the sampling depth belongs, the origins of the aliquots(s) taken in each sample bottle, the time-frame over which measurements are made, etc. These additional considerations fall under the heading of "sample type" in this guidance. Site location and sampling frequency are components of "sampling design," and sample type and sampling design are taken in combination as the overall "sampling methodology."

In order for the data analyst to interpret sampling results appropriately, the sample type, sampling design, and target population must all be identified explicitly. It should be clear from the descriptions of the sampling methodology and target

population whether the data collected are representative of the target population.

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 (USEPA, 1981; Scalf et al., 1981; Brakensiek et al., 1979; USGS, 1977; USEPA, 1978b; USEPA, 1987a; Platts et al., 1983; and Shelley, 1979) and other reference materials (Wetzel and Likens, 1979) describe these various sample types and the equipment used to collect them. Therefore, this document focuses on sampling designs, with only a few illustrative examples of sample type selection. All sampling designs in this guidance are based on probability sampling.

The selection of an appropriate sampling design for nonpoint source M&E efforts can be a very complicated and frustrating experience for the researcher dabbling in statistics. It is highly recommended that all researchers consult with a qualified statistician for guidance in designing statistically-based sampling programs.

3.1. Simple random sampling

In simple random sampling, each unit of the target population has an equal chance of being selected. For example, if the target population is the macroinvertebrate population found on 100-square meters of river bottom, and the population units are one-square meter sections of river bottom, then each unit would have a one percent chance of being sampled under a random sampling program. To select ten one-square meter sections for analysis, the investigator could first number each section (i.e., 1-100), and then use a random number table to identify the first ten sections to be sampled. In some cases, sampling with replacement is possible (e.g., fish surveys), but in many cases sampling without replacement is the only practical option (e.g., collecting aliquots of stream water).

Gilbert (1987, p. 26-43) and Cochran (1977, p. 18-45) both address many aspects of simple random sampling. Included in these texts are estimation of the mean and total for sampling with and without replacement, equations for determining the number of samples required for both independent and correlated data, and the impact of measurement errors.

3.2. Stratified Random Sampling

In stratified random sampling, the target population is divided into separate groups called "strata" for the purpose of getting a better estimate of the mean or total for the entire population" (Gilbert, 1987, p. 21). Simple random sampling is then used within each stratum.

Stratification involves the use of categorical variables to group observations into more homogenous units to reduce the variability of observations within each unit. Cochran (1977, p. 127-131) provides equations and examples to show how to define strata. Much of the introductory discussion in this guidance document (**Section III.A.**) is intended to illustrate the opportunities and needs for creating strata for improving sampling programs and data analyses and interpretations. Strata are useful when the combined data show greater variability than the separate strata.

Stratified sampling is generally used in efforts to estimate parameters (Gaugush, 1986, p. 195). Stratified sampling results can be used to determine either a combined mean or a combined total population estimate (Gaugush, 1986, p. 200). The combined estimate of the population mean is the mean of the weighted strata means (weighted by stratum sample size). However, if proportional allocation is used (i.e., the sampling fraction is the same in all strata), then the sample is "self-weighting" (Cochran, 1977, p. 91).

The sum of the strata estimates is the total population estimate (Gaugush, 1986, p. 200). The population variance estimate is the weighted sum of the strata variances (Gaugush, 1986, p. 200).

As an example, stratified random sampling could be used to monitor water quality in streams below irrigation return flows. Based on a knowledge of irrigation and precipitation patterns for the watershed, the researcher could divide the year into two or more homogeneous periods. Within each period random samples could be taken to characterize the average concentration of a particular pollutant. These random samples could take the form of daily, flow-weighted composite samples, with the sampling dates randomly determined.

3.3. Cluster Sampling

Another sampling design described by Gaugush (1986, p. 204-205) is cluster sampling. 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 (Freund, 1973, p. 453). For one-stage cluster sampling, these selected clusters are sampled totally (Gaugush, 1986, p. 204). In two-stage cluster sampling, random sampling is then performed within each cluster (Gaugush, 1986, p. 204).

Gaugush (1986, p. 205) states that "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." Still, Gaugush believes that the difficulty associated with analyzing cluster samples can be more than compensated for by the reduced sampling requirements and cost.

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, p. 23). 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 may 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.

3.4. Multi-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 (subunits) within each of the selected subsets (Gilbert, 1987, p. 22). All of the random samples from the subunits are measured completely. Two-stage cluster sampling, described above, is one form of two-stage sampling.

Cochran (1977, p. 274-285) 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 his discussion, Cochran (1977, p. 279) notes that if all population units are sampled, then 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 population units treated as strata.

Three-stage sampling would involve measuring randomly selected subunits within each of the subunits selected in two-stage sampling (Gilbert, 1987, p. 71). The reader is referred to Gilbert (1987, p. 71-87) and Cochran (1977, 285-288) for a more complete discussion of three-stage sampling and compositing, including formulas for estimating the population mean, variance of the mean, and total.

3.5. Double Sampling

Double sampling or two-phase sampling involves taking a large preliminary sample to gain information (e.g., population mean or frequency distribution) about an auxiliary variate (x_i) in the context of a larger sampling survey to make estimates for some other variate (y_i) (Cochran, 1977, p. 327). This technique can be used for stratification, ratio estimates, and regression estimates (Cochran, 1977, p. 327).

Double sampling for stratification requires a first sample to estimate the strata weights and a second sample to estimate the strata means (Cochran, 1977, p. 328). The reader should consult Cochran (1977, p. 327-344) for further explanation, equations, and examples of double sampling for stratification, ratio estimates, and regression estimates.

Gilbert (1987, p. 23) discusses a use of double sampling in which two techniques are used in initial sampling, and subsequent sampling is performed using only the cheaper or simpler technique. The initial sampling is used to establish a linear regression between the measurements from the two techniques. This regression is then applied to the subsequent measurements made with the cheaper technique to predict the measurement result which would have been obtained with the better, more expensive technique. Gilbert (1987, p. 106-117) describes double sampling in much more detail, and provides illustrative examples.

3.6. Systematic Sampling

A commonly used sampling approach is systematic sampling, which entails taking samples at a preset interval of time or space, using a randomly selected time/location as the first sampling point (Gilbert, 1987, p. 23).

Systematic sampling is used extensively in water quality monitoring programs, usually because it is relatively easy to do from a management perspective. According to Gaugush (1986, p. 203), "[i]n systematic sampling the first sample placement is generally decided at random within an initial region, and subsequent samples are taken at some constant distance or time from the first."

Cochran (1977, p. 205) points out that the difference between systematic sampling and stratified random sampling with one unit per stratum is that in systematic sampling the sampled unit occurs in the same relative position within each stratum while in stratified random sampling the relative position is selected randomly.

Cochran (1977, p. 207) also states that when a population ($N=k \bullet n$) is divided into k large sampling units for systematic sampling, then "[a] systematic sample is a simple random sample of one cluster unit from a population of k cluster units." For example, if one daily sample is taken each week in the year, then $k=7$ and $n=52$. That is, the systematic sampling can begin at one of seven days (randomly selected), with a total of 52 units (i.e., 52 weeks) per systematic sample. By randomly choosing Tuesday, for example, the entire systematic sample (or cluster unit) has been randomly selected.

Cochran (1977, p. 229-231) recommends systematic sampling for the following situations:

- The ordering of the population is essentially random or it contains at most a mild stratification.
- Stratification with numerous strata is employed and an independent systematic sample is drawn from each stratum.
- Subsampling cluster units.

- Sampling populations with variation of a continuous type, provided that an estimate of the sampling error is not regularly required.

Systematic sampling is generally inappropriate for hypothesis testing due to violation of the assumption of random sampling. The importance of this violation will vary with the situation. Estimation can, however, be achieved through systematic sampling, but the impact of this sampling design on the bias and precision of estimates should be explored. Quantitative procedures for estimating the population mean and variance from systematic sampling data are presented by Gilbert (1987, p. 96-99).

3.7. Sampling for Regression Analysis

Regression analysis is used to predict variable values based on a mathematical relationship between a dependent variable and one or more independent variables (Gaugush, 1986, p. 208). Gaugush points out that regression analysis requires that at least one quantitative independent variable be used, whereas parameter estimation and hypothesis testing can be performed for groups or classes (i.e., only the variable tested needs to be quantitative).

Gaugush (1986, p. 208-211) discusses sampling to support regression analyses in terms of relating variables to either a spatial or a temporal gradient, the latter being for trends over time. Some key points made in this discussion with respect to gradient sampling are:

- The gradient variable is treated as a covariant to the variable of interest.
- If the relationship is linear, only two points need to be sampled; the extreme points are preferred.
- Whenever the relationship 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. However, the investigator should be careful not to fall in step with some natural phenomenon which would bias any data collected.

Some key points made in this discussion with respect to time sampling are:

- Time can be used either as a covariate or as a grouping variable. Grouping by time may be desirable when changes in the variable of interest are either 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 above for gradients, but (a) time is usually only a surrogate for other variables that truly affect the variable of interest, and (b) 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.

3.8. Comparison of Sampling Strategies

3.8.1. Estimation of the Mean

Gilbert (1987, p. 89) states that systematic sampling will, in many cases, "yield more accurate estimates of mean concentrations" than simple random or stratified random sampling plans. Gaugush (1986, p. 204) suggests that systematic sampling may be useful in defining strata for stratified random sampling designs.

Cochran (1977, p. 221) used environmental data sets in comparing systematic sampling to stratified random sampling in estimating the mean, with the result that systematic sampling was superior to stratified random sampling with one or two data per stratum. Gilbert (1987, p. 95) reports that systematic sampling is equivalent to simple random sampling in estimating the mean if the target population has neither trends nor strata nor correlations among population units.

Freund (1973) notes that estimates based on cluster sampling are generally not as good as those based on simple random samples, but they are better per unit cost.

Cochran (1977, p. 214) found that stratified random sampling provides a better estimate of the mean for a population with a linear trend, followed in order by systematic sampling and simple random sampling. For populations exhibiting a linear trend, Gilbert (1987, p. 96) recommends the use of a weighted estimate of the mean to improve the estimate of the mean derived from systematic sampling. Cochran (1977, p. 216) also recommends using weighted means, but specifies that "all internal members of the sample have weight unity (before division by n) but different weights are given to the first and last members." He also notes that centrally located samples could be used as another option for improved estimates of the mean.

Gilbert (1987, p. 96) notes that estimates of the mean for correlated populations are better for stratified random sampling than for simple random sampling. However, the relative quality of the estimate for systematic sampling depends on the sampling interval.

3.8.2. Estimation of Variance

Cochran (1977, p. 99) states that "stratification nearly always results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample." He goes on to note (p. 100) that under optimum allocation of stratified random samples, the variance is decreased relative to simple random samples because both the differences among the stratum means and the effect of differences among the stratum standard deviations are eliminated.

Estimates of variance from systematic samples may differ from that determined from random samples. Cochran (1977, p. 213) notes that "on the average the two variances are equal." However, Cochran (1977, p. 213) also states that for any single finite population for which k (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, p. 100) 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 which do support unbiased estimation of population variance, including multiple systematic sampling, systematic stratified sampling, and two-stage sampling (Gilbert, 1987, p. 96). 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.

Table VI-1 summarizes the conditions under which each of six probabilistic sampling approaches should be used for estimating means and totals (after Gilbert, 1987, p. 20).

3.9. Examples of Sampling Methodology

The following examples are offered to illustrate the relationship between M&E objectives and sampling methodology.

EXAMPLE 1:

Objective: Determine the annual loading of phosphorus (P) from a watershed with no point sources.

Sampling methodology: Assuming that the majority of P is delivered under high flow conditions, the researcher 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.

Table VI-1. Applications of Six Probability Sampling Designs to Estimate Means and Totals (after Gilbert, 1987)

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.
Multistage 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 or they are known a priori or when strictly random methods are impractical.
Double Sampling	Useful when there is a strong linear relationship between the variable of interest and a less expensive or more easily measured variable.

EXAMPLE 2:

Objective: Determine the extent to which bacterial contamination precludes swimming at a lake beach.

Sampling methodology: USEPA (1978, EPA-600/8-78-017, p. 29) recommends sample collection at locations and times of heaviest use, with daily sampling in the afternoon. Weekends and holidays of highest use should receive special attention.

For estuaries USEPA (1978, EPA-600/8-78-017, p. 29) suggests sample collection at high tide, ebb tide, and low tide "in order to determine the cyclic water quality and deterioration that must be monitored during the swimming season." EPA states that composite samples should not be collected for bacterial examination (1978, EPA-600/8-78-017, p. 6). Surface sampling by hand or by sampling devices, several of which are described by USEPA (1978, EPA-600/8-78-017), are recommended.

EXAMPLE 3:

Objective: Determine seasonal concentration distributions for a tributary site in an agricultural watershed.

Sampling methodology: As stated this objective is too general to determine the sample type required. No mention is made of the vertical and/or horizontal gradients which may exist. Assuming that average spatial concentration is needed, samples can be taken at the depth and location at which the average concentration is shown to occur (this may require initial intensive sampling at the site). Since stream depth and width vary with flow, allowances should be made to adjust the sampling point as needed to accommodate changes in stream flow (this may require that intensive sampling is performed at two or more flow conditions to establish the different "average" sampling points needed).

Random samples or stratified random samples can be collected as discrete or composite samples depending upon the averaging time selected. As stated by Sherwani and Moreau (1975, p. 13), "[t]he appropriate length of the averaging interval is determined by the type of parameter." For seasonal concentration distributions it may be appropriate to use weekly, daily, hourly, or instantaneous samples, depending upon the parameter and the stated objectives.

The above objective is insufficient to determine the desired averaging period. A refinement may be to determine seasonal distributions of hourly mean concentrations. In this case, hourly composite samples would be required.

It should be clear from the above three examples that objectives must be explicit in order to lead to the development of a satisfactory M&E program. The examples above only address the relationship of objectives to sampling methodology, but the implications are similar for other aspects of a complete M&E program.

4. Site Location

Site location and establishment are discussed in several existing monitoring guides and texts (USEPA, 1978b; USGS, 1977; Ponce, 1980b); Wetzel and Likens, 1979; Brakensiek et al., 1979; USEPA, 1981). Few differences exist between NPS site location strategies and the approaches discussed in these documents. Within any given budget, site location is a function of water resource type, M&E objectives, and data analysis plans. 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 NPS monitoring stations be located near USGS gaging stations due to the

extreme importance of obtaining accurate flow records for estimating pollutant loads.

Once the number of measurements needed is estimated the number of stations and the frequency of measurement at each station can be determined. Sherwani and Moreau (1975) provide an example in which collection of 100 samples could entail either sampling at 8 stations for a year or sampling at one station for 8 to 9 years. This analysis assumes that the stations are all representative of the conditions to be monitored. It also assumes that the CVs for the parameters of interest do not change over time. This may not be a good assumption for NPS M&E since parameter CVs will probably change over time due to implementation of control practices.

Two techniques are offered by Sherwani and Moreau (1975, p. 56-60) for selecting station locations. These methods are the gradient location technique and the inter-station correlation technique. The former uses parameter decay rates as the key factor in site placement, whereas the latter seeks to maximize the information content derived from all sites by minimizing the information content common to all sites. Both approaches require data that may not be available before monitoring has begun. The authors conclude that it may be wise to collect more data initially in a monitoring program. This information can be used to evaluate and revise the program as needed.

The following overall site location guidance is taken from NURP monitoring guidance (Shelley, 1979, p. D-22 to D-27) which summarizes much of the information generated by a panel of federal and state monitoring professionals (USEPA, 1975).

> *For overall background and problem assessment the following locations are recommended for the chemical and physical sampling of the water column. Biological and sediment stations should also be established at these locations, as appropriate.*
>
> 1. *At critical locations in water quality limited areas. Stations should be located within areas that are known or suspected to be in violation of water quality standards, ideally at the site of the most pronounced water quality degradation. The data from these stations should gage the effectiveness of pollution control measures being required in these areas.*
>
> 2. *At the major outlets from and at the major or significant inputs to lakes, impoundments, estuaries, or coastal areas that are known to exhibit eutrophic characteristics. These stations should be located in such a way as to measure the inputs and outputs of nutrients and other pertinent substances into and from these water bodies. The information from these stations will be useful in determining cause/effect relationships and in indicating appropriate corrective measures.*

3. *At critical locations within eutrophic or potentially eutrophic lakes, impoundments, estuaries, or coastal areas. These stations should be located in those areas displaying the most pronounced eutrophication or considered to have the highest potential for eutrophication. The information form these stations, when taken in combination with the pollution source data, can be used to establish cause/effect relationships and to identify problem areas.*

4. *At locations upstream and downstream of major population and/or industrial centers which have significant waste discharges into flowing surface waters. These stations should be located in such a way that the impact on water quality and the amounts of pollutants contributed can be measured. The information collected from these stations should gage the relative effectiveness of pollution control activities.*

5. *Upstream and downstream of representative land use areas and morphologic zones within the area. These stations should be located and sampled in such a manner as to compare the relative effects of different land use areas (e.g., cropland, mining area) and morphologic zones (e.g., piedmont, mountain) on water quality. A particular concern for these stations is the evaluation of nonpoint sources of pollution and the establishment of baselines of water quality in sparsely populated areas.*

6. *At the mouths of major or significant tributaries to mainstem streams, estuaries, or coastal areas. The data from these stations, taken in concert with permit monitoring data and intensive survey data, will determine the major sources of pollutants to the area's mainstem water bodies and coastal areas. By comparison with other tributary data, the relative magnitude of pollution sources can be evaluated and problem areas can be identified.*

7. *At representative sites in mainstem rivers, estuaries, coastal areas, lakes, and impoundments. These stations will provide data for the general characterization of the area's surface waters and will provide baselines of water quality against which progress can be measured. The purpose of these stations is not to measure the most pronounced areas of pollution, but rather to determine the overall quality of the water. Biological monitoring will be a basic tool for assessing the overall water quality of an area.*

8. *In major water use areas, such as public water supply intakes, commercial fishing areas, and recreational areas. These stations serve a dual purpose: the first is public health protection and the second is for the overall characterization of water quality in the area.*

Determining the presence and accumulation of toxic substances and pathogenic bacteria and their sources are primary objectives of these stations.

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 will be to assess the accumulation of toxic substances, and locations for sediment sampling should be chosen with this in mind. Sediment mechanics and the hydrological characteristics of the water body must be considered.

In general, biological monitoring stations should be established as follows:

1. *At key locations in water bodies that are of critical value for sensitive uses such as domestic water supply, recreation, and propagation and maintenance of fish and wildlife.*

2. *In major impoundments near the mouths of major tributaries.*

3. *Near the mouths of major rivers where they enter an estuary.*

4. *At locations in major bodies potentially subject to inputs of contaminants from areas of concentrated urban, industrial, or agricultural use.*

5. *At key locations in water bodies largely unaffected by man's activities.*

For purposes of biological monitoring, a station will normally encompass areas, rather than points, within a reach of river or area of lake, reservoir, or estuary adequate to represent a variety of habitats typically present in the body of water being monitored. Unless there is a specific need to evaluate the effects of a physical structure, it is advisable to avoid areas that have been altered by a bridge, weir, within a discharge plume, etc. Thus, biological sampling stations may not always exactly coincide with water column or sediment stations.

To the extent possible, all monitoring stations should be located in such a manner as to aid cause/effect analyses. Some station requirements may be such that, with careful station siting, one particular stations could meet the criteria of a number of types of stations. Caution should be exercised, however, to avoid compromising the worth of a station for the sake of false economy. In general, the quality of a monitoring program is not judged solely by the number of stations. A few critically located stations may be extremely valuable, while a large number of randomly selected stations may yield

meaningless data. Resource constraints will limit the total number of stations.

The Rural Clean Water Program (RCWP) includes several examples of NPS M&E strategies. Two project strategies are described here to illustrate slightly different approaches utilized under one overall program with common goals for all projects. In a simple sense, two of the program goals are to document receiving water quality improvements resulting from agricultural NPS control efforts, and to document the water quality effects of specific agricultural pollution control practices.

The Idaho RCWP project had as its major focus the control of sediment from irrigation return flows (Martin, 1984, p.6). Two sets of monitoring stations were utilized for (1) ambient and (2) intensive monitoring (Figure VI-1). The following describes the stations and their purposes (Martin, 1984, p. 10-11):

S-1: near mouth - integrates all pollution sources flowing into Rock Creek and measures the pollutant load that goes into the Snake River (RM 0.75). Water quality, benthic macroinvertebrates and fisheries data being 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 agricultural priority drains and city of Twin Falls urban runoff (RM 7.3). Water quality, benthic macroinvertebrates and fisheries data being 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 being collected.

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

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

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

Intensive monitoring stations were placed on irrigation drains to track changes in sediment load and associated pollutants as close to their source and the solutions (BMPs) as possible. In this way, changes in water quality due to the RCWP program could be more easily and quickly detected. Nineteen stations were located in six subbasins (Figure VI-1). Stations measure the source of

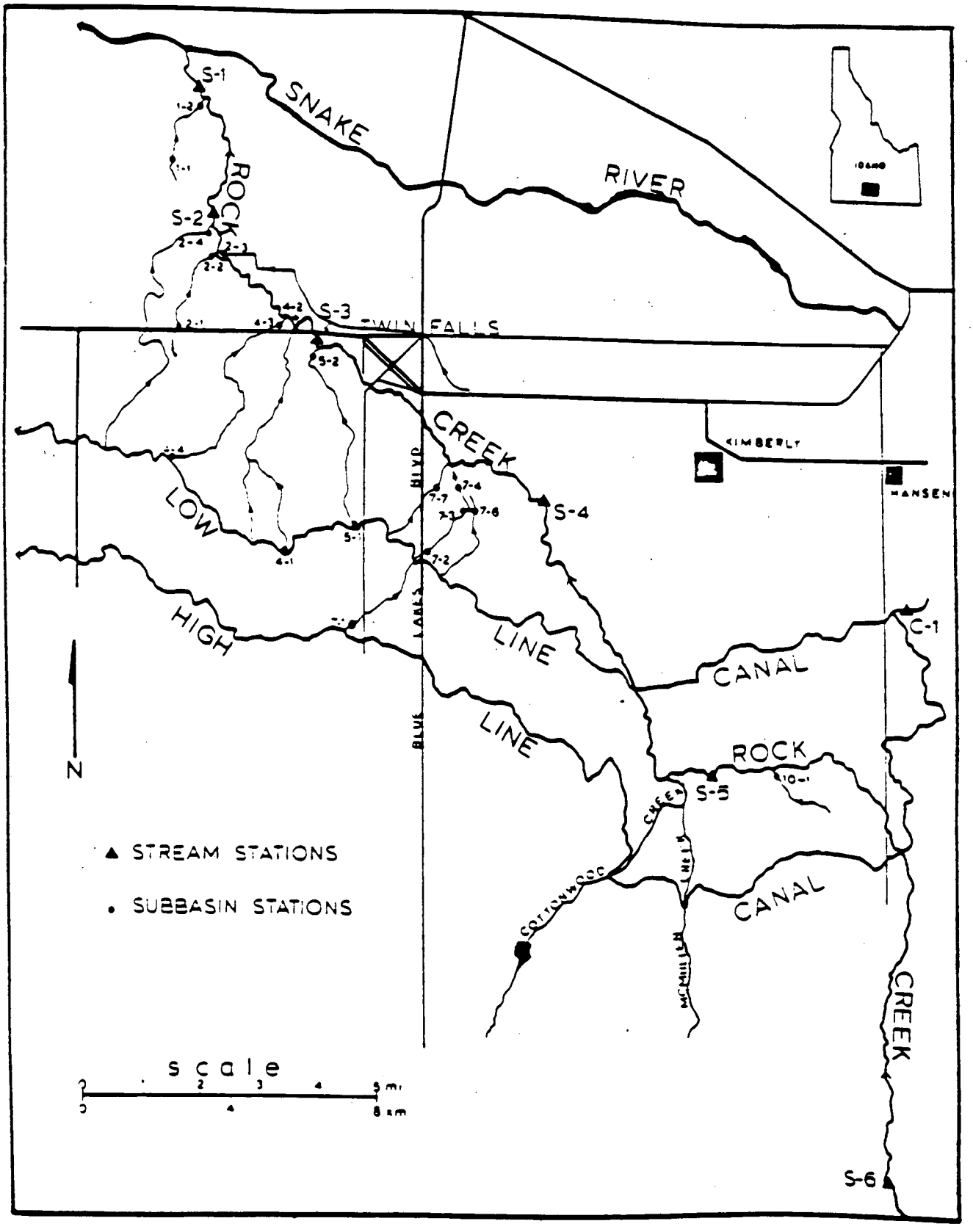


Figure VI-1. Map of the Rock Creek Rural Clean Water Program study area, Twin Falls County, Idaho. (SOURCE: Martin, 1984)

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 utilized a four-level M&E program to meet three objectives (Vermont RCWP Coordinating Committee, 1986, p.III-1):

1. *To document changes in the water quality of specific tributaries within the watershed resulting from implementation of manure management practices.*

2. *To measure the changes in the amount of suspended sediment and nutrients entering St. Albans Bay resulting from implementation of water quality management programs within the watershed.*

3. *To 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 M&E are shown in Figure VI-2. The Level 1 Bay Sampling is designed to determine long-term water quality trends in St. Albans Bay over the life of the project (Vermont RCWP Coordinating Committee, 1984, p. IV-1). The Level 2 Tributary Sampling is 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, p. IV-1). The Level 3 monitoring is directed toward evaluating the effect of best manure management practices on the quality of surface runoff from individual fields, while Level 4 is designed to supplement the Level 2 monitoring by sampling additional tributaries to St. Albans Bay and to isolate sub-units within the Level 2 subwatersheds (Vermont RCWP Coordinating Committee, 1984, p. IV-3).

Additional monitoring efforts include meteorological monitoring, biological monitoring, land use monitoring, wetland sampling, bay and wetland sediment sampling, and a bay circulation study. More about the St. Albans Bay monitoring plan will be discussed under sampling frequencies and approaches.

These two examples should point out the similarities between site location for NPS M&E and the traditional site locations recommended by the National Water Monitoring Panel (USEPA, 1975). Clearly, in order to locate stations effectively the researcher must:

- Have clear, quantitative, M&E objectives.

- Understand the watershed/water body to be monitored.

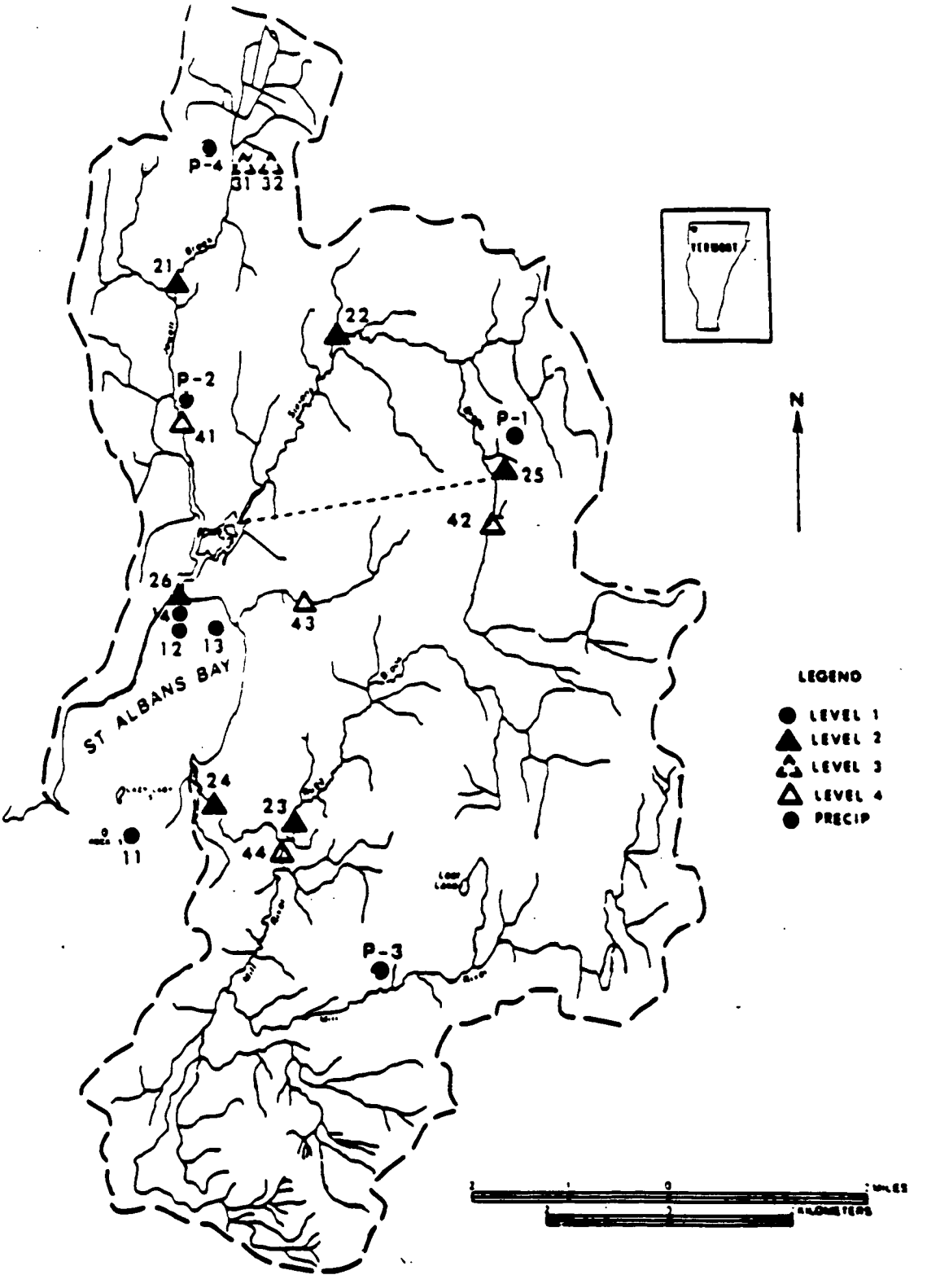


Figure VI-2. St. Albans Bay watershed, Franklin County, Vermont, sampling locations. (SOURCE: Vermont RCWP Coordinating Committee, 1986)

- Know something about the locations of and pollutant transport from point and nonpoint sources.

5. Sample Size and Allocation

The issue of sample size is important from several perspectives, including (1) statistical and/or modeling uses of the monitoring data, (2) logistics of sample collection, and (3) cost. For these reasons it is important to estimate early in an M&E effort the number and frequency of samples required to meet M&E objectives within budget constraints.

5.1. Simple Random Sampling

As discussed by Reckhow (1979, pp. 286-293), Cochran (1963) determines the number of random samples required to achieve a desired level of precision as:

$$n = \frac{t^2 \bullet s^2}{d^2} \qquad (1)$$

where:

n = number of samples

t = Student's t value

s^2 = population variance estimate, and

d = desired precision in units of parameter.

Reckhow concludes that "random sampling design is specified by the variance estimate, the precision, and the sampling program cost (or number of samples)."

To estimate the population variance, Reckhow reiterates Cochran's (1963) four information 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.

5.2. Stratified Random Sampling

For stratified random sampling, the equation is (Reckhow, 1979, p. 290):

$$\frac{n_i}{n} = \frac{w_i \bullet (CV)x_i \bullet (x_i)}{\Sigma w_i \bullet (CV)x_i \bullet (x_i)} \qquad (2)$$

where:

n = total number of samples

n_i = number of samples in stratum i

x_i = magnitude (mean) of characteristic x in stratum i,

w_i = a weight reflecting the size (number of units, for example) of stratum i, and

$(CV)_{x_i}$ = coefficient of variation (standard deviation divided by the mean) of characteristic x in stratum i.

Reckhow (1979, p. 291) reiterates Cochran's (1977, p. 98) rules for determining the composition of a stratified random sampling design. He states that a larger sample should be taken in a stratum if the stratum is:

- more variable (CV).
- larger (w, x), and
- less costly to sample.

Gaugush (1986, p. 196) presents a general equation for sample allocation in a random sampling program. This equation takes into account the size of the strata, the variance of the strata, and the cost of sampling the strata to arrive at <u>proportional</u> allocations. The equation can be simplified if the size, variance, and/or cost are the same across strata. If strata are not needed the equation reduces to Equation 1. The general equation is:

$$n_i = n \bullet \left(\frac{N_i \bullet \sigma_i / \sqrt{c_i}}{N_1 \bullet \sigma_1 / \sqrt{c_1} + N_2 \bullet \sigma_2 / \sqrt{c_2} + \ldots + N_s \bullet \sigma_s / \sqrt{c_s}} \right) \qquad (3)$$

where

s = the number of strata
i = the stratum number (i = 1, 2 ... s)
N_i = the size of the ith stratum
σ_i = the standard deviation of the ith stratum
c_i = the cost of sampling the ith stratum
n = the total number of samples to be allocated
n_i = the number of samples allocated to the ith stratum

The reader is referred to Gaugush (1986, pp. 194-200) for further discussion and illustration of this sample allocation formula.

Cochran (1977, p. 119-124) offers methods for determining optimum sample allocation in cases where more than one item (parameter) is being sampled. For example, initial survey results may indicate that the optimum sample allocation for conductivity measurements is different from that for total phosphorus. Cochran (1977, p. 119-124) provides examples where optimal, compromise, and proportional allocations are compared.

5.3. Cluster Sampling

Cochran (1977, p. 233-272) discusses in great detail single- or one-stage cluster sampling for clusters of either equal or unequal sizes. Cochran (1977, p. 234-237) 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.

5.4. Multi-stage Sampling

Gilbert (1987, p. 58-69) discusses and illustrates the use of two-stage sampling for determining the mean and total in environmental monitoring efforts. The author provides 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.

Gilbert (1987, p. 71-87) shows how to use composite samples of both equal- and unequal-sized units to estimate the mean and total values for the target population. The author also provides equations for calculating the number of composites and composite subsamples needed.

5.5. Double Sampling

Gilbert (1987, p. 109-117) provides methods for estimating the number of samples needed to determine the mean and total for the target population. These methods address fixed cost, fixed variance, and stratified double sampling. A case study is provided to illustrate the use of the sample allocation equations.

5.6. Systematic Sampling

Gilbert (1987, p. 90-94) discusses systematic sampling along a line (space or time) and over space. In sampling along a line, the author notes that the starting point should be randomly selected, and subsequent sampling points should be spaced at regular intervals from the starting point. Procedures are presented for selecting sample points from either a fixed sample size or a fixed sample interval.

Gilbert (1984, p. 92) cautions that any periodic variation in the target population should be known before establishing a systematic sampling program. It is highly likely that sampling intervals equal to or multiples of the target population's cycle of variation will result in biased estimates of the population mean. Cochran (1977, p. 218) states that systematic samples can be designed to capitalize on a periodic structure if that structure can be characterized sufficiently. A simple or stratified random sample is recommended, however, in cases where the periodic structure is not well known (Cochran, 1977, p. 218).

5.7. Sampling for Regression Analysis

According to Gaugush (1986, p. 210), the number of samples required when time is used to create strata should reflect the desired precision. If stratified random sampling is used, then Equations 2 and 3 should be appropriate. If time is used as a covariate, then the time cycle (e.g., annual) should be subdivided, with sample replicates placed randomly within the subunits (Gaugush, 1986, p. 210).

6. Sampling Frequency

Sherwani and Moreau (1975, p. 61) have stated that the desired frequency of sampling is a function of several considerations, including:

- the 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,
- the magnitude of response, and
- variability of the inputs.

NOTE: While this guidance focuses on nonpoint source pollution, it is important to note that effluent from wastewater treatment plants can show considerable fluctuation in both quantity and quality throughout the day and has a weekly cycle (Sherwani and Moreau, 1975, p. 62).

Sherwani and Moreau (1975, p. 67) state that for a given confidence level and margin of error, the frequency of sampling is proportional to the variance. Since the variance of water quality parameters may differ considerably over time, the frequency requirements of a monitoring program may vary depending on the time

of the year. Sampling frequency will need to be greater during periods of greatest variance.

One approach to determine sampling seasons is to plot variance over time, dividing the year into periods of relatively constant variance (Sherwani and Moreau, 1975, p. 68). Sampling frequency requirements can then be determined for each period.

The authors calculate the longest acceptable sampling interval using Equation 4 (Sherwani and Moreau, 1975, p. 73).

$$t = \frac{d}{S \bullet (1/n_1 + 1/n_2)^{1/2}} \qquad (4)$$

where t is Student's t value for the level of confidence desired (d.f.=n_1-1), d is the specified precision (in parameter units if normal distribution), n_1 and n_2 are the number of observations per year, and S is the standard deviation of the larger number of observations (n_1). To use this equation, data must already exist for a sampling program with shorter sampling intervals than for the desired program (i.e., the equation is useful for decreasing sampling frequency from n_1 to n_2, but should not be used for increasing sampling frequency). For example, daily sampling would result in 365 samples (n_1) with a corresponding standard deviation (S) for the parameter of interest. The minimum number of samples per year (i.e., the longest sampling interval) is then calculated as n_2 for a desired confidence (t) and precision (d) with Equation 4.

The following is taken from Sherwani and Moreau (1975, p. 69) to illustrate the effect of parameter selection on sampling frequency.

The sampling interval of a constituent should be related to the temporal behavior of the constituent that is being measured. Here, the dichotomy between a conservative and non-conservative substance is quite useful. The conservative substances have relatively long half-lives in the sense that they are not degraded or their rates of uptake by organisms are low. They are transported relatively unchanged in form over long distances. Problems caused by these substances in a natural aquatic system are generally those that result from build-up over relatively long periods of time. Their instantaneous concentrations are less important than their long-term average concentrations. For conservative substances, monitoring can be carried out over relatively coarse intervals of time. For some constituents annual averages may be sufficient; for others, seasonal or monthly averages may be necessary to characterize the mass flux through the system.

Non-conservative substances are characterized by relatively short half-lives. The adverse effects of these substances are

realized over short intervals of time and space. The temporal coverage can be limited to critical periods. Mean annual concentrations of these substances are of little value in water quality management. The half-lives of non-conservative constituents may be small relative to the time of travel through the basin, especially under low-flow conditions.

Sherwani and Moreau (1975, p. 71) discuss the serial correlation between parameter values taken over time and warn against wasting monitoring funds on too frequent sample collection. However, they also note that a sufficient number of samples is needed "to describe adequately the significant information in the high frequencies."

Noting that most water quality parameters exhibit a periodic structure, Sherwani and Moreau (1975, p. 71) state that "the least favorable sampling interval is equal to the period or an integral multiple of a period." They say that the best case occurs when the sampling interval is an odd multiple of half-period. As an example for a weekly cycle, they recommend that sampling is staggered such that every week day is equally represented.

In both random sampling and stratified random sampling Reckhow (1979, p. 288) points out that systematic sampling is acceptable only if "there is no bias introduced by incomplete design, and if there is no periodic variation in the characteristic measured." He goes on to say that such systematic sampling is even better if the starting point is based on a randomly selected day, site, and/or depth, with systematic sampling from then on.

When time is used as a stratification variable, the samples within the time strata should be randomly allocated, with the number of samples reflecting the desired precision (Gaugush, 1986, p. 210). If time is used as a covariate, the annual cycle should be subdivided, with sample replicates randomly placed within the subunits. Gaugush (1986, p. 210) reports that "monthly sampling is usually adequate to detect the annual pattern of changes with time." More frequent sampling would be required for the detection of short-lived phenomena.

Finally, two-stage sampling incorporates 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, then the primary units could be daily average concentrations (n=365). A subset of these daily concentrations (e.g., n=24) could be selected at random for further systematic sampling of hourly concentrations (four hourly composites per day). Thus, 96 hourly composite samples would be taken for calculation of the population mean and variance. Four systematic, hourly samples would 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. For example, if hour six (6:00 AM) was selected randomly as the first hour, then hourly composite samples would also be taken at noon, 6:00 PM, and midnight.

C. Ground Water Monitoring

Ground water monitoring and evaluation is addressed in a peripheral manner in this NPS M&E guidance. Other documents address ground water resources, subsurface sampling, and ground water data analysis in much greater detail (USGS, 1980; Scalf et al., 1981; USEPA, 1987a; Everett, 1980), and should be reviewed by those investigating ground water contamination from NPS. The following are general comments taken from these very useful documents.

1. Data Needs

As for surface water, the data to be collected is determined by the objectives of the study. It is recommended that "[b]efore the field is ever visited, a thorough search should be made of files and the literature for pertinent information" (USEPA, 1987a, p. 22). The types of information to be collected include: soil, geologic, topographic, county, and state maps; geologic cross sections; aerial photographs; satellite imagery; location of pumping centers and discharge rates; well logs; climatological and stream discharge records; chemical data; and the locations of potential sources of ground water contamination.

Most of the above data sources are the same as those discussed for surface waters, however, the applications of the data may be quite different. For example, soil information can be used "to evaluate the potential for movement of organic and inorganic compounds through the unsaturated zone" (USEPA, 1987a, p. 23). Aerial photographs and satellite imagery are helpful in locating potential drilling sites and indicating the presence of faults or joints.

Well logging is the description of collected samples and factors involved in drilling the sampling wells (USGS, 1980, p. 2-80). The types of well logs include geologic, lithologic, driller's, construction, drilling rate, and fluid loss. These logs are essential in ground water investigations, providing information on types and characteristics of rocks in the subsurface, drilling conditions, well construction details, and water levels (USEPA, 1987a, p. 23).

The USEPA (1987a, p. 109) recommends candidate parameters to be measured for both detective and assessment monitoring programs. Scalf et al., (1981, p. 87) provide a summary of sampling methods and well construction considerations, keyed to the nature of the monitored pollutant. As is the case for surface water monitoring, the parameter lists for ground water monitoring programs should be tailored to match the objectives of the M&E effort.

2. Site Location

Ground water research has shown that "it is impossible to prescribe standard locations for sampling points that would be applicable to all sites" (Scalf et al., 1981, p. 7). Furthermore,

"[t]he placement and number of wells will depend on the complexity of the hydrologic setting and the degree of spatial and temporal detail needed to meet the goals of the program" (USEPA, 1987a, p. 110).

The importance of investigating existing data sources cannot be overemphasized with regard to site location. For example, water-level data can be used to develop contours for use in selecting sampling sites (Scalf et al., 1981, p. 6). Background sampling is recommended to help situate wells and/or define sampling points. In addition, when sampling to track pollutant plumes, "it is extremely important to locate subsequent monitoring wells one at a time, sample, and base succeeding well locations on results of previous sampling" (Scalf et al., 1981, p. 11).

While it is sometimes possible to consider surface water monitoring to involve only two dimensions, it is always important to consider three dimensions when locating ground water monitoring sites. Four general categories reflecting the depth of monitoring are the land surface, the top soil, the vadose zone, and the zone of saturation (Everett, 1980, p. 130). Considerations made in sampling from each of these depth categories are discussed in detail in several documents (Everett, 1980; USEPA, 1987a; Scalf et al., 1981; USGS, 1980).

3. Sampling Frequency

It is generally recommended that monthly, seasonal, or annual sampling is sufficient for most ground water monitoring programs. However, more frequent sampling may be required in very shallow aquifers or under conditions of rapidly changing water quality (USGS, 1980, p. 2-88).

It is important to note that irrigation, precipitation, drainage control devices, and other natural and artificial events may influence the water level and/or ground water quality in such a way that storm-event, weekly, or some other unique sampling frequency may be needed to track ground water quality adequately.

D. Example Monitoring Programs

Below are three examples of nonpoint source monitoring programs which are considered by many NPS 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. Furthermore, cost will be a major factor in determining the scope and intensity of a monitoring program, so these examples should be put into a cost perspective. The St. Albans Bay, Vermont monitoring and evaluation program was funded at $1.6 million over eleven years (1980-1990), a price well beyond reach for many watersheds of similar size (33,344 acres) (Smolen et al., 1986, p. 35).

1. EXAMPLE 1: St. Albans Bay, Vermont RCWP Project

The monitoring and evaluation objectives for St. Albans Bay have already been given under **Section III.C.2.a**. The complete monitoring strategy is given below (Vermont RCWP Coordinating Committee, 1984, p. IV-1 to IV-11).

Site Locations

The monitoring strategy for the St. Albans Bay Watershed includes long-term water quality monitoring, related long-term monitoring, and short-term intensive studies:

- *Long-Term Water Quality Monitoring*

Level 1: Bay Sampling

Level 1 monitoring is designed to determine long-term water quality trends in St. Albans Bay over the life of the project. Four sampling locations are located in St. Albans Bay (Figure VI-2):

Station 11: Outer Bay
Station 12: Inner Bay
Station 13: Beach
Station 14: Off Bridge (new)

Level 2: Trend Stations

Level 2 monitoring is 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 continuous monitoring stations are located throughout the watershed (Figure VI-2):

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: Mobile Monitoring

Level 3 monitoring is designed to evaluate the effect of best manure management practices on the quality of surface runoff from individual fields. Two continuous monitoring stations are located in the headwaters of the Jewett Brook subwatershed (Figure VI-2):

Station 31: Larose ditch - below site
Station 32: Larose ditch - above site

Level 4: Random Sampling

Level 4 monitoring is designed to supplement the Level 2 monitoring by sampling additional tributaries to St. Albans Bay and to isolate sub-units within the Level 2 subwatersheds. Randomly timed grab samples are collected at four locations (Figure VI-2):

Station 41: Jewett Brook
Station 42: Stevens Brook
Station 43: Guayland Brook
Station 44: Mill River

- *Related Long-Term Monitoring*

Meteorological Monitoring

Precipitation and other climatological data are being collected in the St. Albans Bay Watershed at four locations (Figure VI-2):

P-1: St. Albans Radio Station
P-2: Dunsmore Farm
P-3: Franklin Ford Tractor
P-4: LaRose Farm

Biological Monitoring

The monitoring of biological characteristics is being conducted in St. Albans Bay and in Bay tributaries. Bay biological monitoring is conducted at the Level 1 stations (Figure VI-2). Tributary biological monitoring is conducted at five locations (Figure VI-2):

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

Land Use Monitoring

A cooperative program has been developed to collect land use and agricultural activity information from each farm in the watershed. Both baseline information and daily field log data are collected. These data are then input to a computerized Geographic Information System (GIS).

- *Short-Term Intensive Studies*

Wetland Influences

Stevens Brook wetland has been sampled to determine the effects of the wetland on water entering St. Albans Bay from point and nonpoint sources. Fifteen sampling stations are located within

the wetland along the brook channels. Station 26 serves as a continuous monitoring of the wetland outlet.

Bay and Wetland Sediments

Fifteen sampling stations have been located within the Bay and contiguous wetland to determine the chemical and physical properties of the sediments. Six of these stations have been used for within-year temporal studies. Sediment phosphorus release studies have also been conducted at these sites.

Bay Circulation

Wind, water current, and concentration data have been collected in St. Albans Bay to determine the effect of bay circulation on water quality. A model has been developed to predict phosphorus concentrations in the Bay under different loading rates and meteorological conditions.

Parameters Monitored

- *Long-Term Water Quality Monitoring*

Table VI-2 lists the parameters monitored for each level of the long-term monitoring.

- *Related Long-Term Monitoring*

Table VI-3 lists the parameters monitored for each of the related long-term monitoring efforts.

- *Short-Term Intensive Studies*

Table VI-4 lists the parameters monitored for each of the short-term studies.

- *Data Collection Schedule*

Table VI-5 summarizes the frequency and type of sample collected for the long-term water quality monitoring. However, as indicated in Table VI-6, not all stations are operational each year of the project.

- *Sampling and Analytic Techniques*

The sampling years for each monitoring level are summarized in Table VI-6. The methods used to determine each water quality parameter are given in Table VI-7.

2. EXAMPLE 2: Bellevue, Washington NURP Project

The Nationwide Urban Runoff Program (NURP) was designed in such a way that each project met minimum monitoring requirements to support national analyses. Thus, the monitoring programs were very similar for all NURP projects. The Bellevue, Washington monitoring

Table VI-2. Long-Term Monitored Parameters for Each Level in the St. Albans Bay Watershed.

Level 1 - Bay Monitoring

Turbidity	Chlorophyll a
Total Suspended Solids	Fecal Coliform
Volatile Suspended Solids	Fecal Streptoccoccus
Total Phosphorus	Temperature
Ortho-Phosphorus	Dissolved Oxygen
Ammonia-Nitrogen	pH
Total Kjeldahl Nitrogen	Conductivity
Nitrite + Nitrate - Nitrogen	Secchi Disc

Level 2 - Subwatershed Trend Stations

Turbidity	Fecal Coliform
Total Suspended Solids	Fecal Streptococcus
Volatile Suspended Solids	Temperature
Total Phosphorus	Dissolved Oxygen
Ortho-Phosphoru	pH
Ammonia-Nitrogen	Conductivity
Total Kjeldahl Nitrogen	
Nitrite + Nitrate - Nitrogen	

Level 3 - Mobile/Field Monitoring Stations

Total Suspended Solids	Ammonia-Nitrogen
Volatile Suspended Solids	Total Kjeldahl Nitrogen
Total Phosphorus	Nitrite + Nitrate - Nitrogen
Ortho-Phosphorus	

Level 4 - Random Grab Sampling

Turbidity	Fecal Coliform
Total Suspended Solids	Fecal Streptococcus
Volatile Suspended Solids	Temperature
Total Phosphorus	Dissolved Oxygen
Ortho-Phosphorus	pH
Ammonia-Nitrogen	Conductivity
Total Kjeldahl Nitrogen	Flow
Nitrite + Nitrate - Nitrogen	

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

A.6.2 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. Specialized software includes SAS, Statistica, and Statgraphics that cover a wide range of statistical and graphical support. Numerous spreadsheets, data base management packages, and other graphics software can also be used to perform many of the needed analyses. In addition, the following programs, written specifically for environmental analyses, are also 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 Agricultural and Chemical Engineering, Colorado State University, Fort Collins, CO 80524.

B. Exploratory Analysis

The first step of data analysis is exploratory. The purpose of this step is to perform a preliminary assessment of the data and "develop a feeling" for the data. This section describes approaches for data inspection, data stratification, data normalization, and tests of assumptions. If the data analyst is familiar with the data set, data inspection and stratification is fairly straightforward and usually intuitive. Data normalization involves the process of transforming data into a usable form and assumption testing includes evaluating hypothesis about the statistical characteristics of the data. The last portion of this section discusses probability distributions in detail and is considered optional reading.

B.1 Data Inspection

Researchers generally agree that data inspection is necessary before data analyses are performed on a data set (Gaugush, 1986; Spooner et al., 1986, p. 3). Gaugush (1986) recommends graphical displays as a means for inspecting data as follows:

> *It is good practice in statistical analysis to begin with a study involving graphical displays of the data. That is,*

The other type of trend commonly tested for is 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.

Determining only the existence of a trend is sometimes not sufficient for decision makers. It is also necessary to estimate the magnitude of the trend. The Seasonal Hodges-Lehman estimator is recommended for estimating the trend magnitude when comparing two or more random samples. The Seasonal Kendall slope estimator is recommended when estimating the magnitude of monotonic trends.

A.5.3 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 ran-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. This may result in selecting portions of data sets for analysis. It is also important to make sure that length of record is appropriate.

A.6 Recommended Reading List and Available Software

A.6.1 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 and are strongly recommended for those who need a more in-depth discussion than can be 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.* Wile, New York, 493 pp.

A.5.2 Changing Conditions

One of the most frequently asked questions related to the evaluation of monitoring is whether conditions have improved or degraded. The approaches used to select monitoring station locations range from random sampling to fixed stations (see Section VI.A.3). Both types of monitoring designs can be used to evaluate trends. In random sampling, each location along a stream has an equal probability of being selected and is common to intensive surveys. Random sampling (and variations) is generally most appropriate when estimating means and totals if there is a low likelihood of major patterns of contamination or trends. Fixed station sampling is commonly based on additional knowledge about where impairments are likely to be or is based on professional judgment.

In the case of random sampling, data are typically collected in a "before and after," "upstream/downstream," or "reference and test site" setting. The resulting data usually come as two or more sets of random samples. In the case of upstream/downstream sampling, the upstream sampling locations typically represent a reference or background condition, whereas the downstream locations typically represent impaired (targeted) conditions. The goal might be to determine whether there is a significant difference in phosphorus concentrations between the background and impaired (targeted) areas. If so, it might be possible to infer that significant phosphorous loadings occurred between the two locations. Similarly, when performing a biological assessment the goal would be to determine whether there is a significant difference in the biological metric between the reference and test (targeted) sites.

In either case, at least two random samples have been collected. 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. (Several researchers have 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.)

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 indexes between the different seasons (index periods).

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.

Table VIII-2. Methods for Routine Data Analysis (adapted from Ward et al., 1990).

INFORMATION NEED	METHOD (P = parametric, N = nonparametric)		
	GRAPHICAL	**ESTIMATION**	**HYPOTHESIS TESTING**
Average conditions	Box-and whiskers plots (B.1) Time series plots (B.1)	**(P)** Sample mean (or geometric mean) and sample standard deviation (or geometric standard deviation) with confidence limits (C.1, C.2) **(N)** Sample median and interquartile range with confidence limits (C.1, C.2) **(N)** Seasonal Hodges-Lehman estimator (D.2.2.2)	**Two random samples** **(P)** Student's *t* test (D.2.1.1) **(N)** Mann-Whitney test (D.2.2.2) **Matched samples** **(P)** Paired *t* test (D.2.3.1) **(N)** Wilcoxon signed rank test (D.2.1.2) **Three or more random samples** **(P)** ANOVA (D.2.4.1) **(N)** Kruskal-Wallis test (D.2.4.2)
Changing conditions	Annual box-and-whiskers plots (B.1) Time series plots (B.1)	**(P)** Linear regression (D.3.4) **(N)** Seasonal Kendall slope estimator (D.3.4)	**(P)** *t*-test for significance of slope (D.3.4) **(N)** Seasonal Kendall test (D.3.4)
Extreme conditions	Time series plots with excursion limit	**(P,N)** Proportion (frequency) of Excursions (E.3)	**(P)** Test for equality of proportions (E.3) **(N)** Confidence limits on proportions (E.3) **(N)** Tolerance intervals (E.3)

Site Selection
Sampling Frequency Determination

A.5 Recommendations for Selecting Statistical Methods

Information goal assessment typically involves the evaluation of the average, changing, and extreme conditions of environmental variables (Ward and Loftis, 1986). This section provides recommendations for selecting statistical methods that can be applied on a routine basis (Table VIII-2, adapted from Ward et al., 1990). In some instances, more appropriate methods might be available depending on the specific information needs. One decision facing many data analysts is the selection of parametric versus nonparametric methods. For routine analyses, nonparametric methods are recommended. Nonparametric tests also have underlying assumptions and violation of these assumptions can lead to incorrect conclusions about the collected data.

A.5.1 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 in Section VIII.A.3, 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. Since these parameters are only point estimates, the uncertainty should also be indicated by reporting parameter estimates with confidence limits.

The selection of the mean (and standard deviation) versus the median (and interquartile range) is sometimes a subjective decision. If the data are symmetrically distributed, either set of parameters is acceptable. 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.

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 of what might be considered in an expectations report (the data analyst should modify this outline to suit:

1. *Evolution of Water Quality Management Program*
   - *Geographical/Hydrological Setting*
   - *Water Quality Problems*
   - *Water Quality Laws*
   - *Management Program Structure*
   - *Management Procedures*

2. *Information "Expected" by Management Program*
   - *Implications of the Law Establishing the Program*
   - *Legal Goals*
   - *Management Powers and Functions*
   - *Directly State Monitoring Requirements*
   - *Information Needs of Management Operations*
   - *Water Quality Criteria*
   - *Water Quality Standards*
   - *Permits*
   - *Compliance*
   - *Enforcement*
   - *Construction Loans*
   - *Planning*
   - *Water Quality Assessment*

3. *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*

4. *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*

5. *Resulting Monitoring Network Design Criteria*
   - *Variable Selection*

Table VIII-1. Methods for characterizing data (adapted from Ward et al., 1990).

DATA CHARACTERISTIC	METHOD	METHOD TYPE	SECTION
Central tendency	Sample mean	P	C.1.1
	Sample median	N	C.1.1
	Sample geometric mean	P	C.1.1
	Box-and-whiskers plot	G	B.1
Spread	Sample standard deviation	P	C.1.2
	Range, maximum-minimum	P,N	C.1.2
	Interquartile range	N	C.1.2
	Box-and-whiskers plot	G	B.1
Distribution shape	Histogram	G	B.1
	Percentiles	N	C.2.3
	Sample skewness	P	C.1.3
	Sample kurtosis	P	C.1.4
	Chi-squared test	P	B.5.2.3
	Kolmogorov-Smirnov test	N	B.4.1.4
	Shapiro-Wilk test	N	B.4.1.3
Seasonal variation	Time series plots	G	B.1
	Seas. box-and-whiskers plot	G	B.1
	ANOVA	P	D.2.4.1
	Kruskal-Wallis test	N	D.2.4.2
Serial correlation	Sample autocorrelation	P	B.4.4
	Spearman's rho	N	B.4.4

Key to Method Type:
P = Parametric N = Nonparametric G = Graphical

6. 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 expectation report, the data analyst may 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 expectation report, the presentation of results should be selected depending on the audience reviewing the information and the objectives of the

they 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). Section VIII.G describes data record limitations in more detail and provides suggestions for performing routine data analyses that minimize the impact of such limitations.

Common statistical characteristics include location (central tendency), variability (scale or spread), distribution shape, seasonality, and serial correlation. While these characteristics do not cause computational problems, they can violate the underlying assumptions of statistical methods. "All statistical methods which produce probability statements, such as confidence intervals, are based on assumptions about the data used in the analysis" (Ward et al., 1990). Common assumptions include distribution shape, seasonality, correlation, and applicability of a linear model. Table VIII-1 presents a variety of methods for characterizing data that are helpful for selecting appropriate statistical methods and providing a general understanding of water quality data. Sections VIII.B and VIII.C provide a detailed discussion of each data characteristic and method.

A.4 Communicating Results

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 have bought into 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:

1. Perform a thorough review of the legal basis for the management effort and define the resulting "implications" to monitoring.

2. Review administrative structure and procedures developed from the law in order to define information expectations of management staff.

3. Review ability of monitoring program to supply information.

4. Formulate information expectation report for monitoring system.

5. Present information expectation report to all users of the information.

statistical inference: estimation and hypothesis testing (Remington and Schork, 1970).

In estimation, summary statistics are computed from sample information to make a reasonable conclusion regarding the value of an unknown parameter. For example, the sample mean and standard deviation can be used to estimate a range within which it is likely that the population mean falls. This sort of estimation can be useful in developing baseline information, developing or verifying models, estimating the NPS contributions in a total maximum daily load (TMDL), or determining the load of a single NPS runoff event. Means, percentages, proportions, and variations are all parameters that can be estimated (Freund, 1973). This chapter addresses many of these parameters, and includes numerous examples.

In the broad arena of environmental management and protection, it is more the rule than the exception that monitoring and evaluation are performed for the purpose of making decisions regarding remedial and/or preventive actions. For example, confidence intervals for the mean concentration of a pollutant before and after an NPS control program might tell very little about the success of the program. Granted, if the means were so different that the confidence intervals had no overlap it would be reasonable to state that a change has occurred (note that a cause and effect relationship can not be established with water quality data alone). However, if the confidence intervals overlapped, a very likely scenario, what could be said about changes in water quality? Statistical tests would need to be performed which establish rules for deciding whether a change occurred and whether this change is significant. This type of statistical analysis falls under the general category of hypothesis testing.

A.3 Characteristics of Environmental Data

The statistical methods selected 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 10th or 15th percentile during summer months over the same time period might have a greater impact on subsequent management decisions. Environmental managers and data analysts must collectively determine which statistical methods will result in the most useful information for decision makers. As a result, any data analysis protocol should include a suite of techniques to meet the monitoring and evaluation objectives.

The selection of statistical methods is further complicated by the presence of data attributes (Harcum, 1990). Two main types of attributes are commonly found in environmental monitoring: 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, and

VIII. DATA ANALYSIS

This chapter provides detailed information on the statistical analysis of environmental monitoring data. The first section of this chapter is intended for both the manager and data analyst. The goal of this section 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 data analysis and can be used to guide the reader in selecting additional portions of the chapter for more detailed reading.

The organization of the remainder of this chapter follows the normal progression of a typical statistical analysis: exploratory analysis, summary statistics, trend testing, extreme events, and advanced procedures. The final section of this chapter has been specifically designed to address special issues related to environmental monitoring such as how to treat missing values, outliers, and censored data (e.g., "less-thans").

A. Introduction

A.1 Purpose of Data Analysis

Data analysis begins at the monitoring program design phase. Those responsible for monitoring should specifically identify the reasons for monitoring and the methods to be used for analyzing the collected data. Preliminary assessment of the specific reasons for monitoring (i.e., what will the project leaders try to say based on the data collected?) is critical to the success of any monitoring and evaluation program. Chapter II provides an overview of commonly encountered monitoring objectives.

Once goals and objectives have been clearly established, data analysis approaches can be explored. Usually, monitoring data by themselves do not provide the necessary information to support decision making. This chapter provides guidance for selecting and applying statistical methods. By selecting and applying suitable methods, the data analyst responsible for evaluating the data can prevent the "data rich–information poor syndrome" described by Ward et al. (1986).

A.2 Estimation and Hypothesis Testing

Ponce (1980b, p. 6) states that monitoring objectives "should be specific statements of measurable results to be achieved within a stated time period." He notes that monitoring objectives "should be specific enough so that the [data analyst] can convert them into statistical hypotheses which can be tested with the data obtained from the water quality monitoring program." MacDonald et al. (1991) state that "statistics [provide] the scientific basis and procedures for studying numerical data and making inferences about a population based on a sample of that population (Mendenhall, 1971; Sokal and Rohlf, 1981)." There are two major types of

CHAPTER 8

DATA ANALYSIS

Table VII-10e. Selected biomonitoring program components, Vermont Department of Environmental Conservation (Vermont DEC). [Contact: Steve Fiske, 802-244-4520]

	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" x 8" (460 x 200 mm) 500 μm mesh	Rectangular kick net 18" x 8" (460 x 200 mm) 500 μm mesh	electro-shocking
Sampling Method	2 min timed composite; 30 sec in ea 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 vert 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
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 info ave. stream velocity, depth	qualitative or quantitative transect method; depth, velocity, substrate
Comments	also record temperature, pH, alkalinity, cond and characterization of site, elevation, DO; also will be looking 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		have modified the IBI to fit Vermont's wadable streams

Table VII-10c. Selected biomonitoring program components, Montana Department of Health and Environmental Sciences (Montana DHES). [Contact: Steve Trallis/Bob Bukantis, 406-444-2406]

	Macroinvertebrates
Region of State	entire
Site Selection	riffles
Sample Gear	D-frame (12") 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 VII-10d. Selected biomonitoring program components, North Dakota Department of Environmental Health (North Dakota DEH). [Contact: Mike Ell, 701-221-5214]

	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 next year	in development

Table VII-10b. Selected biomonitoring program components. Florida Department of Environmental Protection (Florida DEP). [Contact: Russ Frydenborg, 904-487-2245]

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 VII-10a. Selected biomonitoring program components, Delaware Department of Natural Resources and Environmental Conservation (DNREC). [Contact: John Maxted, 302-739-4590]

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 & 700-900 micron mesh), Sieve bucket (600 micron mesh), 70% ethanol, storage containers (1-2 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.

Table VII-9. 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 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.

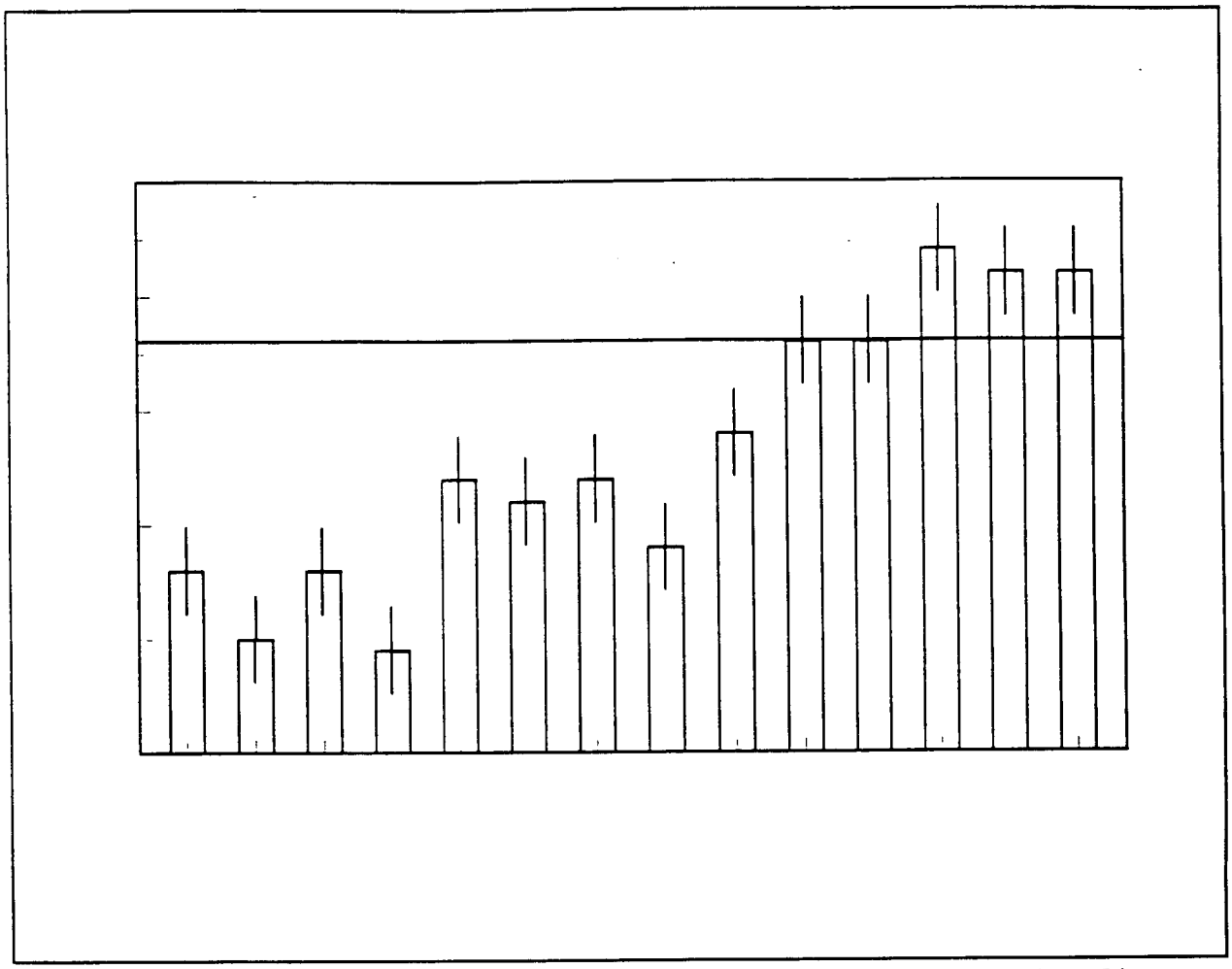


Figure VII-6. Trend in biological condition (based on the multimetric ICI) of the Cuyahoga River at Independence, Ohio (river mile 15.6) (Ohio EPA 1992).

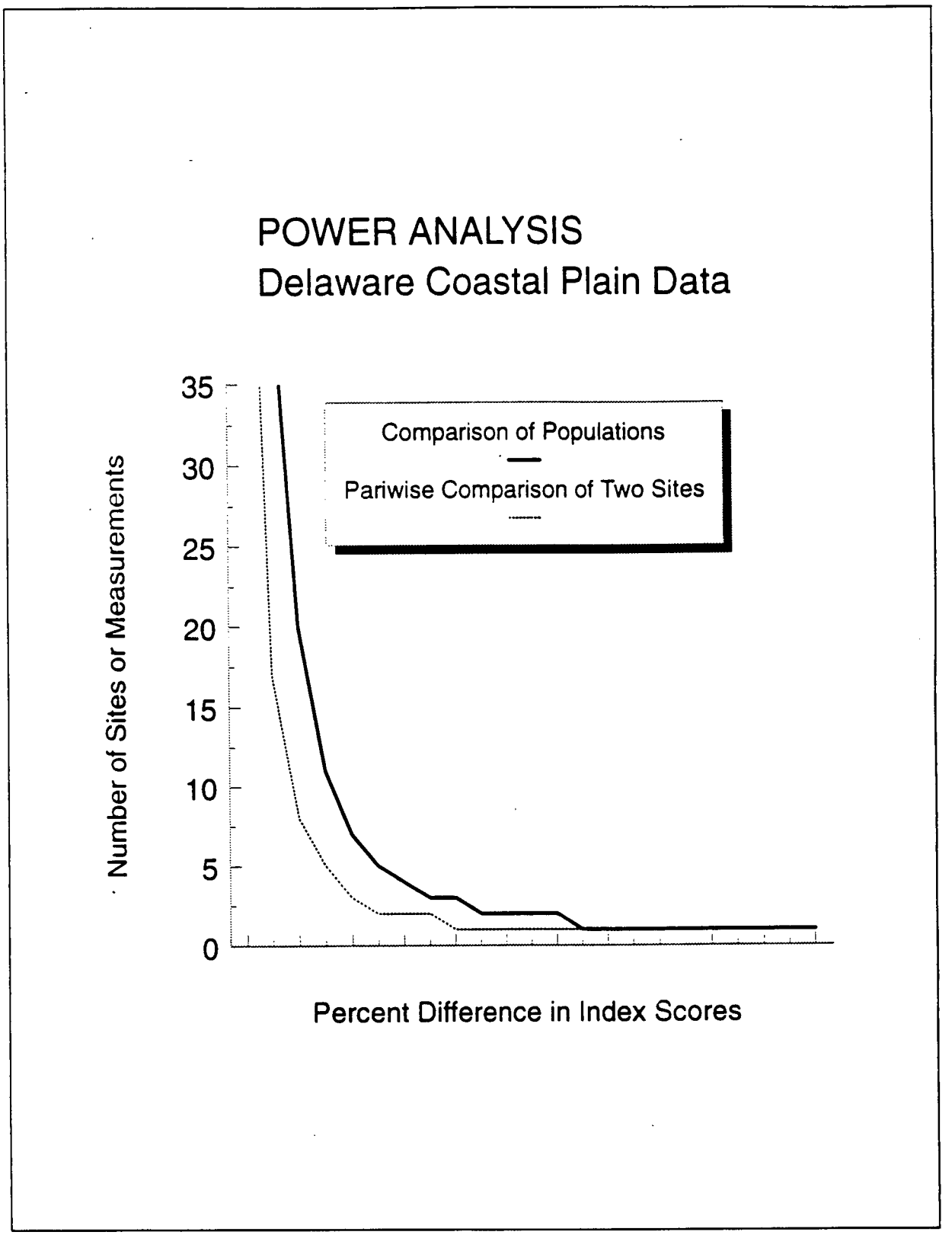
gradient streams, and benthic macroinvertebrates from multihabitats in low gradient streams.

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 VII-6 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.

J. Overview of Some State Programs

Biological monitoring programs of five states are highlighted as an overview of technical program components (Table VII-9). 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 (MACS 1993 (draft)). They were able to determine that the most appropriate method for obtaining a representative 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 VII-10a). The Florida Department of Environmental Protection has adopted this method for its nonpoint source monitoring program (Table VII-10b) 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. They have adapted a traveling kick method (Table VII-10c) for macroinvertebrates and is developing reference conditions. The health of the fish assemblage is the primary biological monitoring indicator for the North Dakota Department of Environmental Health (Table VII-10d). They are 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 VII-10e): fish, benthic macroinvertebrates from riffles in high



Box VII-4. (continued)

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, in reality, does not exist. These false positives are called Type I errors. It is traditional to accept a five percent probability of false positives as occurring (see page VIII-47 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 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: (i) Taxa richness, (ii) EPT index, (iii) percent EPT abundance, (iv) percent Chironomidae abundance, (v) percent dominant family, and (vi) 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 we are characterizing the site. The power analysis estimated 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).

Box VII-4.

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; Box VII-4). 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 alternatively, one wants to detect a relatively small change in assessment score over time with a high degree of confidence, then a greater number of monitoring sites within a strata (i.e., stream order, urban versus agricultural) will need to be sampled.

The Prince George's County case study reviewed earlier illustrates how this power analysis example 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 (Text Box VII-4) 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 six 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 lead 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

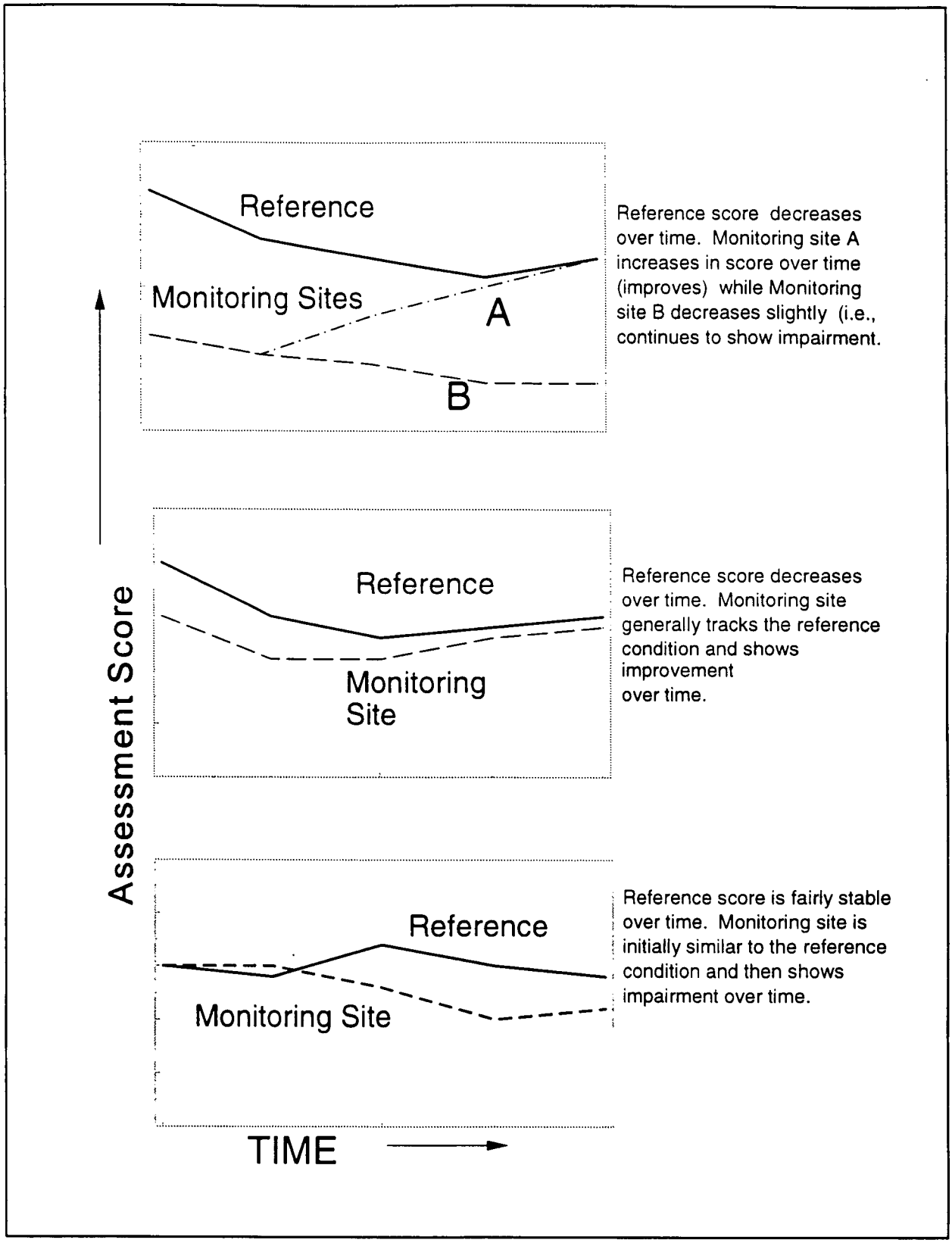


Figure VII-5. Some trends that might be observed during the course of a biological monitoring program.

a biological monitoring nonpoint source program is accomplished, in part, by using a comparative analysis of reference site and monitoring site data. Figure VII-5 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 Chapter VIII, Section 2 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. 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 VII-5). 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 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 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.

sites. These data allow estimation of annual variation and trends in the biological characteristics of the reference sites. These estimates are critical for determining biological status of test sites.

Forty-one watersheds have been identified to address status and trends for the county. Site 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 six to ten 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. 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 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, D.C. 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.

I. 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. Chapter VIII, Section 2 outlines statistical considerations regarding this issue and Figure VIII-33 illustrates several common types of data patterns that may 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 to have 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

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 their 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 lay-person involvement.
- Evaluate the effectiveness of environmental management and mitigation activities.

The program design includes both probabalistic and targeted elements that will allow specific questions to be addressed at three spatial scales: county-wide, watershed-wide, and stream-specific. There are approximately 100 sampling locations that have been selected for the initial year of monitoring (1995). In two separate sampling periods (early spring and fall), there will be approximately 50 probability sites, 20 known problem sites, 20 reference sites, and another 15 sites that will be 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.

Box VII-3.

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

geographic area of concern. NPDES permits, urban storm water 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. Integrated Network Design

Integrated network design consists of integrating multiple monitoring subprograms to effectively meet various monitoring objectives and improve the applicatibility of data. An emerging biological monitoring program in Prince George's County, Maryland, is used here at a useful example demonstrating the design components reviewed above (Box VII-3).

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. Judgement of impairment or nonimpairment results from comparison to reference conditions (see Section VII-D). Two major components of this program are volunteer monitoring of selected sites and nonvolunteer 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 nonvolunteer 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 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 if 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

Table VII-8. Waterbody stratification hierarchy.

Population of Interest	Streams and Rivers [1]	Lakes [2,3]	Reservoirs [4,2]	Estuaries [5]	Wetlands [6]	Groundwater [7]
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 (homogenous, heterogenous, isotropic, anisotropic)
Habitats within SUs	macrohabitat (pool/riffle; shorezone vegetation; submerged aquatic macrophytes)		longitudinal zone (riverine; transitional; lacustrine; tailwaters [may 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	

[1] Frissell et al. 1986; [2] Gerritsen et al. 1994 [Draft]; [3] Wetzel 1975; [4] Thornton et al. 1990; [5] Day et al. 1989; [6] Cowardin et al. 1979; [7] Heath 1989.

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 should be stratified by stream order (first, second, third, etc.; Strahler [1957]), or specific sections of a bay or lake may be identified as separate strata.

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

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

As 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 among-group variance (Gilbert 1987). Table VII-8 summarizes a waterbody stratification hierarchy for streams and rivers, lakes, reservoirs, estuaries, wetlands, and groundwater. 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 EMAP-Estuaries program 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, post-stratification of the dataset may be required. If stratification to the level of microhabitat is needed, it may indicate that the sampling and analysis methods used at higher levels are inappropriate or inadequate.

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, multi-stage; Chapter VI, Section 3). There are three major steps involved in selecting sampling sites using a probabilistic design:

(1) Identify the level of site classification. Monitoring program objectives will dictate the geographic extent at which monitoring is to be done, for instance state-wide, 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 may 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 VII-8). If there are clear clusters of differing land use among

Table VII-7. Comparison of probabalistic and targeted monitoring designs.

	Advantages	Disadvantages
Probabilistic Design	• 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 among-strata variance from within-strata variance.	• 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 that 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.

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 may or may 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 be whole stream, river, 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.

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 VII-7). 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.

Network Design refers to the array, or network, of sampling sites selected for a monitoring program; usually taking one of two forms:

- *Probabilistic Design* Network that includes sampling sites selected randomly in order 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 are 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.

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.

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).

2.1. 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).

2.2. 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.

H. 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, thereby 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

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 (MACS, 1993 [draft]). 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).

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 20-40 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)

Example of transforming benthic macroinvertebrate data into biological metrics

Raw data

Taxa	Number	FFG	TV
Oligochaeta	2	col	8
Valvata	4	scr	6
Fossaria	3	col	6
Gyraulus parvus	15	scr	8
Pisidium	32	fil	8
Hydracarina	1	pre	6
Caecidotea	6	col	6
Hyallela azteca	35	col	8
Arthroplea	2	fil	3
Ameletus	34	col	0
Amphinemura	1	shr	3
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
Culicodes	1	pre	2
Similium	40	fil	6
Prosimilium	4	fil	2
Chrysogaster	1	col	10
Molophilus	1	shr	4
Pseudolimnophia	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**		

Calculated Metrics

Metrics	Values
Taxa Richness	30
Hilsenhoff Biotic Index	5.7
EPT Index	5
Percent Dominance	17
Percent Pisidium	14
Percent Simuliidae	19
Percent Isopoda	3
Percent Diptera	38
Percent EPT	18
Percent Filterers (fil)	33
Percent Collectors (col)	45

FFG funcitonal feeding group
TV tolerance value
SCR scrapers
PRE predators
SHR shredders
FIL filterers
COL collectors

Box VII-2. (continued)

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 (MACS 1993 [draft]) for low gradient streams. The eleven metrics calculated from the data 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.

Box VII-2.

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 200-300 organisms. 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 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

Table VII-6. 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		

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 montane 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.

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 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

Table VII-4. (continued)

	Alternative benthic metrics	Metric Category*	ICI[a]	RBP[b]	RBP[c]	RBP[d] ID	RBP[d] OR	RBP[d] WA	B-IBI[e]
21.	Shannon-Weiner Diversity Index	A					X		
	Equitability	A				X			
	Index of Community Integrity	A				X			

* 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, Kansas et al. (1992).

Table VII-5. 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		

Table VII-4. 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 Abund.	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				
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			

Table VII-3. (continued)

	Alternative IBI metrics	Metric Category	Midwest	New England	Ontario	Central Appalachia	Colorado Front Range	Western Oregon	Sacramento-San Joaquin	Wisconsin
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 catchable wild trout	D							X	
10.	Number of individuals	A	X		X	X	X	X	X	X[a]
	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.

[a]Metric suggested by Moyle or Hughes as a provisional replacement metric in small western salmonid streams.
[b]Excluding individuals of tolerant species.

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

Table VII-3. Fish IBI 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					
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		

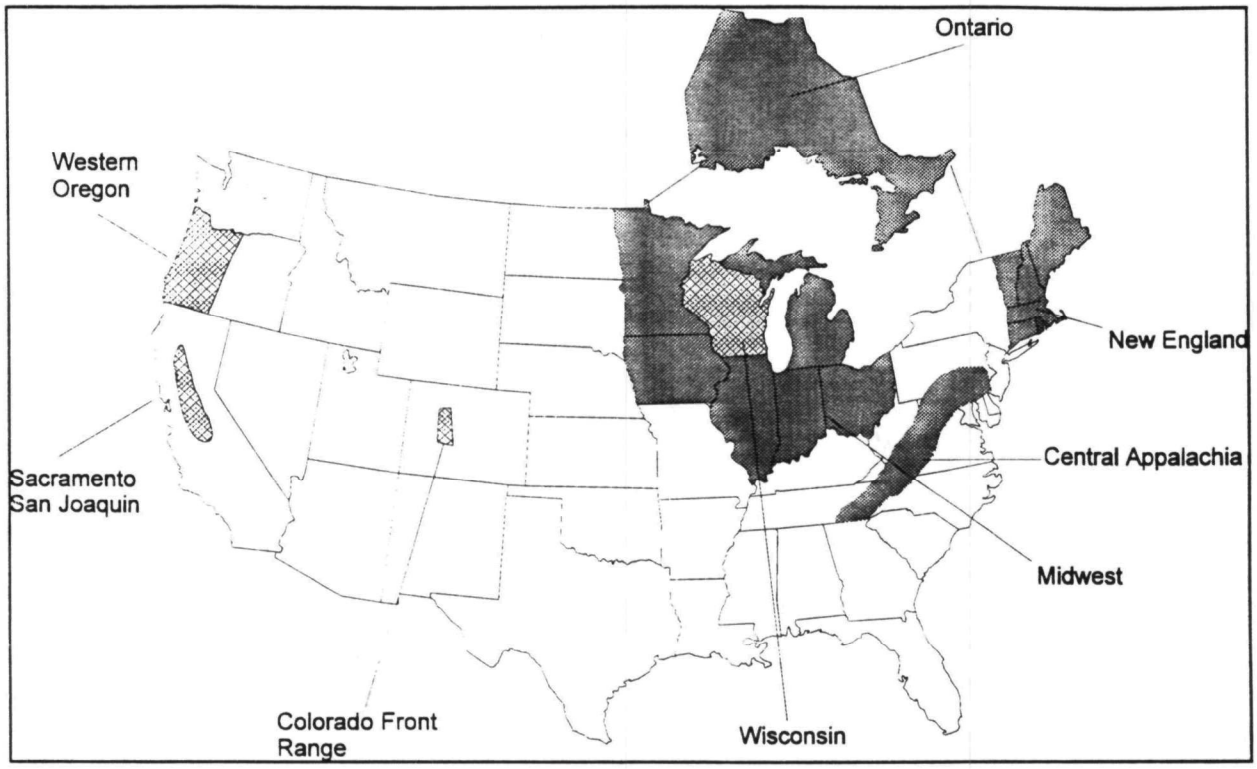


Figure VII-4. Regions in which various fish IBI metrics (see Table VII-4) have been used.

correlation between the two for a particular ecoregion. Obviously, the more correlative data that are collected, the more useful they will become 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 (e.g., 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 and among stations or among data from a single station at different times.

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

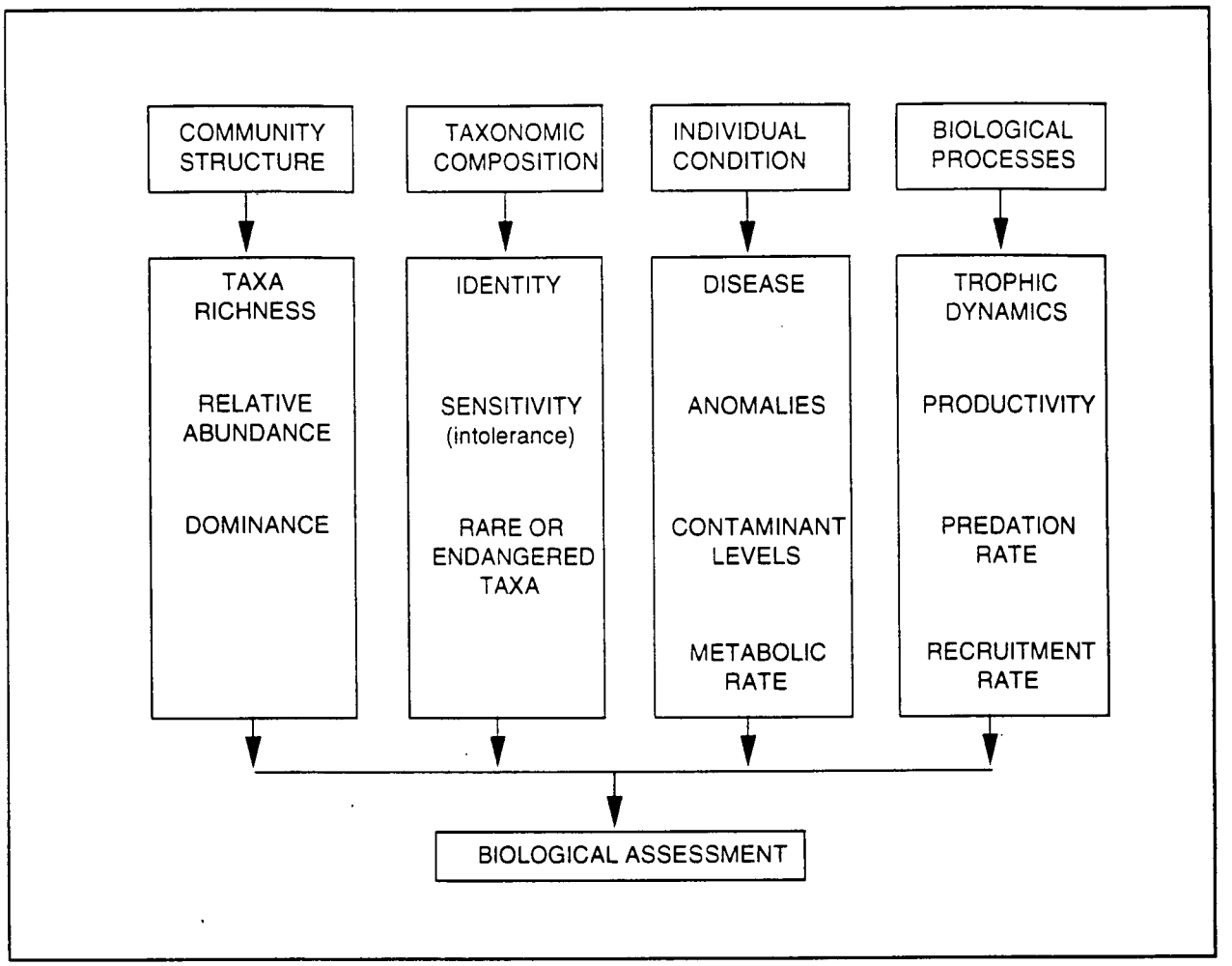


Figure VII-3. Organizational structure of attributes that can serve as metrics.

G. 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

Figure VII-3 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 (A) 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 its diversity. 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 (B) refers to the types of taxa in the sample. *Sensitivity* 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 (C) is more easily measured with 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.

Biological processes (D) 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 VII-4), primarily for fish and benthos. Summaries of those used have recently been presented in tabular form (Gibson et al., 1994; Barbour et al., 1995) and are reproduced in Tables VII-3 and VII-4. The letter (A, B, C, D) following each metric or group of metrics in Tables VII-3 and VII-4 indicates to which category of metric listed above it (or they) belong. Examples of metrics that have been tested and have had scoring criteria established are those for the montane region of Wyoming (Table VII-5) and the plains streams of Florida (Table VII-6). Box VII-2 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.

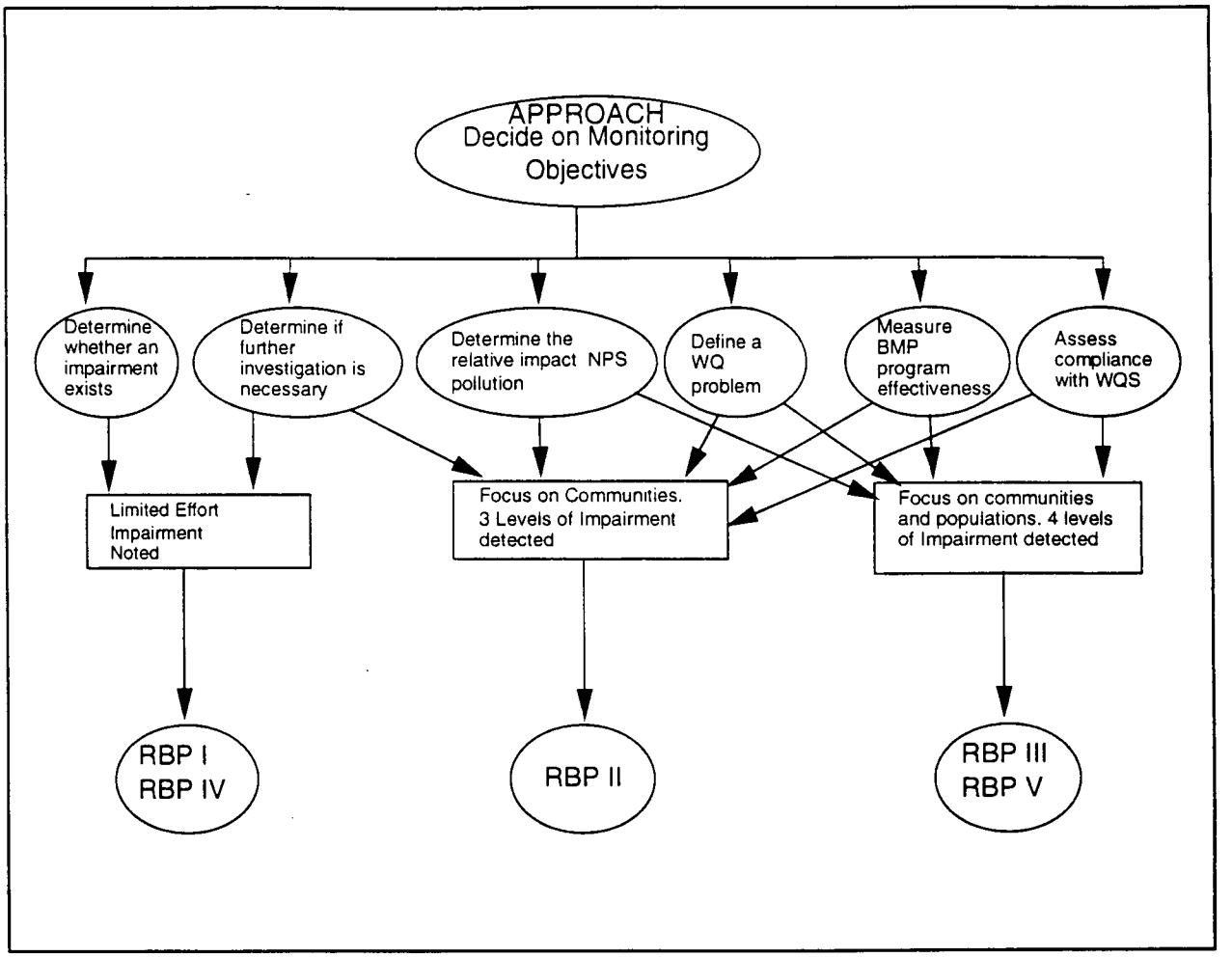


Figure VII-2. Selection and application of the different tiers of RBP depend on monitoring objectives.

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 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.

Table VII-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

* Mayflies, stoneflies, and caddisflies, respectively.

approaches and are useful for setting priorities for more intensive study. RBP IV is a screening technique used to survey persons knowledgeable about the fish in an area. It is not described here; 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 VII-2). 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).

F. 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 data collected during a biological 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,

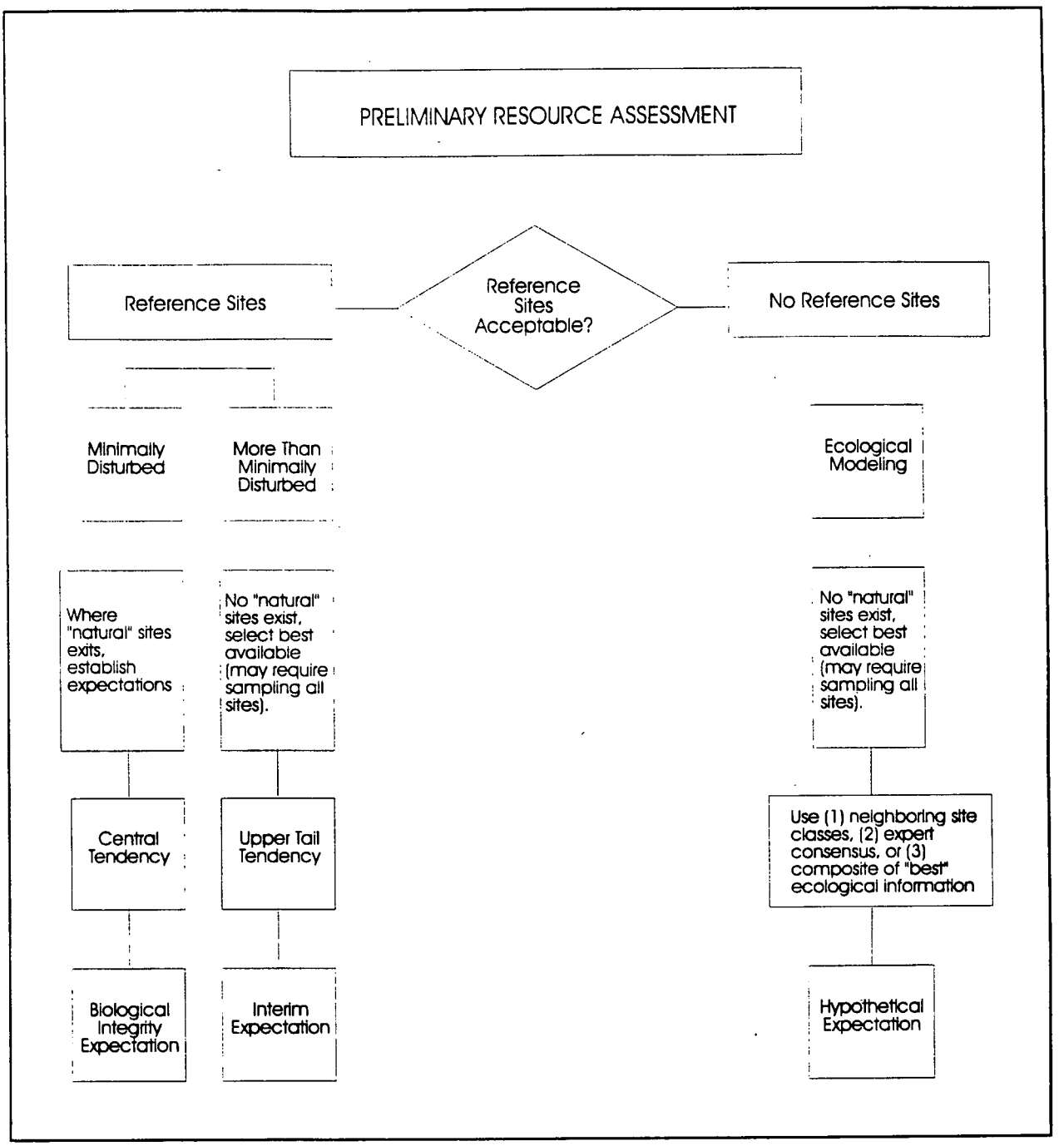


Figure VII-1. Approach to establishing reference conditions (Gibson et al. 1994).

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 VII-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

would not be considered candidates for reference sites because they are not representative of the norm of the region.

Reference conditions are established for ecologically homogeneous regions. Ecoregion boundaries delineated by Omernik (1987) are appropriate in many cases for the establishment of regional boundaries for reference condition applicability. Omernik used the perceived patterns of four combined causal and integrative factors—land use, land surface form, potential natural vegetation, and soils—to delineate 76 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 may 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* (1994) 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 judgement 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 VII-1). If the use of reference sites is not an alternative because "natural" sites do not exist and "minimally impaired sites" are unacceptable, then 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 or temporary until more reliable information is obtained.

E. 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.

region created from information gathered at multiple reference sites. The reference condition accounts for natural variability in the biological communities within a region because it is established using data from reference sites from different streams in the region. Between-region differences in reference conditions can be large, so reference conditions established for a particular ecoregion should be used for the interpretation of data from that region. For instance, reference conditions for a mountainous region should not be used as a basis for analyzing sites monitored in a lowland plains area.

Reference site - a specific location on a waterbody that is minimally impaired and is representative of the expected ecological integrity of other localities on the same waterbody or nearby waterbodies

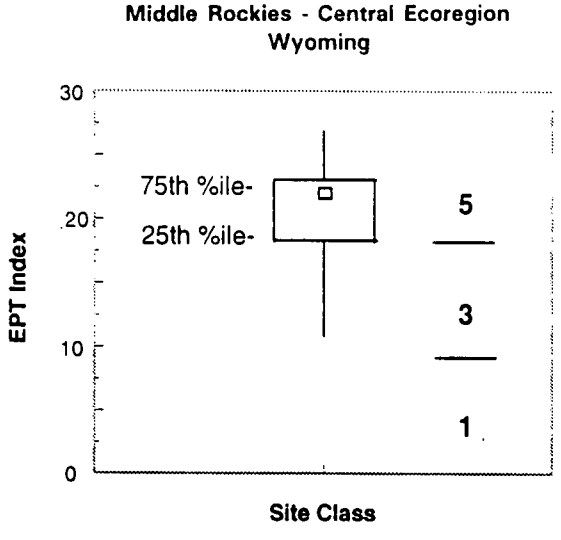
Reference condition - a set of selected measurements or conditions of minimally impaired waterbodies characteristic of a

The overall goal of establishing a reference condition is to describe the natural potential of the biota for the waterbody and habitat types characteristic of the region, independent of the extent of human degradation. Reference conditions account for environmental variability and thus minimize background "noise" as a factor when making comparisons among reference and monitored sites. As described above, comparison of the biological communities and physical habitats at monitored sites to the appropriate regional reference condition indicates the degree and possible cause of biological impairment. For instance, if the habitat at a monitored site appears to be in good condition but the benthic macroinvertebrate or fish species differ from those of the reference condition, poor water quality resulting from point or nonpoint source pollution might be responsible. If habitat degradation is noted and the biological condition is impaired, physical habitat restoration might be necessary to improve the biological condition. Refer to the summary of the *Process of Metric Selection, Validation, and Development of Reference Conditions* in Box VII-1 (p. VII-11) for a step-by-step explanation of the process of developing reference conditions.

A reference condition can be formulated from historical data sets, from extrapolation from ecological or other information, or from multiple reference sites (Gibson et al., 1994). However, the availability of candidate reference sites, the preferred approach, dictates which method might be most appropriate. There are two primary criteria for selecting candidate reference sites: minimal impairment and representativeness (Gibson et al., 1994). The minimal impairment criterion acknowledges that pristine sites are nonexistent in a region and are not likely to become available. Sites with the least amount of impairment, therefore, are used as reference sites. The representativeness criterion refers to the requirement that reference sites be representative of a particular region or class of sites. Surface waters that are unique in some way or unusual within the particular region

Step 5. ***Scoring of Metrics.*** Develop bioassessment scoring criteria for each of the core metrics, within each site class. Using scoring criteria, normalize metrics.

Metrics vary in their scale; they are integers, percentages, and ratios. Prior to developing an integrated index for assessing biological condition, it is necessary to normalize the core metrics via a transformation to unitless scores. This is accomplished by selecting the lower quartile of the range in reference metric values, which assumes that only the upper 75% of the reference values are representative of natural conditions for the site class. Therefore, the upper 75% of the values are given the highest score, and the remainder of the range is bisected to give progressively lower scores. The figure below demonstrates this approach.



Assignment of unitless scores (5, 3, 1)
to reference metric values

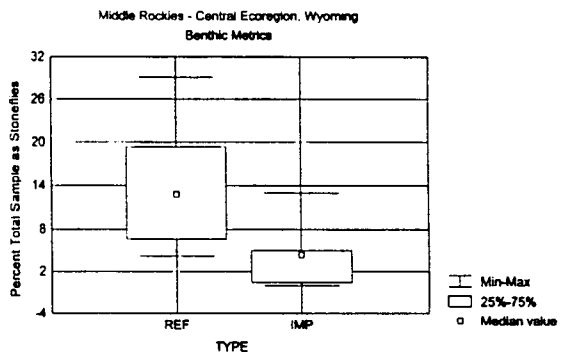
Step 6. ***Index Development.*** Following development of scoring criteria for all metrics, score metric values from all sites, reference and impaired, and sum bioassessment points. The reference condition is the distribution of total bioassessment scores from multiple reference site representing an individual site class.

The index is a means of integrating information from the various measures of biological attributes (or metrics). In monitoring, the "tracking" of an index value that integrates all of the core metrics will enable an interpretation of improvement or degradation of the biological assemblage. Aggregation of metric scores simplifies management and decisionmaking so that a single index value is used to determine whether action is needed.

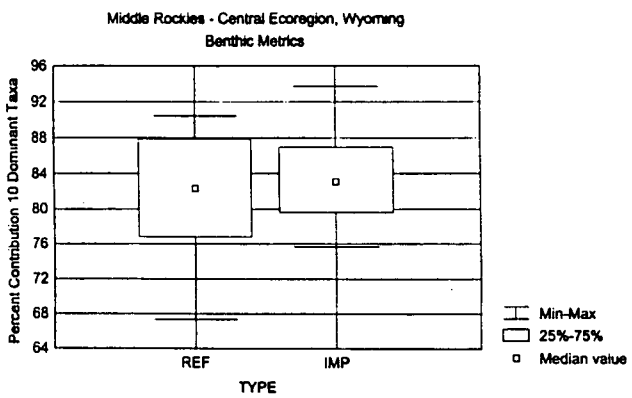
Box VII-1. (continued)

Step 4. ***Determination of Core Metrics***. Calculate metrics based on biosurvey data (see Box VII-2, page VII-27). 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. 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 distributions, then the metric can be considered to be a strong discriminator between reference and impaired conditions. The following two figures, (a) and (b), 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

Box VII-1. (continued)

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

Step 1. ***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.

Step 2. ***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 text (p. VII-10) 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 landuse maps and confirmed through site reconnaissance.

Step 3. ***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 non-impairment, 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 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 range 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.

Box VII-1.

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 that section (C.1) for a description of what is involved.

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

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.

Metric - an enumerated or calculated term representing some aspect of biological assemblage structure, function, or other measurable feature and which changes in a predictable way in response to environmental (including human) influences

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. Box VII-1 describes the development of metrics and associated reference conditions in a step-by-step manner.

D. 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 closedly 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

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 are chosen that respond differently to environmental factors, 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. 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 unimpaired status, 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. It also requires the greatest amount of time, specialized expertise, and field and laboratory effort to perform, largely because in order to conduct a composited reference site biological assessment and monitoring program, it is necessary to initially establish a reference database for the region in which monitoring will be conducted.

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, and thus, 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 variation. This approach can include more than one index period (as in the Florida nonpoint source program) but is usually based on a single index period established to optimize

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 their nonpoint source pollution control program.

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. It is useful for the detection of changes due to changes in water quality, habitat quality, or land use features. A 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 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. Type 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 prior to 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

The first element of the screening-level approach, as with all biological assessment approaches, is a habitat assesement. 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. If 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 Box VII-2 (page VII-27). 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, etc. 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 or similar to that where monitoring is done.

Since a screening-level bioassessment is done without the benefit of comparison to unimpaired sites, a judgement of biological condition is made based solely on the presence or absence of indicator taxa, dominance of nuisance or sensitive taxa in the sampled habitats, or evenness of taxonomic distribution. A trained biologist will be able to determine

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 and 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: (1) the biological community varies directly with habitat quality—water quality is not the principal factor affecting the biota; (2) 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; and (3) 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 (Barbour and Stribling, 1991). 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.

C. Overview of Biological Assessment Approaches

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 condition of a site is a screening-level, or reconnaissance, bioassessment (USEPA, 1994, Plafkin et al., 1989). 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 1994).

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

STRENGTHS	LIMITATIONS
Properly developed methods, metrics, and reference conditions provide 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	Biological assessment information does not provide information on sources of pollution or cause-effect relationships
Biological community condition reflects both short- and long-term effects	There may be a lag time between pollution abatement or BMP installation and community recovery, so monitoring over time is required for trend detection
Biological assessment data can be interpreted based on regional reference conditions where reference sites for the immediate area being monitored are not available	The optimal season for biological sampling season varies regionally, and sampling during multiple seasons may be required in some areas
Bioassessments involving 2 or more organism groups at different trophic levels provide a reasonable assessment of ecosystem health	Biological assessment does not distinguish between the effects of different stressors in a system impacted by more than one stressor

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 either 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, wherein habitat structural features that 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

assessments (including reference conditions and metrics, which are discussed later). Establishing these protocols can require intensive training and a considerable level of effort, and therefore can be quite expensive. Also, detailed biological assessments require some knowledge of the aquatic communities specific to the region where the assessments are being conducted. The biological assessment methods and the means of interpreting the results from assessments are not available for many habitats and regions of the country. Establishing a biological assessment and monitoring program, therefore, can require significant amounts of time, staff, and money.

Establishment of a biological assessment program is further complicated by the fact that for a given biological community, one season may be most appropriate for sample collection. It is often not known which season is best for obtaining reliable data, or whether it might be necessary in a particular region to take samples during one optimal season per year or during multiple seasons. Annual variations in the status of biological communities at any given time of year, for instance due to the severity of winter, also complicate data interpretation. The seasonality element of a biomonitoring program is best determined with some well-designed investigative preliminary sampling.

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. As such, 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 VII-1 summarizes the strengths and limitations of the biomonitoring approach.

B. Habitat Assessment

Habitat assessment is an important component of biological assessment and monitoring. 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

Some NPS pollutant inputs such as wet-weather runoff of urban 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 determine the expected natural 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., 1994). Other levels of biological assessment involve comparisons of impacted sites to control sites, the latter being sites that are similar to monitored sites but which 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).

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 and their sources responsible for biological or water quality impairments. Chemical monitoring and toxicity tests are also necessary to design appropriate pollution control programs.

A detailed biological monitoring program requires the development of sampling and analysis protocols for biological

and excessive sediment loading—and thus provide an overall measure of the aggregate impact of the stressors. While 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. Biosurvey techniques useful for detecting aquatic life impairments and assessing their relative severity are discussed below.

A small number of factors are often key in determining the structure of a community and its response to stress, e.g., 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 and fish 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 the 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.

Natural biological communities are generally 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).

VII. BIOLOGICAL MONITORING OF AQUATIC COMMUNITIES

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.

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. Many state agencies, such as Ohio Environmental Protection Agency (EPA), 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. 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.

Community - any group of organisms of different species that co-occur in the same habitat or area

Biological survey approaches differ depending on the waterbody, i.e., stream, river, lake, estuary, or wetland. EPA is currently developing bioassessment survey methods appropriate for use in these different waterbody types.

1. Rationale and Strengths of Biological Assessment

The central purpose of assessing biological condition is to determine how well a waterbody supports aquatic life. Biological communities integrate the effects of different pollutant stressors—such as excess nutrients, toxic chemicals, increased temperature,

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 detection of changes in biological condition

CHAPTER 7

BIOLOGICAL MONITORING AND ASSESSMENT

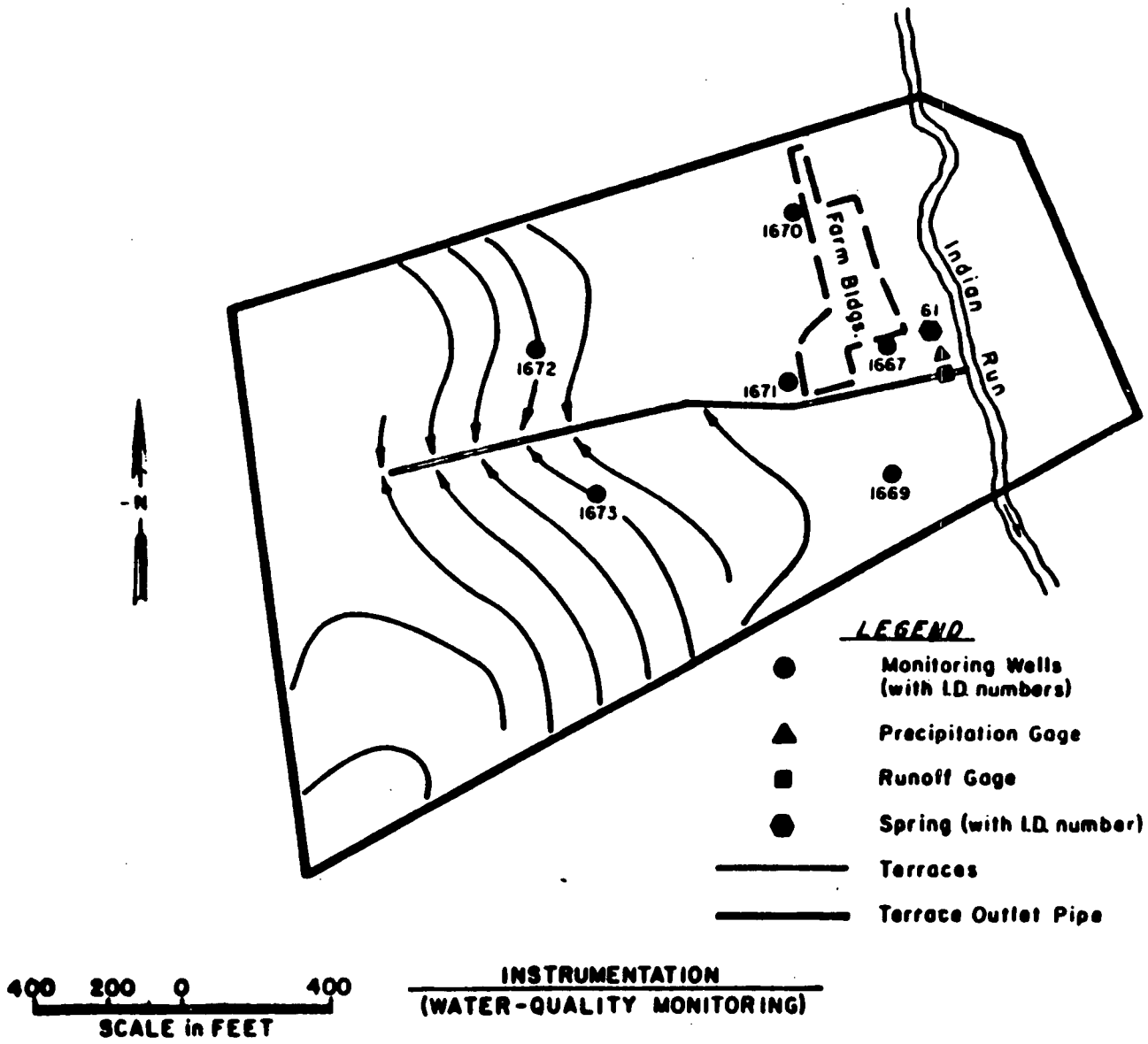


Figure VI-10. Field site 2, Lancaster County, Pennsylvania. (SOURCE: Pennsylvania RCWP Coordinating Committee, 1984)

Field Site 2:

...The planned schedule, approach, and data collection are similar to those for field site 1 (Table VI-12). The only significant difference is that data collection will concentrate on nutrients at field site 2; nutrient management BMPs are planned for this component, so current sediment and pesticide levels are not expected to change...

Table VI-12. Monitoring plan for field site 2.

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 GROUND-WATER 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

BI-WEEKLY REPORT FROM FARMER, INCLUDING INFORMATION ON:

PLOWING
PLANTING
HARVESTING
FIELD CONDITIONS
LIMING
FERTILIZING
MANURE SPREADING

(SOURCE: Pennsylvania, 1984)

The location[s] of the data collection facilities [at this field site] are shown in Figure VI-10. Field site 2 is in a carbonate area and will be monitored for two years for pre-BMP monitoring...

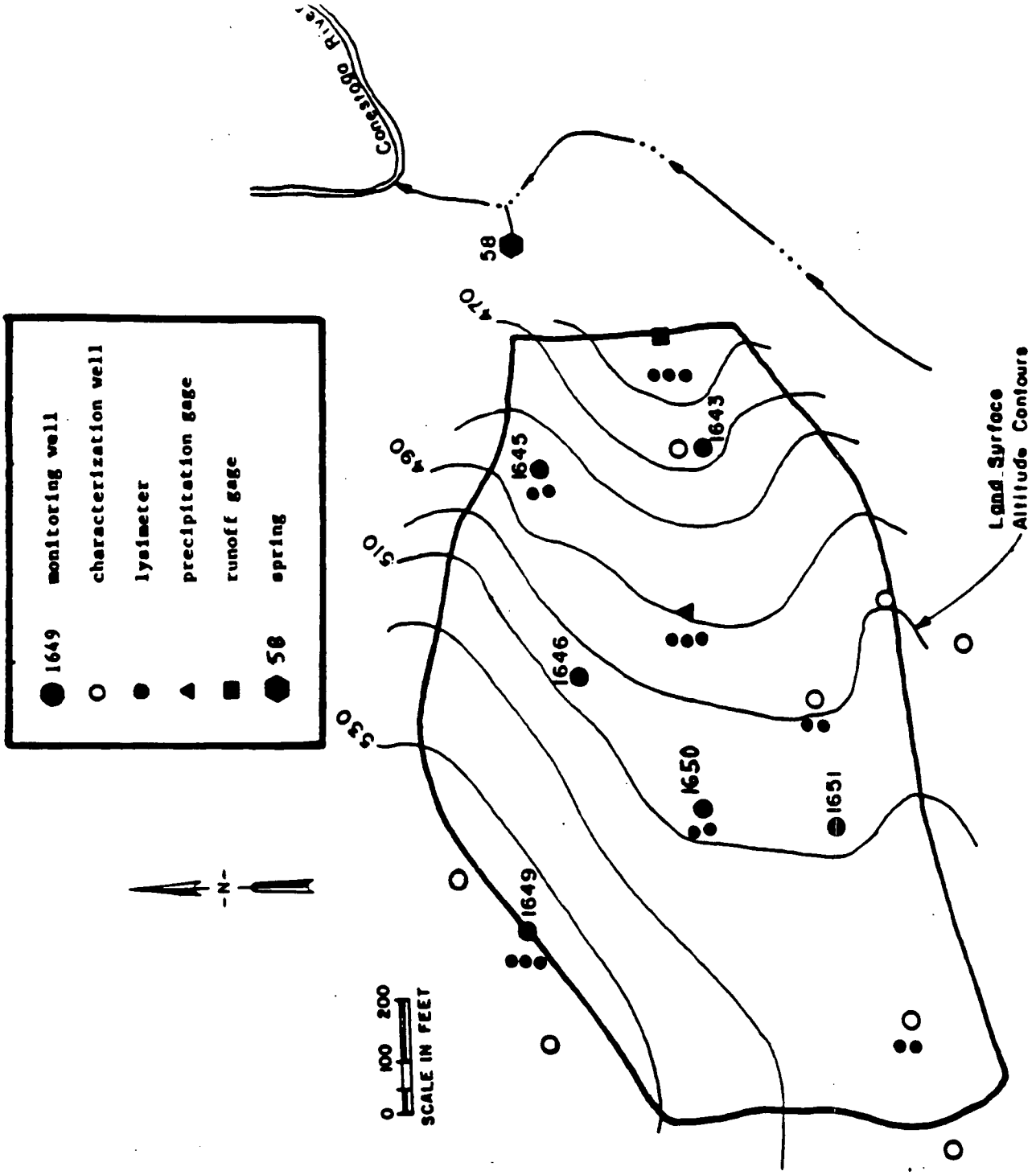


Figure VI-9. Location of monitoring facilities for field site 1. (SOURCE: Pennsylvania RCWP Coordinating Committee, 1984)

Table VI-11. Monitoring plan for field site 1.

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 GROUND-WATER 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

BI-WEEKLY REPORTS FROM FARMER, INCLUDING INFORMATION ON:

PLOWING
PLANTING
HARVESTING
FIELD CONDITIONS
LIMING
FERTILIZING
MANURE SPREADING
PESTICIDE APPLICATION

(SOURCE: Pennsylvania, 1984)

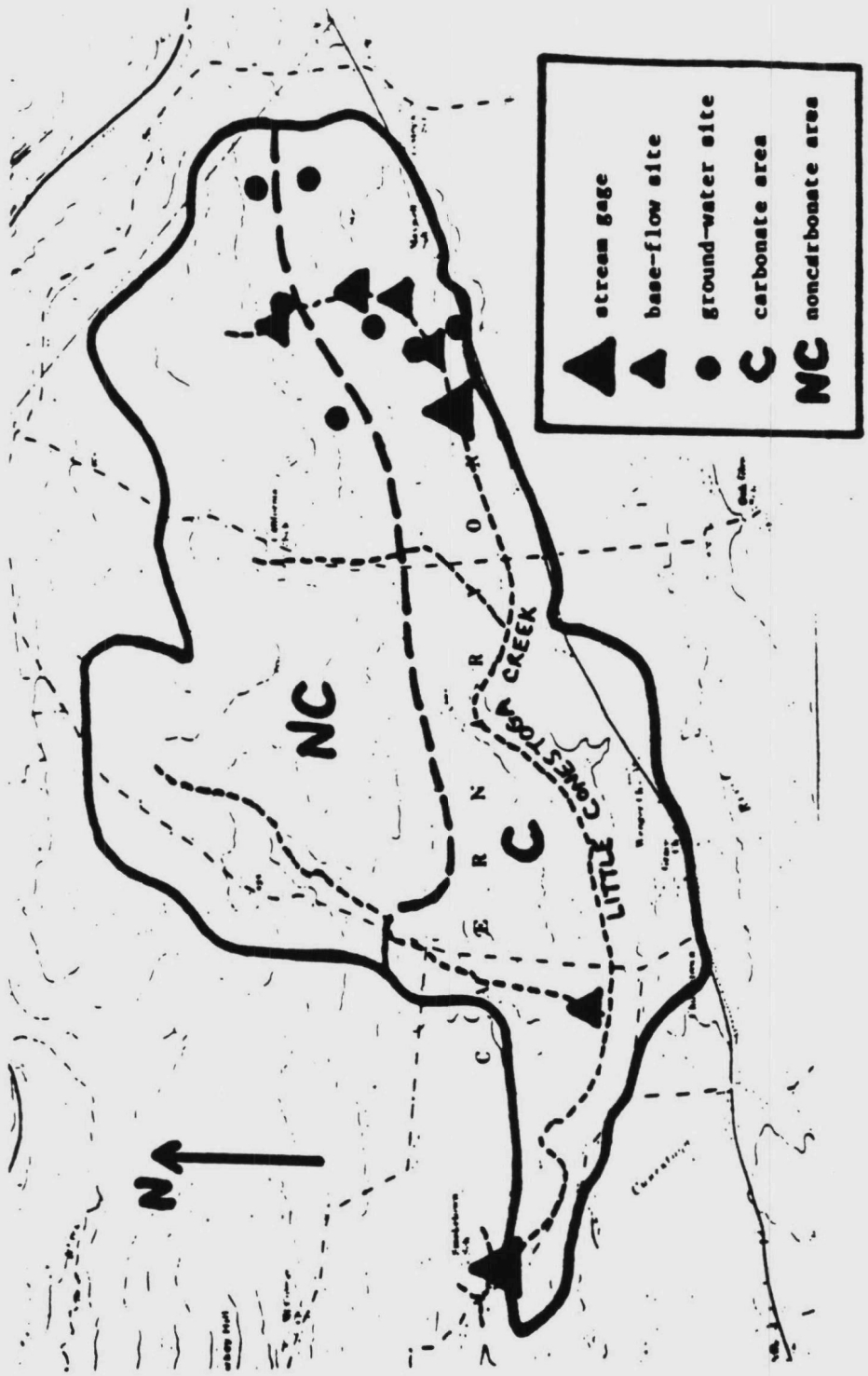


Figure VI-8. Location of monitoring facilities for small watershed sites. (SOURCE: Pennsylvania RCWP Coordinating Committee, 1984)

Table VI-10. Monitoring plan for small watershed site.

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, 1984)

...One gage, four additional base-flow sites, and all the ground water sites are located in the eastern end of the basin where BMP implementation is projected to be greatest (Figure VI-8)...

Field Site 1:

...In addition to the same data that is being collected at the small watershed site (with the exception of base-flow data), data are being obtained from lysimeters (Table VI-11)...The locations of the data collection facilities at [one of the field sites] are shown in Figure VI-9 ...

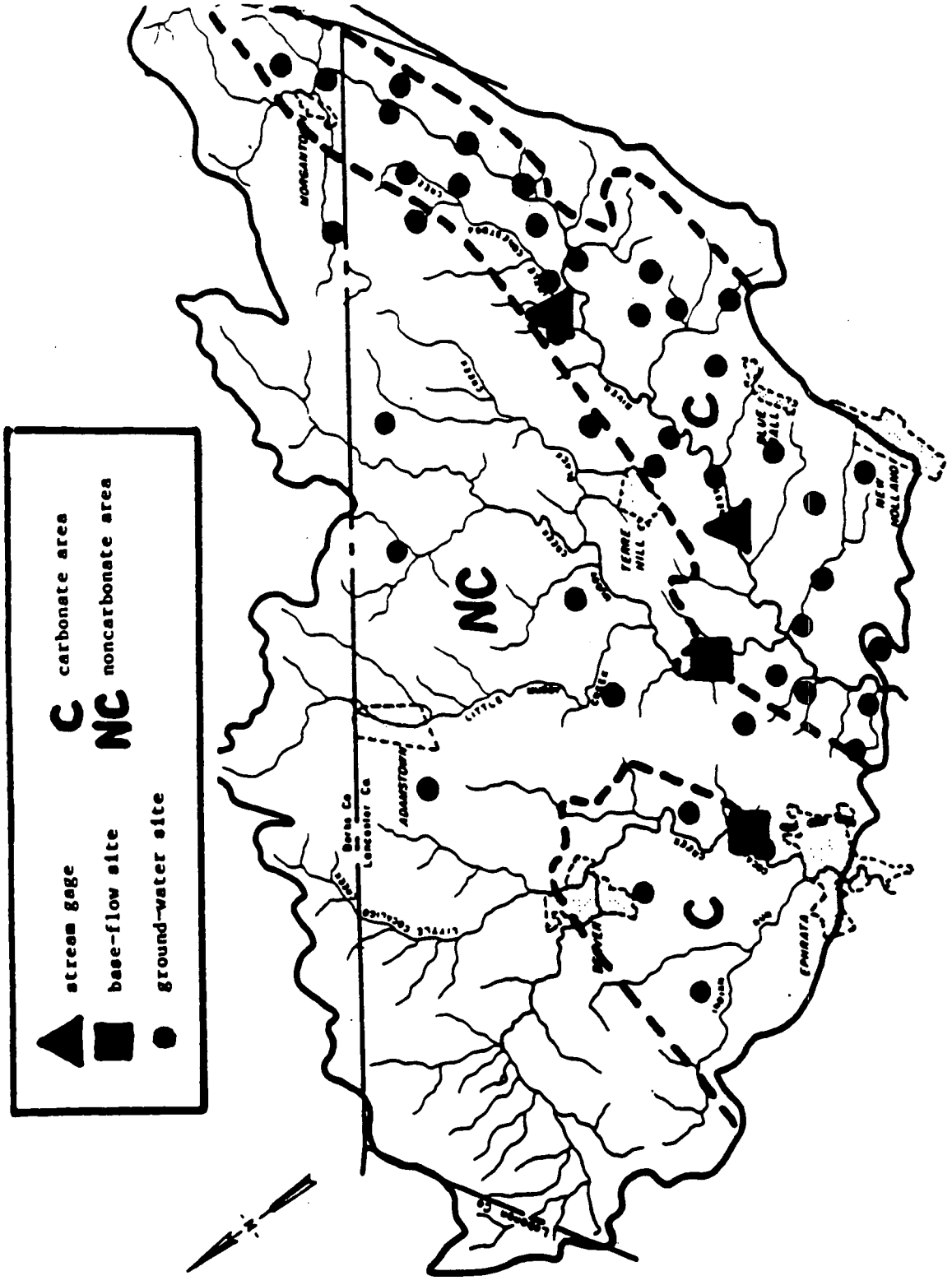


Figure VI-7. Location of monitoring facilities for regional network. (SOURCE: Pennsylvania RCWP Coordinating Committee, 1984)

Table VI-9. Monitoring plan for regional network. (SOURCE: Pennsylvania, 1984.)

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 - NUTIRENTS 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 BMP'S IMPLEMENTED.

...[The 43 ground water sites] include 42 private domestic, and farm wells and one spring....Of the 43 ground water sites, 33 are located in carbonate rocks (Figure VI-7). This difference in geology [carbonate vs. noncarbonate areas] plays an important role in the quality of water. The areas having the highest levels of nutrients, sediment, pesticides, and bacteria in water consistently coincide with areas underlain by carbonate rocks. This is 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 in the carbonate areas, the regional network, as well as the other three monitoring components, are concentrated in the carbonate areas.

Small Watershed Site:

...In addition to ground water, surface water, and precipitation data, data on nutrients in soils and manure also are being collected [at this site] (Table VI-10). Also, more detailed land-use data are being collected.

components consist of intensive monitoring on a field scale. [Both field sites are on farms].

All four components are designed to permit the comparison of water quality before and after the implementation of BMPs. As such, they include pre-BMP and post-BMP monitoring periods. Although the basic strategy is the same, each component addresses a slightly different problem (Table VI-8). All four components address the problem of nutrients in ground water and surface water; three address the sediment problem in surface water; and two address the pesticide problem in ground and surface waters...Because the single most important source of water quality problems appears to be the excess of nutrients, nutrient management BMPs are emphasized in most of the monitoring components.

Table VI-8. Specific problems to be addressed for each monitoring component. (SOURCE: Pennsylvania, 1984.)

Regional Network (188 MI^2)
PROBLEM: WHAT IS THE COMPOSITE EFFECT OF ALL IMPLEMENTED BMP'S ON SEDIMENT, NUTRIENTS, AND PESTICIDES IN STREAMS AND NUTRIENTS AND PESTICIDES IN GROUND WATER?
SMALL WATERSHED SITE-LITTLE CONESTOGA CREEK (5.8 MI^2)
PROBLEM: WHAT IS THE EFFECT OF A TYPICAL COMBINATION OF BMP'S (WITH EMPHASIS ON NUTRIENT MANAGEMENT) ON SEDIMENT AND NUTRIENTS IN THE CREEK AND NUTRIENTS IN GROUND WATER?
FILED SITE 1 - ZIMMERMAN 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 GROUND WATER?
FIELD SITE 2 - ? (<100 ACRES)
PROBLEM: WHAT IS THE EFFECT OF NUTRIENT MANAGEMENT ON NUTRIENTS IN RUNOFF AND GROUND WATER?

Regional Network:

The monitoring schedule for the regional network consists of three one-year monitoring periods: (1) a pre-BMP period; (2) an early post-BMP period; and (3) later post-BMP period (Table VI-9)... The quality of ground and surface waters, as well as precipitation, is being monitored.

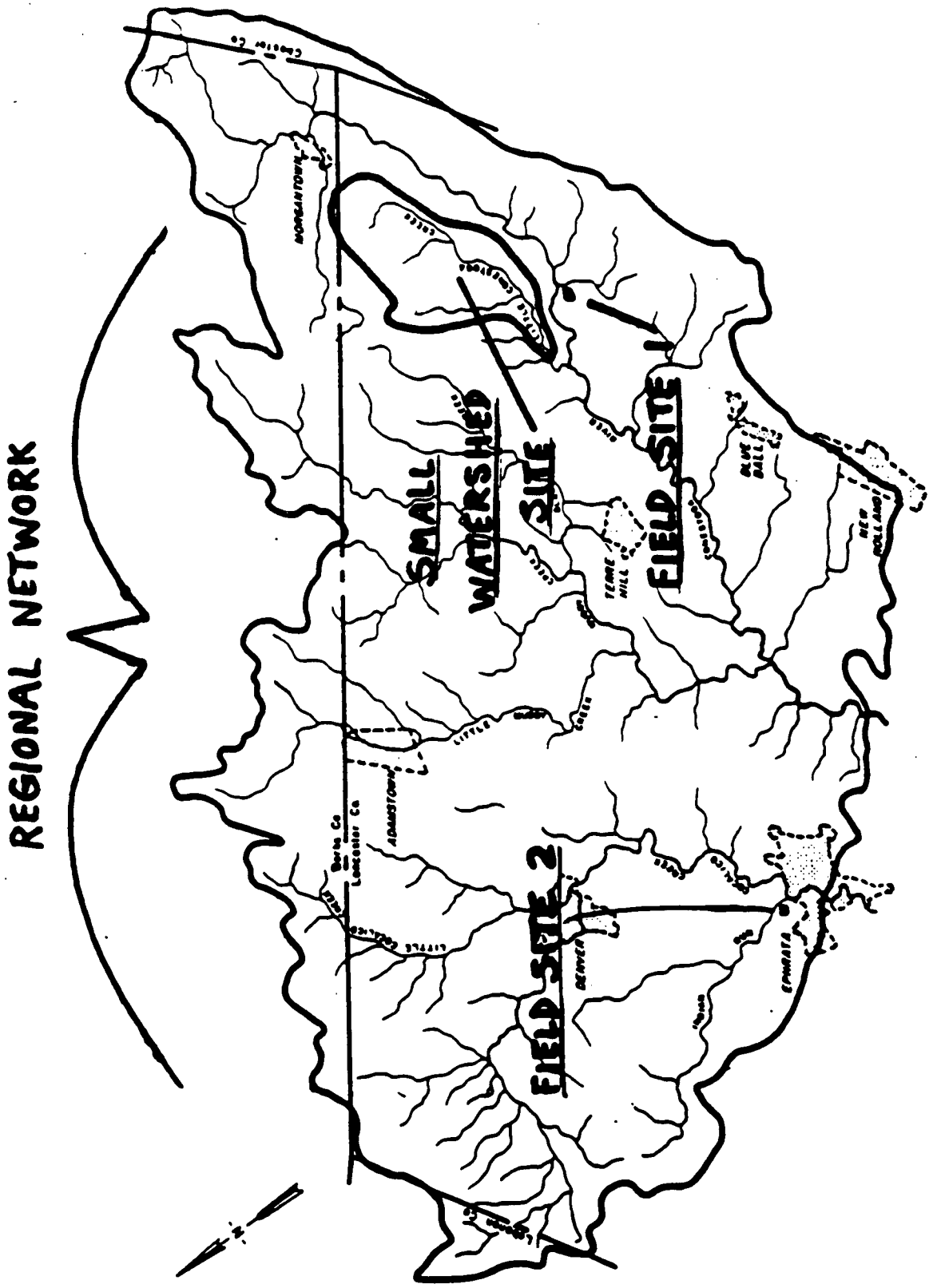


Figure VI-6. Location of Conestoga River headwaters area and four monitoring components. (SOURCE: Pennsylvania RCWP Coordinating Committee, 1984)

precipitation in up to the three rain gages, and (4) up to two stages for computing discharge.

For the City's street sampling task various vacuu, hose and gulper attachment combinations 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.

A pick-up truck was used to cary 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.

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 smaples. 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. A wand and a gulper attachment were also used.

Controls

Alternate streetsweeping in the Lake Hills and Surrey Downs basins, using the unswept basin as a control, was described earlier. The other control being evaluated is a unique, small, short-term detention basin at 148th St. The storm sewer system consists of a main trunk line parallet to the street, which is fed by short laterals that connect to catchbasins in 148th Avenue and in adjacent lands. The sewer has a complex system of gates and valves in five junction boxes that permit the storm water to be backed up into five grassy swales which serve as detention basins.

3. EXAMPLE 3: Conestoga Headwaters, Pennsylvania RCWP Project

The following are the key elements to the ground water monitoring program of the Pennsylvania Rural Clean Water Program (Pennsylvania RCWP Coordinating Committee, 1984, p. 24-26).

> *There are four components to the monitoring strategy and they include three scales or levels of monitoring (Figure VI-6). The first component, the regional network, consists of general monitoring on a regional scale and includes the entire 188 square-mile area. The second component involves more detailed monitoring in a small watershed area of about 5.8 square miles (Little Conestoga Creek basin). The third and fourth*

record the dimensions of each catch basin. Sediment volume could then be 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 are scheduled for 1981. Since December, 1980, spot checks of fifteen to twenty-five selected catch basins in each study area have been made after each significant storm event. This information, along with storm and street loading data should allow characterization of flushing and deposition within the sewerage system.

For the toxicant inventory portion of the study, stormwater runoff samples are collected as flow proportioned composites using Manning S300T automated samplers -- all teflon and glass contact surfaces -- activated by ultrasonic flowmeters, except for the volatile samples which are collected as grabs early in the storm events. Samplers and containers are cleaned between events according to USEPA protocols using "Micro" brand soap and nitric acid; the hydrochloric acid and methylene chloride rinses are not used. Deionized distilled water blanks are taken through each sampler before use and have proven to be completely clean of organic and metal contaminants. Street surface dust samples are collected as described above using a stainless steel vacuum and PVC flexible hose. No special cleaning protocal has been applied to the vacuum. Some sample contamination could occur 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 are 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 is in conjunction with the "Ecological Impacts of Stormwater Runoff in Urban Streams" project of the University of Washington.

Equipment

The equipment used by the City of Bellevue at the Lake Hills and Surrey Downs sites for flow-weighted composite stormwater monitoring consists of a Manning composite samples (S-3000), a Manning flowmeter with an ultrasonic stage sensor (UF-1100) and a 12 volt power converter. The samplers were factory modified for priority pollutant sampling. All surfaces contacting the sampler are either glass or teflon.

For the USGS sampling effort at Lake Hills, Surrey Downs and 148th St., a walk-in instrument shelter was constructed near the mouth of each catchment for housing a data recording system and sample control and collection system. A digital paper punch recorder records: (1) clock time, (2) a number code which indicates if a sample was taken by the automatic sampler, (3) accumulated

The 148th Avenue catchment contains 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 can be used as detention basins. A little over one-fourth of the catchment area is taken up by the 148th Avenue street surface.

USGS sample collection and management procedures are essentially the same at all three sites. A digital paper punch recorder records: (1) clock time, (2) a number code which indicates 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 are recorded at 5-minute intervals whenever the gage exceeds a [preset] threshold or whenever there is measurable precipitation. Precipitation is measured with tipping-bucket rain gages. Three gages are operated for the Surrey Downs catchment and two each are operated for the Lake Hills and 148th Avenue catchments. Rainfall and dry deposition quality samples are collected at one location in each catchment. Discrete runoff samples are taken during storms for defining the temporal variation of water quality during storm hydrographs. Samples are taken at a preset time interval (5 to 50 minutes) once the stage exceeds a preset threshold.

The procedures and techniques used by Bellevue for collecting composite flow and proportional stormwater runoff samples are as follows. The sampler is triggered at pre-determined increments of flow by the flowmeter (300 and 500 cubic feet, the former to obtain more subsamples when small events were expected). The flowmeters use an ultrasonic transducer to sense relative stage. Stage is 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 is programmed from a stage/discharge rating developed by the USGS. Storm samples are removed from the samplers as soon as possible after storms, typically within two or three hours. Samples are kept on ice until pH, conductivity and turbidity are measured in-house. Subsamples are 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 (the width of the gulper) from one side of the street to the other (curb-to-curb). In heavily traveled streets wher traffic was a problem, some subsamples consisted of two separate half-street strips (curb-to-crown).

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

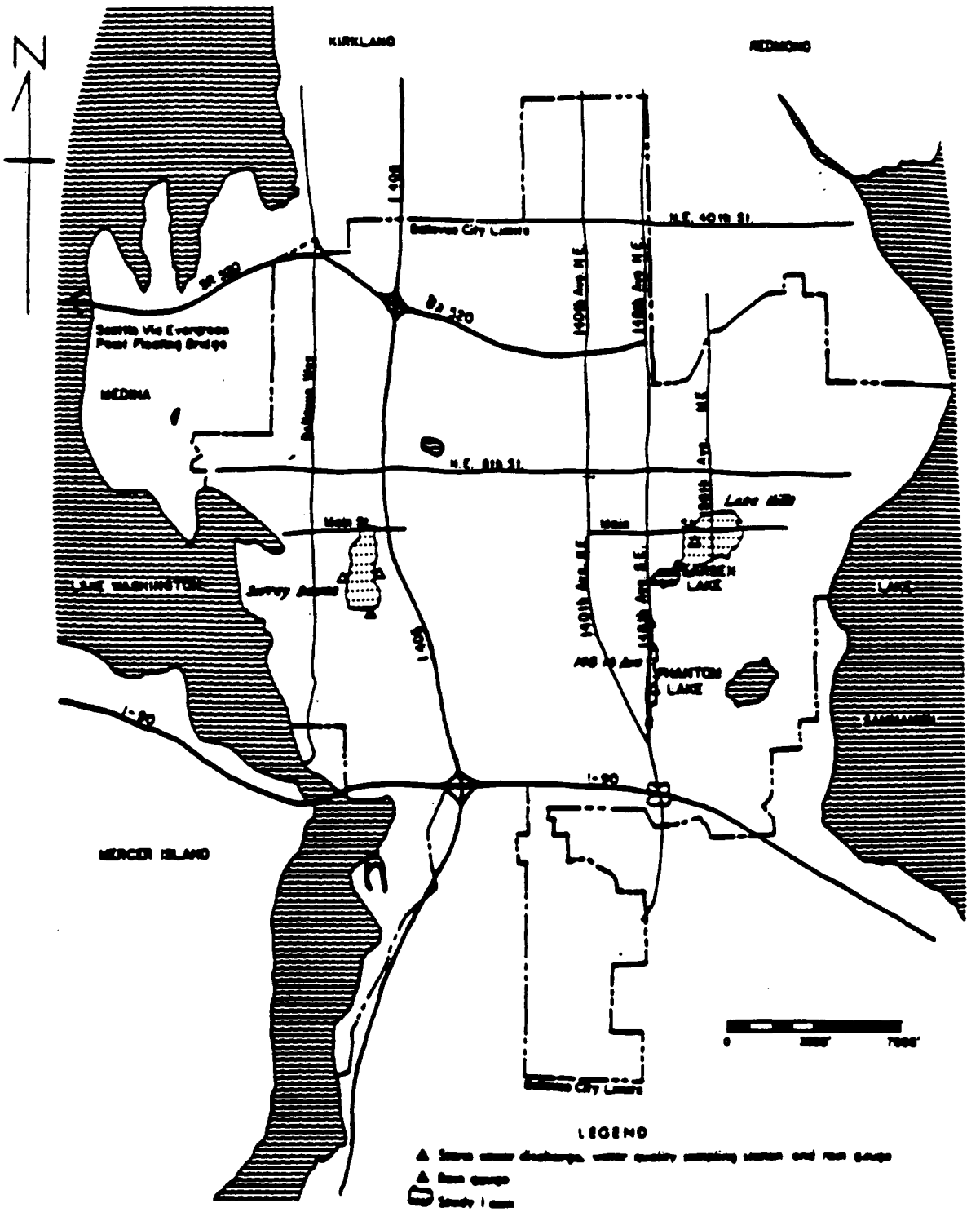


Figure VI-5. Bellevue sampling sites. (SOURCE: USEPA, 1982c)

- *To evaluate methods of transfering the data to ungaged watersheds in other regions.*

Monitoring activities took place in three catchment areas (Figure VI-5) (USEPA, 1982c, p. G27-7): (1) Surrey Downs, (2) Lake Hills, and (3) 148th Avenue. Three groups participated in the monitoring efforts: the USGS, the City of Bellevue, and the municipality of Metropolitan Seattle (Metro). Objectives of the monitoring activities included:

- *To define pollutant hydrographs for each of the three catchments during approximately 12 storms per year.*

- *To determine the effectiveness of street cleaning equipment for various levels of effort under the actual conditions encountered.*

- *To describe the quantities and characteristics of sewerage system particulates in the study area.*

- *To obtain a continuous mass balance relationship between total runoff yields and all the sources of urban runoff pollution.*

- *To analyze samples for the 129 EPA toxic or "priority" pollutants.*

The following, taken from NURP project summaries (USEPA, 1982c, pp. G27-14 to G27-16), describes in some detail the sampling approaches and equipment used to meet the monitoring objectives.

Two of the three study catchments, Surrey Downs and Lake Hills, are single-family residential areas of similar size. These two basins are used to investigate the effectiveness of street sweeping for reducing the amount of pollutants in storm runoff. The third catchment, 148th Avenue, consists mainly of a divided 4-lane arterial street. The data from this site are used to investigate the effects of detention basins on the quality of runoff.

The area comprising the Surrey Downs catchment consists of single family homes and the Bellevue Senior High School. Slopes in the basin are generally moderate, with the exception of the steep slopes on the west side. Surrey Downs is relatively isolated from neighboring communities by the general lack of easy vehicular access and convenient "short cuts" through this residential neighborhood.

The Lake Hills catchment contains single family residences and the St. Louise Parish Church and School. Although there are relatively isolated residential areas within the catchment, two through-streets, which carry more traffic than a typical residential street, cross the area.

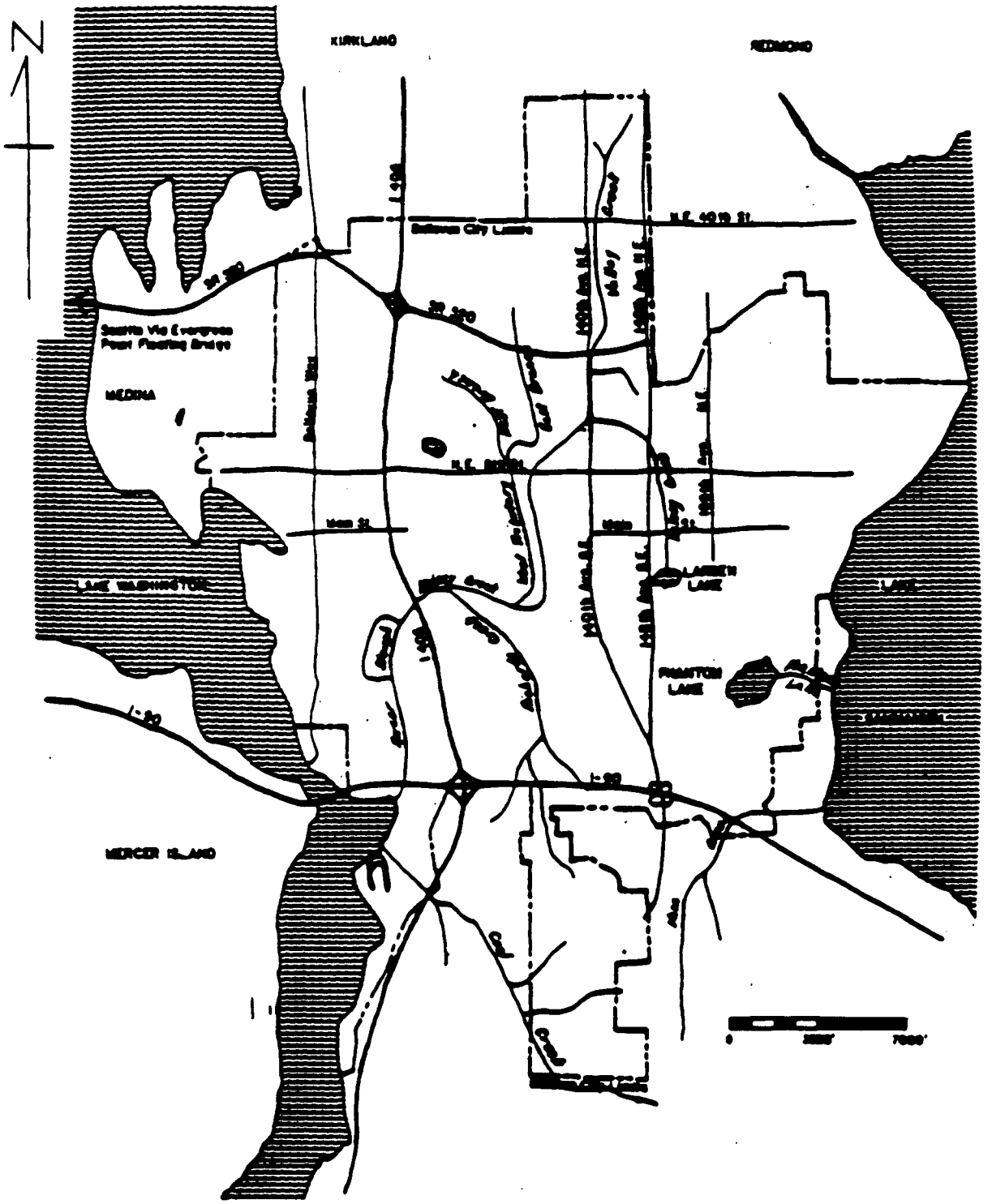


Figure VI-4. Bellevue stream systems. (SOURCE: USEPA, 1982c)

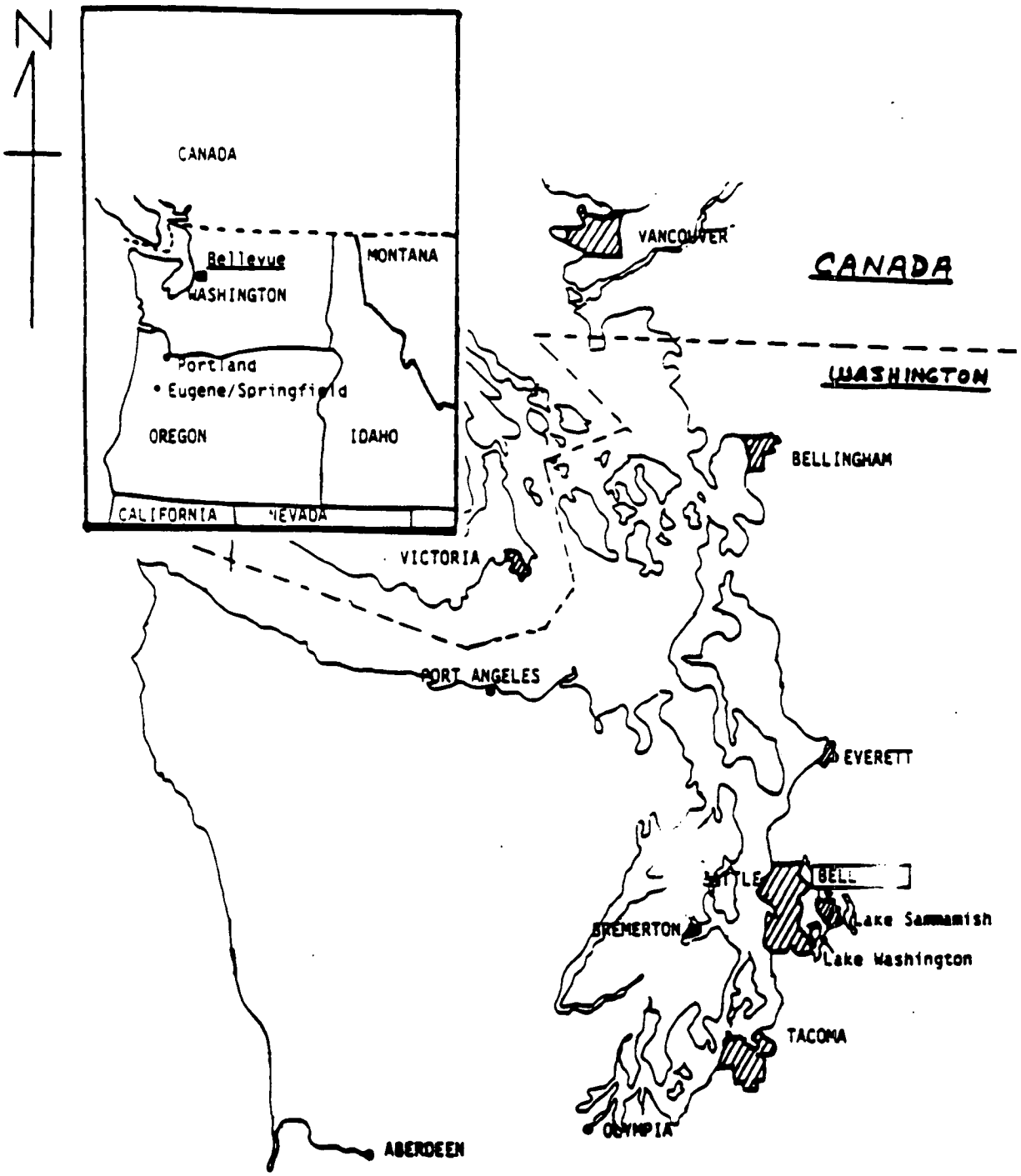


Figure VI-3. State locus of Bellevue NURP. (SOURCE: USEPA, 1982c)

program is described to illustrate the type of monitoring used in NURP.

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 VI-3) (USEPA, 1982c, p. G27-4). Land use is primarily residential and mean annual precipitation, mostly rain, is about 42 inches.

The land surface in Bellevue is mostly hilly, with moderate slopes predominant. Drainage is carried by a system of totally separate storm sewers, open channels and streams largely to the west into Lake Washington through Mercer Slough (Figure VI-4) (USEPA, 1982c, p. G27-6). Some drainage flows east into Lake Sammamish through Phantom Lake and another stream.

The project objectives were as follows (USEPA, 1982c, p. G27-12):

- *To apply uniformly, in selected drainage basins, a variety of management practices which 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 data base 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, and*

Table VI-7. Methods of Water Quality Analysis

Characteristic	Storage	Preservation/ Method of Analysis	Reference
pH	in-situ	Orion 399A meter	Method 150.1[2]
Dissolved Oxygen	in-situ	YSI 51B meter	Method 360.1[2]
Conductance/Temperature	in-situ	YSI 33 S-C-T meter	Method 120.1[2]
Turbidity Turbidimeter (NTU)	cool, 4°C	Hach Ratio	Method 180.1[2]
Total Suspended Solids[1] 103-105°C filtration	cool, 4°C	Gravimetric	Method 160.2[2]
Volatile Suspended Solids	cool, 4°C	Filter ignition at 550°C	Method 160.4[2]
Total Phosphorus acid (pH≤2)	cool, 4°C	Persulfate digestion	Method 365.2[2]
Orthophosphate[1]	cool, 4°C	Ascorbic acid	Method 365.2[2]
Nitrate + Nitrite-N[1] acid (pH≤2)	cool, 4°C	Cd-Cu reduction sulfanilamide	Method 353.3[2]
Ammonia-N acid (pH≤2)	cool, 4°C	Boric Acid distillation	Method 350.2[2]
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[3]
Fecal Streptococci	cool, 4°C	Membrane filtration	Method 909C[3]

[1] Filtration through glass fiber filter.
[2] USEPA, 1974.
[3] American Public Health Administration, 1975.

Table VI-6. Years of Sampling for Each Monitoring Level in the St. Albans Watershed

Long-Term Monitoring	1981 YEAR 1990 1	2	3	4	5	6	7	8	9	10
LEVEL 1: Bay Sampling										
Sta. 11	———	———	———	———	———				———	
Sta. 12	———	———	———	———	———				———	
Sta. 13	———		———		———				———	
Sta. 14	———									
LEVEL 2: Trend Stations										
Sta. 21	———	———	———	———	———	———	———	———	———	———
Sta. 22	———	———	———	———				———	———	
Sta. 23	———	———	———					———	———	
Sta. 24	———	———	———	———						
Sta. 25	———	———	———					———	———	
Sta. 26	———	———	———	———	———					
LEVEL 3: Mobile Sampling										
Sta. 31	———	———	———	———	———				———	
Sta. 32	———	———	———	———	———				———	
LEVEL 4: Random Sampling										
Sta. 41-44	———	———	———	———	———					

Table VI-5. Water Quality Sampling Schedule for the St. Albans Watershed

Level	Parameters	Sample Type	Season	Sampling Frequency
1	All	grab	Oct. - Apr.	monthly
			May - July	biweekly
			Aug. - Sept.	weekly
2	FC, FS	grab	All	weekly
	temperature	in-situ	All	biweekly
	turbidity, pH, O_2, solids	composite	All	2-48 hr. and 1-72 hr.
	nitrogen			weekly
	phosphorus			weekly
3 Larose	All	composite	All	4-hr.
4	All	grab	May - Feb	20 days(average)
			Mar. - Apr.	weekly

Table VI-3. Related Long-Term Monitored Parameters in the St. Albans Bay Watershed.

Meteorological Monitoring
- Continuous precipitation
- Continuous streamflow
- Continuous temperature (Level 3 only)
- Wind speed and direction (Bay only)

Biological Monitoring
- Fish species and abundance
- Benthic invertebrates
- Periphyton
- Macroinvertebrates

Land Use Monitoring
- 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 VI-4: Monitored Parameters for the Short-Term Intensive Studies.

Wetland Influences

- Temperature
- Dissolved Oxygen
- pH
- Discharge
- Chloride
- Total Suspended Solids
- Volatile Suspended Solids
- Total Phosphorus
- Ortho-phosphate
- Total Kjeldahl Nitrogen
- Ammonia-Nitrogen
- Nitrate + Nitrite - Nitrogen

Bay and Wetland Sediments

- Redox potential
- Porosity
- Grain size distribution
- % organic matter
- Total phosphorus
- NH_4Cl phosphorus
- NH_4Ac phosphorus
- NaOh phosphorus
- HCl phosphorus
- Iron

Bay Circulation

- Total phosphorus
- Chloride
- Wind speed and direction
- Water velocity and direction

before descriptive statistics are calculated from a data set, and before analyses such as correlation, analysis of variance, or linear regression are performed, it is wise to look at various displays of the raw data. The graphs recommended for this task [time series, histograms, stem and leaf displays, box plots, and scatter plots] are useful in identifying the need to edit or transform the data prior to conducting the statistical analysis. Most procedures in statistics (e.g., regression analysis, hypothesis testing) derive summary values (e.g., mean and standard deviation) from a data set. Thus, if the inferences drawn from the statistical procedures are to be valid for the entire data set, it is important that the summary statistics represent the entire data set. The graphical displays help guide the choice of any necessary manipulations of the data set and the selection of appropriate statistics to summarize the data.

The reader is referred to Gaugush (1986) or Chambers et al. (1983) for more detailed discussions regarding these graphical displays of data. 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 may 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 appropriate methods.

Figures VIII-1 to VIII-4 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 VIII-1) demonstrates a seasonal nature to the data. In this case the time series includes data over a 10-year time span. Similar plots can also be made over shorter time periods such as intensive data collection during a storm event. In the case of a storm event, the investigator may plot precipitation and runoff volume together with pollutant concentrations. It is also apparent from Figure VIII-1 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 summer 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 VIII-2 and VIII-3 are a DO histogram and stem and leaf plot, respectively. These figures demonstrate that most of the observations fall between 6.0 and 10.0 mg/L. The difference between these plots is that the stem and leaf plot displays the actual data rather than a bar. In typical applications the analyst might display histograms from several stations or different groupings of data.

Figure VIII-4 is a box and whisker plot. For each month along the horizontal axis, the box indicates the middle 50 percent of the

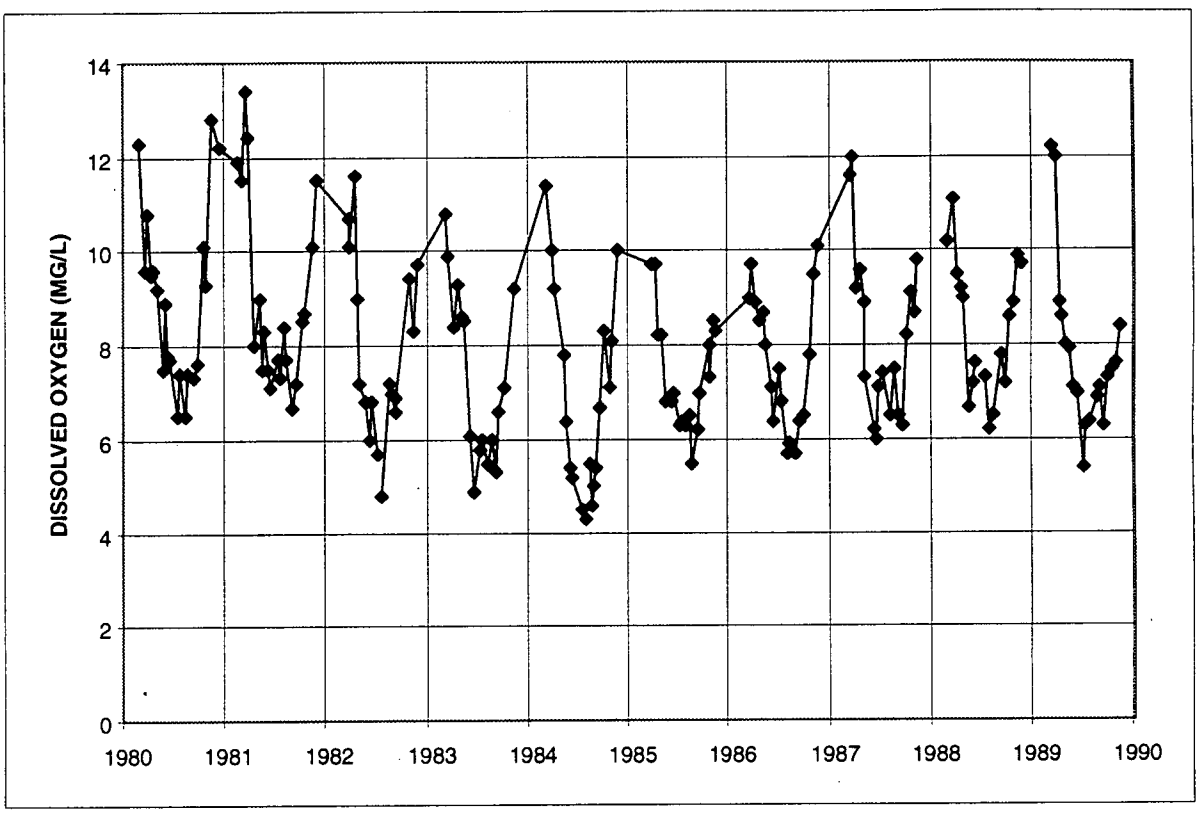


Figure VIII-1. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware using a time series plot.

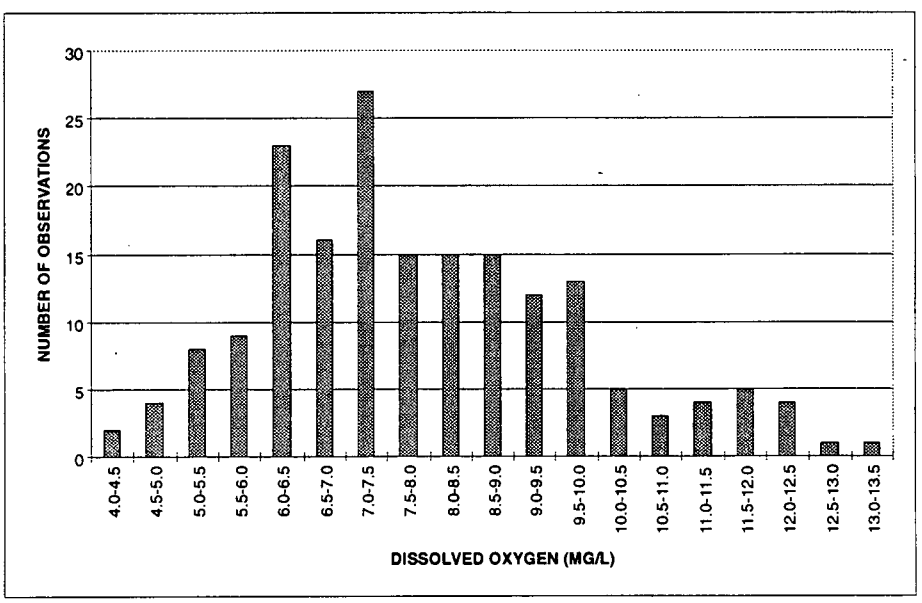


Figure VIII-2. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware using a histogram.

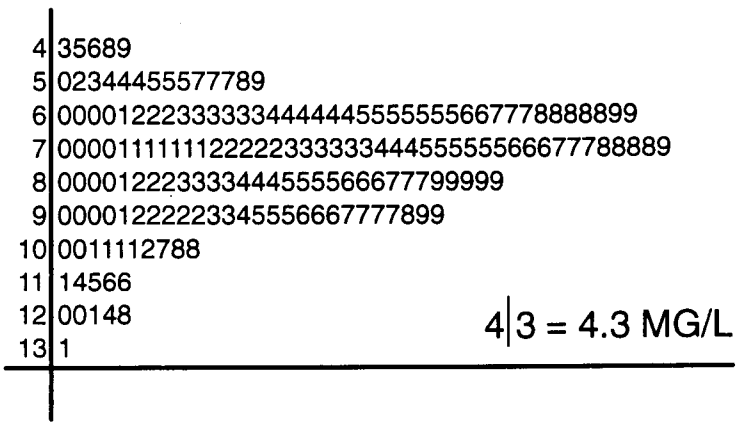


Figure VIII-3. Stem and leaf plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

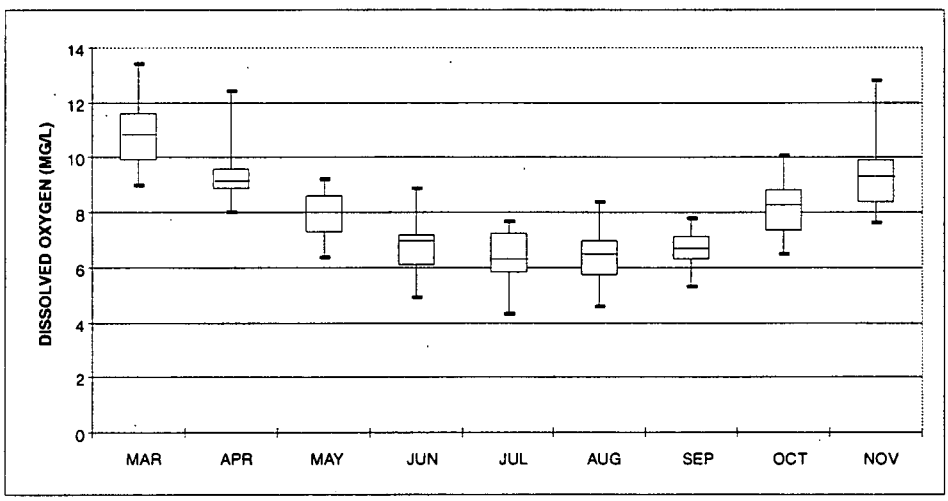


Figure VIII-4. Box and whisker plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

data. The lower and upper ends of the box represents the 25th and 75th percentiles, respectively. The horizontal line inside the box represents the 50th percentile or median. Depending on the convention used, the whiskers extending from the box represent the range of the remaining observations. In this case, the short horizontal lines at the end of the whiskers represent 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 these differences when mixing and matching analyses from different packages. The expected seasonal nature of DO is strongly depicted in Figure VIII-4 confirming the suspicions developed from visual inspection of Figure VIII-1. 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.

If the analyst is comparing biological metrics, it may be appropriate to summarize the data by station classification. Figure VIII-5 is a box and whisker plot of mean abundance per grab from the Louisiana Province of the Environmental Monitoring and Assessment Program (EMAP) for 1991 and 1992. Based on visual inspection, it appears as though mean abundance per grab between large rivers (LR) and small rivers (SR) might be similar. A statistical test is required however. In other cases, it may be helpful to plot water quality data as a function of other explanatory variables such as flow. Figure VIII-6 is an example 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 may be positive (as in Figure VIII-6) or could be negative. Typically, a negative correlation (decreasing concentrations with increasing flows) are indicative of constant pollutant sources (e.g., traditional point sources).

Figure VIII-7 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 on September 23, 1991 can be observed. In this particular case, the analyst should be alerted to the potential of erroneous data. In this case, one potential cause may be due to unit conversions. The data were stored as mg/L of P; however, another set of units for orthophosphate is mg/L of PO_4. If one were to multiply the data collected on September 23, 1991 by one-third (approximate conversion from PO_4 to P) the data will 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. These types of errors also occur while converting data from ppm to ppb, converting from wet to dry weight basis, normalizing for organic carbon, etc. Data visualization is a good method picking out gross errors; however, cannot be relied

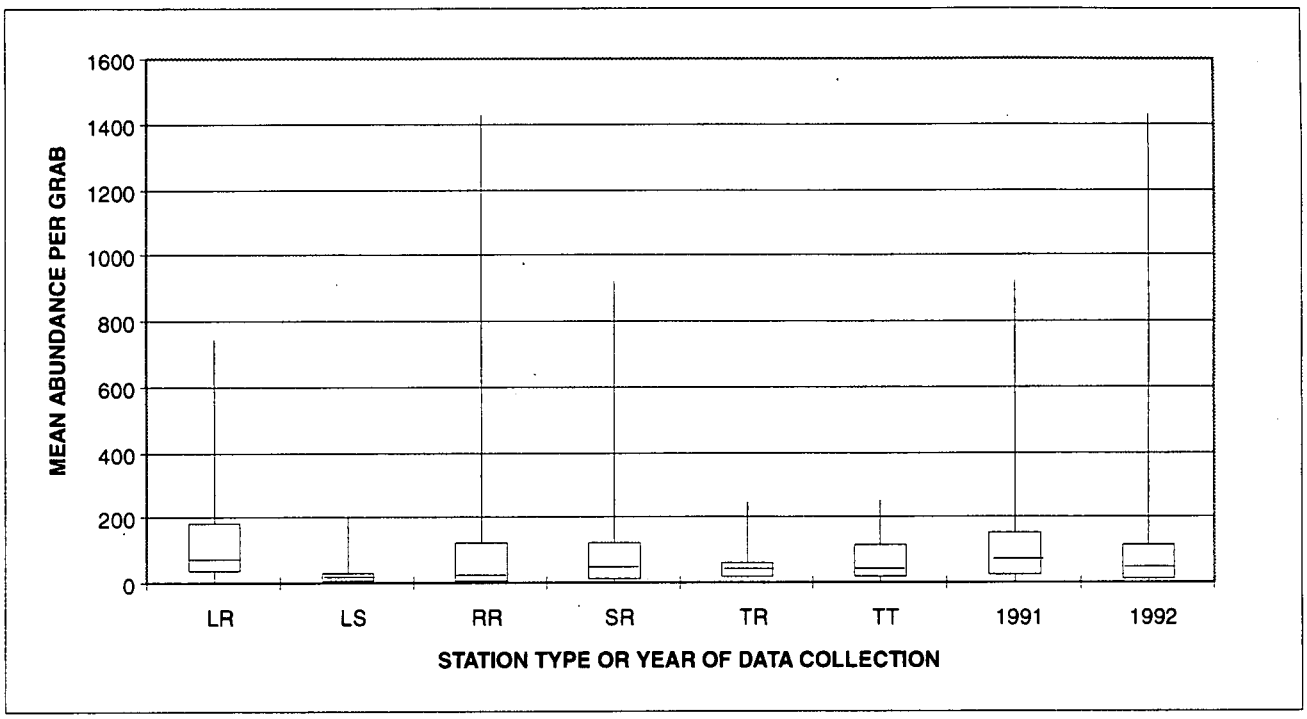


Figure VIII-5. Box and whisker plot of mean abundance per grab from 1991 through 1992 for the Louisiana Province of the Environmental Monitoring and Assessment Program (EMAP).

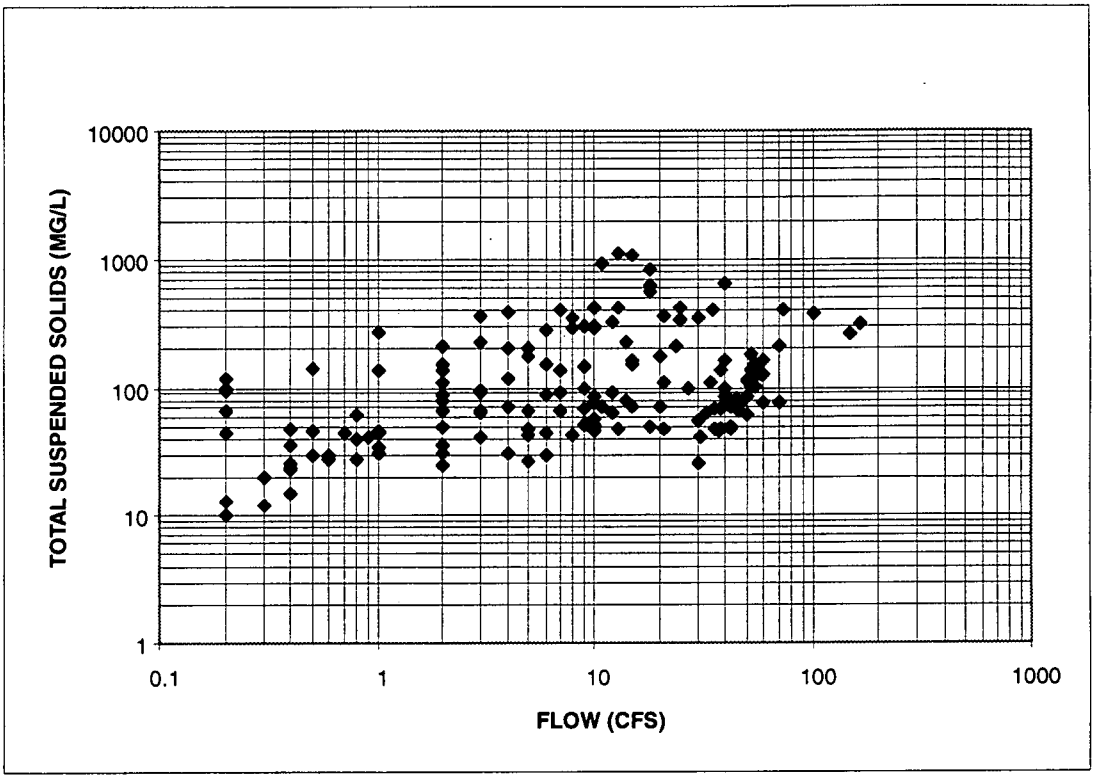


Figure VIII-6. Bivariate scatter plot of total suspended solids and flow at 36th Street storm sewer in Denver, Colorado.

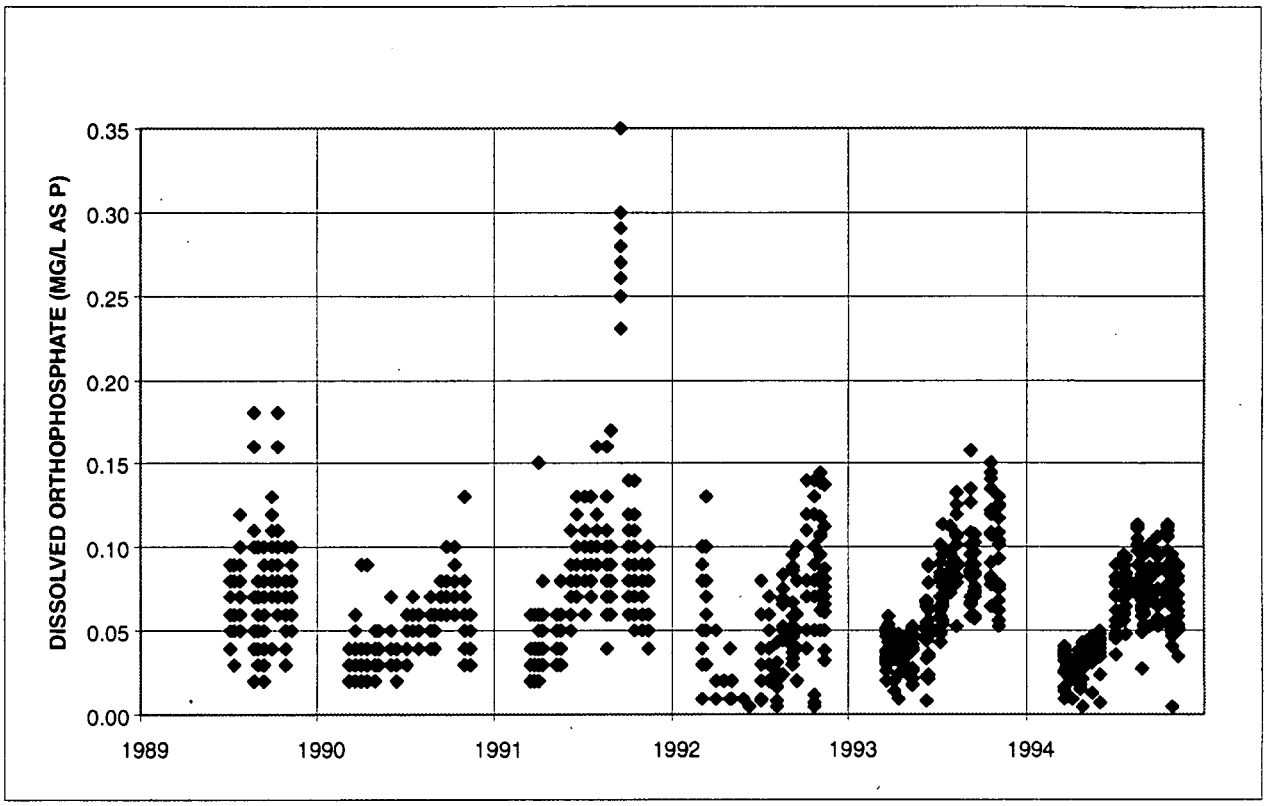


Figure VIII-7. Time series plot of dissolved orthophosphate from 1989 through 1994 for portions of the Delaware River.

upon for more subtle errors. *The likelihood of correcting data errors goes down significantly with time.*

B.2 Data Stratification

An important component in data analysis is data stratification. Lumping measurements over a period of time has limited use in water quality evaluation, unless the period of time is defined in more specific terms and 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 masks the effectiveness of the management measure.

In urban areas the periods should be set to correspond to the pollutant of concern and urban activities. Depending on the monitoring objectives, it may be necessary to consider periods of activity and non-activity. 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, then periods that correspond to the construction season should be considered. For irrigated agriculture, two periods should be established to correspond to irrigation time and non-irrigation time.

In non-irrigated 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. 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).

Some variation in period length from year to year can be expected due to the climate's impact upon agricultural activities. There must be enough flexibility within your data analysis to account for this variation.

Once temporal stratification has been completed, and if sufficient data are available, the water quality variable being examined can be categorized by initiation/transport mechanism. For a sediment related problem, three categories have been devised (Davenport, 1983) related to the principal detachment process of sediment particles:

- Baseflow (no rainfall or overland runoff to the stream). This consists of non-precipitation induced flow and is considered

as the normal day-to-day flow (Viessman et al., 1977). Sediment concentrations are dependent upon available material in the channel network and the carrying capacity of the stream.

- Rainfall and snow-melt runoff. This category consists of runoff events whose rainfall is less than Wischmeier and Smith's (1965) definition of a storm and snow-melt. Sediment concentrations are dependent upon flowing water detachment or re-entrainment of previously detached soil particles, together with sufficient overland flow to transport them to the stream network.

- Event. This consists of rainfall-runoff events whose rainfall characteristics meet Wischmeier and Smith's definition. Sediment concentrations are dependent upon the detachment of soil particles due to the impact of raindrops and flowing water detachment or re-entrainment 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 is to be noted that data stratification results in smaller data sets. You must check these new data sets for normality before performing any statistical analyses upon them. It is also important to note that due to the smaller data set size the differences between data sets must be pronounced to be significant.

B.3 Data Normalization

Data normalization is the process or technique used to convert data into a production value; i.e., a rate. The normalization of data results in the factoring out of inherent factor(s) of the data base that are related to the data collection or generation process. Data normalization is most commonly utilized in comparison of loading data.

The most common normalization of data are time period (kg/yr), unit area (kg/ha), or combination of unit area per time period (kg/ha/month) basis. The other major type of normalization is related to parameter generation or transport factors such as rainfall and runoff; examples are kg/cm of precipitation or kg/l^3 of streamflow.

Data normalization is extremely useful in evaluating the impact of implementation activities, source control, and determination of implementation priorities. An example of raw and normalized data is provided in Tables VIII-3 and VIII-4.

Table VIII-3. Raw data by time period. The watershed is 20 ha and has three consecutive years of pre- and post-implementation sediment loading, precipitation, and runoff data.

1970-1973	
Sediment loading	48 kg
Precipitation	120 cm
Runoff	15 l^3
1974 implementation of terraces and conservation tillage	
1975-1978	
Sediment loading	45 kg
Precipitation	180 cm
Runoff	18 l^3

Table VIII-4. Normalized data. The watershed is 20 ha and has three consecutive years of pre- and post-implementation sediment loading, precipitation, and runoff data.

1970-1973	average annual loading	16 kg/year
	average annual loading	0.10 kg/cm/year
	average annual loading	1.07 kg/l^3/year
1974-1978	average annual loading	15 kg/year
	average annual loading	0.08 kg/cm/year
	average annual loading	0.83 kg/l^3/year

Review of Table VIII-4 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, irrespective of flow and precipitation, sediment loading has decreased. Based on this exploratory analysis, a more detailed frequency analysis would be required to test for statistical significance.

Another useful normalization of data is converting or summarizing storm event data. The flow-weighted mean concentration (FWMC) and time-weighted mean concentration (TWMC) are calculated as (EPA, 1990)

$$FWMC = (\text{Sum of } C_i T_i Q_i) \; / \; (\text{Sum } T_i Q_i) \qquad (8.1)$$

$$TWMC = (\text{Sum of } C_i T_i) \; / \; (\text{Sum } T_i) \qquad (8.2)$$

where

C_i = concentration of the ith sample,
T_i = time period for which the ith sample is used to characterize the concentration, and
Q_i = instantaneous discharge at the time of the ith sample.

The numerator of Equation 8.1 is equal to the total loading. Figure VIII-8 presents a summary of the 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 respective curves in Figure VIII-8 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. The raw nitrogen are provided in Table VIII-5 together with the example calculations for calculating total loading and flow-weighted mean concentration.

Table VIII-5. Total nitrogen runoff concentrations for a single storm event in Florida.

	Q_i	C_i	T_i	T_iQ_i	$C_iT_iQ_i$
TIME (HRS)	FLOW (CFS)	TN (MG/L)	INT (SEC)	(FT3)	
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 of T_iQ_i = 2080.44 FT3
SUM of $C_iT_iQ_i$ = 2768.23 MG-FT3/L (=78.40 grams of TN)

Flow-weighted mean concentration = 2768.23 / 2080.44 = 1.33 MG/L

The first column in Table VIII-5 is the time since the beginning of the storm. The fourth column is the time interval represented by each sample. For example the first entry of 540 seconds is computed as (0.24 hours - 0.09 hours) • 3600 seconds/hour. The value of 0.24 is one-half way between 0.20 and 0.28 hours. Using conversions, the total loadings of nitrogen for this storm are 78.40 grams. Using Equation 8.1, the FWMC for total nitrogen is 1.33 mg/L. Because different analysts use different

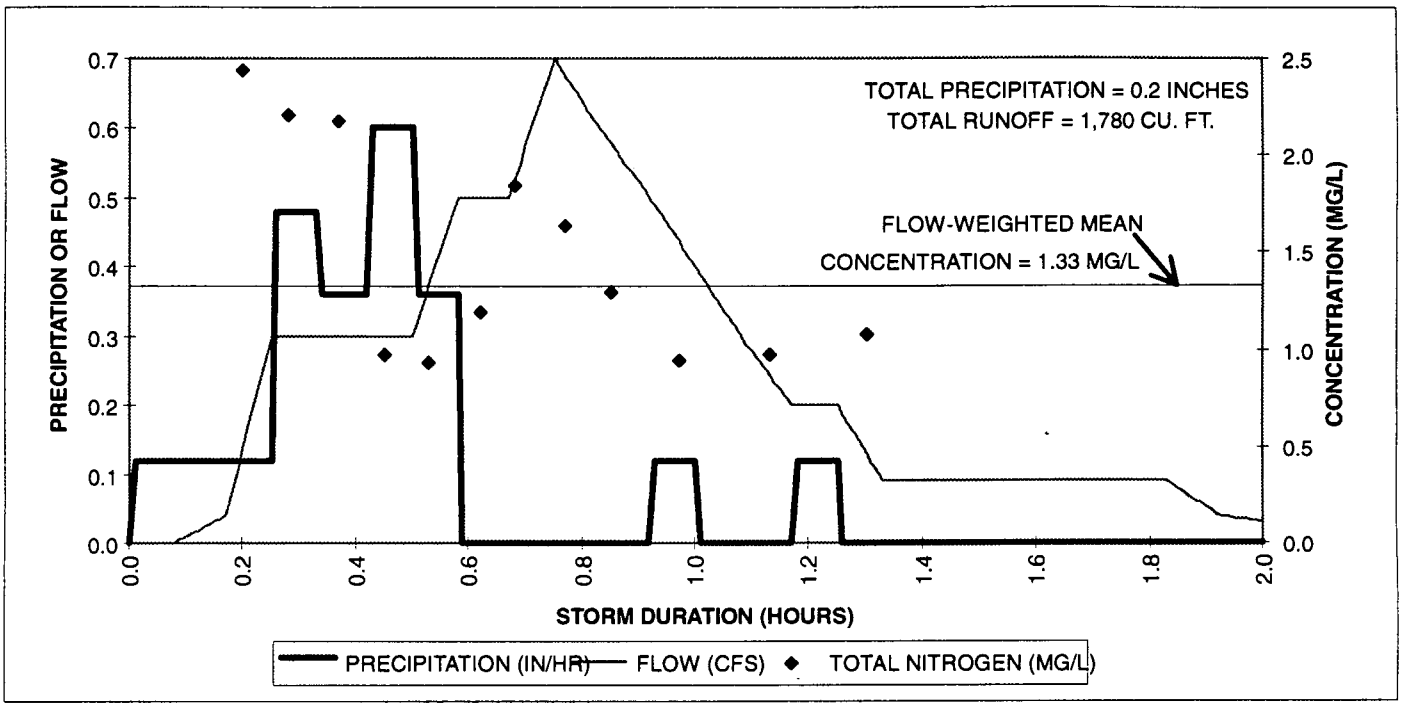


Figure VIII-8. Precipitation, runoff, total nitrogen, and total phosphorous from a single storm event in Florida.

conventions for analyzing storms, it is important that the analyst exercise care when comparing the storm summaries computed by different analysts.

Equation 8.3 provides an approach for estimating the event mean concentration (EMC):

$$EMC = \frac{\text{Sum } (T_{i+1} - T_i) \bullet 0.5 \bullet (C_iQ_i + C_{i+1}Q_{i+1})}{\text{Sum } (T_{i+1} - T_i) \bullet 0.5 \bullet (Q_i + Q_{i+1})} \quad (8.3)$$

Table VIII-6 demonstrates the use of this equation using the same storm event presented in Figure VIII-8 and Table VIII-5. The first three columns of Table VIII-6 correspond to Table VIII-5. The next four columns correspond to intermediate calculations needed for Equation 8.3. 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 • 0.000, 0.14 • 2.44, and 0.00+0.14, respectively. The last two columns correspond to intermediate calculations for the numerator and denominator of Equation 8.3, respectively. Finally, the EMC can be calculated as 0.6722 / 0.4981 or 1.35 mg/L as shown in Table VIII-6.

Table VIII-6. Total nitrogen runoff concentrations for a single storm event in Florida and example calculations for the EMC.

TIME (HRS)	FLOW (CFS)	TN (MG/L)	T_{i+1} $-T_i$	$C_i \bullet Q_i$	$C_{i+1} \bullet Q_{i+1}$	$Q_i + Q_{i+1}$	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 = 0.6722 / 0.4981 = 1.35 MG/L

B.4 Tests of assumptions

In section VIII.A, the need to evaluate test assumptions was discussed. Ponce (1980) outlined the basic assumptions when using parametric statistical tests as:

- The observations are independent.
- The distribution of the population is known.
- The variances of the populations being compared are equal or of known ratio.

He also states that parametric statistical tests require that the data:

- Were collected in a random manner.
- Have error variation independent of the means, normally distributed, and homogeneous.
- Have variance components which are additive.

This section describes tests which can be used to determine whether or not a data set satisfies some of the assumptions and requirements of parametric tests. Researchers are referred to statistics texts such as Snedecor and Cochran (1980) for further information regarding tests of assumptions.

B.4.1 Tests of normality

Remington and Schork (1970) state that "the consequences of violating the assumption [of normality] vary from relatively mild in the case of inferences on means to relatively severe for inferences on variances." The following tests can be used to determine whether a data set satisfies the normality assumption.

B.4.1.1 Skewness

Remington and Schork (1970) recommend a standardized measure of skewness (a_3) such that:

$$a_3 = \frac{m_3}{m_2^{3/2}} \quad \text{where} \quad m_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \quad \text{and} \quad m_3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n} \qquad (8.4)$$

The approach used in testing for normality using a measure of skewness is that a non-normal distribution may be skewed, whereas a normal distribution is not skewed. Sample sizes as small as n=25 can be tested for skewness at confidence levels of 0.05 or 0.01 using the table in Appendix F. The calculated value of a_3 is compared with the appropriate critical value in the table; if a_3 is greater than or equal to the critical value then the hypothesis that the observations are a random sample from a normally distributed population is rejected (Remington and Schork, 1970). Otherwise, the hypothesis is accepted.

Using the mean abundance per grab data from Figure VIII-5 for large rivers (LR) and small rivers (SR), selected statistics were computed and are summarized in Table VIII-7. Selected statistics are also calculated for log-transformed data. Using Equation 8.4, a_3 for large rivers is equal to (465.0E+06/77)/(1.5E+06/77)$^{1.5}$ or 2.22. The critical value from Appendix F for a sample size of 77 is approximately 0.44 (0.05 confidence level). Therefore, it is assumed that the data do not come from a normally distributed population.

Table VIII-7. Selected summary statistics for mean abundance per grab from 1991 and 1992 for the Louisiana Province of the Environmental Monitoring and Assessment Program (EMAP) from large rivers (LR) and small rivers (SR).

	LR	SR	log(LR)	log(SR)
Number	77	42	77	42
Sum	10,416.07	5,098.00	144.34	68.53
Mean	135.27	121.38	1.87	1.63
Variance	20,295.22	39,031.80	0.31	0.52
St. Dev.	142.46	197.56	0.55	0.72
$\Sigma\lvert x_i-x\rvert$	7,701.40	4,994.28	32.92	23.83
$\Sigma(x_i-x)^2$	1.5E+06	1.6E+06	23.24	21.46
$\Sigma(x_i-x)^3$	465.0E+06	942.5E+06	-10.38	-7.50
$\Sigma(x_i-x)^4$	258.1E+09	736.8E+09	24.11	31.37

The SAS Institute (1985a, Basics) calculates sample skewness as:

$$n/(n-1)(n-2) \bullet \sum_{i=1}^{n} (x_i-\bar{x})^3/s^3 \qquad (8.5)$$

Skewness values using this statistic can range from negative to positive infinity, with a value of 0 for normal distributions. Using the intermediate calculations provided in Table VIII-7, the sample skewness for large rivers is 77/(76 • 75) • 465.0E+06/ 142.46^3 or 2.17.

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 (<25) the test is particularly unreliable. That is, because of the small sample size, very large departures of normality are required before statistical tests will reject the null hypothesis of normality.

Using Fisher's measure of skewness (G_1) Cochran (1963) proposed a general rule for determining how large n must be 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 is most common for environmental data.

$$n > 25G_1^2 \tag{8.6}$$

where G_1 is

$$G_1 = \frac{E(x_i - X)^3}{\sigma^3} = \frac{1}{Ns^3} \bullet \sum_{i=1}^{N} (x_i - \bar{x})^3 \tag{8.7}$$

Applying Equations 8.6 and 8.7, G_1 can be computed as 2.09 for the large river data from Table VIII-7 and therefore more than 109 (=25 • 2.09^2) samples are needed. The example data set only contains 77 samples. Therefore, there are not sufficient data to allow safe use of the normal approximation in computing confidence limits for the mean.

B.4.1.2 Kurtosis

The test for kurtosis is as limited as 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 to measure kurtosis:

$$g = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{s\sqrt{n(n-1)}} \tag{8.8}$$

where $|x_i - \bar{x}|$ is the absolute value of $x_i - \bar{x}$. For a normally distributed population g would be 0.7979, with smaller and larger values indicating leptokurtic and platykurtic distributions, respectively. (Probability distributions are discussed in section B.5 in more detail.) The table in Appendix G permits upper or lower one-sided kurtosis tests of the hypothesis that the observations are a random sample from a normally distributed population (Remington and Schork, 1970). If the calculated value of g falls outside of the values in the table for the selected level of confidence then there is evidence of non-normal kurtosis.

The SAS Institute (1985a) calculates sample kurtosis as:

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)s^4} \sum_{i=1}^{n} (x_i - \bar{x})^4 - \frac{3(n-1)(n-1)}{(n-2)(n-3)} \tag{8.9}$$

Measures of kurtosis using this statistic range from -2 to positive infinity with a value of 0 for normal distributions. Using the intermediate calculations provided in Table VIII-7 and Equation 8.9, the sample kurtosis for large rivers is 5.8.

B.4.1.3 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 or equal to 50 (SAS Institute, Inc., 1985a). 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., 1985a). 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.

Anderson and McLean (1974) recommend this 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 $d = \Sigma(x_i - \bar{x})^2$.

3. Compute k. If n is even, k=n/2. If n is odd, k=(n-1)/2.

4. Compute

$$b = \sum_{i=1}^{k} a_i (x_{n-i+1} - x_i) \qquad (8.10)$$

where the values of a_i appear in Appendix H.

5. Compute $W = b^2/d$. (8.11)

6. Compare W to the quantile given in Appendix I. If W is less than the quantile, reject the null hypothesis of normality.

For further information regarding this test the reader is encouraged to consult statistics texts (Gilbert, 1987, p. 158) and the publication by Shapiro and Wilk (1965). Table VIII-8 presents the mean abundance per grab data for small rivers in a format ready for analysis. The results for step 2 are provided in Table VIII-7. Since there are 42 observations, k is equal to 21. The first column in Table VIII-8 indicates the value of i for each row in the table. The second column corresponds to the values of a_i from Appendix H. The third and fourth column, x_i and x_{n-i+1}, represent the raw abundance data. The third column represents the first half of the observations and the fourth column represents the last half of data in reverse order (e.g., 923.33 is the largest sample observation). The fifth and sixth columns correspond to the log-transformed data from columns 3 and 4. For example, log(0.67) is equal to -0.17. The last two columns provide intermediate calculations associated with Equation 8.10 (e.g., $a_i(x_{n-i+1} - x_i)$) for the raw and log-transformed data, respectively.

Table VIII-8. Example analysis of the Shapiro-Wilk W test using mean abundance per grab data for the Louisiana Province of the Environmental Monitoring and Assessment Program (EMAP) from 1991 and 1992 for small rivers.

i	a_i	x_i	x_{n-i+1}	log x_i	log x_{n-i+1}	Intermediate Calculations	
1	0.3917	0.67	923.33	-0.17	2.97	361.41	1.23
2	0.2701	1.67	869.67	0.22	2.94	234.45	0.73
3	0.2345	2.00	374.00	0.30	2.57	87.23	0.53
4	0.2085	3.33	343.00	0.52	2.54	70.82	0.42
5	0.1874	3.67	260.33	0.56	2.42	48.10	0.35
6	0.1694	4.00	219.67	0.60	2.34	36.53	0.29
7	0.1535	7.00	215.00	0.85	2.33	31.93	0.23
8	0.1392	8.67	204.00	0.94	2.31	27.19	0.19
9	0.1259	14.00	153.00	1.15	2.18	17.50	0.13
10	0.1136	14.67	147.33	1.17	2.17	15.07	0.11
11	0.1020	14.67	123.00	1.17	2.09	11.05	0.09
12	0.0909	22.67	118.00	1.36	2.07	8.67	0.07
13	0.0804	25.00	99.67	1.40	2.00	6.00	0.05
14	0.0701	25.67	99.33	1.41	2.00	5.16	0.04
15	0.0602	27.33	90.67	1.44	1.96	3.81	0.03
16	0.0506	29.00	90.33	1.46	1.96	3.10	0.02
17	0.0411	33.33	87.33	1.52	1.94	2.22	0.02
18	0.0318	37.33	75.33	1.57	1.88	1.21	0.01
19	0.0227	43.33	66.33	1.64	1.82	0.52	0.00
20	0.0136	49.67	59.00	1.70	1.77	0.13	0.00
21	0.0045	53.67	58.33	1.73	1.77	0.02	0.00
					SUM	972.13	4.56

Applying Equation 8.11,

$W = 1/1.6E+06 \times 972.13^2 = 0.59$ for raw data
$W = 1/21.46 \times 4.56^2 = 0.97$ for log-transformed data

Summing the last two columns results in completing the computations for Equation 8.10. The W statistic may now be computed using Equation 8.11 as shown in Table VIII-8. From Appendix I, the quantile for 42 observations (95 percent confidence level) is 0.942. As a result, it can be concluded that the raw data are not normally distributed and the log-transformed data are normally distributed. In other words, the data are lognormally distributed.

B.4.1.4 Kolmogorov-D statistic

The Kolmogorov-D statistic is one of a class of goodness-of-fit statistics referred to as EDF (empirical distribution function) statistics (Stephens, 1974). When compared to other EDF statistics, Stephens (1974) found that the D statistic was inferior to other EDF statistics such as the Cramer-von Mises statistic and

the Anderson-Darling statistic, but still roughly comparable to the W statistic (above). The reader is referred to advanced statistics texts and a publication by Stephens (1974) for more detailed discussion. The SAS Institute (1985a) uses this test statistic instead of the Shapiro-Wilk W statistic for sample sizes greater than fifty. The major reason for not using the W statistic for n greater than 50 is that Shapiro-Wilk coefficients (Appendix H) are not available for these larger sample sizes (Stephens, 1974).

The Kolmogorov-D statistic tests the sample data against a normal distribution with mean and variance equal to the sample mean and variance, i.e., a comparison of the actual sample distribution against what the sample distribution would be if it was normal. A D statistic is computed and tested against critical values provided by Stephens (1974). Ponce (1980a) and Spooner et al. (1986) apply this test to actual data using SAS (Statistical Analysis System) procedures. Ponce notes that both the Kolmogorov D and Shapiro-Wilk W tests are also available in SPSS (Statistical Package for the Social Sciences).

B.4.1.5 Graphic method

Ponce (1980a) recommends a graphic method for testing the distribution of a data set. In this method one simply plots the data in a cumulative manner on arithmetic graph paper; for continuous data you will have to select intervals and plot interval frequencies. The cumulative frequency can be plotted on normal probability graph paper. To test for normality using this method one then compares the resulting plots against the examples plotted on probability graph paper in Figure VIII-9 (Ponce, 1980a, p. 28).

B.4.1.6 Transformations

What does the researcher do when the sample data set fails to pass the normality tests? There are several options (including proceeding in violation of the normality assumption--which is not always so bad), but the one most appreciated by statisticians is to perform data transformation. According to Gaugush (1986, p. 30), data transformations may:

1. Straighten (linearize) a nonlinear relationship between two variables.

2. Reduce skew (achieve symmetry) in a data set for a single value.

3. Stabilize variance (create constant variance) for a particular variance across two or more data sets.

It is generally agreed that log transformations are the most useful for water quality and hydrologic variables (Spooner et al., 1986; Gaugush, 1986; Ponce, 1980a; USEPA, 1983b). Ponce (1980a) states that log transformations are particularly useful when the data are skewed to the right. Gaugush (1986) agrees with Ponce because "the effect of a log transformation is to stretch out data

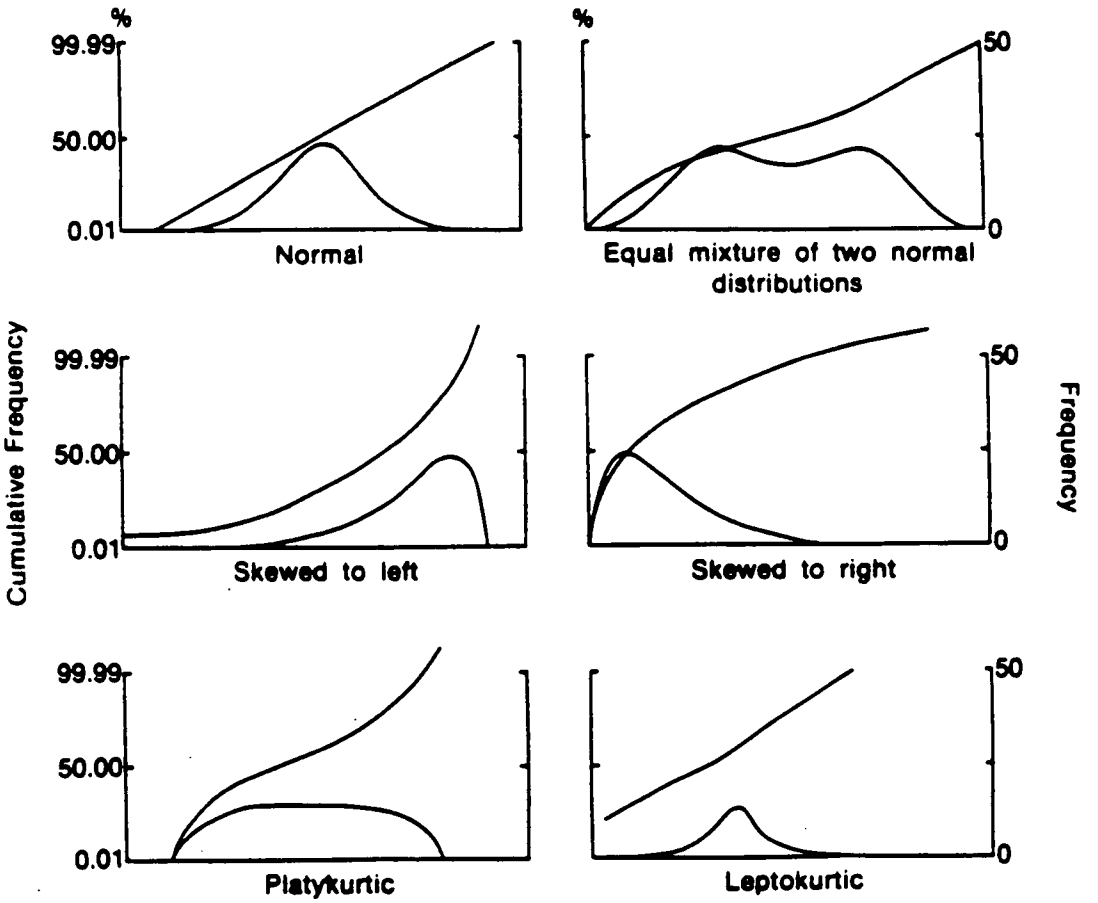


Figure VIII-9. Examples of several frequency distributions and their respective cumulative frequency distributions. (SOURCE: Ponce, 1980a)

on the left side of a plot and to pull in data on the right side of a plot." Ponce (1980a) says that the transformation to use when zero values exist in the data set is the log(x+1). This eliminates all zeros; there should be no negative numbers either, but this is unlikely in water quality investigations (an example of an exception is redox potential).

The reader is encouraged to study the examples of log transformations presented by Ponce (1980a) and USEPA (1983b). Additional information regarding possible transformations is provided by Snedecor and Cochran (1980, p. 287-292). After the data have been transformed it makes sense to take another look at them before proceeding with statistical analyses. As stated by Spooner et al. (1986), the transformed data should be tested for normality to determine if the transformation is adequate. The example analysis provided in section B.4.1.3 (see Table VIII-8) demonstrates that the log-transformed data were normally distributed while the raw were not normally distributed.

B.4.2 Tests of equal variance

Using the F distribution we can test the assumption of equal variance of two populations based upon the two sample variances (s_a^2 and s_b^2) (Freund, 1973). In this test F is equal to the ratio of the two sample variances, s_a^2/s_b^2 or s_b^2/s_a^2. The null hypothesis in this test is that the ratio is 1, and the alternative hypothesis is that the ratio is not 1. The calculated F statistic is compared versus critical F values (Appendix E) which are based upon the sample sizes (n_a and n_b) and the selected level of significance (α). The following equation is used to calculate the region of rejection for the test (Winer, 1971, p. 40):

$$\Pr\left[\frac{s_a^2}{s_b^2} > F_{1-\alpha} \bullet (n_a-1,\ n_b-1)\right] \qquad (8.12)$$

Recalling the data from Table VIII-7, the F statistic can be computed as 39,031.8/20,295.22 or 1.92 with 41 and 76 degrees of freedom. Using Appendix E, the critical F value (95 percent confidence level) is approximately 1.56. Therefore the null hypothesis is rejected and variances of the mean abundance per grab from large and small rivers have different variances.

B.4.3 Tests of constant variance of the residuals

The following taken from Spooner et al. (1986) describes tests for constant variance:

> *The assumption of constant variance of the residuals can [also be] tested. There are many diagnostic tools available. In the case where the Y variable is regressed on a continuous X variable, the two most common methods are: 1) the plot of the residuals vs. the predicted Y values, and 2) the plot of the*

residuals vs. their X values. For constant variance and constant mean of the residuals, both of these plots should look like a "constant horizontal band." For residuals that are calculated from ANOVA-type [analysis of variance] analyses, equality of variance tests between the groups should be performed (Snedecor and Cochran, 1967, p. 196-198). A rough estimate is that variances need to differ by a factor of two to be considered statistically significant.

When the residuals do not exhibit constant variance, then 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 (i.e., constant CV). Square root transformations are used when the variance is proportional to the mean of Y. 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 which helps spread out the values near 0 and 1 to increase their variance (Snedecor and Cochran, 1967).

B.4.4 Tests of randomness

Freund (1973, p. 382) describes a test of randomness in which the total number of runs (u) above and below the median are analyzed. A run is a string of values all above or all below the median. 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 series of sample values:

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 ten 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 is:

$$z = \frac{u - \mu_u}{\sigma_u} \qquad (8.13)$$

where

$$\mu_u = \frac{2n_1n_2}{n_1 + n_2} + 1 \qquad (8.14)$$

and

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \qquad (8.15)$$

Using the example data set above:

$$\mu_u = 11 \quad (n1=n2=10)$$

$$\sigma_u = 2.1764$$

$$z = (8 - 11)/2.1764 = -1.38$$

Using a level of significance (α) of 0.05 in a two-tailed test, the z values ($\alpha/2$) are 1.96 and -1.96 (Appendix B). Since -1.38 falls within this range we accept the null hypothesis that the sample is random.

Spearman's rho is another common test for evaluating correlation. If the data does not have ties, the following equation is recommended (Conover, 1980):

$$p = \frac{\sum_{i=1}^{n}\left[R(X_i) - \frac{n+1}{2}\right]\left[R(Y_i) - \frac{n+1}{2}\right]}{n(n^2-1)/12} \qquad (8.16)$$

where $R(\cdot)$ is the rank of the data. If there are ties in the data set, the following equation should be used:

$$p = \frac{\sum_{i=1}^{n} R(X_i)\,R(Y_i) - n\left[\frac{n+1}{2}\right]^2}{\left[\sum_{i=1}^{n} R(X_i)^2 - n\left[\frac{n+1}{2}\right]^2\right]^{0.5}\left[\sum_{i=1}^{n} R(Y_i)^2 - n\left[\frac{n+1}{2}\right]^2\right]^{0.5}} \qquad (8.17)$$

Table VIII-9 presents an example analysis using the same data as above. The first two columns correspond to the chronological order of sampling and the observations, respectively. The first column could also correspond to another variable such as rainfall while the second column could correspond to runoff. As a result, the first column may not always be ordered. $R(X_i)$ and $R(Y_i)$ correspond to the ranks of X_i and Y_i, respectively. If observations are tied, they are assigned the average of the associated ranks. For

example, an observation of 5 is the 6th and 7th smallest observations. As a result, $R(X_i=5)$ is equal to 6.5.

Table VIII-9. Example application of Spearman's rho test for evaluating independence.

Y_i	X_i	$R(X_i)$	$R(Y_i)$	$R(X_i)R(Y_i)$	$R(X_i)^2$	$R(Y_i)^2$
1	5	6.5	1	6.5	42.3	1
2	5	6.5	2	13.0	42.3	4
3	6	8.5	3	25.5	72.3	9
4	9	12.5	4	50.0	156.3	16
5	13	18.5	5	92.5	342.3	25
6	12	16.5	6	99.0	272.3	36
7	2	2.5	7	17.5	6.3	49
8	3	4	8	32.0	16.0	64
9	2	2.5	9	22.5	6.3	81
10	8	11	10	110.0	121.0	100
11	14	20	11	220.0	400.0	121
12	13	18.5	12	222.0	342.3	144
13	11	14.5	13	188.5	210.3	169
14	20	21	14	294.0	441.0	196
15	4	5	15	75.0	25.0	225
16	6	8.5	16	136.0	72.3	256
17	9	12.5	17	212.5	156.3	289
18	1	1	18	18.0	1.0	324
19	7	10	19	190.0	100.0	361
20	11	14.5	20	290.0	210.3	400
21	12	16.5	21	346.5	272.3	441
			SUM	2661.0	3307.5	3311

$$n\left[\frac{n+1}{2}\right]^2 = 21\ (22/2)^2 = 2,541$$

Applying Equation 8.17,

$$p = \frac{2,661 - 2,541}{(3,307.5-2,541)^{0.5}\ (3,311-2,541)^{0.5}} = 0.156$$

The last three columns provide intermediate calculations necessary to apply Equation 8.17. The test statistic is computed as 0.156. Using Appendix S, the appropriate quantile (95 percent confidence level) is 0.3688. The null hypothesis is accepted that the data are independent.

The analyst may also use the correlation coefficient, r. When the analyst is comparing a time series of data such as that displayed in Figure VIII-1, r is referred to as the autocorrelation coefficient. Salas et al. (1980) provides the formula for the autocorrelation coefficient as:

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})\ (x_{t+k} - \bar{x})}{\sum_{t=1}^{n-k} (x_t - \bar{x})^2} \qquad (8.18)$$

Anderson (1941) gave the limit

$$r_k = \frac{-1 \pm 1.96\ (n-k-1)^{0.5}}{n-k} \qquad (8.19)$$

for the 95 percent probability levels where n is the sample size.

The correlation coefficient ranges from -1 to 1. Values close to 1 mean that there is a strong positive correlation (e.g., high values follow high values). Values close to -1 mean that there is a strong negative correlation (e.g., high values follow low values).

To apply Equation 8.18 to the dissolved oxygen data from Figure VIII-1, some preprocessing is necessary. Due to the lack of data in December, January, and March, these data are dropped from the analysis. Therefore, there nine seasons (i.e., months). In addition, only the first observation from each month and year are used. Figure VIII-10 presents the autocorrelation coefficient (correlogram) calculated for lags 1 through 20. The figure shows a strong cyclic pattern that repeats every 9 lags. This corresponds to the number of months used in the analysis. There is also a very strong correlation at lag 1 which suggests that there is a tendency for the dissolved oxygen in the following month to be similar to the current months dissolved oxygen data. The negative correlation at lag 4 suggest also suggests a strong correlation with data 4 months apart.

B.5 Probability Distributions (optional reading)

Before discussing probability distributions, it is important to review some of the central concepts upon which our statistical analyses are based. The following definitions describe briefly some of the terms that will be used in the following sections (Remington and Schork, 1970, p. 98-109):

> *A <u>random variable</u> is a quantity that takes various values or sets of values with various probabilities.*
>
> *A table, graph, or mathematical expression giving the probabilities with which a random variable takes different values or sets of values is called the <u>distribution of the random variable.</u>*

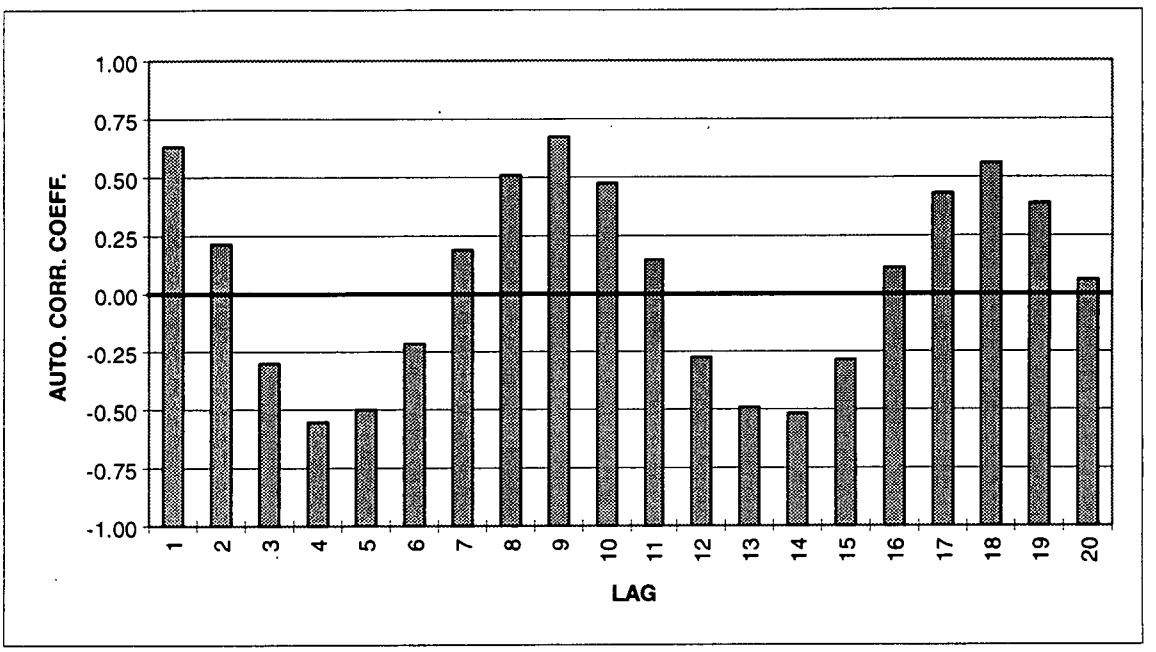


Figure VIII-10. Autocorrelation (correlogram) for dissolved oxygen data from 1980 through 1989 (March-November) for the Delaware River at Reedy Island, Delaware.

The distribution of a statistic over all random samples of size n from an underlying population is called the sampling distribution of the statistic for random samples of size n.

The standard deviation of the sampling distribution of a statistic for random samples of size n is called the standard error of the statistic.

A simple random sample of size n from a population of size N is a sample selected in such a way that every group of n different units has the same probability of being selected as the sample.

Suppose a population is divided into subpopulations or strata. If a simple random sample is selected from each of these strata, the aggregated sample is called a stratified random sample.

As stated by Ponce (1980a) theoretical probability density functions can be used as models for samples from a population. Thus, by utilizing selected sample statistics in conjunction with a probability density function we can make generalizations about the population from which the sample was taken. Freund (1973) noted that "whenever possible, we try to express probability functions by means of formulas which enable us to calculate the probabilities associated with the various values of a random variable." The sum of probabilities across all values must equal 1, and the probability of any value must be either positive or zero.

Ponce (1980a) stated that if a sample can be fitted to a theoretical probability density function, then we can estimate errors of the population parameters, compare temporal and spatial changes in frequency, and examine the effect of environmental factors and management practices. What are these theoretical probability density functions?

B.5.1 Discrete data distributions

For discrete data there are the binomial and Poisson distributions (Ponce, 1980a). The binomial distribution is appropriate for data sets in which there are two possible outcomes (e.g., heads or tails, true or false, success or failure). Freund (1973) determines the probability of getting x successes in n independent trials with:

$$f(x) = np^{x}(1 - p)^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots, n \qquad (8.20)$$

where p is the constant probability of a success for each individual trial. For more information about binomial distributions consult a statistical text such as the ones referenced in this manual.

Freund (1973) noted that binomial probabilities with a large n and a small p are often approximated with:

$$f(x) = \frac{(np)^x e^{-np}}{x!} \quad \text{for } x = 0, 1, 2, 3, \ldots \qquad (8.21)$$

where x! is factorial notation and e is "approximately" 2.71828. This equation describes the Poisson distribution. Remington and Schork (1970) replace np with the mean of the distribution (μ); thus, while the binomial distribution requires knowledge of n and p, the Poisson distribution requires only knowledge of μ. The reader is again referred to statistics texts for additional information.

B.5.2 Continuous data distributions

Ponce (1980a) refers to the normal, t, chi-square, and F distributions as the most important theoretical probability density functions for continuous or measured data. The following discussion is intended to introduce the reader to the t, chi-square and F distributions while devoting more attention to the normal distribution. The t, chi-square, and F distributions will be discussed in further detail in subsequent sections.

B.5.2.1 Normal distribution

The normal distribution plays a major role in most statistical analyses performed by researchers. Most of the descriptive and inferential statistics we use are based on the assumption of a normally distributed sample data set. Remington and Schork (1970) explain the importance of the normal distribution in terms of its:

1. convenient mathematical properties leading directly to much of the theory of statistics available as a basis for practice

2. availability as an approximation to other distribution

3. direct relationship to sample means from virtually any distribution

4. application to many random variables that either are approximately normally distributed or can be easily transformed to approximate normal variables.

The graph of a normal distribution is a bell-shaped curve that extends indefinitely in both directions (Freund, 1973). Figure VIII-11 illustrates the single peak or mode at $x=\mu$ and the symmetry about $x=\mu$ (Remington and Schork, 1970). Figures VIII-12 and VIII-13 show that changes in μ and σ, respectively, will alter the normal distribution, which can have an unlimited number of different appearances. It should be noted that the normal distribution is completely determined by its mean and standard deviation (Freund, 1973). The equation for the normal curve is (Ponce, 1980a):

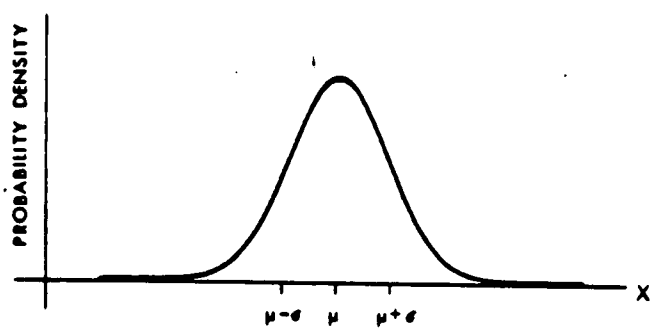


Figure VIII-11. A normal distribution. (SOURCE: Remington and Schork, 1970)

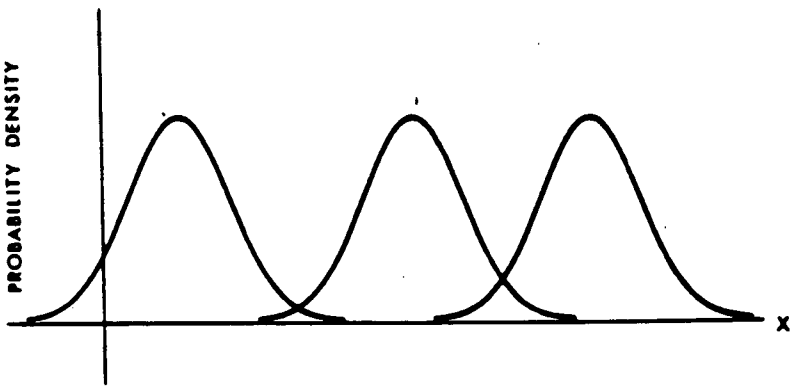


Figure VIII-12. The effect of changing μ in the normal distribution while keeping σ constant. (SOURCE: Remington and Schork, 1970)

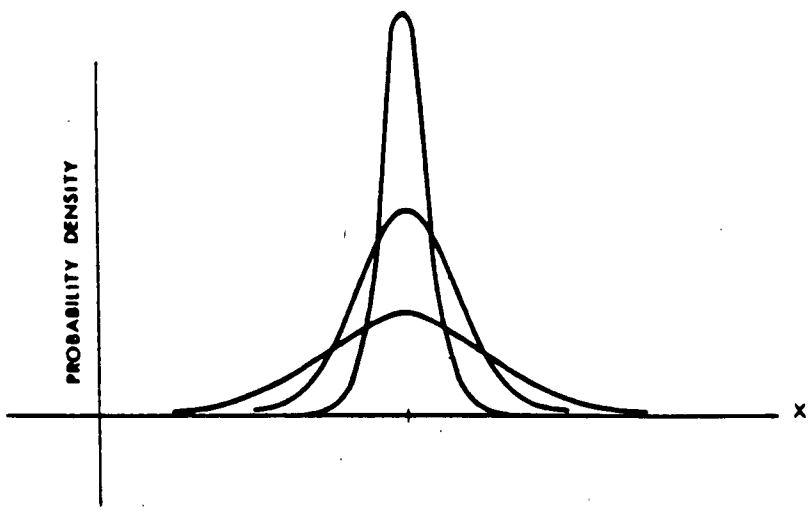


Figure VIII-13. The effect of changing σ in the normal distribution while keeping μ constant. (SOURCE: Remington and Schork, 1970)

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2(x-\mu/\sigma)^2} \qquad (8.22)$$

where y is the frequency of observations of a given x, π is 3.1416, e is 2.71828, μ is the population mean, and σ is the population standard deviation.

Due to the unlimited number of possible normal distributions, a convention was decided upon for defining the standard normal curve. Hence, as shown in Figure VIII-14, the standardized normal distribution is a bell-shaped curve with a mean (μ) of zero and a variance (σ^2) of one (Gaugush, 1986). To create this standard normal distribution, the z statistic was defined as $(x-\mu)/\sigma$ (Remington and Schork, 1970). Substituting z into the above equation and using a standard deviation (σ) of 1, we obtain:

$$y = \frac{1}{\sqrt{2\pi}} \bullet e^{-z^2/2} \qquad (8.23)$$

In Figure VIII-14 the z-scale shows the number of standard units from the mean of zero. For the standard normal curve 68.3 percent of the observations lie within plus or minus one standard unit (z=1 and z=-1) of the mean, while 95.5% of the observations lie within two standard units (z=2 and z=-2) of the mean (Ponce, 1980). Tables such as that in Appendix B can be used to calculate the area under the curve (probability) for other values of z. This table can be used for normal distributions where the mean is not 0 and the standard deviation is not one by converting x values into z values as described above for the z statistic (Freund, 1973). To calculate the area between (i.e., the probability that a random variable will assume a value between) any two values of x, one simply has to convert the x values to z values (using μ and σ), determine the areas corresponding to the z values (Appendix B), and add (one z positive, one z negative) or subtract (each z positive) these areas. For further discussion of the normal and standard normal distributions consult statistics texts; Ponce (1980a) demonstrates applications of the normal distribution.

B.5.2.2 t-distribution

In the above equation for the standard normal distribution we used the population mean (μ) and the population standard deviation. What do we do when we don't have population statistics, i.e., when we have only sample statistics? The logical approach is to use our sample statistics to approximate the population statistics.

To understand how to apply sample statistics in estimating population statistics it is first useful to discuss sampling distributions. The following two theorems provide nearly all we need to know about the sampling distribution of the mean (Freund, 1973):

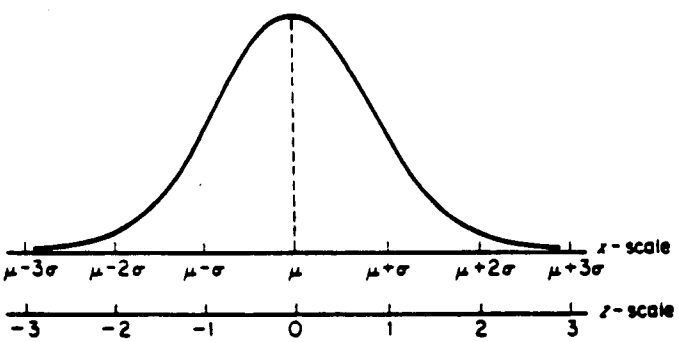


Figure VIII-14. Change of scale. (SOURCE: Freund, 1973)

For random samples of size n from a population having the mean μ and the standard deviation σ, the (theoretical) sampling distribution of x has the mean $\mu_x = \mu$, and its standard deviation is given by

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \text{ or } \sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{(N-n)/(N-1)} \qquad (8.24)$$

depending on whether the population is infinite or finite of size N. If n is large, the (theoretical) sampling distribution of the mean can be approximated closely with a normal distribution.

Taking this a step further, if n is large (greater than 30), the sampling distribution of the z statistic can be approximated closely with the standard normal distribution (Freund). The need for use of the t-distribution arises when one attempts to estimate the error associated with using x to estimate μ, a subject to be discussed in more detail in subsequent sections. For this application we must know the population standard deviation (σ), but only have an estimate, namely the sample standard deviation (s). The equation used to estimate the error associated with an estimate of the mean involves the quantity

$$t = \frac{x-\mu}{s/\sqrt{n}} \qquad (8.25)$$

which does not follow the standard normal distribution (Remington & Schork, 1970). Instead, it follows the t-distribution (Figure VIII-15) which depends solely on the sample size as reflected in the number of degrees of freedom (n-1) (Freund, 1973). The t-distribution is similar to the normal distribution in that it is symmetrical, has a mean of zero, and approximates the normal curve as n gets large (>30). However, as shown in Figure VIII-15, the peak is not as high and the tails are larger than for the standard normal curve. Appendix C is a table of t values to be used in calculations using the t statistic.

B.5.2.3 Chi-square distribution

The chi-square distribution (Figure VIII-16) is commonly used to estimate the confidence intervals for the standard deviation (σ) of a population with a normal distribution based upon a sample standard deviation (s). The equation for the chi-square (X^2) statistic is (Freund, 1973):

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \qquad (8.26)$$

where (n-1) is the mean of the distribution and is referred to as the number of degrees of freedom. As the equation indicates, this distribution does not include negative numbers. Calculations using

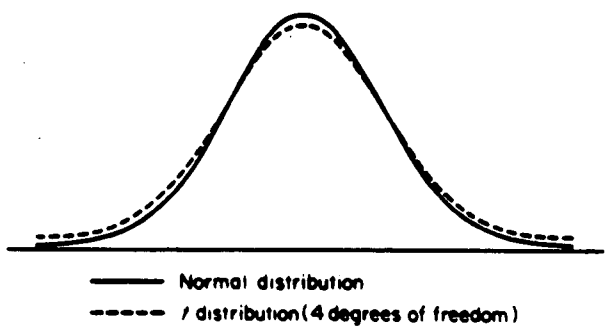


Figure VIII-15. Standard normal distribution and example of t-distribution. (SOURCE: Freund, 1973)

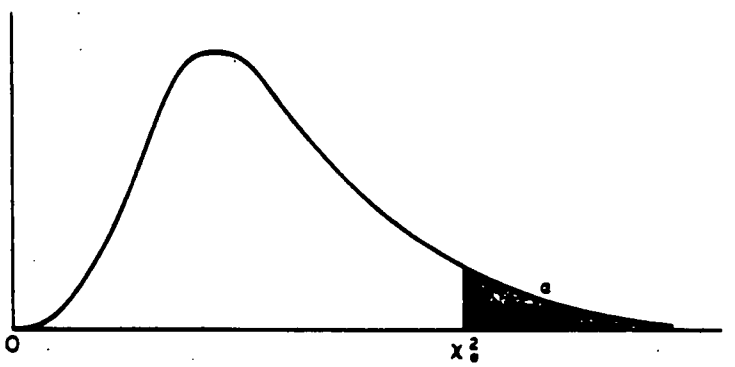


Figure VIII-16. Chi-square distribution. (SOURCE: Freund, 1973)

the chi-square distribution will be discussed in subsequent sections; Appendix D is a table of values for this distribution.

B.5.2.4 F distribution

The F distribution (Figure VIII-17) is a continuous distribution which represents the sampling distribution of the ratio of the sample variances (s^2) of two normal distributions (Freund, 1973). The F distribution depends upon two parameters called the numerator (n_1-1) and denominator (n_2-1) degrees of freedom (Remington and Schork, 1970). The F distribution is similar to the chi-square distribution in that both contain no negative numbers and both are positively skewed. Applications of the F distribution will follow; Appendix E contains F values.

C. Summary (Descriptive) Statistics

Rather than present every data element within a data set, the data analyst usually summarizes the major numeric characteristics of the data set with a few descriptive statistics. As stated by Freund (1973, p.3) descriptive statistics "includes any treatment designed to summarize, or describe, important features of a set of data without going any further; that is, without attempting to infer anything that pertains to more than the data themselves." In the case of water quality monitoring we almost invariably use descriptive statistics of samples to formulate conclusions regarding populations. Thus, the analyst carries the investigation a step further by making generalizations which go beyond the data. A conclusion concerning a population of observations made on the basis of a sample of observations is called a statistical inference (Remington and Schork, 1970, p. 91). This section describes methods that can be used to characterize the average conditions. Sections VIII.D and VIII.E describe methods for characterizing changing and extreme conditions, respectively.

A point estimate is a single number representing the unknown parameter (Freund, 1973). Such point estimates for central tendency could be the mean, median, mode, or geometric mean from a sample. The sample standard deviation and interquartile range could likewise be used as point estimates of spread or variability of the population. This manual recommends summarizing data using the sample mean, median, geometric mean (if applicable), standard deviation, and the interquartile range (25th percentile subtracted from the 75th percentile) for characterizing average conditions in most situations.

The use of point estimates is warranted in some cases, but in NPS analyses point estimates should be coupled with an interval estimate because of the large spatial and temporal variability of NPS pollution. In estimating the true population mean and standard deviation from any one sample or combination of samples we have to account for the uncertainty associated with use of sample statistics. As stated by Freund (1973), "point estimates do not tell us anything about the intrinsic reliability or precision of the method of estimation which is being used." This manual

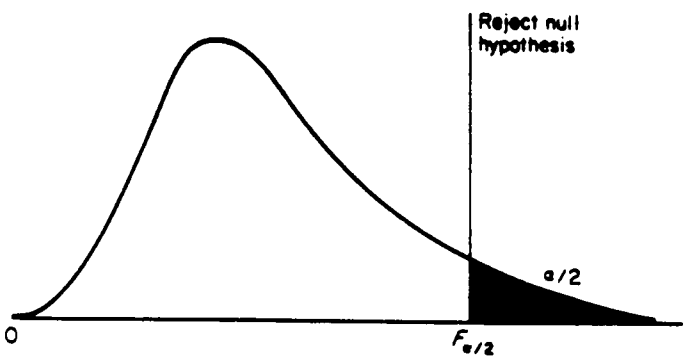


Figure VIII-17. F distribution. (SOURCE: Freund, 1973)

recommends that at a minimum, the data analyst should provide confidence intervals for the point estimates of central tendency (i.e., mean or median).

After the methods for point estimates and confidence intervals have been implemented, the data analyst may be discouraged with interval estimates that are too large to allow for effective environmental management. This section concludes with a discussion on sample size calculations. This analysis is usually performed during the monitoring program design; however, it is sometimes necessary to modify an existing program in order to achieve the desired monitoring objectives.

C.1 Point Estimation

C.1.1 Central Tendency

The central tendency of a data set is the most important statistic and it is widely used (Gaugush, 1986, p. 21; Ponce, WS1980a, p. 5). Remington and Schork (1970) emphasize that it is important to realize that measures of central tendency apply to groups rather than to individuals. The mean, median, and mode are the three major measures of central tendency. The arithmetic mean is the sum of the observations divided by the number of observations (n). The median (i.e., 50th percentile or middle quartile) is the middle value when all observations are ordered by magnitude. The median is the horizontal line inside the box of a box and whisker plot. When there is an even number of observations, the median is the arithmetic mean of the two middle observations. 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 (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 mean and median event may be very different. It is important that the data analyst considers the ramifications of relying on just one of these statistics when reporting results. This manual recommends that when presenting tabular information, the mean and the median should be computed.

Other measures of central tendency include the midrange, geometric mean, harmonic mean, and weighted mean (Remington and Schork, 1970). The midrange is the arithmetic mean of the smallest and largest values and, like the mean, is influenced by extreme values. Gilbert (1987) also describes the trimmed mean and the winsorized mean for when there are censored (e.g., non-detects) in the data set.

The geometric mean of n observations is the n^{th} root of their product. It can also be calculated as the antilog of the means of the logarithms of the individual observations (natural or base 10). All observations must be positive to calculate this statistic.

Gaugush (1986) states that the geometric mean is "a reasonable measure of central tendency for a set of data that exhibit a lognormal distribution." It is common to report the geometric mean for coliform data.

$$\text{Geometric mean} = \text{antilog} \frac{\sum_{i=1}^{n} \log (x_i)}{n} \qquad (8.27)$$

The harmonic mean is "the reciprocal of the arithmetic mean of the reciprocals of the observations" (Remington and Schork, 1970). Freund (1973, p. 46) refers to the harmonic mean as n divided by the sum of the reciprocals of the n numbers. It has become common practice to estimate the harmonic mean flow for performing chronic risk assessments.

$$\text{Harmonic mean} = \frac{n}{\sum_{i=1}^{n} 1/x_i} \qquad (8.28)$$

The weighted mean is a mean for which all observations do not have equal importance. For example, when we want to determine the overall mean of several sets of data on the basis of their individual means and the number of observations in each we calculate the weighted mean as the sum of the products of the individual means and number of observations divided by the overall sum of the observations (Freund, 1970, p. 42). In this case the number of observations is the weight variable (w_i). This may occur when the monitoring program has used a stratified sampling strategy and the strata have different sample sizes.

$$\text{Weighted mean} = \frac{\sum_{i=1}^{n} w_i \bullet x_i}{\sum_{i=1}^{n} w_i} \qquad (8.29)$$

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). (Censored observations are results that are presented as less-than or not-detected.) The trimmed mean is equal to the arithmetic mean after equal proportions of the smallest and largest observations are 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, then no more than 30 percent of the data should be dropped (Mosteller and Rourke, 1973).

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. Gilbert (1987, p. 180) provides detailed instructions on how to compute the associated standard deviation and confidence limits.

C.1.2 Dispersion

Measures of dispersion or measures of variation describe extent to which the data are dispersed, spread out, or bunched (Freund, 1970, p. 63). The measures of dispersion described in this manual are the range, variance, standard deviation, and interquartile range.

The range of a set of observations is simply the difference between the largest and smallest values. Ponce (1980, WSDG-TP-00001, p. 6) argues that the range should only be considered as a rough estimate of dispersion due to its dependence upon extreme values. Remington and Schork (1970, p. 30) agree and point out that the range is also dependent upon the number of observations in a sample (larger samples have a greater chance of containing extreme values), an undesirable characteristic for a measure of variability. Gaugush (1986, p. 26) states clearly that "the range should not be relied upon as the sole indicator of variability."

The variance (s^2) is the sum of the squares of the deviations from the mean divided by n-1 (Freund, 1973, p. 67). Deviations from the mean are the differences between individual observations and the arithmetic mean of all observations. The standard deviation (s) is the square root of the variance (Remington and Schork, 1970, p. 34). For a normal distribution, about 68 percent of the data are within ± one standard deviation of the mean.

$$\text{Variance} = s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} \qquad (8.30)$$

Standard deviation = s

In cases where it is necessary to compare standard deviations (e.g., comparing the precision of measurements) we need a measure of relative variation. Freund (1973) recommends the coefficient of variation (CV) as the "most widely used measure of relative variation." The CV is defined as:

$$CV = 100\ (s/\bar{x}) \quad \text{or} \quad CV = 100\ (\sigma/\mu) \qquad (8.31)$$

Freund (1973) states that the CV can be used to compare the dispersions of two or more sets of data that are measured in

different units. The analyst should be careful in presenting the results for the CV, since some analysts do not multiply by 100.

The interquartile range is a robust (changes little if any particular observation is deleted from the sample) alternative to the standard deviation (like the median is to the mean) for situations where the standard deviation is not an appropriate measure of dispersion (Gaugush, 1986, p. 26). It is the difference between the observation at the upper quartile, Q_3, and the observation at the lower quartile, Q_1. 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. For n equal to 9, the lower quartile is equal to the 2.5th ordered observation or the interpolated value between the second and third observation. Similar to the CV, the coefficient of quartile variation can be used to compare different data sets.

$$V = 100 \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (8.32)$$

C.1.3 Skewness

Skewness is a measure of the symmetry of a distribution (Remington and Schork, 1970, p. 219). The normal distribution is only one of many symmetric distributions (Figure VIII-18). A positively and negatively skewed distribution are shown in Figure VIII-19. Section B.4.1.1 provides computational formulas for skewness.

C.1.4 Kurtosis

The kurtosis of a distribution describes its peakedness relative to the length and size of its tails (Remington and Schork, 1970, p. 223). It has been argued, however, that kurtosis measures tail heaviness, not the peakedness of a distribution (SAS Institute, Inc., 1985a, p. 741). The normal distribution is considered to have intermediate kurtosis (mesokurtic). Flat distributions with short tails have low kurtosis (platykurtic), while distributions with sharp peaks and long tails have high kurtosis (leptokurtic). These types of distributions are shown in Figure VIII-20. Section B.4.1.2 provides computational formulas for kurtosis.

C.2 Interval Estimation

In taking random samples from a fixed population the sampler will be obtaining fixed values for sample parameters (mean,

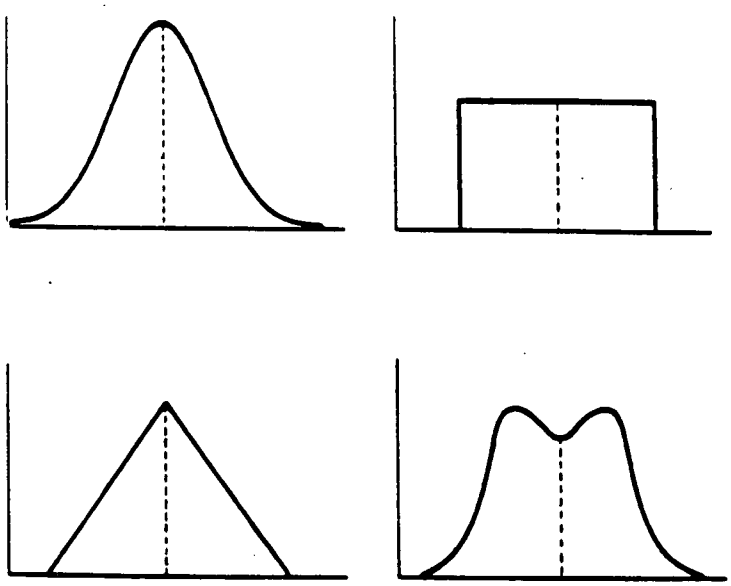


Figure VIII-18. Some examples of symmetric distributions. (SOURCE: Remington and Schork, 1970)

Figure VIII-19. Examples of skewed distributions. (SOURCE: Remington and Schork, 1970)

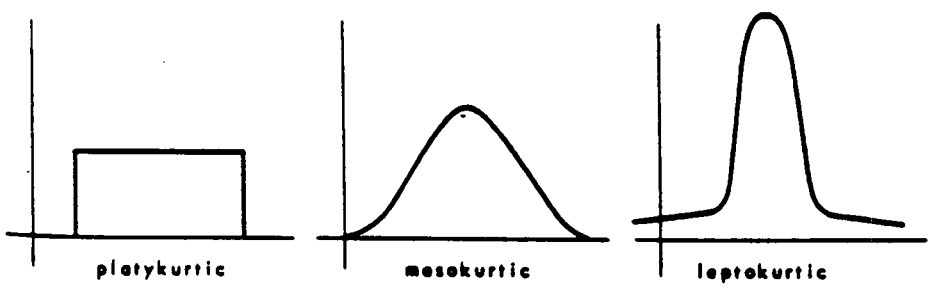


Figure VIII-20. Symmetric distributions of varying kurtosis. (SOURCE: Remington and Schork, 1970)

standard deviation, etc.) that fall within a fixed distribution of sample parameter values (Remington and Schork, 1970). In estimating the true population mean and standard deviation from any one sample or combination of samples we have to account for the uncertainty associated with use of sample statistics.

C.2.1 Interval Estimation for the Mean

To estimate the confidence intervals for the mean, it is necessary to estimate the standard deviation of the sample mean:

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \text{ or } \sigma_x = \frac{\sigma}{\sqrt{n}} \bullet \sqrt{(N-n)/(N-1)} \qquad (8.33)$$

The standard deviation is a function of the sample standard deviation, the sample size n, and whether the population is infinite or finite of size N.

For large n (>30), the sampling distribution of the mean can be approximated closely with the normal curve (Freund, 1973). Thus, by knowing the sample mean and population standard deviation we can make an interval estimate for the population mean. By using the above equation, the confidence interval can be computed as:

$$\mu = \bar{x} \pm z_{\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} \qquad (8.34)$$

where $z_{\alpha/2}$ is the statistic from the standard normal curve (Appendix B). Figure VIII-21 illustrates how values of z and the confidence $(1-\alpha)$ of the interval estimate are related (Freund, 1973). By selecting a desired confidence level (0.90, 0.95, and 0.99 are the most common) for the estimate, a z value from Appendix B (column C) is defined.

We can obtain a value of x from the sample, but the standard deviation of the population is generally unknown. Hence, the above equation applies only to the confidence interval for the mean of a normal distribution with known variance (Remington and Schork, 1970).

When the population variance is not known we have to use the sample variance (s^2) as a substitute. Thus, the confidence interval for the mean of a normal distribution with unknown variance can be determined as (Freund, 1973, p. 255):

$$\mu = \bar{x} \pm t_{\alpha/2} \bullet \frac{s}{\sqrt{n}} \qquad (8.35)$$

where $t_{\alpha/2}$ is determined by the sample size and the desired confidence $(1-\alpha)$ for the interval estimate (Appendix C). Because one additional statistic (i.e., standard deviation) was estimated from the data, the degrees of freedom used in Appendix C is equal

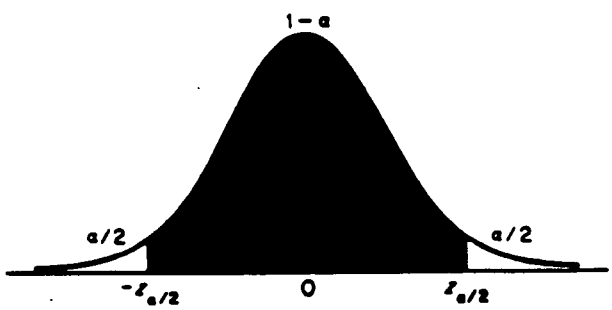


Figure VIII-21. Normal distribution. (SOURCE: Freund, 1973)

to the n-1. For large sample sizes the value of $t_{\alpha/2}$ approaches $z_{\alpha/2}$ in Appendix B.

The following example of setting confidence limits for the population mean is based upon an example given by Ponce (1980, 00001, p. 48-49).

Problem:

Fifty-four (54) 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 limits for the population mean, μ.

Solution:

For the 95 and 99 percent confidence limits, $\alpha/2$ is equal to 0.025 and 0.005, respectively. The t-value is then estimated by interpolation between the values for 50 and 60 degrees of freedom (Appendix C) using the columns $t_{0.975}$ and $t_{0.995}$, respectively. We estimated t-values of 2.0061 and 2.6726.

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

$$\mu = X \pm t_{.025(53)} \bullet s/\sqrt{n}$$

$$\mu = 0.0265 \pm 2.0061 \bullet (0.0040/\sqrt{54}) = 0.0265 \pm 0.0011$$

$$0.0254 < \mu < 0.0276$$

There is a 95 percent chance that the population mean, μ, will be covered by the interval 0.0254 to 0.0276.

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

$$\mu = X \pm t_{.005(53)} \bullet s/\sqrt{n}$$

$$\mu = 0.0265 \pm 2.6726 \bullet (0.0040/\sqrt{54}) = 0.0265 \pm 0.0015$$

$$0.0250 < \mu < 0.0280$$

There is a 99 percent chance that the population mean, μ, will be covered by the interval 0.0250 to 0.0280.

C.2.2 Interval estimation for the standard deviation

The chi-square distribution is used to estimate the confidence intervals for the standard deviation of a normal distribution. Referring to Figure VIII-16, we establish the small sample confidence interval for the standard deviation of a normal distribution as (Freund, 1973, p. 290):

$$\left[\frac{(n-1)\bullet s^2}{X^2_{\alpha/2}}\right]^{1/2} < \sigma < \left[\frac{(n-1)\bullet s^2}{X^2_{1-\alpha/2}}\right]^{1/2} \qquad (8.36)$$

Note that since the chi-square distribution is asymmetric the above inequality requires a different chi-square value for each end of the confidence interval, i.e., values for $\alpha/2$ and $(1-\alpha/2)$. Remember that $1-\alpha$ is the confidence level for the estimation interval.

For large samples ($n \geq 30$) Freund (1973) notes that the sampling distribution of s approximates a normal distribution with a mean of σ and a standard deviation of $\sigma/\sqrt{2n}$. Thus, for large samples the following formula may be used (Freund, 1973, p. 291):

$$\frac{s}{1 + \frac{z_{a/2}}{\sqrt{2n}}} < \mu < \frac{s}{1 - \frac{z_{\alpha/2}}{\sqrt{2n}}} \qquad (8.37)$$

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

C.2.3 Interval estimation for the median and quartiles

Although several approaches exist to estimate confidence levels for any percentile, many rely on assuming a normal or lognormal distribution. The approach presented here (Conover, 1980, p. 111) for more than 20 observations does not rely on these assumptions. Conover (1980) provides a procedure for smaller sample sizes.

- Order the data from smallest to largest observation such that

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

  where x_p corresponds to the corresponding quantile (e.g., median = $x_{0.50}$).

- 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 Appendix B.

- Round r^* and s^* up to the next highest integers r and s. The $1-\alpha$ 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 dissolved oxygen data presented in Figure VIII-3.

Solution:

- From Figure VIII-3, the median observation can be computed as 7.6 mg/L. There are 182 observations. Note that Figure VIII-3 already presents the raw data from smallest to largest observation.

- r' and s' may then be computed as

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

$$= 182 \times 0.5 - 1.645 (182 \times 0.5 \times 0.5)^{0.5}$$

$$= 79.9$$

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

$$= 182 \times 0.5 - 1.645 (182 \times 0.5 \times 0.5)^{0.5}$$

$$= 102.1$$

- r and s are therefore 80 and 103, respectively. From Figure VIII-3, x_{80} and x_{103} can be estimated as 7.3 and 8.0 mg/L, respectively.

C.3 Sample Size Calculation

A natural extension of Section C.2 is to estimate the number of samples needed estimate the mean within a prescribed relative error, d_r when the observations are independent (Gilbert, 1987).

$$n = (z_{\alpha/2} \, CV / d_r)^2 \qquad (8.38)$$

$$d_r = | \bar{x} - \mu | / \mu \qquad (8.39)$$

Table VIII-10 summarizes the sample sizes required for estimating the true mean for a variety of CV and relative error. The analyst can use these formulas for suggesting modifications to an existing monitoring program if the current relative errors are too large.

Table VIII-10. Sample Sizes Required for Estimating the True Mean (after Gilbert, 1987).

Confidence Level $(1-\alpha)$	Relative Error d_r	Coefficient of Variation				
		0.10	0.50	1.00	1.50	2.00
0.80	0.10	2	42	165	370	657
	0.25	-	7	27	60	106
	0.50	-	2	7	15	27
	1.00	-	-	2	4	7
	2.00	-	-	-	-	2
0.95	0.10	4	97	385	865	1,537
	0.25	-	16	62	139	246
	0.50	-	4	16	35	62
	1.00	-	-	4	9	16
	2.00	-	-	-	3	4

D. Trend Testing

Evaluating the effectiveness of controls and changing environmental conditions is one of the key objectives of monitoring programs. For this reason statistical analysis of monitoring data usually involves hypothesis testing for trends. In general, there are two basic types of trend testing: step and monotonic. Step trends are typically evaluated when comparing at least two different sample populations such as site and reference biological data or one sample population to an action level. Monotonic trends are evaluated when the analyst is investigating long-term gradual changes over time.

D.1 Hypothesis Testing

The null hypothesis (Ho) is the root of hypothesis testing. Traditionally, null hypotheses are statements of "no change," "no effect," or "no difference." Remington and Schork (1970) prefer the term "tested hypothesis" since these hypotheses can take the form of expected changes, effects, or differences. The alternative hypothesis (Ha) is the counter to the null hypothesis, traditionally being statements of change, effect, or difference. That is, upon rejection of Ho stating no change we would accept the Ha of change. We could, however, state an Ho of the type "change of at least 10 percent" with an Ha of the type "no change of at least 10 percent." The choice is left to the researcher.

The following are examples of hypotheses formulated by participants in the joint USDA/USEPA Rural Clean Water Program (RCWP) (Jamieson, 1986):

- Best Management Practices (BMPs) will reduce total suspended solids (TSS) concentrations to 25 mg/L.

- BMPs will reduce phosphate (PO_4) by 50 percent from pre-implementation levels.

- Reduced surface application of fertilizer will not reduce nitrate (NO_3) concentrations in the soil profile.

For further discussion of basic concepts in hypothesis testing the reader is encouraged to consult text books such as Winer (1971) and Snedecor and Cochran (1980).

D.1.1 Type I and II Errors

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 that is acceptable in the analysis. There are two types of errors in hypothesis testing:

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

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

Table VIII-11 summarizes these two error types, 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 at most equal to the level of significance (α) of the test. Winer (1971) emphasizes that the probability of making a Type I error is "controlled by the level of significance." In most cases, managers or analysts will define $1-\alpha$ to be in the range of 0.90 to 0.99 (e.g., a significance level of 90 to 99 percent) although there have been environmental applications where $1-\alpha$ was set to 0.80.

Table VIII-11. Errors in Hypothesis Testing

Decision	State of affairs in the population	
	Ho True	Ho False
Reject Ho	Type I error (α)	No error ($1-\beta$) "Power"
Accept Ho	No error ($1-\alpha$) "Signif. level"	Type II error (β)

Type II error "depends in part upon the significance level and in part upon which one of the possible alternative hypotheses is actually true." He notes that each alternative hypothesis has an associated Type II error of different magnitude.

The power of a test ($1-\beta$) is defined as the probability of rejecting the tested (null) hypothesis when it is false (Remington and Schork, 1970). This probability of correctly rejecting Ho should be as close to unity as possible, meaning that β should be minimized.

It is evident that there can be tradeoffs between minimizing α (maximizing the probability of accepting a true Ho) and minimizing β (maximizing the power). Remington and Schork (1970) state that, in general, for a fixed sample size, α and β vary inversely. For a fixed value of α, we can reduce β by increasing the sample size.

Figure VIII-22 illustrates the relationship between α and β using a one-tailed test that will be discussed later. Note that as α is decreased (fixed sample size) from 0.10 to 0.01 the area of overlap between Ha_1 and the acceptance region for Ho (i.e., β) increases. As the alternative hypothesis distribution moves farther away from Ho (i.e., move from Ha_1 to Ha_2) β decreases.

D.1.2 One-sided Versus Two-sided Tests

Depending on the type of null hypothesis, it is appropriate to select either a one- or two-sided test. For example, if the analyst knew that TSS could only go down as a result of BMP implementation then a one-sided test could be formulated.

Ho: TSS(Post) = TSS(Pre)
Ha: TSS(Post) < TSS(Pre)

Alternatively, if data have been collected in a pre- and post-BMP implementation and the analyst does not know whether TSS will go up or down then a two-sided test is necessary.

Ho: TSS(Post) = TSS(Pre)
Ha: TSS(Post) ≠ TSS(Pre)

When the researcher selects a one-sided test instead of a two-sided test there is an increase in the power with respect to Ha (Winer, 1971). That is, a corresponding one-sided test is more powerful, and therefore desirable, than a two-sided test given the same α and sample size. The manager and analyst should take great care in selecting one- or two-sided tests.

D.1.3 Selection of Nonparametric Procedures

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). These methods involve population parameter estimation and hypothesis testing and are appropriate when the underlying distributions are known and can be analyzed analytically. For data with unknown distributions, nonparametric methods might have to be used since these methods do not require

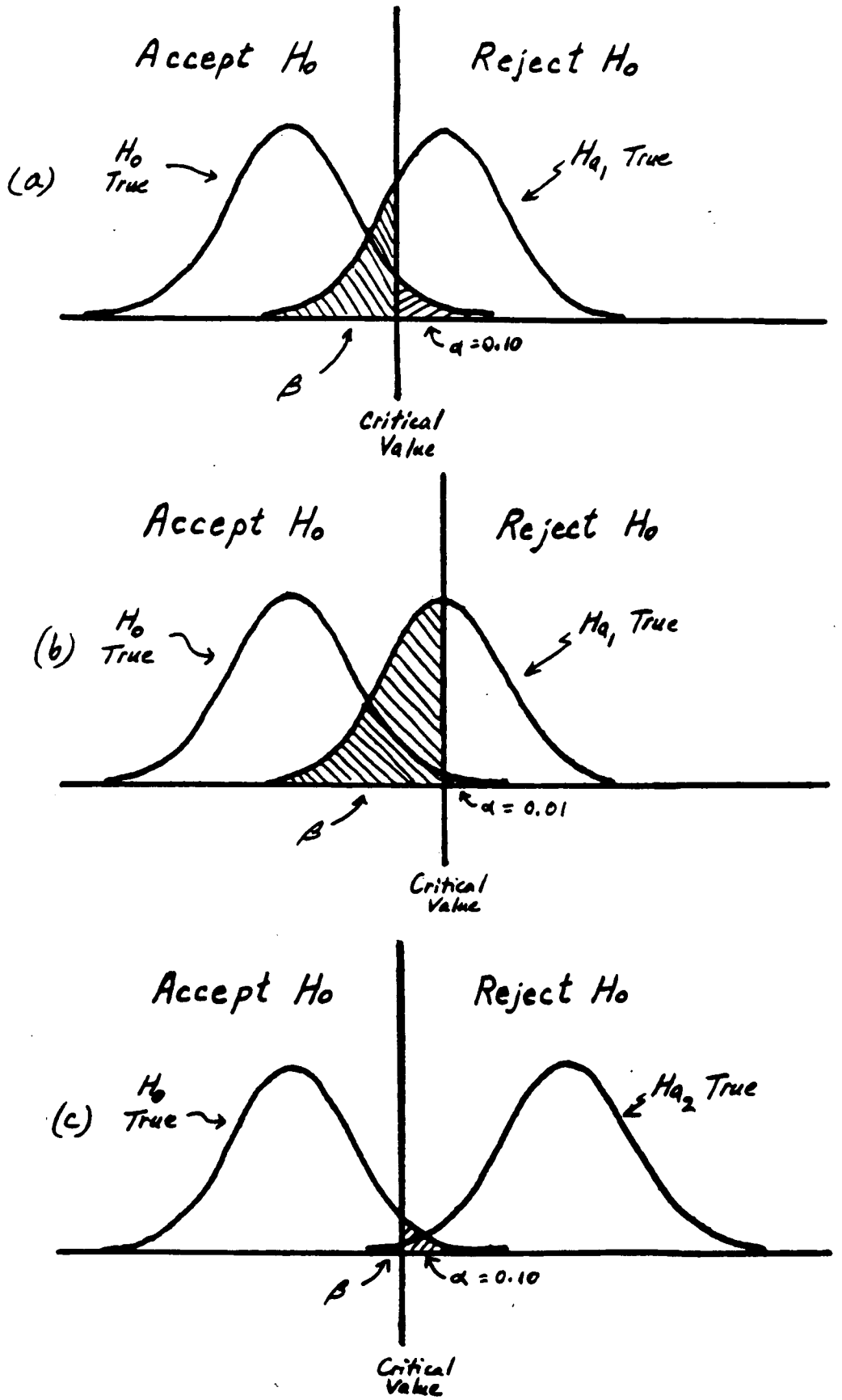


Figure VIII-22. Relationship between α and β.

that the data have a defined distribution; thus, these are often referred to as distribution-free methods.

Theoretically, nonparametric methods and distribution-free methods are different. In NPS analyses, however, nonparametric methods often stand for all distribution-free methods, and thus the terms "distribution-free statistics" and "nonparametric statistics" are used interchangeably. A more detailed discussion on the scope of nonparametric methods is provided by Hipel (1988).

Nonparametric methods can also handle special data commonly found in the NPS area, such as missing and censored data. Missing data can occur when an automatic sampling station is out of order for a short period of time, when samples are improperly handled, or when an error occurs in the laboratory analysis. Censored data are those without an exact numerical value, such as a value of "less than 10 ug/l" or "below detection limit." 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. Missing and censored data can cause problems in parametric methods because these methods often require all data with 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 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 detection limit" is unknown.

While nonparametric methods have several advantages, they are not as powerful as the parametric methods when the assumptions of the parametric procedure are met. Thus, when the underlying distributions of the data being analyzed are known or can be 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 small (Hipel, 1988).

D.2 Step Trends

D.2.1 One-sample tests

D.2.1.1 Student's t-test

Using the TSS hypothesis formulated by the RCWP participants from above, a one-sided hypothesis is developed to evaluate the post-implementation mean TSS concentrations as compared against a hypothesized value of 25 mg/L (Gaugush, 1986). A formalized statement of hypotheses would be:

$$H_o: \mu = 25$$
$$H_a: \mu < 25$$

In this case we are using the mean TSS concentration as the best measure for evaluation. We have also chosen to assume that any change in TSS is due solely to BMPs; this will not be true in most cases as will be shown in subsequent sections. Note that we are stating the alternative hypothesis (Ha) such that we can apply a one-sided test since we care specifically about whether the TSS is lower than 25 mg/L.

Data from the Highland Silver Lake RCWP project will be used to evaluate the above null hypothesis. The TSS data shown in Table VIII-12 are from May 21, 1981 through October 31, 1984. The period before March 16, 1983 will be called the pre-implementation period and the period after April 12, 1983 will be called the post-implementation period for this analysis.

Table VIII-12. Highland Silver Lake TSS Data for Site 1

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
3/16/83	42		

Overall: n = 31 mean = 24.77 s = 14.93 median = 20
Pre-Implementation: n = 16 mean = 29.38 s = 16.15 median = 23
Post-Implementation: n = 15 mean = 19.87 s = 12.17 median = 14

Before testing Ho with a statistical analysis we must inspect the data and test the assumptions of randomness and normality. We will perform these tests separately on the pre-implementation and post-implementation data sets. Using the SAS Univariate procedure (SAS Institute, Inc., 1985a), stem and leaf, boxplot, and normal probability plots were generated for the two data sets (Figures VIII-23 and VIII-24).

The values for skewness (0.82) and kurtosis (-0.42) indicate that there is positive skew (compare stem and leaf plot versus Figure VIII-19) and low kurtosis (leptokurtic) in the pre-BMP sample distribution. The Shapiro-Wilk W statistic (0.893) and associated probability (0.063) show that the null hypothesis that

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)

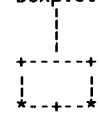
100% Max	60	99%	60
75% Q3	40.75	95%	60
50% Med	23	90%	60
25% Q1	17	10%	11.7
0% Min	11	5%	11
		1%	11
Range	49		
Q3-Q1	23.75		
Mode	20		

Extremes

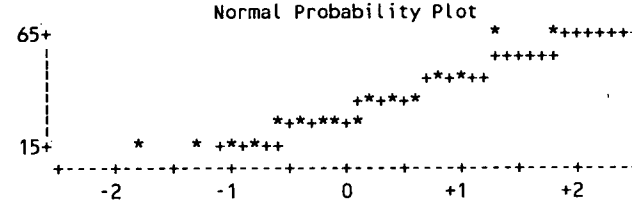
Lowest	Obs	Highest	Obs
11(	1)	37(	6)
12(	4)	42(	16)
13(	3)	48(	13)
16(	12)	60(	8)
20(	11)	60(	9)

```
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	Cum	Value	Count	Percents Cell	Cum	Value	Count	Percents Cell	Cum	Value	Count	Percents Cell	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 VIII-23. Pre-implementation 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)

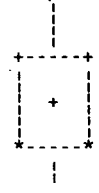
100% Max	42	99%	42
75% Q3	32	95%	42
50% Med	14	90%	40.8
25% Q1	10	10%	6.6
0% Min	6	5%	6
		1%	6
Range	36		
Q3-Q1	22		
Mode	10		

Extremes

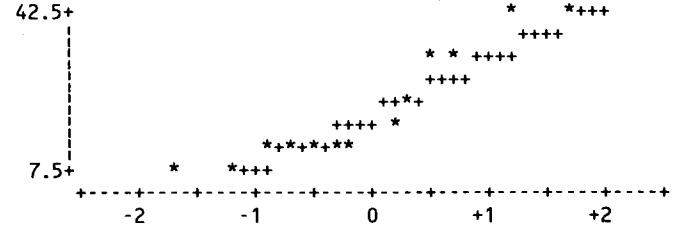
Lowest	Obs	Highest	Obs
6(	2)	30(	3)
7(	4)	32(	5)
10(	13)	32(	7)
10(	8)	40(	11)
11(	15)	42(	9)

```
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	Cum	Value	Count	Cell	Cum	Value	Count	Cell	Cum	Value	Count	Cell	Cum
6	1	6.7	6.7	11	1	6.7	33.3	16	1	6.7	60.0	32	2	13.3	86.7
7	1	6.7	13.3	12	1	6.7	40.0	22	1	6.7	66.7	40	1	6.7	93.3
10	2	13.3	26.7	14	2	13.3	53.3	30	1	6.7	73.3	42	1	6.7	100.0

Figure VIII-24. Post-implementation data set.

the sample is from a normal distribution 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 was indeed taken from a normal distribution. Hence, we reject the assumption of a normal distribution and take the alternative hypothesis that the distribution is nonnormal.

The values for skewness (0.70) and kurtosis (-0.99) indicate that there is also positive skew and low kurtosis in the post-BMP sample distribution. 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. We also reject 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, we again ran the SAS Univariate procedure 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 VIII-25 and VIII-26. 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 we cannot reject the hypothesis that the transformed data are normally distributed.

For the log-transformed post-BMP data set the skewness (0.072) is also reduced and the kurtosis (-1.23) more negative than for the raw data set. The W statistic (0.939) and associated probability (0.367) indicate that we cannot reject the hypothesis that the transformed data are normally distributed. 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, we used the test proposed by Freund (section VIII.B.4.4). Since the test requires only the number of runs and the number of values above and below the median it doesn't matter whether the raw or transformed data are used. Using the raw data in Table VIII-12, the number of runs for the pre-implementation data set is 6 while for the post-implementation data set there are 9 runs. The resulting z statistics for the pre-implementation and post-implementation data sets are 1.5526 and 0.8971, respectively. We cannot reject the null hypothesis that the samples are random (α=0.05) for either data set, but we did not use the special tables needed for n_1 and n_2 values less than 10 (Freund, 1973). Without having special tables (probably a common situation), we do a further test on the overall data set (pre- and post- combined) and obtain a z statistic of 0.0276 which is well within the 95 confidence levels we seek (z=1.96). This result, combined with the results for the tests on the two data subsets is sufficient to conclude that we are working with random samples.

Satisfied that we now have randomly sampled, normally distributed data sets (after log-transformations), we can perform

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)

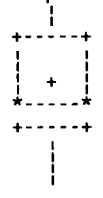
100% Max	1.778151	99%	1.778151
75% Q3	1.609487	95%	1.778151
50% Med	1.36008	90%	1.778151
25% Q1	1.228347	10%	1.067845
0% Min	1.041393	5%	1.041393
		1%	1.041393
Range	0.736759		
Q3-Q1	0.38114		
Mode	1.30103		

Extremes

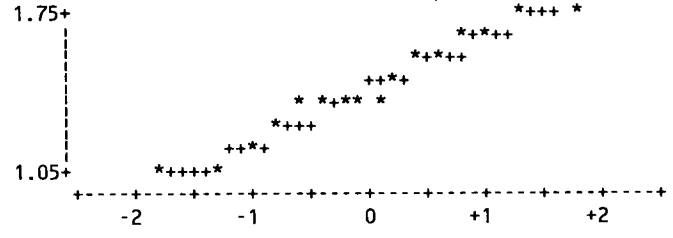
Lowest	Obs	Highest	Obs
1.041393(	1)	1.568202(	6)
1.079181(	4)	1.623249(	16)
1.113943(	3)	1.681241(	13)
1.20412(	12)	1.778151(	8)
1.30103(	11)	1.778151(	9)

```
Stem Leaf                 #
  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
```

Boxplot



Normal Probability Plot



Frequency Table

Value	Count	Percents Cell	Cum	Value	Count	Percents Cell	Cum	Value	Count	Percents Cell	Cum	Value	Count	Percents Cell	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 VIII-25. Log-transformed pre-implementation 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)

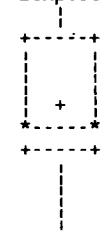
100% Max	1.623249	99%	1.623249
75% Q3	1.50515	95%	1.623249
50% Med	1.146128	90%	1.610536
25% Q1	1	10%	0.818319
0% Min	0.778151	5%	0.778151
		1%	0.778151
Range	0.845098		
Q3-Q1	0.50515		
Mode	1		

Extremes

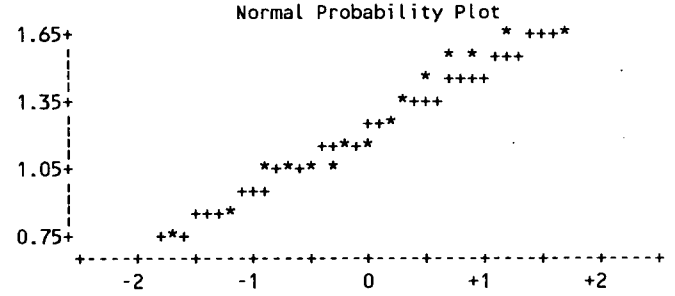
Lowest	Obs	Highest	Obs
0.778151(	2)	1.477121(	3)
0.845098(	4)	1.50515(	5)
1(	13)	1.50515(	7)
1(	8)	1.60206(	11)
1.041393(	15)	1.623249(	9)

```
Stem Leaf                 #
  16 02                   2
  15 11                   2
  14 8                    1
  13 4                    1
  12 0                    1
  11 55                   2
  10 0048                 4
   9                      0
   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 VIII-26. Log-transformed post-implementation data set.

the one-sample hypothesis test using the post-BMP data set. As shown in Figure VIII-26 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. Using the t-statistic (section VIII.B.5.2.2) we determine whether or not the post-BMP TSS concentration is less than or equal to 25 mg/L. The one-sided t-statistic is calculated using Equation 8.25:

$$t = \frac{1.21969 - 1.3979}{0.273571/\sqrt{15}}$$

or

$$t = -2.53$$

The schematic representation of this test is shown in Figure VIII-27, where the critical t value (-1.761) for the one-sided test (n=15, d.f.=14, α=0.05) is taken from Appendix C. Our t value falls to the left of the critical value so we reject the null hypothesis. In turn, we accept the alternative hypothesis that the post-BMP TSS concentration is less than 25 mg/L.

The power of this test can be computed using the noncentral t distribution with respect to various alternative hypotheses (Winer, 1971). The statistic t' is defined as (Winer, 1971, p. 21):

$$t' = t + \delta \bullet (\sigma^2/s^2)^{1/2} \qquad (8.40)$$

where

$$\delta = (\mu' - \mu) \bullet \sqrt{n}/\sigma \qquad (8.41)$$

and δ is called the noncentrality parameter. Using the t test and the log-transformed data from our one-sample hypothesis test above (μ=1.3979, μ'=1.21969, s=0.273571, n=15), we calculate (using s as estimate of σ) δ using Equation 8.41 as:

$$\delta = (1.21969-1.3979)(\sqrt{15}/0.273571)$$

$$= -2.5229$$

Consulting the table for α=0.05 in Appendix J, we find that β (the type II error associated with Ha) is approximately 0.23 (f = d.f. = 14). Thus, the power with respect to Ha is 0.77.

D.2.1.2 Wilcoxon Signed Rank Test

Alternatively, if the log transformation did not result in normally distributed data, the analyst could have selected to analyze the raw data using the Wilcoxon Signed Rank test. Table VIII-13 shows the calculations used to evaluate the hypotheses

Ho: $\mu \geq 25$
Ha: $\mu < 25$

For convenience the post-implementation data are sorted from smallest to largest observation. The difference, d_i, is computed

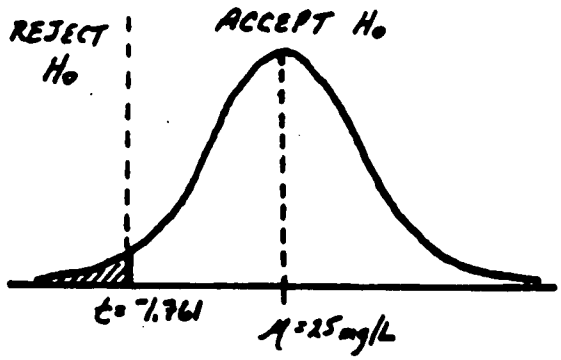


Figure VIII-27. One-sided t-test.

as 25-TSS_i. For example, the first entry is equal to 25 - 6 or 19. The third column is the absolute value of the difference, $|d_i|$. The last column is the rank of the $|d_i|$. The test statistic, T, is normally distributed and is given by Conover (1980) as

$$T = \frac{\sum_{i=1}^{n} \text{Rank}|d_i|}{\left[\sum_{i=1}^{n} \text{Rank}|d_i|^2\right]^{0.5}} \qquad (8.42)$$

Table VIII-13. Nonparametric evaluation of post-implementation data using the Wilcoxon Signed Ranks test.

TSS (mg/L)	d_i=25-TSS_i	$\|d_i\|$	rank$\|d_i\|^*$
6	19	19	15
7	18	18	14
10	15	15	11
10	15	15	11
11	14	14	9
12	13	13	8
14	11	11	6.5
14	11	11	6.5
16	9	9	5
22	3	3	1_*
30	-5	5	-2_*
32	-7	7	-3.5_*
32	-7	7	-3.5_*
40	-15	15	-11_*
42	-17	17	-13

* = Assign the negative of the rank if d_i is negative
Sum of rank$|d_i|^*$ = 54 Sum of rank$|d_i|^2$ = 1237

where the rank is assigned a negative value if d_i is negative. T is equal to $54/\sqrt{1237}$ or 1.54. Since 1.54 is less than 1.64, we reject the null hypothesis that the mean concentration is less than 25 mg/L. El-Shaarawi and Damsleth (1988) provides a modified version of the Wilcoxon Signed Rank test for use with serially correlated data.

In this case, the t-test and Wilcoxon Signed Rank test result in different conclusions. We propose that the results from the t-test (Section VIII.D.2.1.1) are more appropriate for this example since all of the assumptions of the parametric test were met. Had we elected to perform the analysis in spite of assumption violations, we would have selected the results from the Wilcoxon

Signed Rank test as being more appropriate. That is, if all assumptions are met, parametric procedures are more powerful (Section VIII.D.1.3) than their nonparametric alternative.

D.2.2 Two-sample tests

D.2.2.1 Student's t-test

Suppose that we now wish to compare the pre-BMP and post-BMP data sets to see if the BMPs have had an effect on TSS levels in Highland Silver Lake. Remembering the simplifying assumptions made earlier, we can use these two data sets in a one-sided hypothesis:

$$\text{Ho: } TSS(Post) = TSS(Pre)$$
$$\text{or} \quad \text{Ho: } TSS(Post) - TSS(Pre) = 0$$

$$\text{Ha: } TSS(Post) < TSS(Pre)$$
$$\text{or} \quad \text{Ha: } TSS(Post) - TSS(pre) < 0$$

Note that in this case we test the Ho that the post-BMP TSS equals the pre-BMP TSS concentration with an Ha that post-BMP TSS is lower. We would interpret this in a very simple sense as a test of whether the BMPs have worked; other factors must be considered. We could also have set this up as a two-sided test where Ha would simply be:

$$\text{Ha: } TSS(Post) \neq TSS(Pre)$$
$$\text{or} \quad \text{Ha: } TSS(Post) - TSS(Pre) \neq 0$$

We test this Ho using a one-sided t-test. Both the pre-BMP and post-BMP 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 are equal (Gaugush, 1986). Since a major effect of many NPS 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 NPS loads. For this and other reasons equal variance should be tested, not assumed. Thus, we perform an F test of the assumption of homogeneity of variance before proceding with the t-test. It should be noted, however, that Winer (1971) notes that the t-test is robust with respect to moderate departures from homogeneous variance.

Since we are using the log-transformed data (Figures VIII-25 and VIII-26) we must use the variance of the transformed data in the F test. The resulting F statistic is computed from Equation 8.12:

$$F = 0.075/0.057 = 1.32$$

The variances are substituted into equation 8.12 so that the F statistic is greater than unity to account for the fact that the F tables in Appendix E cover only the right side of the curve. The critical F value from Appendix E ($f1=14$, $f2=15$, $\alpha/2=0.025$) is 2.86.

We compare 1.31 versus 2.86 and accept the null hypothesis of equal variance.

Satisfied now that our data meet all of the assumptions required of the two-sample hypothesis test, we can test the null hypothesis that the pre-BMP and post-BMP mean log-transformed TSS concentrations are the same. The two-sample t statistic is (Remington and Schork, 1970):

$$t = \frac{(x_1 - x_2) - \Delta_o}{s_p \bullet \sqrt{(1/n_1 + 1/n_2)}} \qquad (8.43)$$

where s_p is the square root of the pooled variance which is defined by:

$$s_p^2 = \frac{(n_1-1) \bullet s_1^2 + (n_2-1) \bullet s_2^2}{(n_1-1) + (n_2-1)} = \frac{(n_1-1) \bullet s_1^2 + (n_2-1) \bullet s_2^2}{(n_1 + n_2 - 2)} \qquad (8.44)$$

The difference quantity (Δ_o) can be any value, but in our case we test for a difference of zero. Using the above transformed data for pre-BMP (n1=16, variance=0.057087, mean=1.407) and post-BMP conditions (n2=15, variance=0.074812, mean=1.21969), we calculate the pooled variance:

$$s_p^2 = \frac{(16-1)(0.057087) + (15-1)(0.074812)}{(16 + 15 - 2)}$$

$$= .06564$$

and t statistic:

$$t = \frac{(1.407 - 1.21969) - 0}{.0921}$$

$$= 2.034$$

Comparing this t statistic in a one-tailed test versus the t value from the table (α=0.05, d.f.=n1+n2-2=29) in Appendix C we find that the 2.034 exceeds the table value of 1.6991. We therefore reject the null hypothesis and say that the post-BMP mean log-transformed TSS concentration is lower than the pre-BMP level, i.e., the BMPs worked (given earlier assumptions). Note that if we had used a two-tailed test we would have accepted Ho (t=2.0452).

Remington and Schork (1970, p. 210-214) give test statistics for other cases in which we are testing the difference between means. These cases and corresponding equations are given below (Table VIII-14). Ponce (1980a) contains examples of several of these cases.

Table VIII-14. Summary of parametric tests used to evaluate difference between means (Remington and Schork, 1970).

Case 1: Difference Between Means When Variances Are Known (test statistic is standard normal distribution)

Null Hypothesis	Test Statistic	Assumptions
Ho: $\mu_1-\mu_2 = \Delta_o$	$t = \frac{(x_1-x_2) - \Delta_o}{(\sigma^2_1/n_1 + \sigma^2_2/n_2)^{0.5}}$	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)

Ho	Test Statistic	Assumptions
$\mu_1-\mu_2 = \Delta_o$	$t = \frac{(x_1-x_2) - \Delta_o}{s_p(1/n_1 + 1/n_2)^{0.5}}$	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)

Ho	Test Statistic	Assumptions
$\mu_1-\mu_2 = \Delta_o$	$t = \frac{(x_1-x_2) - \Delta_o}{(s^2_1/n_1 + s^2_2/n_2)^{0.5}}$ $df' = \frac{(s^2_1/n_1 + s^2_2/n_2)^2}{\frac{(s^2_1/n_1)^2}{n_1+1} + \frac{(s^2_2/n_2)^2}{n_2+1}}$ $df = df' - 2$	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)

Ho	Test Statistic	Assumptions
$\mu_d = \Delta_o$ $\mu_d = \mu_2 - \mu_1$	$t = \frac{d - \Delta_o}{s_d/n^{0.5}}$	random sample of size n paired differences from a normally distributed populations of differences

The above test showed that the difference between pre-BMP and post-BMP log-transformed TSS concentrations is not zero, but it would be useful to describe in more detail what the difference really is. One way to describe this difference more completely is to determine the confidence interval about the difference. The 95% confidence interval for equal variance conditions can be calculated by (Winer, 1971):

$$(post\text{-}pre) - t \bullet s_p \sqrt{(1/n1 + 1/n2)} \leq \Delta_o \leq (post\text{-}pre) + t \bullet s_p \sqrt{(1/n1 + 1/n2)} \qquad (8.45)$$

For $t=2.05$ ($\alpha=.025$, d.f.$=29$) this becomes:

$$(1.21969 - 1.407) - .18876 \leq \Delta_o \leq (1.21969 - 1.407) + .18876$$

$$-0.18731 - .18876 \leq \Delta_o \leq -0.18731 + .18876$$

or

$$-0.3761 \leq \Delta_o \leq 0.0014$$

Note that this interval includes 0, a result of the fact that if we had used a two-tailed test above we would have accepted Ho.

Since we have selected the one-tailed test above, we will determine the one-tailed confidence interval as:

$$(post\text{-}pre) - t \bullet s_p \sqrt{(1/n_1 + 1/n_2)} \leq \Delta_o \qquad (8.46)$$

where $t = 1.6991$

Thus, $$(1.21969 - 1.407) - .1564 \leq \Delta_o$$

or $$-0.3437 \leq \Delta_o$$

This difference in the means of log-transformed data can be translated into TSS concentration data by taking into account the fact that the difference of two logs is simply the ratio of the antilogs. So, using the one-sided confidence interval above for the difference between log-transformed pre-BMP and post-BMP means, we can express the actual difference in TSS concentrations as:

$$antilog(-0.3437) \leq TSS(post)/TSS(pre) \qquad (8.47)$$

$$0.453 \leq TSS(post)/TSS(pre)$$

or $$TSS(post) \leq 0.453 \bullet TSS(pre)$$

rearranging and solving for the difference in terms of TSS(pre),

$$Difference \geq 0.453 \bullet TSS(pre) - TSS(pre)$$

Since the mean TSS(pre) concentration is 29.38 mg/L this becomes:

Difference ≥ -16.07 mg/L

In other words, there is no more than a 16 mg/L reduction in mean TSS from the pre-BMP to the post-BMP periods. Using the two-tailed test, we would say that the difference is between 17 and 0 mg/L TSS.

The power of this test with respect to Ha can be estimated using the noncentrality parameter (Equation 8.41) and by approximating σ with the pooled variance (Winer, 1971). Assuming that the difference between the means is actually 5 mg/L (log=0.69897), the noncentrality parameter is (Winer, 1971):

$$\delta = \frac{0.69897 - 0}{0.2562\sqrt{(1/16 + 1/15)}} = 7.59$$

From Appendix J we find that β is (d.f.=29=f, α=0.05) less than 0.01, so the power is greater than 0.99. If we were to select an Ha with a difference of 15 mg/L (log=1.1761) the resulting noncentrality parameter would be 12.8, and the power would again be greater than 0.99. For an Ha with a very small difference (0.5 mg/L, log=-0.3010) the noncentrality parameter would be negative 3.27 and the power would be about 0.94. This group of scenarios for Ha emphasizes the point made earlier that for a given situation (n, α) an increase in the difference between Ha and Ho parameters (e.g., mean) results in an increased power of the test with respect to Ha.

D.2.2.2 Mann-Whitney (Wilcoxon's Rank Sum) Test

The Mann-Whitney test can also be used to evaluate the one-sided hypotheses

Ho: TSS(Pre) ≤ TSS(Post)
Ha: TSS(Pre) > TSS(Post)

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. In NPS data analyses, this test can be used to compare data obtained from different locations—or for example, whether the BOD, concentration at one location in the stream is different from that at other location. The test can also be used to compare data collected at two different periods of time at the same location (before and after a management practice).

The Mann-Whitney test accomplishes the same objective as the Student's t-test. The Student's t-test requires that the two populations be normally distributed, whereas Wilcoxon's rank sum test requires only that the two populations have the same distribution, no matter what the distribution.

The test statistic can be computed from the following equation which allows for ties (Conover, 1980)

$$T_1 = \frac{T - n\frac{N+1}{2}}{\left[\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}\right]^{0.5}} \qquad (8.48)$$

where

n = number of samples in first random sample (e.g., pre-implementation,
m = number of samples in second random sample (e.g., post-implementation,
N = n + m,
T = Sum of ranks for first random sample, and
R_i = rank for the ith ordered sample.

Table VIII-15 has been prepared showing 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.

Table VIII-15. Nonparametric evaluation of post-implementation data using the Mann-Whitney test.

Rank	Pre-Implementation TSS (mg/L)	Post-Implementation TSS (mg/L)	Rank	Pre-Implementation TSS (mg/L)	Post-Implementation 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 pre-implementation, T = 302.5
Sum of ranks squared, ΣR_i^2 = 10409.5

Applying Equation 8.48 yields

$$T_1 = \frac{302.5 - 16\,\frac{31+1}{2}}{\left[\frac{16 \cdot 15}{31(31-1)}\left(10409.5 - \frac{16 \cdot 15(31+1)^2}{4(31-1)}\right)\right]^{0.5}}$$

$$T_1 = 46.5/25.26 = 1.84$$

T_1 is normally distributed and Appendix B can be used to determine the appropriate quantile. Since the hypothesis was one-sided and we wish to use an α equal to 0.05, the appropriate quantile from Appendix B is 1.645. T_1 is greater than 1.645 and we reject the null hypothesis. The post-implementation mean TSS concentration is less than the pre-implementation mean TSS concentration.

The Hodges-Lehman 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 compute the difference between all n and m observations. Using the data in the above example, there are 16·15 or 240 differences to compute. The Hodges-Lehman estimator is the median of these differences or 8 mg/L.

Conover (1980) provides an approach for estimating the confidence limits for the difference. First, the analyst determines the value $w_{\alpha/2}$ statistics for n and m samples from Appendix T. Then k may be calculated as

$$k = w_{\alpha/2} - n(n+1)/2 \qquad (8.49)$$

The upper and lower confidence limits on the difference between the two samples is equal to the k^{th} smallest and largest differences. Using the data from above where n=16 and m=15, $w_{\alpha/2}$ is equal to 207 from Appendix T. Applying Equation 8.49, k is equal to 71. From the 240 differences computed for the Hodges-Lehman estimator, it can be said that the confidence interval is -1 ≤ difference ≤ 19 mg/L.

D.2.3 Matched sample tests

Paired observations are a series of data collected as pairs at a given time or location. For example, if BOD5 is sampled at two stream locations at a regular time interval, the result each time is a pair of BOD5 data. Several statistical tests are available for comparing paired observations, depending on the study objectives. If samples are taken for total nitrogen and total phosphorus concentrations regularly at the same location, a pair of data for each sample collected is obtained.

For the BOD5 data at two locations, we may want to know whether location 1 has a higher or lower concentration of BOD5 than location 2, or whether the BOD5 concentrations at the two locations

are the same. In this case, the paired t-test or the Wilcoxon Signed Rank test are good candidates.

For the paired total nitrogen and phosphorus concentrations, we may want to compare whether there is a correlation between the two sets of concentration data. In other words, is there a tendency for high total nitrogen concentration to occur with a higher total phosphorus concentration, or are the two concentrations unrelated? Spearman's test is a good test to answer such questions. The Kendall tau test is similar to Spearman's test and could be used; however, has found most of its applications in time series data discussed in Section VIII.D.3.

D.2.3.1 Paired t-test

The null hypothesis and test statistic is presented in Table VIII-14, Case #4.

D.2.3.2 Wilcoxon Signed Rank Test

The Wilcoxon Signed rank test is presented in section VIII.D.2.1.2 of this chapter.

D.2.3.3 Sign Test

The sign test is the oldest and simplest test among the tests discussed here. It compares paired data to test whether one data set presents higher values than the other or whether the two sets are the same. The data being analyzed consist of paired observations (X1, Y1), (X2, Y2)....(Xi, Yi),...(Xm, Ym), where m is the number of pairs. Generally, we want to know whether the sets X and Y have the same mean. The assumptions and test procedures are summarized in the box on the next page.

Table VIII-16 presents water samples collected daily at two locations from the same stream and analyzed for BOD5 and the intermediate calculations used for the sign test. In this case, we want to test the hypothesis that there is no difference in BOD5 concentrations between the two locations.

Solution:

1. Construct hypothesis.

   Ho: BOD5 concentrations are the same at the two locations.
   Ha: BOD5 concentrations are different at the two locations. (This is a two-sided test).
   Use confidence level $\alpha = 0.05$.

2. Take the difference between the BOD5 of location 2 and location 1 and assign a sign to it, as indicated in column 4 of Table VIII-16. If the two data are equal, assign a "tie."

3. From the table, calculate the number of pairs with a "+" sign (B=8).

Sign Test - Procedure

Assumptions

- The variables in each pair of data (Xi, Yi) are mutually independent.
- The measurement in each pair of data is at least ordinal. That is, the pair can be compared and assigned a "+", "-", or "tie" sign.
- The pairs (Xi, Yi) are internally consistent. That is, if the probability for "+" is greater than "-" for one pair of data, it is also true for all pairs of data.

Test Steps

1. Construct test hypotheses.
   - Null hypothesis (Ho): the median of the population of all possible difference is zero. In other words, there is the same likelihood that Xi is greater than Yi as Yi is greater than Xi.
   - Alternative hypothesis (Ha): Select one of the following:
     a. Two-sided test: Either Xi is more likely greater than Yi, or vice versa.
     b. One-sided test 1: Xi is more likely greater than Yi.
     c. One-sided test 2: Xi is more likely less than Yi.
2. Compute the following:
   - The difference for each pair (Yi - Xi) and assign a sign to it if ($Yi > Xi$), the pair is assigned a "+", if ($Yi < Xi$), a "-" sign is assigned. A "tie" is assigned if Xi = Yi. Let B equal to the number of pairs that have a "+" sign, and let n equal the total number of pairs with either a "+" or "-" sign (ties are excluded).
3. Compute the total number of pairs that have a "+" sign.
4. Draw conclusion.
   a. Two-sided test: From Appendix P, find out the l and u values for the corresponding significant level (alpha) and n. Reject Ho and accept Ha if $B \geq u$ or $B < l\text{-}1$
   b. One-sided test 1: Multiply the alpha by 2 and use the new alpha and n to find out u in Appendix 1. If $B \geq u$, reject Ho and accept Ha.
   c. One-sided test 2: Multiply the alpha by 2 and use the new alpha and n to find out l in Appendix 1. Reject Ho and accept Ha if $B < l\text{-}1$.

Table VIII-16. Sign test for comparing daily BOD5 concentrations at two locations.

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

4. Calculate the total number of pairs excluding ties. (n=15-2=13).

From Appendix P, using n=13 and alpha=0.05, we find l=3, u=11. Since B≤l-1 and B≥u, Ho cannot be rejected. So we conclude that the BOD5 concentrations at location 1 are not different from those of location 2 at the 0.05 confidence level.

D.2.3.4 Spearman's Test

Spearman's test is used to detect whether there is a correlation between the paired data; that is, whether the two data in the pair are independent, or when one datum in the pair is high and the other datum tends to be high (or low). The procedure for Spearman's Test is summarized in the box on the next page. This test can be used in NPS data analysis to find the cause for a certain pollutant in the river. Suppose we observe that when a location has high nitrogen concentration, the phosphorus concentration at the same location is also high, and when the nitrogen concentration is low, the phosphorus concentration is also low. We also know that fertilizers with nitrogen and phosphorus are applied to upstream lands and it is likely that the nitrogen and phosphorus come from the applied fertilizer. On the other hand, if the concentrations of nitrogen and phosphorus are unrelated, we might suspect that the nitrogen is coming from sources other than fertilizer applications.

A monitoring program was designed to see whether there is a correlation between the total nitrogen and total phosphorus concentrations. Water samples were collected and analyzed on a

Spearman's Test - Procedure

1. Arrange ranks.

   For n number of paired data (X1, Y1), (X2, Y2)....(Xn, Yn), let R(Xi) be the rank of Xi among X1 to Xn, with the smallest number being rank 1 and the largest number being rank n. In the same manner, let R(Yi) be the rank of Yi among Y1 to Yn. In case of ties, assign the average value of the rank that would have been assigned if there were no ties. For example, if rank 6, 7, and 8 have the same value, assign 7 to all three ranks.

2. Compute the Spearman's ρ value.
   If there are no ties or moderate number of ties,

$$\rho = 1 - \frac{6\sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2}{n(n^2-1)} = 1 - \frac{6T}{n(n^2-1)} \qquad (8.50)$$

   where T is the entire sum of square of difference between the two ranks. If there are many ties,

$$\rho = \frac{\sum_{i=1}^{n} [R(X_i) - R(Y_i)] - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\sum_{i=1}^{n} R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2}\sqrt{\sum_{i=1}^{n} R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2}} \qquad (8.51)$$

   where n is the number of paired observations.

3. Construct hypothesis.
   Ho: There is no correlation between Xi and Yi.
   H1: Use one of the following
   a. Two-sided test:
      There is a correlation between Xi and Yi; that is, either a large Xi value tends to pair with a large Yi value or a large Xi value tends to pair with a small Yi value.
   b. One-sided test 1:
      A large Xi value tends to pair with a large Yi value, and a small Xi value tends to pair with a small Yi value.
   c. One-sided test 2:
      A large Xi value tends to pair with a smaller Yi value, or vice versa.

4. Draw Conclusion.
   From Appendix Q, obtain the quantile for the corresponding confidence level and n.
   For two-sided test:
      get two quantile values, w1 and w2, from Appendix 2. Use p=1-(alpha/2) for w1 and p=alpha/2 for w2. (For example, if n=20, alpha=0.05, w1=0.4451 based on p=0.975, w2=-0.4451, based on the equation provided in footnote of the table.) If $r > w1$ or $r < w2$, then reject Ho and accept Ha.
   For one-sided test 1:
      use p=1-alpha and n to get w. For example, if n=20 and alpha=0.05 p=0.95, and w is 0.3789. Reject Ho if $r > w$.
   For One-sided test 2:
      obtain w using the same procedure as one-sided test 1, reject Ho if $p < w$.

monthly base for one year. The total nitrogen and total phosphorus concentrations are listed in columns 2 and 3 of Table VIII-17. The objective is to determine whether higher phosphorus concentration tends to pair with higher nitrogen concentration.

Table VIII-17. Spearman's test for total nitrogen and total phosphorus correlation.

Month	Total N	Total P	R (N_i)	R (P_i)	$[R(N_i)-R(P_i)]^2$
Jan	2.31	0.21	2	2.5	0.25
Feb	2.13	0.29	1	1	0.0
Mar	3.05	0.2	4	7	9.0
Apr	3.67	0.35	7	4	9.0
May	4.5	0.36	9	8	1.0
Jun	4.98	0.45	11	10	1.0
Jul	5.14	0.49	12	12	0.0
Aug	4.91	0.48	10	11	1.0
Sep	3.45	0.32	6	6	0.0
Oct	4.10	0.39	8	9	1.0
Nov	3.11	0.31	5	5	0.0
Dec	2.52	0.21	3	2.5	0.25

Solution:

1. Arrange data ranking.
   The ranking of total nitrogen and phosphorus are made in columns 4 and 5 of Table VIII-17.

2. Calculate λ value.
   Since there is only one tie in the data, Equation 8.50 can be used. The square of difference between the two ranks is calculated in column 6 of Table VIII-17.

$$\rho = 1 - \frac{6(23.0)}{12(12^2-1)} = 0.08$$

3. Construct hypothesis.
   Ho: There is no correlation between total nitrogen and total phosphorus concentrations.
   Ha: Higher total nitrogen concentration tends to pair with higher phosphorus concentration. (This is a two-sided test.)
   Use α=0.05 as the confidence level for the test.

4. Draw conclusion.
   From Appendix Q, using p=1-0.05=0.95 and n=12, we find w=0.4965. we cannot reject Ho because ρ<w. So we conclude that higher total nitrogen concentration tends to come in pair with higher total phosphorus concentration.

For the one-sided test, use p=1-0.05=0.95 and n=12. From Appendix Q, we find w=0.4965. We cannot reject Ho because ρ < w. So we conclude that higher total nitrogen concentration tends to come in pair with higher total phosphorus concentration.

D.2.4 Comparison of More Than Two Means

Comparison of more than two means requires that analysis of variance (ANOVA) and post-ANOVA hypothesis testing are performed or a nonparametric method applied. The ANOVA is a general procedure which can be applied to any number of treatments to determine, with an F test, if differences exist among treatment means. Post-ANOVA hypothesis testing incorporates such procedures as Duncan' multiple range test, Tukey's honestly significant difference, Scheffe's method, the Newman-Keuls test, and least significant difference test (Anderson and McLean, 1974, p. 10), as well as the Bonferroni t tests, Gabriel's multiple-comparison procedure, the Ryan-Einot-Gabriel-Welsch multiple range test, and the Waller-Duncan k-ratio test (SAS Institute, Inc., 1985b, p. 117-118).

D.2.4.1 Analysis of Variance

Analysis of variance (ANOVA) is a procedure for expressing in a general linear model the total variability of a set of data as a sum of terms, each of which can be attributed to a specific source, or cause, of variation (Freund, 1973, p. 340). The technique is basically a method for studying variation among two or more sample means, this variation being measured by a variance (Remington and Schork, 1970, p. 283). Snedecor and Cochran (1980, p. 216) note that ANOVA is the only simple and reliable way of computing the appropriate pooled error variance for data under complex classifications, and that ANOVA provides an F test of the null hypothesis that population means for different classes are identical. The components of an ANOVA are determined by the experimental design (Gaugush, 1986, p. 94), and researchers should, therefore, take care in planning monitoring activities to assure that the proper data are available for meeting the needs of the ANOVA.

The discussion of ANOVA will begin with text book definitions, followed by example applications to water quality data. A similar pattern will be followed for analysis of covariance. Neither discussion will cover all of the many possible applications to water quality investigations, and the reader is encouraged to study the references used in this section, as well as other statistics texts (e.g., Snedecor and Cochran (1980) and Cochran (1977)) and reference materials.

The basic assumptions made in using ANOVA are (Remington and Schork, 1970, p. 284):

1. Each sample is a random sample from the corresponding population, and observations from different populations are independent.

2. The measurement variable X is normally distributed in each of the k populations.

3. The populations all have the same variance (homoscedasticity).

An additional assumption stated by Ponce (1980a, p. 74) is that "the treatment (group) and environmental effects are additive." Ponce goes on to note that if the error terms for each variate are not independent the F-test of significance can be "seriously impaired." He also states that "the assumption of homogeneity of variances is very important in the ANOVA because each sample variance is considered to be an estimate of the same parametric error variance." Tests for homogeneity of variance are discussed in section VIII-B.4.2.

It should be noted that an ANOVA and its F test do not lead to the inference that all means are different from each other (Gaugush, 1986, p. 116). A significant F test simply indicates that differences exist, but additional tests (e.g., tests for differences among means) are needed to answer questions about particular means and differences among means. Gaugush describes two tests for differences among means, the Bonferroni method and Scheffe's method (Gaugush, 1986, p. 135). The reader is encouraged to seek statistics texts which discuss these and related methods since this guidance addresses them in a very limited fashion.

D.2.4.1.1 One-way analysis of variance

D.2.4.1.1.1 Fixed Effects - Model I

To perform the calculations involved in ANOVA the following definitions are required (Freund, 1973, p. 341-343):

$$\text{total sum of squares} = SST = \sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij}-\bar{x}..)^2 \qquad (8.52)$$

(equal sample sizes)

$$= SS(Tr) + SSE$$

$$\text{(for unequal sample sizes)} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x^2_{ij} - \frac{1}{N} \bullet T^2.. \qquad (8.53)$$

treatment sum of squares = SS(Tr)
(equal sample sizes)

$$= n \bullet \sum_{i=1}^{k} (x_{i.} - x..)^2 \qquad (8.54)$$

$$\text{(for unequal sample sizes)} = \sum_{i=1}^{k} \frac{T^2_{i.}}{n_i} - \frac{1}{N} \bullet T^2.. \qquad (8.55)$$

$$\text{error sum of squares} = SSE = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i.})^2 \qquad (8.56)$$

(equal or unequal sample sizes)

$$\text{treatment mean square} = MS(Tr) = \frac{SS(Tr)}{k - 1} \qquad (8.57)$$

$$\text{error mean square} = MSE = \frac{SSE}{k(n - 1)} \qquad (8.58)$$

$$F \text{ statistic} = F = \frac{MS(Tr)}{MSE} \qquad (8.59)$$

number of means = k (8.60)

sample size for each of k means = n (8.61)

grand total of all the data = T.. (8.62)

$$\text{grand mean} = \bar{x}.. = \frac{T..}{nk} \qquad (8.63)$$

total of observations for ith treatment = Ti (8.64)

mean of the ith sample = $\bar{x}_{i.}$ (8.65)

According to Snedecor and Cochran (1980, p. 218), in the above notation subscript i denotes the class, and takes on values from 1 to k, where k is the number of classes. Similarly, subscript j denotes the member of the class, from 1 to n.

The observations (x_{ij}) within each class (i) are assumed normally distributed about the mean μ_i with variance σ^2 (Snedecor and Cochran, 1980, p. 218). As stated earlier, the variance is the same for all classes, but the mean can vary among classes. The mean of the k values of μ_i is denoted as μ, and the linear model is expressed as (Snedecor and Cochran, 1980, p. 219):

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad [i=1,\ldots k;\ j=1,\ldots n;\ \epsilon_{ij} = N(0,\sigma^2)] \qquad (8.66)$$

This "fixed effects model" (or model I) shows that each observed value is the sum of an overall mean (μ), a treatment or class deviation (α_i), and a random element (ϵ_{ij}) from a normally distributed population with mean 0 and standard deviation σ (Snedecor and Cochran, 1980, p. 219). The model is referred to as "fixed" because the α_i are considered to be fixed, but unknown quantities to be measured. The random element represents variations due to such things as unit-to-unit variation in treatment effect, measurement errors, or individual characteristics of the unit (Snedecor and Cochran, 1980, p. 219).

According to Snedecor and Cochran (1980, p. 221) the F value is a "good criterion for testing the null hypothesis that the population means are the same in all classes." An F value of 1 represents the condition where Ho is true, and large F values indicate differences among the μ_i. Snedecor and Cochran (1980, p.228) note that the F test is more affected by nonnormality and heterogeneity of variances with unequal sample sizes (n_i) than with equal sample sizes.

As an example of one-way ANOVA consider the situation where the trout populations (Platts et al., 1983, p.28) of three streams are measured by the multiple-step Zippin approach for electrofishing at five randomly selected sites in each of three geographic regions. The data from this monitoring effort are shown in Table VIII-18.

Using the SAS ANOVA procedure (SAS Institute, Inc., 1985b) we modeled trout population as a function of stream (i.e., MODEL TROUT=RIVER) to test the null hypothesis that stream has no effect on trout population (i.e., treatment effect is zero). The results of this test are shown in Table VIII-19. Note that the F VALUE (2.60) is equal to MS(Tr) (109.9) divided by MSE (42.3). (NOTE: SAS uses MODEL instead of "Treatment".) The probability of obtaining a greater F value if the null hypothesis is true (PR > F) is 0.0865. We can therefore reject with 90% confidence the null hypotheis and accept the alternative hypothesis that stream does affect trout population (i.e., trout population varies across streams). However, we would have to <u>accept</u> the null hypothesis at the 95% level of confidence.

The SAS ANOVA procedure also allows the analyst to perform several post-ANOVA hypothesis tests, including tests of differences between means (SAS Institute, Inc., 1985b, p. 117). For this example we used the Bonferroni t tests, Duncan's multiple range test, Gabriel's multiple-comparison procedure, the Ryan-Einot-Gabriel-Welsch multiple F test, the Ryan-Einot-Gabriel-Welsch multiple range test, Scheffe's multiple-comparison procedure, Tukey's studentized range test, and the Waller-Duncan k-ratio test. The reader should consult statistics texts (e.g., Snedecor and Cochran, 1980, p. 232-236) to learn more about these procedures. The Waller-Duncan k-ratio test and Duncan's multiple range test showed that (at 95 percent confidence) the mean trout population for Black Stream (67.1) is statistically different from that of Red Stream (61.8), but that the mean trout population for Blue Stream

Table VIII-18. Stream Trout Population

Stream	Region	Site 1	2	3	4	5
		- - Trout Population - - (Pounds/Acre/Year - Year Class 2)				
Black	Mountain	75	70	65	72	68
	Piedmont	68	72	70	70	67
	Coastal Plain	60	65	64	63	58
Blue	Mountain	70	76	69	67	74
	Piedmont	64	66	60	69	62
	Coastal Plain	49	60	54	58	57
Red	Mountain	68	70	63	65	70
	Piedmont	62	66	58	69	67
	Coastal Plain	50	56	51	60	52

Table VIII-19. One-way ANOVA

1-Way ANOVA With River As Treatment
Analysis of Variance Procedure

Dependent Variable: TROUT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F	R-SQUARE	C.V.		
MODEL	2	219.733	109.9	2.60	0.0865	0.110021	10.1331		
ERROR	42	1777.467	42.3					ROOT MSE	TROUT MEAN
CORRECTED TOTAL	44	1997.200						6.5054	64.2

SOURCE	DF	ANOVA SS	F VALUE	PR>F
RIVER	2	219.733	2.60	0.0865

is not statistically different from that of either Black Stream or Red Stream. All other tests showed (at 95 percent confidence) no statistical differences between the mean trout populations of the three streams. Table VIII-20 shows results from both the Waller-Duncan k-ratio test and the Bonferroni t test, which are representative of all tests performed on differences between means. Note that the two tests determined different minimum significant differences, which caused the different test results.

D.2.4.1.1.2 Random Effects - Model II

Model II ANOVA conforms to the following equation for one-way classification (Snedecor and Cochran, 1980, p. 238):

$$x_{ij} = \mu + A_i + \epsilon_{ij} \quad [i=1,...a;\ j=1,...n) \qquad (8.67)$$

where:

$$A_i = N(0,\sigma_A^2),\ \epsilon_{ij} = N(0,\ \sigma^2)$$

A clear difference between the model II and model I equations is the term A_i which represents the difference between the value of the ith class and the average value over the population ($N=n \bullet a$) (Snedecor and Cochran, 1980, p. 238). The value A_i varies from class to class, whereas α_i (model I) is a fixed class deviation. Snedecor and Cochran (1980, p. 238) notes that for the simplest version of model II the A_i are assumed normally distributed with variance σ_A^2 and mean 0. Further assumptions are the ϵ_{ij} and A_i are independent, and that ϵ_{ij} is normally distributed with variance σ^2 and mean 0. The variance of an observation is $\sigma_A^2+\sigma^2$, with the two terms called the "components of variance" (Snedecor and Cochran, 1980, p. 238).

Snedecor and Cochran (1980, p. 239-254) discuss the random effects model (model II) in considerable detail, including a comparison between model I and model II, use of model II in measurement problems, intraclass correlation coefficients, confidence limits related to the components of variance, samples of unequal sizes, and tests of homogeneity of variance. The reader is referred to this and other statistics texts for more on the random effects model.

In a simple sense, the model I and model II ANOVA can be applied in the following general situations.

Model I	Model II
class means differ	class means differ
class deviation fixed	class deviations differ

As an example, a study to investigate the phosphorus content of runoff from corn fields, city streets, and mining areas may require application of model II ANOVA since both the mean phosphorus content and variability are likely to vary from class to class. However, a split-plot study to explore the phosphorus content of three corn plots receiving different levels of phosphorus

Table VIII-20. Waller-Duncan k-ratio and Bonferroni t test results

1-Way ANOVA With River As Treatment
Analysis of Variance Procedure

Waller-Duncan K-ratio t test for variable: TROUT
NOTE: This test minimizes the Bayes Risk under additive loss and certain other assumptions.

K-Ratio = 100 DF = 42 MSE = 42.3206 F = 2.59605
Critical Value of t = 2.21
Minimum Significant Difference = 5.2403

Means with the same letter are not significantly different.

WALLER GROUPING		MEAN	N	RIVER
	A	67.133	15	Black
B	A	63.667	15	Blue
B		61.800	15	Red

Bonferroni (Dunn) t test for variable: TROUT
NOTE: This test controls the Type I experimentwise error rate, but generally has a higher Type II error rate than REGWQ.

ALPHA = 0.05 DF = 42 MSE = 42.3206
Critical Value of t = 2.49367
Minimum Significant Difference = 5.9236

Means with the same letter are not significantly different.

BONFERRONI GROUPING	MEAN	N	RIVER
A	67.133	15	Black
A	63.667	15	Blue
A	61.800	15	Red

fertilizer may support application of model I ANOVA since the class means should differ but the class deviations may be comparable.

D.2.4.1.2 Two-way analysis of variance

In two-way ANOVA we quantify the variation due to one variable (treatments), the variation due to a second variable (blocks), and the variation due to chance (experimental error) (Freund, 1973, p. 348). We then calculate F statistics for both the treatments and the blocks. The two null hypotheses are that the treatment effect and block effect are zero. Snedecor and Cochran (1980, p. 258) note that the post-ANOVA hypothesis tests discussed for one-way ANOVA are also applicable to two-way ANOVA.

The terminology used in two-way ANOVA varies with the application, "blocks" being used in agriculture and "replicates" or "groups" used in other applications. As such, in a "randomized block" experimental design sample units are randomly selected from blocks under each treatment (Snedecor and Cochran, 1980, p. 255).

The following additional definitions are required for two-way analysis of variance with no interaction (Freund, 1973, p. 350):

$$\text{block sum of squares} = SSB = \frac{1}{k} \cdot \sum_{j=1}^{n} T_{\cdot j}^{2} - \frac{1}{kn} \cdot T_{\cdot\cdot}^{2} \quad (8.68)$$

$$\text{block mean square} = MSB = \frac{SSB}{n-1} \quad (8.69)$$

$$\text{F statistic for treatments} = \frac{MS(Tr)}{MSE} \quad (8.70)$$

$$\text{F statistic for blocks} = \frac{MSB}{MSE} \quad (8.71)$$

$$\text{number of treatments} = k \quad (8.72)$$

$$\text{number of blocks} = n \quad (8.73)$$

Also, the error sum of squares and error mean square are redefined as (Freund, 1973, p. 350):

$$SSE = SST - SS(Tr) - SSB \quad (8.74)$$

$$MSE = \frac{SSE}{(n-1)(k-1)} \quad (8.75)$$

The linear model for two-way classification with one observation per cell is (Snedecor and Cochran, 1980, p. 259):

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad [i=1,...k;\ j=1,...n) \qquad (8.76)$$

where:

x_{ij} = measurement for the ith treatment and jth block
μ = the overall mean
α_i = treatment effects ($\Sigma\alpha_i = 0$)
β_j = block effects ($\Sigma\beta_j = 0$)
ϵ_{ij} = random element of error

The two assumptions made in using this model are that (1) the treatment and block effects are additive, and (2) the residuals ϵ_{ij} are independent, random variables normally distributed with mean 0 and variance σ^2 (Snedecor and Cochran, 1980, p. 259-260).

Snedecor and Cochran (1980, p. 264-272) discuss two-way ANOVA in further detail, covering such topics as least squares estimation, blocking efficiency, more than one observation per cell, balancing the order in which treatments are given, and Latin squares.

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. For the case of no interaction the above definitions are used. The ANOVA table for a two-way analysis with no interaction is shown as Table VIII-21 (Freund, 1973, p. 350).

In the above example, the researcher has anticipated some regional effect on the trout populations for each of the streams. The data in Table VIII-18 point out this fact since samples were collected in the mountains, piedmont, and coastal plains region of each stream. A two-way ANOVA with stream as the treatments and region as the blocks was performed to test the hypotheses that both treatment and block effects are zero. The results of this SAS ANOVA procedure (SAS Institute, Inc., 1985b) are shown in Table VIII-22. Comparing Table VIII-22 with Table VIII-19 it is clear that the addition of region to the ANOVA has increased the F VALUE, the R-SQUARE, and the significance of the F statistic. Note that the F VALUE for river is higher in the two-way (7.80) than in the one-way (2.60) ANOVA. This is due to the fact that the MSE (the measure of variability due to chance) is lower in the two-way ANOVA.

The results of this two-way ANOVA cause us to reject with 95% confidence both null hypotheses in favor of the alternative hypotheses that stream and region each have an effect on trout population. Remembering that this analysis assumes no interactions between the two variables, we performed both the Waller-Duncan K-ratio test and the Bonferroni t test on differences between means for both stream and region. The results of this test for stream and region are shown in Tables VIII-23 and VIII-24, respectively. The results in Table VIII-23 differ from those in Table VIII-20 since the MSE is different. Table VIII-24 shows that mean trout population is different for each region.

Table VIII-21. Two-way ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	$k - 1$	$SS(Tr)$	$MS(Tr) = \frac{SS(Tr)}{k - 1}$	$\frac{MS(Tr)}{MSE}$
Blocks	$n - 1$	SSB	$MSB = \frac{SSB}{n - 1}$	$\frac{MSB}{MSE}$
Error	$(n-1)(k-1)$	SSE	$MSE = \frac{SSE}{(n-1)(k-1)}$	
Total	$nk - 1$	SST		

Table VIII-22. Two-way ANOVA

2-Way ANOVA With River and Region As Treatments
Analysis of Variance Procedure

Dependent Variable: TROUT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F	R-SQUARE	C.V.	
MODEL	4	1433.47	358.37	25.43	0.0001	0.717738	5.8475	
ERROR	40	563.73	14.09				ROOT MSE	TROUT MEAN
CORRECTED TOTAL	44	1997.20					3.7541	64.2

SOURCE	DF	ANOVA SS	F VALUE	PR>F
RIVER	2	219.733	7.80	0.0014
REGION	2	1213.733	43.06	0.0001

Table VIII-23. Waller-Duncan k-ratio and Bonferroni t test results

2-Way ANOVA With River and Region As Treatments
Analysis of Variance Procedure

Waller-Duncan K-ratio t test for variable: TROUT
NOTE: This test minimizes the Bayes Risk under additive loss and certain other assumptions.

K-Ratio = 100 DF = 40 MSE = 14.0933 F = 7.79565
Critical Value of t = 1.94
Minimum Significant Difference = 2.6586

Means with the same letter are not significantly different.

WALLER GROUPING	MEAN	N	RIVER
A	67.133	15	Black
B	63.667	15	Blue
B	61.800	15	Red

Bonferroni (Dunn) t test for variable: TROUT
NOTE: This test controls the Type I experimentwise error rate, but generally has a higher Type II error rate than REGWQ.

ALPHA = 0.05 DF = 40 MSE = 14.0933
Critical Value of t = 2.49886
Minimum Significant Difference = 3.4254

Means with the same letter are not significantly different.

BONFERRONI GROUPING	MEAN	N	RIVER
A	67.133	15	Black
B	63.667	15	Blue
B	61.800	15	Red

Table VIII-24. Waller-Duncan k-ratio and Bonferroni t test results

2-Way ANOVA With River and Region As Treatments
Analysis of Variance Procedure

Waller-Duncan K-ratio t test for variable: TROUT
NOTE: This test minimizes the Bayes Risk under additive loss and certain other assumptions.

K-Ratio = 100 DF = 40 MSE = 14.0933 F = 43.0605
Critical Value of t = 1.81
Minimum Significant Difference = 2.4835

Means with the same letter are not significantly different.

WALLER GROUPING	MEAN	N	REGION
A	69.467	15	Mountain
B	66.000	15	Piedmont
C	57.133	15	Coastal

Bonferroni (Dunn) t test for variable: TROUT
NOTE: This test controls the Type I experimentwise error rate, but generaly has a higher Type II error rate than REGWQ.

ALPHA = 0.05 DF = 40 MSE = 14.0933
Critical Value of t = 2.49886
Minimum Significant Difference = 3.4254

Means with the same letter are not significantly different.

BONFERRONI GROUPING	MEAN	N	REGION
A	69.467	15	Mountain
B	66.000	15	Piedmont
C	57.133	15	Coastal

For the case in which there are interactions between the two variables the ANOVA table shown as Table VIII-25 is used (Freund, 1973, p. 353).

The following definitions apply to Table VIII-25 (Freund, 1973, p. 353):

$$\text{total sum of squares} = SST = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{h=1}^{r} x^2_{ijh} - \frac{1}{knr}\bullet T^2 \quad (8.77)$$

$$\text{treatment sum of squares} = SS(Tr) = \frac{1}{nr}\bullet\sum_{i=1}^{k} T^2_{i..} - \frac{1}{knr}\bullet T^2 \quad (8.78)$$

$$\text{block sum of squares} = SSB = \frac{1}{kr}\bullet\sum_{j=1}^{n}\bullet T^2_{.j.} - \frac{1}{knr}\bullet T^2 \quad (8.79)$$

$$\text{interaction sum of squares} = SSI = \frac{1}{r}\bullet\sum_{i=1}^{k}\sum_{j=1}^{n}\bullet T^2_{ij.} - \frac{1}{knr}\bullet T^2_{...} - SS(Tr) - SSB \quad (8.80)$$

$$\text{error sum of squares} = SSE = SST - SS(Tr) - SSB - SSI \quad (8.81)$$

$$\text{number of treatments} = k \quad (8.82)$$

$$\text{number of blocks} = n \quad (8.83)$$

$$\text{number of observations for each k \& n combination} = r \quad (8.84)$$

As an example of two-way ANOVA with interactions we ran the SAS ANOVA procedure (SAS Institute, Inc., 1985b) on the data in Table VIII-18 checking for interactions between stream and region. 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). The results of this test are shown in Table VIII-26. Note that although R-SQUARE is greater than in Table VIII-22, the F VALUE is smaller (yet still significant). The interaction of stream and region has not improved the model noticeably since the interaction F statistic (F=1.82) is not significant at the 95% confidence level (PR>F is 0.1458).

D.2.4.1.3 More on analysis of variance

The above discussion of ANOVA is simple in many respects. For example, in all cases there were no holes in the data set. That is, there was a data value for each combination of variables, a complete block design. In situations where multiple variables are examined the complete block design is not likely to be either

Table VIII-25. Two-way ANOVA with interaction

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	k - 1	SS(Tr)	$MS(Tr) = \frac{SS(Tr)}{k - 1}$	$\frac{MS(Tr)}{MSE}$
Blocks	n - 1	SSB	$MSB = \frac{SSB}{n - 1}$	$\frac{MSB}{MSE}$
Interaction	(n-1)(k-1)	SSI	$MSI = \frac{SSI}{(n-1)(k-1)}$	$\frac{MSI}{MSE}$
Error	kn(r-1)	SSE	$MSE = \frac{SSE}{kn(r - 1)}$	
Total	rkn - 1	SST		

Table VIII-26. Two-way ANOVA with interaction

2-Way ANOVA With River and Region As Treatments
Interaction Between River and Region

Analysis of Variance Procedure

Dependent Variable: TROUT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F	R-SQUARE	C.V.		
MODEL	8	1528.40	191.05	14.67	0.0001	0.765271	5.6209		
ERROR	36	468.80	13.02				ROOT MSE	TROUT MEAN	
CORRECTED TOTAL	44	1997.20					3.6086	64.2	

SOURCE	DF	ANOVA SS	F VALUE	PR>F
RIVER	2	219.733	8.44	0.0010
REGION	2	1213.733	46.60	0.0001
RIVER*REGION	4	94.933	1.82	0.1458

feasible or economical. Gaugush (1986, p. 99-102) discusses both completely randomized and randomized block designs as "the two designs upon which all designs are based." This text and other statistical texts such as those by Winer (1971) and Cochran (1963) should be reviewed to gain a greater understanding of the relationships between ANOVA and sampling design.

Snedecor and Cochran (1980, p. 274-297) discuss the consequences of and remedies for failures in meeting the assumptions of ANOVA. This chapter addresses missing data points, extreme values, correlations between errors, transformations, and Tukey's test for nonadditivity of the effects of different factors. In the discussion of transformations the authors describe variance-stabilizing transformations, square-root transformations for counts, arc sine transformations for proportions, and logarithmic transformations.

More than two variables can be used in ANOVA, but it is advisable that computer resources be made available to facilitate these analyses. The SAS ANOVA procedure (SAS Institute, Inc., 1985b, p. 113) used in the above examples is recommended only for balanced data sets (with some exceptions), but the SAS GLM (general linear models) procedure is capable of handling both balanced and unbalanced data (SAS Institute, Inc., 1985b, p. 114).

Gaugush (1986, p. 114) includes in his clear chapter on ANOVA a discussion of fixed and random treatment-population effects, where treatment-populations are the "populations under consideration in sampling designs or the controlled manipulations employed by the researcher in experimental designs." The following statements taken from Gaugush (1986, p. 114-118) summarize the major points made in distinguishing between fixed and random treatment-population effects.

> Fixed effects refer to a set of treatment-populations which is finite either by nature or by an "a priori" selection process; more importantly, inferences are limited to only those treatment-populations that are sampled.
>
> Due to the nature of fixed effects, the inferential process conventionally involves either or both estimation of treatment-population means or mean differences and tests of hypotheses about significant differences among means.
>
> The ANOVA for fixed effects models is called Model I ANOVA.
>
> Random effects refer to a random sample of treatment-populations; inferences can be made about the entire treatment-populations sampled.
>
> The inferential process involves tests of hypotheses in terms of the variance component being estimated.
>
> The ANOVA for random effects models is called Model II ANOVA.

Gaugush also notes that ANOVA models can be "mixed" when the structure or nature of the treatment-population is partially random (1986, p. 117).

As a final note on ANOVA, Gaugush discusses in some detail factorial and hierarchical arrangements for fixed effects models (1986, p. 118-133). In this section he discusses main effects as effects for which questions addressing differences among means at each level of a single factor can be asked. For example, a main effect question might ask whether significant differences exist among the mean chlorophyll a concentrations at three stations on a lake. In this case station is the factor and each station is a level. Interaction effects involve relationships among levels of different factors. Using the above example, an interaction effects question might ask whether similar differences exist among the mean chlorophyll a concentrations at three stations at three different depths. The two factors involved are station and depth, with the levels being each station and each depth. Again, the reader is referred to Gaugush (1986) and statistics texts for further discussion of these concepts.

D.2.4.2 Analysis of Covariance

Analysis of covariance is a procedure which combines the features of analysis of variance (ANOVA) and regression (Snedecor and Cochran, 1980, p. 365). A typical analysis of covariance model (corresponding to one-way ANOVA) can be represented as (Snedecor and Cochran, 1980, p. 365):

$$Y_{ij} = \alpha_i + \beta(X_{ij} - X..) + \epsilon_{ij} \qquad (8.85)$$

where Y_{ij} = value of jth observation in the ith class

α_i = population mean of the ith class

β = the regression coefficient of Y on X

X_{ij} = value of jth observation in the ith class of another measured variable

$X..$ = mean of all X_{ij} values

ϵ_{ij} = residual of jth observation in the ith class

The analysis of covariance model which corresponds to two-way ANOVA is (Snedecor and Cochran, 1980, p. 365):

$$Y_{ij} = \mu + \alpha_i + p_j + \beta(X_{ij} - X..) + \epsilon_{ij} \qquad (8.86)$$

where μ = the mean of the Y_{ij} observations
p_j = intraclass correlation

In one-way ANOVA where there are two additional, or auxiliary, variables related linearly to Y_{ij}, the model is (Snedecor and Cochran, 1980, p. 365):

$$Y_{ij} = \alpha_i + \beta_1(X_{1ij} - X_{1..}) + \beta_2(X_{2ij} - X_{2..}) + \epsilon_{ij} \quad (8.87)$$

where β_1 = the regression coefficient of Y on X_1

X_{1ij} = the value of the jth observation in the ith class of X_1

$X_{1..}$ = the overall mean of X_{1ij}'s

β_2 = the regression coefficient of Y on X_1

X_{2ij} = the value of the jth observation in the ith class of X_2

$X_{2..}$ = the overall mean of X_{2ij}'s

Analysis of covariance can be used to (Snedecor and Cochran, 1980, p. 365-366):

1. increase the precision of randomized experiments
2. adjust for sources of bias in observational studies
3. examine the nature of treatment effects in randomized experiments
4. study regressions in multiple classifications

The following discussion of analysis of covariance is adapted from a pair of articles written for a technical nonpoint source newsletter (Spooner, 1986a and Spooner, 1986b).

Analysis of covariance can be used to test for differences in the average value for a meteorologic or hydrologic parameter Y (e.g., sediment concentration) between levels of a group variable (e.g., seasons or years) after adjusting for the parameter X (e.g., flow or upstream concentration). In this case X serves as a regression covariate. For this discussion, periods of time in years will be used as the group variable. Hence, the annual average value of Y will be compared for different time periods after adjusting for the effects of X on Y. This type of analysis has merit when the Y and X parameters are correlated (i.e., exhibit a significant linear relationship).

There are several steps in this analysis. First, linear regression of a parameter (Y) versus another parameter (X) should be performed and tested for significance. Paired observations (i.e., collected on the same date, from the same storm, or matched samples representing the same parcel of water as it passes above and below the area of interest) are required.

There are several uses for analysis of covariance; those listed above, as well as to determine if the same relationship of Y and X exists from year to year. In both cases, the added information from pairing X with Y should decrease the residual error as compared to an analysis of variance that did not include X. Therefore, a calculation of an adjusted yearly mean in Y would

allow a more precise comparison between yearly means of parameter Y.

One statistical model for analysis of covariance can be expressed (Snedecor, 1967) as:

$$Y_{ij} = \mu_i + \beta(X_{ij} - X..) + \epsilon_{ij} \qquad (8.88)$$

where:

Y_{ij} = j^{th} observation for Y in year i

μ_i = mean for year i

β = regression coefficient of Y on X

X_{ij} = j^{th} observation for X in year i

X.. = mean value for all X_{ij}'s

ϵ_{ij} = residuals or experimental error for the j^{th} observation for Y in year i

Computation of an estimate for β was performed by least squares (Snedecor, 1967). Note that this version of the covariance model forces the slope of the regression of Y on X to be the same for each year.

Figure VIII-28 shows a general case in which each time period is represented by separate lines which have a commmon (equal) slope.

Adjusted means which correct for the bias in X between years can be calculated by (Snedecor, 1967):

$$\text{Adj } Y_i = Y_i. - b(X_i. - X..) \qquad (8.89)$$

where:

$Y_i.$ = sample mean of paramete Y for year i

b = estimate of the regression coefficient of Y on X

$X_i.$ = sample mean of parameter X for year i

X.. = sample mean value for all X_{ij}'s, the value to which each $Y_i.$ is adjusted.

The result of this calculation is that each yearly mean for Y is adjusted to an overall common value of X. For this model, the common value of X is arbitrary.

The F-test of the adjusted means can be performed to determine if there is sufficient evidence that at least one of the yearly adjusted means is different from the other yearly adjusted means. Simply, the F-test determines if there is still a difference

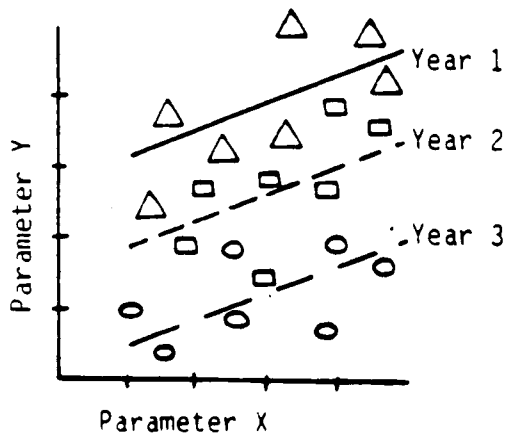


Figure VIII-28. Parameter Y vs. X for three time periods.

between years after adjusting for the covariance. The F-ratio can be calculated by:

$$\text{F-ratio} = \frac{(SSE_R - SSE_F)/(k-1)}{MSE_F} \qquad (8.90)$$

where:

SSE_R = residual sum of squares for the reduced model with a common yearly mean (i.e., $\mu_i = \mu$)

SSE_F = residual sum of squares for the full model which allows for separate means for each year (i.e., equation 1)

k = number of years

MSE = mean square error from the full model with (N-K-1) d.f. (i.e., equation 1)

This F-statistic is compared to an F distribution with (k-1) and (N-K-1) degrees of freedom (d.f.), where K is the number of years and N is the total sample size. Given a significant F-ratio for differences between adjusted yearly means, one may want to compare which adjusted means are significantly different. The confidence interval of a difference between two adjusted means (CI_D) can be obtained by:

$$CI_D = (\text{Adj } Y_i. - \text{Adj } Y_j.) \pm t\sqrt{(S_D^2)} \qquad (8.91)$$

where:

$(\text{Adj } Y_i. - \text{Adj } Y_j.)$ = the difference between the adjusted means for the ith and jth year

t_{2n-2} = Student's t statistic with (2n-2) d.f.

n = sample size per year

S_D^2 = variance of the difference between the adjusted mean of the ith and jth year given by:

$$S_D^2 = MSE \cdot 2/n + \left[\frac{(X_i - X_j)^2}{\sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - X_i.)^2} \right] \qquad (8.92)$$

where:

MSE is the mean square error from the full covariance model and k is the number of years.

If this interval contains the value zero, there is no evidence of a significant difference between the means of the ith year and the jth year.

The above covariance model fits a single pooled or equal slope for all years. An extension to this model which allows separate regression lines for each year is (Snedecor, 1967):

$$Y_{ij} = \mu_i + \beta_i(X_{ij}) + \epsilon_{ij} \qquad (8.93)$$

where:

Y_{ij} = the jth observation for Y in year i

μ_i = the mean for year i

β_i = the regression coefficient of Y on X for year i

X_{ij} = the jth observation for X in year i

ϵ_{ij} = the residuals or experimental error

The F-test for the homogeneity of slopes involves testing if there is a significant reduction in the residual sum of squares (SSE) by allowing separate slopes for each year as compared to the model with pooled slopes. The F statistic for testing the homogeneity of slopes is

$$\text{F-ratio} = \frac{(SSE_R - SSE_F)/(k-1)}{MSE_F} \qquad (8.94)$$

where:

SSE_R = residual sum of squares for the reduced model with a common slope

SSE_F = residual sum of squares for the full model which allows for separate slopes for each year

k = number of years

MSE_F = mean square error from the full model

This F statistic is compared to an F distribution with (k-1) and (N-2k) d.f. Figure VIII-29 shows a general case in which each time period is represented by its individual regression line.

If there is no evidence for separate slopes then the model with common slopes (Equation 93) should be used.

In summary, the homogeneity of slopes, midpoints, and adjusted means between time periods can be tested separately, but simultaneously for the individual time periods (e.g., years). The homogeneity test then indicates if one or more signficant differences exist among the slopes or adjusted means.

If there are not significant differences between the slopes, then all of the periods of time can be represented by a common slope and the relationship between parameters Y and X is constant over the tested period. If a slope is significantly different from

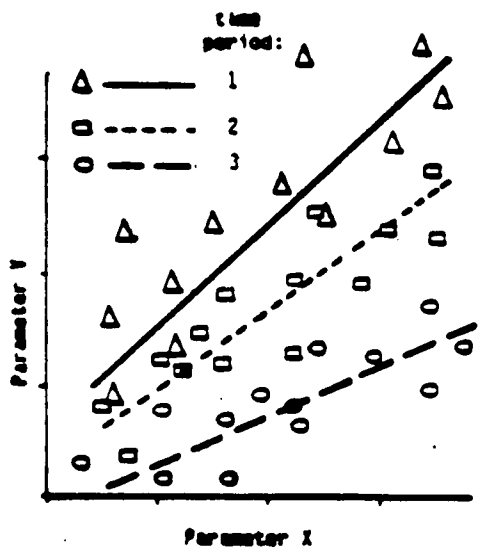


Figure VIII-29. General representation of regression between paired observations of parameters, Y and X. Each unit of time has its own slope.

the common slope, then it is represented by its own slope when testing the homogeneity of the adjusted means. The appropriate model, i.e., either a common slope or unique, individual slopes, should be used for this test.

A significant difference among the adjusted means indicates that at least one mean of the adjusted parameter Y values is different from the other means of adjusted parameter Y values. Examination of which time periods are different indicates if a trend is occurring over the entire time period and what is the direction and magnitude of such a trend. Further trend analyses in which time is a continuous variable may be useful.

Caution must be used when interpreting the results for the comparisons of adjusted means when individual slopes are used. When the slopes are not parallel, the comparisons of adjusted means may not be the most meaningful analysis to perform. One may be more interested in the behavior over the entire range of X. In this case a picture may be worth a thousand words.

Spooner (1986b) goes on to offer tips on how to use the Statistical Analysis System (SAS) computer program (SAS Institute) for analysis of covariance. SAS is a handy program for water quality professionals to use since it handles very large data bases with many variables, and is directly available to all users of EPA's STORET system. The following is adapted from Spooner's SAS tips.

The SAS program statements that generate a covariance model with unique slopes for each year are given by:

```
PROC GLM; CLASS YEAR;
MODEL Y = X YEAR YEAR*X/
SOLUTION SS1 SS3;
LSMEANS YEAR/PDIFF;
```

A summary of the useful information generated by the Type I SS from the above SAS program is given by Freund and Littell (1980, p. 202):

1. X: SS due to a single regression of Y on X, ignoring the grouping.

2. Year: SS due to adjusted yearly differences, assuming a single regression relationship.

3. Year*X: An additional SS due to different regression coefficients for the groups specified by the factor YEAR.

If there is no evidence for separate slopes (i.e., the YEAR*X interaction term in the previous SAS program is not significant) then the model with common slopes (Equation 93) should be used. The SAS model should be rerun without the interaction term.

The adjusted means and the test for the significance of the difference between all possible pairs can be obtained from the PDIFF operations in the LSMEANS statement.

D.2.4.3 Kruskal-Wallis Test

The Kruskal-Wallis test is an extension of the Mann-Whitney test described earlier. The procedure presented below is adapted from Conover (1980).

- Let k equal the number of random samples of possibly different sizes. N is the total number of observations and is computed as

$$N = \sum_{i=1}^{k} n_i \tag{8.95}$$

- Rank all of the data from lowest to highest.

- Compute the sum of ranks, R_i, as the sum of the ranks for the ith random sample. Compute R_i for all k random samples.

- Compute the test statistic, T

$$T = \frac{1}{S^2}\left[\sum_{i=1}^{k}\frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right] \tag{8.96}$$

where

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} R_{ij}^2 - N\frac{(N+1)^2}{4}\right] \tag{8.97}$$

- Reject Ho at the level α, if T exceeds the quantile found in Appendix D with k-1 degrees of freedom.

D.3 Monotonic Trends

In NPS data analysis, we often want to know whether there is a tendency that a pollutant concentration increases or decreases 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 the downstream water quality. To identify a trend, we need to analyze the data on a long-term basis because observed data often fluctuate randomly over the short term. A trend can be visually detected by plotting the observed data versus time, although a statistical test might be required to analyze the trend more accurately. Gilbert (1987) provided a list of time series types (Figure VIII-30). Figure VIII-30(a) represents random data. We can see that the data fluctuate randomly along their average value. The data do not tend to increase or decrease over

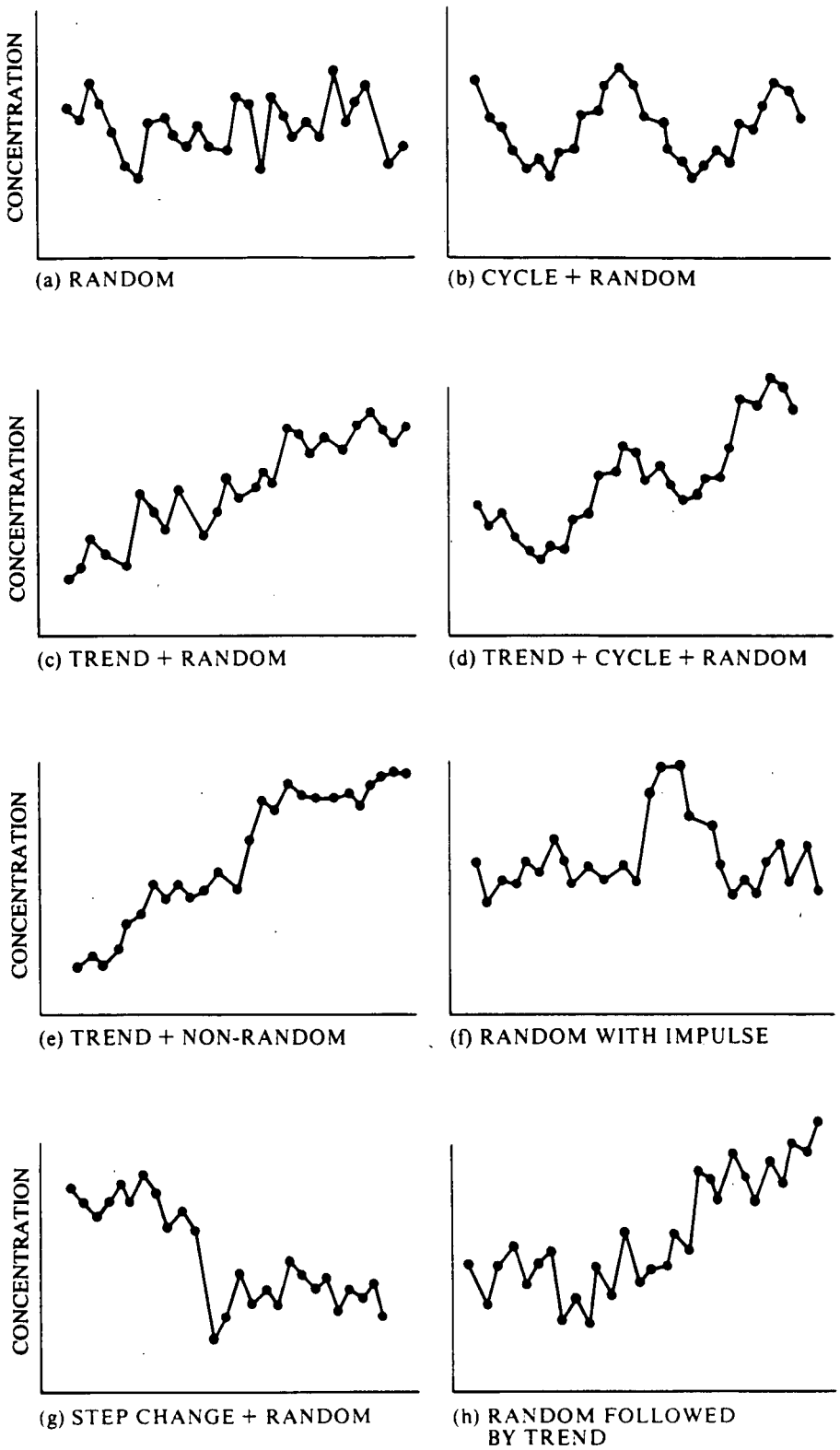


Figure VIII-30. Types of time series (after Gilbert, 1987).

a long term. Figure VIII-30(b) represents a cyclical time series. As with Figure VIII-30(a), the data do not tend to increase or decrease over the long term but follow a cyclical pattern. Many time series have a cyclical nature, such as the mean monthly temperature, or mean monthly flow rate in a stream, observed over many years.

Figure VIII-30(c) represents an increasing-trend time series. The data tend to increase over time in the long term although there is a short-term fluctuation. Figure VIII-30(d) represents an increasing trend with a cyclical pattern. The data cycle during a short period but tend to increase over the long term. The monthly mean temperature over many years under a global warming environment (if it exists) might follow this pattern. Figure VIII-30(e) represents an increasing trend with data dependency. We can see that large data tend to be followed by large data and small data by small data and the overall trend is increasing. Figure VIII-30(f) represents a random time series with an impulse. There is a spike in the data, but it does not affect the long-term tendency of the data, which is fluctuating along the average without a tendency of increase or decrease. This may happen to a time series of pollutant concentration in a stream affected by a spill. The spill causes a spike in the concentration but does not affect the long-term patten.

Figure VIII-30(g) represents a step change of the time series. There is no trend before or after the step change; that is, before or after the sudden decrease, the time series tends to fluctuate along its average. The only difference is that the average value after the step change is less than that before the change. This type of time series might reflect stream pollutant measurements over time, when an upgrade of a wastewater treatment plant is implemented. Figure VIII-30(h) represents a random time series followed by an increasing trend. The first part of the time series has no trend while the second part tends to increase over time. In NPS data, the first part might represent a measurement of stream pollutants before industries were established in the upstream area. The industries might release pollutants during their operation, causing a gradual increase of stream pollutants, which is indicated by the second part of the time series.

The Mann-Kendall Test (Mann, 1945; Kendall, 1975) can be used to detect trends when the data available contain such value as "below detection limit." For trends with a cyclical nature or those interrupted by a sudden change (step trend), the seasonal Kendall test (Hirsch et al., 1982) can be used. When a trend is detected, the slope of the trend (change of data per time unit) can be estimated by the Sen Slope Estimator (Sen, 1968). The slope can then be used to predict future data. In the following, we discuss some tests commonly used in NPS trend analysis and provide examples. Berryman et al. (1988) provided a comprehensive review of nonparametric trend analysis. After discussing these nonparametric procedures, regression analysis is reviewed.

D.3.1 Cox and Stuart Test

The Cox and Stuart test can be used to detect whether there is a trend for a time series. For a series of observed data, X_1, X_2, ..X_i,.. X_m, where m is the total number of data, the data are arranged in the order of time they are obtained. We need to know whether there is an upward or downward trend for these data. In other words, we need to know whether the data tend to increase or decrease with time, or just fluctuate randomly along their average value. The assumptions and test procedure are as follows:

Assumptions:
1. The random variables X_1, X_2, ...X_i, ...X_m are mutually independent.
2. The measurement scale the data is at least ordinal (true for most NPS data)
3. Either the data are identically distributed or there is a trend.

Example for Cox and Stuart Test

The annual total rainfall for 21 years is given in Table VIII-27. Determine whether there is a tendency that the rainfall tends to decrease or increase over time.

Solution:

1. Construct hypothesis.
   Ho: There was no trend for the annual rainfall
   Ha: There was either an upward or downward trend for the annual rainfall. (This is a two-sided test.)
   Use alpha=0.05 as confidence level.

2. Calculate number of signs.
   Group the data as in columns 1 and 2 in Table VIII-28. As the total number of data is odd, the value at the middle point (X_{11}=36.4) is omitted. Assign the sign for each pair of data, as in column 3 of Table VIII-28. From the table we get B=6, n=10.

3. Draw Conclusion
   As the hypothesis is a two sided test, use alpha = 0.05 and n = 10 to get the l and n values. From appendix P, we find l = 2, u = 9. Because B is less than u and is greater than l -1, Ho is rejected. We conclude that there is no trend for the annual rainfall data at the confidence level of 0.05.

Procedure for the Cox and Stuart Test

1. Construct hypotheses:
   Two-sided test:
   Ho: No trend exists
   Ha: There is a trend, either upward or downward
   One-sided test 1:
   Ho: No upward trend exists
   Ha: There is an upward trend
   One-sided test 2:
   Ho: There is no downward trend
   Ha: There is a downward trend

2. Calculate number of signs.
   Group the data into pairs (X1, X1+c), (X2, X2+c)...(Xi, Xi+c)... (Xm-c, Xm), where c=m/2 if m is a even number, and c=(m+1)/2 if m is an odd number. If m is odd, the data at the middle of the time series is excluded. Assign a sign for each pair. If Xi<Xi+c, a "+" sign is assigned. If Xi>Xi+c, a "-" sign is assigned. A "tie" is assigned when Xi=Xi+c. Let n=total number of pairs with either a "+" sign or "-" sign (ties are excluded), and B=total number of pairs with a "+" sign.
3. Draw conclusion.
   Two-sided test:
   From Appendix P, obtain the l and u values for the corresponding significant level (alpha) and n. Reject Ho and accept Ha if B≥u or B<l-1.
   One-sided test 1:
   Multiply the alpha by 2 and use the new alpha and n to find out u in Appendix P. If B≤u, reject Ho and accept Ha.
   One-sided test 2:
   Multiply the alpha by 2 and use the new alpha and n to find out l in Appendix P. Reject Ho and accept Ha if B>l-1.

Table VIII-27. Annual total rainfall for 21 years

Year	Rainfall (in.)	Year	Rainfaill (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		

Table VIII-28. Sign calculation for Cox and Stuart test

X	X_{i+c}	Sign
40.2	51.2	+
53.4	54.3	+
43.5	41.5	-
37.7	44.8	+
50.2	46.7	-
38.7	51.8	+
47.8	49.5	+
39.5	34.1	-
44.9	33.2	-
41.7	53.7	+

D.3.2 Mann-Kendall Test

Like the Cox and Stuart test, the Mann-Kendall test also analyzes the sign of the difference between later-measured data and the earlier-measured data. As discussed above, the Cox and Stuart test breaks the time series into two groups of data from the middle point, then pairs the data in each group based on the order of time and takes the sign of difference for the pair. The Mann-Kendall test, however, pairs a later-measured datum to all data measured earlier. This results in a total of n(n-1)/2 possible pairs of data, where n is the total number of data in the time series. In the Cox and Stuart test, there are only n/2 pairs of data for a time series of X_1, X_2, ...X_i, ...X_n, where n is the total number of measurements and the data are arranged in the order of time they are measured. The assumptions and test procedure for the time series are as follows:

Assumptions.

1. The random variables X1, X2, ...Xi, ...Xn are mutually independent.

2. The measurement scale of the data is at least ordinal (true for most NPS data).

3. Either the data are identically distributed or there is a trend.

Procedure for the Mann-Kendall Test

1. Construct hypotheses.
   Two-sided test:
   Ho: No trend exists
   Ha: There is a trend, either upward or downward
   One-sided test 1:
   Ho: No upward trend exist
   Ha: There is an upward trend
   One-sided test 2:
   Ho: There is no downward trend
   Ha: There is a downward trend

2. Calculate signs.
   Take the difference between the later-measured values to all earlier-measured values, $(X_j - X_k)$, where $j > k$. Thats (X_2-X_1), (X_3-X_1), (X_3-X_2)....(X_n-X_1).. (X_3-X_2), (X_4-X_2),... (X_n-X_2).... (X_n-X_{n-2}), (X_n-X_{n-1}), as listed in Table VIII-26. There are a total of n(n-1)/2 possible differences. Assign a sign integer, sgn, to each difference according to the following criteria:

Xk-Xj	sgn
>0	1
0	0
<0	-1

See Table VIII-29 for calculation scheme.

Table VIII-29. Data difference calculation for Mann-Kendall test.

X_1	X_2	X_3	.....	X_{n-1}	X_n
	$X_2 - X_1$	$X_3 - X_1$ $X_3 - X_2$	....	$X_{n-1}-X_1$ $X_{n-1}-X_2$ .... $X_{n-1}-X_{n-2}$	X_n-X_1 X_n-X_2 .... X_n-X_{n-2} X_n-X_{n-1}

3. Sum all the integers.

$$S=\sum_{k=1}^{n-1}\sum_{j=k+1}^{n} sgn(X_j-X_k) \qquad (8.98)$$

The S value is basically the sum of all positive and negative signs. It is easily seen that when s is a large positive number, it indicates that 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, the 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 statistical hypothesis test for the trend is done in the following step.

4. Draw conclusion.

From Appendix R, use the n and S value to obtain the probability value p, which is the probability that the calculated S is greater than the specified S value in the table when no trend is present. Compare the probability value p to the confidence level (alpha) as follows.

Two-sided test:
Multiply the p value from appendix R by 2. If the multiplied value is less than alpha, reject Ho and accept H1.

One-sided test 1:
Reject Ho if the probability value from appendix R is less than alpha.

One-sided test 2:
Use the absolute value of the calculated S value to get the probability value in appendix R. Reject Ho when the probability value in appendix R is less than alpha.

The calculation for the Mann-Kendall test becomes tedious when n is large. Gilbert (1987) provided a FORTRAN program to alleviate the computation effort. When n is greater than 40, Gilbert recommends using the normal approximation method, which approximates the test to a normal distribution problem. For data with seasonal cycles, the modified Mann-Kendall, also referred to as the Seasonal Kendall test, should be used (Hirsch et al., 1982). Basically, the test computes the S and variance of S as in the Mann-Kendall test for each season, then sums them and computes Z statistics, and then uses the standard normal distribution table to test statistically significant trends. An example for the test was provided by Gilbert (1987).

D.3.3 Trend Test with Seasonal and Autocorrelated Data

Both the Mann-Kendall and seasonal Kendall tests assume no correlations among the data being analyzed. In NPS area, many data follow seasonal patterns (Figures VIII-30(b) and VIII-30(d)) and might be autocorrelated (Figure VIII-30(e)). While the seasonal Kendall test works with seasonal data, both tests require independence of data. Autocorrelation data can cause errors in both results. Therefore, in general trend analysis, we should test for seasonality and autocorrelation first. In particular, we need to ensure that the data are not correlated, or, if they are, to make some corrections before applying the Mann-Kendal test or the seasonal Kendall test. Reckhow et al. (1988) provide a test procedure for this purpose as follows.

1. Remove all the known deterministic effects.

If we find a deterministic pattern in the data, we try to remove the deterministic portion. For example, when analyzing a time series of total phosphorus concentration in the river, if we know the concentration is a function of the flow rate, we should try to deterministically model the relationship between the flow and concentration, then subtract the modeled concentration data from the original data. We can then use the residual for the following step.

2. Remove seasonal and trend effects.

To see whether there is a seasonal pattern in the data, make a correlogram using equations provided by Pankratz (1983). A correlogram is a plot of autocorrelation values versus the time lag, and seasonality can be found in the correlogram. For example, a significant correlation value for the time lag of 12 for a monthly time series indicates that an annual cyclical pattern exists in the data. When a seasonal pattern is found in the data, we can deseasonalize and detrend the data using the seasonal Sen slope estimator with the seasonal median. We can then use the deseasonalized and detrended data for the following step.

3. Test for autocorrlation and select appropriate trend analysis methods.

Create a correlogram again with the deseaonalized and detrended data from step 2. If the autocorrelation is significant, we must correct the autocorrelation. This can be done with the method provided by Hirsch and Slack (1984), in conjunction with the Mann-Kendall Test or seasonal Kendall test. If correlation is not significant, a standard Mann-Kendall test or standard seasonal Kendall test can be applied.

D.3.4 Regression

D.3.4.1 Overview

The following overview of regression is adapted from an article by Spooner and Maas (1984).

Regression analyses are used to test mathematical expressions that relate one or more dependent variables (response variables, Y's) to one or more independent variables (X's, usually considered to be measured without error) (Spooner and Maas, 1984). There are two primary purposes for utilizing regression techniques: 1) to model the behavior of a dependent variable, and 2) to predict the future behavior of the dependent variable. In the first approach, one may wish to examine cause and effect relationships or establish past trends. In prediction applications, one may use the information collected to date to predict the future behavior of one variable based on the expected behavior of another variable. Trend analyses such as time series, slope and intercept estimations, and confidence intervals become the key ingredients in the latter applications. In all cases, regression techniques attempt to

explain as much of the variation in the Y variable(s) as possible. As such, the goal in regression analysis is to minimize the sum of the squared distances of the Y values from the regression curve (i.e., the unexplained variation). Hence, the techniques used are often referred to as "least squares."

Linear univariate regression, or linear regression, relates the behavior of a single dependent variable (Y) to one or more independent variables by a linear function. This is in contrast to multivariate analysis which incorporates simultaneous consideration of several dependent variables. The "linear" term reflects the fact that the coefficients (βi's) of the X's in the regression equation are linear. The X's may be linear or non-linear (e.g., X can represent X2, X3, X-1, etc.). The general form of a linear univariate regression model, where ϵ is an error term, is shown as Equation 52, whereas Equation 53 is a linear regression model containing just one independent variable.

$$Y = \beta_o + \beta_1 \bullet X_1 + \ldots + \beta_i \bullet X_i + \epsilon \qquad (8.99)$$

$$Y = \beta_o + \beta_1 \bullet X_1 + \epsilon \qquad (8.100)$$

If the coefficients of the X's are not linear, then the model is a "non-linear" model. Non-linear models are commonly applied to physical systems, but are more difficult to analyze because iterative least squares techniques are involved. Thus, many non-linear equations are linearized before analyses.

Multiple linear regression (also referred to as multiple univariate linear regression, and multiple regression) analyses are used when a single independent variable does not explain sufficiently the behavior of the dependent variable (Y). Two or more independent variables (X's) are utilized in a linear function to describe the behavior of Y. Equation 52 (general linear univariate model) depicts a multiple regression model, while Equation 53, with one independent variable, is just a simplified form of the general linear univariate model.

In NPS research multiple (univariate) linear regression analyses are most often used to determine the extent to which the value of a water quality parameter (Y) is influenced by land use or hydrologic factors (X's) such as crop type, soil type, percentage of land treatment, amount of rainfall, etc. Practical applications of these regression results include the ability to predict the water quality impacts of changes in the independent variables.

The most common multiple regression equations are those that involve a list of independent variables that may affect the behavior of Y (i.e., time, treatments, rainfall, etc.). These X's can be continuous or discrete variables. Polynomial equations are common models for empirical descriptions of the behavior of the Y's.

In both modeling and predictive uses of multiple regression, the user may wish to select the best "subset" of the independent

variables. There is a point where useful information is no longer gained by adding new variables. Also, correlations among X's (multicollinearity) 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 ANOVA 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. An NPS applications might be to examine the effect of different BMP implementation programs on several water quality parameters.

In summary, the types of regression models discussed in this guidance are:

REGRESSION MODEL	DEFINITION
Univariate	One dependent variable (Y)
Multivariate	More than one Y and X
Linear regression	Coefficients (β_i's) on the X's are linear
Non-linear regression	Coefficients (β_i's) on the X's are non-linear
Multiple	One Y as a function of several X's

Researchers are strongly encouraged to read the detailed discussion of regression in statistics texts such as Snedecor and Cochran (1980), Cochran (1979), and Srivastava and Khatri (1979) for a more complete discussion of this important statistical procedure.

D.3.4.2 Linear regression

Regression is a statistical method often used in water quality analyses. Referring back to our list of water quality monitoring objectives, we find that regression analysis can be used in developing baseline information (relationships between variables of interest), trend analysis (parameter versus time), and model development and/or verification.

D.3.4.2.1 The model

In this guidance we will not delve into the origins and derivations of regression techniques, but will instead just discuss the uses of, assumptions made in, and equations used for regression analyses. Regression analysis, in general, is "a means of studying the variation of one quantity (dependent variable) at selected

levels of another quantity (independent variable)" (Remington and Schork, 1970). A key to regression is that we express and quantify this relationship with a mathematical model. The model to be discussed in this section is the linear model of the general form (Freund, 1973):

$$y = a + bx \qquad (8.101)$$

where y is the dependent variable, x is the independent variable, and a and b are numerical constants representing the y-intercept and slope, respectively. For a general theoretical approach to regression the reader is referred to Winer (1971) and other statistics texts.

The first step in applying linear regression is to examine the data to see if linear regression "makes sense." That is, use a bivariate scatter plot (section III.D.1) to see if the points approximate a straight line. If they do, then linear regression "makes sense," but if they don't fall in a straight line it may be that data transformation is needed or that a non-linear relationship should be used.

To illustrate the use of linear regression we use the data in Table VIII-30 which are a selected (biased) subset of calibration data for a plot-size runoff sampler (Dressing et al., 1987). In this data set the sampling percentage (split) was measured for a range of flow rates. The scatter plot in Figure VIII-31 shows that linear regression can be applied to the data.

Table VIII-30. Runoff Sampler Calibration Data

X Flow Rate (gpm)	Y Split (%)	X Flow Rate (gpm)	Y Split (%)
52.1	2.65	25.8	2.83
19.2	3.12	17.6	2.84
4.8	3.05	37.6	2.60
4.9	2.86	41.4	2.54
35.2	2.72	40.1	2.58
44.4	2.70	47.4	2.49
13.2	3.04	35.7	2.60
13.9	3.19		
n = 15	$\bar{x}$ = 28.89	Σx^2 = 15,940.33	
	Σx = 433.3	$\bar{x}^2$ = 834.44	
	Σy = 41.81	y^2 = 117.25	
	Σxy = 1,166.93	$\Sigma(x-\bar{x})^2$ = 3,423.74	

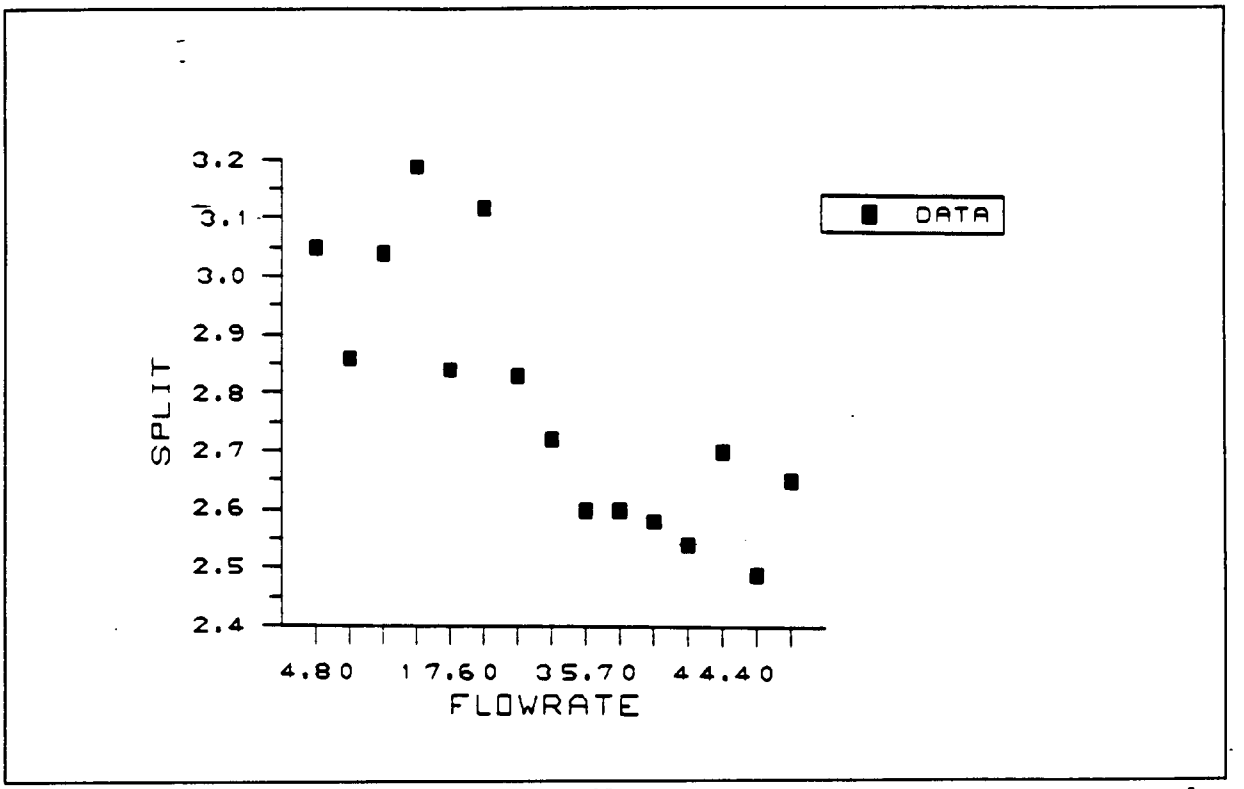


Figure VIII-31. Split versus flowrate.

D.3.4.2.2 Least squares method

To develop the regression line we use the method of least squares which "demands that the line which we fit to our data be such that the sum of the squares of the vertical deviations (distances) from the points to the line is a minimum" (Freund, 1973). To determine the values of a and b in the above general linear model, the following normal equations can be used (Freund, 1973, p. 394):

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \qquad (8.102)$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \qquad (8.103)$$

For the above data we used the SAS REG procedure (SAS Institute, Inc., 1985b) to determine the regression line parameters. Table VIII-31 shows that the intercept (INTERCEP) is 3.132 and the slope (FLOWRATE) is -0.012. Thus, the linear model for predicting split versus flow rate is:

$$\text{Split} = 3.132 - 0.012 \bullet \text{Flowrate}$$

DEP VARIABLE: SPLIT

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	0.48662350	0.48662350	28.410	0.0001
ERROR	13	0.22266984	0.01712845		
C TOTAL	14	0.70929333			

ROOT MSE	0.1308757	R-SQUARE	0.6861
DEP MEAN	2.787333	ADJ R-SQ	0.6619
C.V.	4.695373		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER=0	PROB > |T|
INTERCEP	1	3.13171795	0.07291415	42.951	0.0001
FLOWRATE	1	-0.01192192	0.002236705	-5.330	0.0001

TEST: NUMERATOR: 58.5125 DF: 2 F VALUE: 3416.1012
DENOMINATOR: .0171284 DF: 13 PROB >F : 0.0001

OBS	ACTUAL	PREDICT VALUE	STD ERR PREDICT	LOWER95% MEAN	UPPER95% MEAN	LOWER95% PREDICT	UPPER95% PREDICT	RESIDUAL
1	2.6500	2.5106	0.0619	2.3768	2.6444	2.1978	2.8234	0.1394
2	3.1200	2.9028	0.0401	2.8161	2.9895	2.6071	3.1986	0.2172
3	3.0500	3.0745	0.0636	2.9371	3.2119	2.7601	3.3888	-0.0245
4	2.8600	3.0733	0.0634	2.9363	3.2103	2.7591	3.3875	-0.2133
5	2.7200	2.7121	0.0366	2.6329	2.7912	2.4185	3.0057	.0079337
6	2.7000	2.6024	0.0484	2.4977	2.7070	2.3009	2.9039	0.0976
7	3.0400	2.9743	0.0487	2.8691	3.0796	2.6727	3.2760	0.0657
8	2.8300	2.8241	0.0345	2.7496	2.8986	2.5317	3.1165	.0058677
9	2.8400	2.9219	0.0422	2.8308	3.0130	2.6248	3.2190	-0.0819
10	2.6000	2.6835	0.0390	2.5992	2.7677	2.3884	2.9785	-0.0835
11	2.5400	2.6382	0.0439	2.5434	2.7329	2.3399	2.9364	-0.0982
12	2.5800	2.6536	0.0421	2.5627	2.7446	2.3567	2.9506	-0.0736
13	2.4900	2.5666	0.0534	2.4512	2.6821	2.2612	2.8720	-0.0766
14	2.6000	2.7061	0.0371	2.6260	2.7862	2.4122	3.0000	-0.1061
15	3.1900	2.9660	0.0476	2.8632	3.0688	2.6651	3.2669	0.2240

SUM OF RESIDUALS	2.37588E-14
SUM OF SQUARED RESIDUALS	0.2226698
PREDICTED RESID SS (PRESS)	0.3112981

Table VIII-31. Linear regression output.

Now that we have a simple linear regression line the obvious questions to ask are:

(1) How well does the line fit the data?

(2) How can the line be used?

Before we address these questions we should review the assumptions made in performing linear regression. These assumptions are (Gaugush, 1986):

1. The x values are constants measured without error.

2. The units in the sample with any one value of x are randomly chosen from all units in the target population with that value of x.

3. The values of y at any particular value of x are normally distributed about the regression line.

4. The variance of y is the same at each value of x, i.e., the variance is homogenous.

5. The deviations of the observed points from the line are independent. Multiple observations in a single unit will not, in general, be independent.

The first assumption is made because no consideration is given to variation in x. The third assumption is necessary only if hypothesis tests are performed or confidence intervals are determined. The fourth and fifth assumptions are needed for the parameter estimates to be the most precise possible (Gaugush, 1986).

Tests for some of the above assumptions are described by Ponce (1980a). To test the fourth assumption we plot the residuals against the predicted y values. As shown in Figure VIII-32 this plot should be a band around a residual of 0 (Ponce, 1980a. Figure VIII-33 shows this plot for the above data set (Table VIII-30). This plot, although not having a well-defined band around zero, does indicate that the variance is fairly homogenous.

To test the independence of the deviations of the observed points from the regression line, Ponce (1980a recommends that we plot the residuals in time-sequence (Figure VIII-34). Assuming that the data in Table VIII-30 are in time-sequence we plotted the ordered residuals as shown in Figure VIII-35. This plot illustrates that the residuals are independent.

Gaugush (1986) notes that regression analysis is generally robust against violation of several assumptions. In other words, the estimates of a and b may be useful even if assumptions are violated somewhat. However hypothesis tests and confidence intervals will not be correct if the normality assumption is violated.

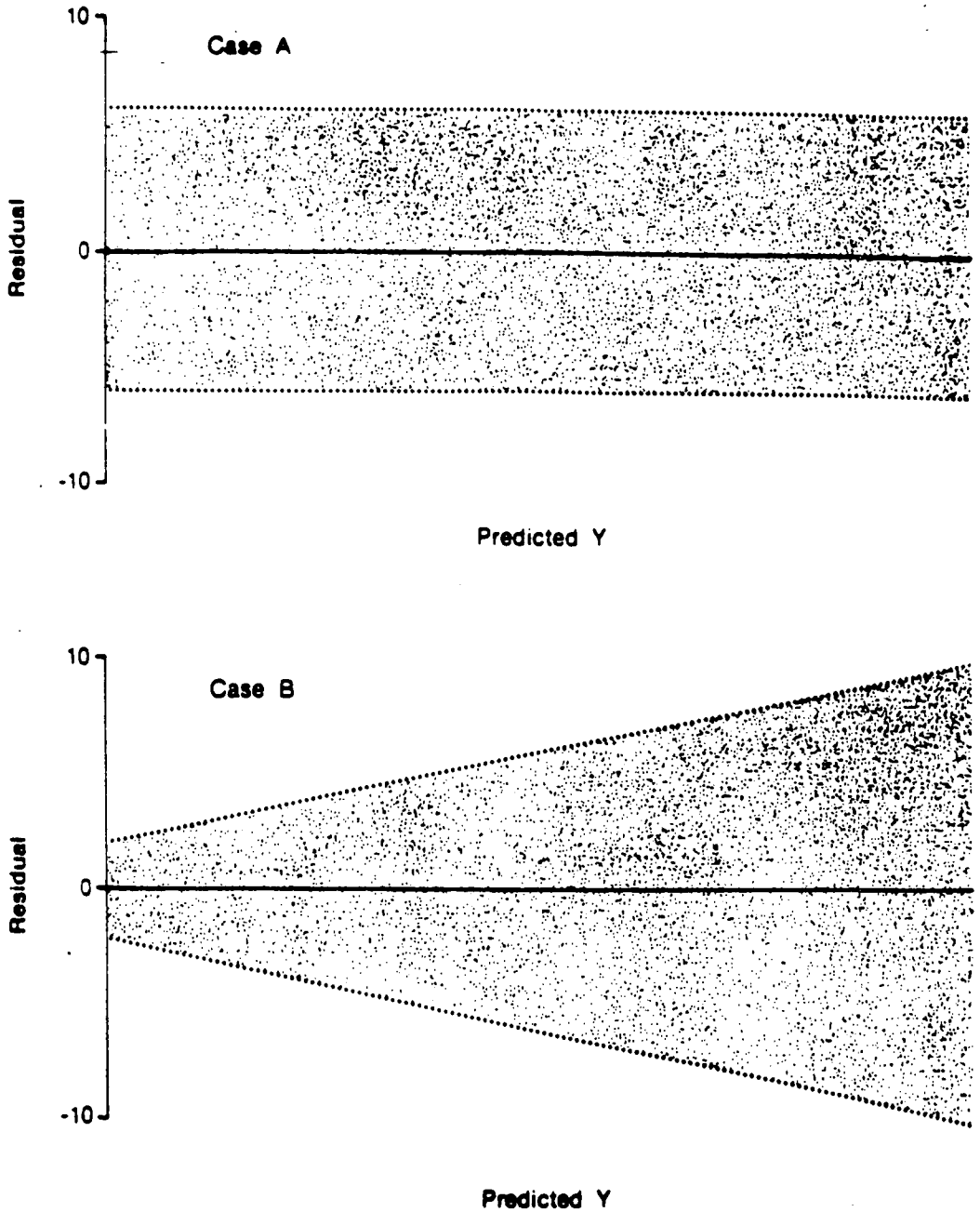


Figure VIII-32. Plot of residuals versus predicted values. (SOURCE: Ponce, 1980a)

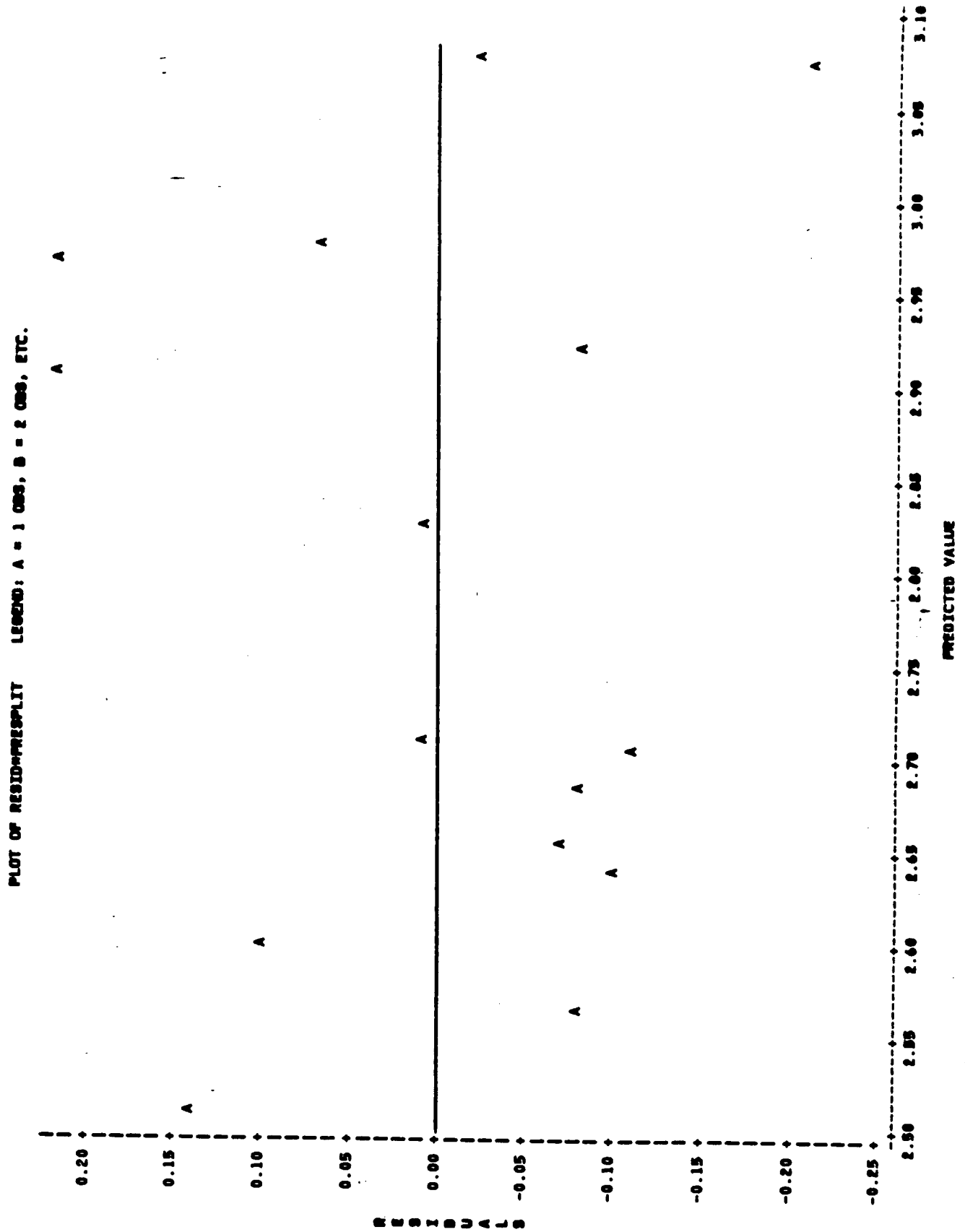


Figure VIII-33. Plot of residuals versus predicted values.

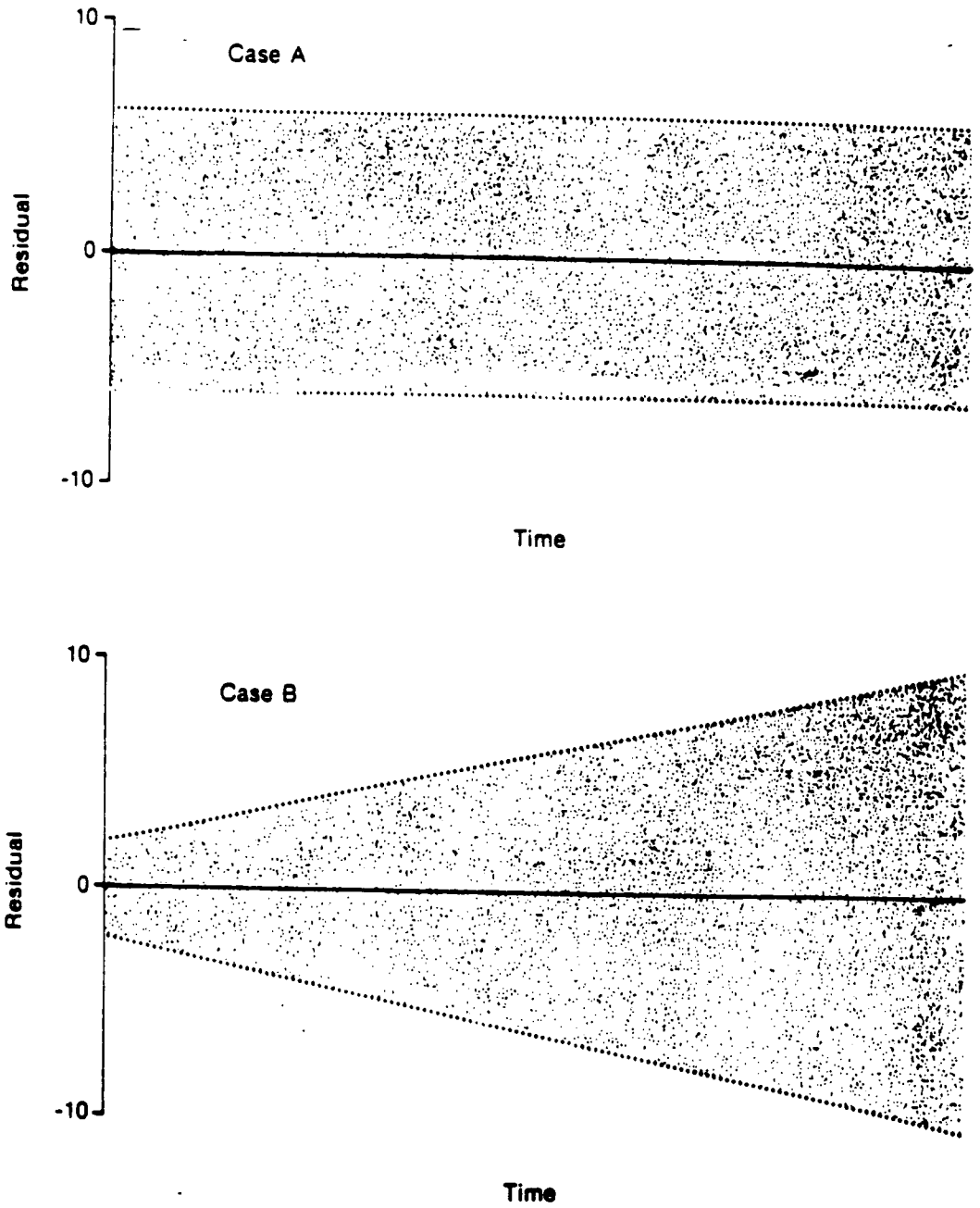


Figure VIII-34. Plot of residuals versus time. (SOURCE: Ponce, 1980a)

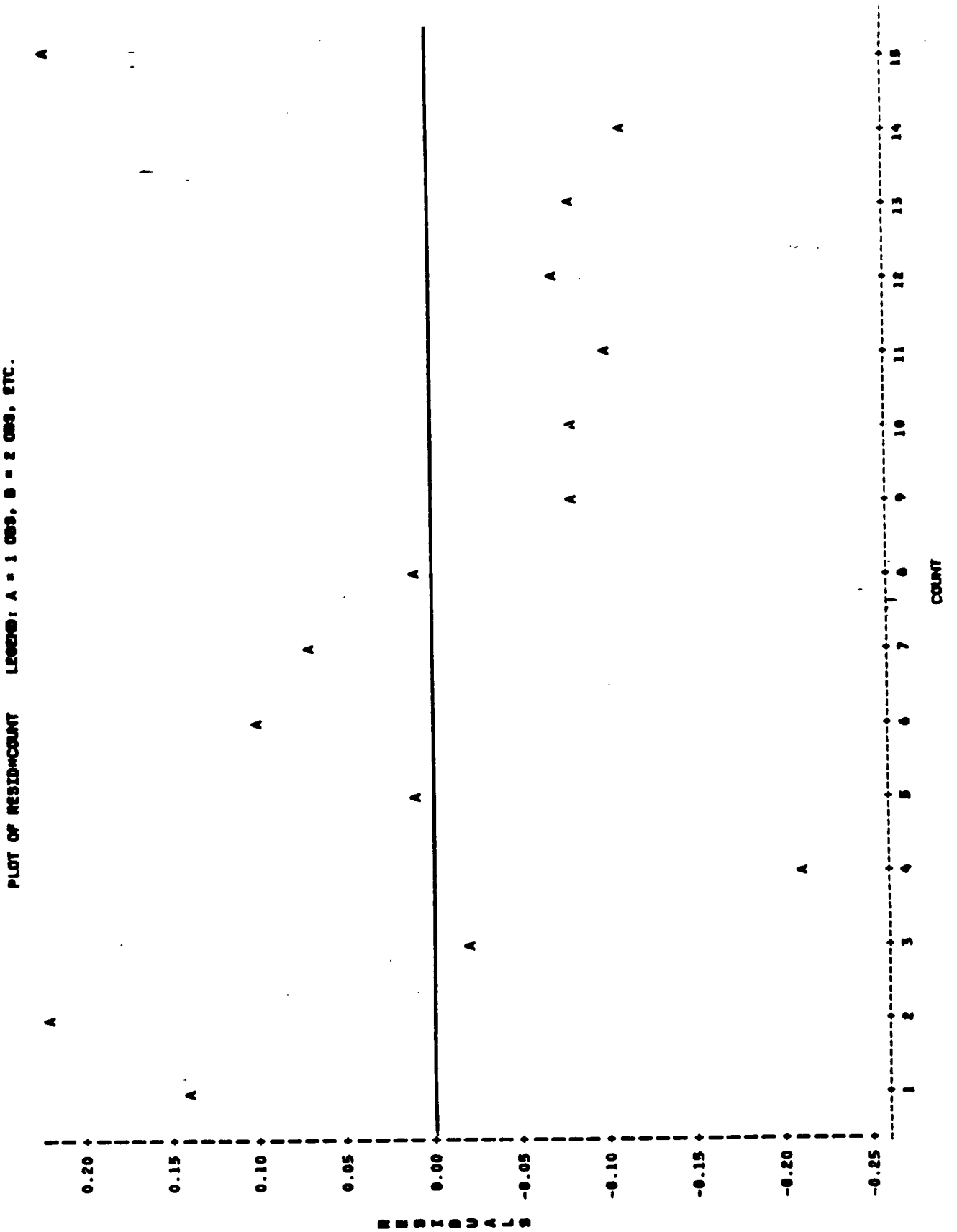


Figure VIII-35. Time sequence plot of residuals.

D.3.4.2.3 Tests of fit

To determine how well the regression line fits the data we can test several things:

1. Test whether the slope is zero.

2. Test whether the intercept is zero.

3. The confidence interval for a.

4. The confidence interval for b.

5. The proportion of variation in y explained by the model.

Before discussing tests for the above hypotheses and estimates we need to define several mathematical terms used in the calculations to follow. The following terms and definitions are taken from Ponce (1980a) and Gaugush (1986):

1. The corrected total sum of squares is the total amount of variation in y and is calculated by computing the sum of the squared deviations $(Y_i - \bar{Y})$:

$$SS_{corr.\ total} = \Sigma(Y_i - \bar{Y})^2 = \Sigma y^2 \qquad (8.104)$$

2. The model (regression) sum of squares is the amount of variation in y that is associated with the regression on x, and is calculated as:

$$SS_{model} = \Sigma(\hat{Y}_i - \bar{Y})^2 = (\Sigma xy)^2/\Sigma x^2 \qquad (8.105)$$

where the $\hat{Y}_i$ (reads as "Y hat") values are the Y values generated from the regression equation

$$\hat{Y}_i = a + b \bullet X_1 \qquad (8.106)$$

3. The error (residual) sum of squares is the amount of variation in y not associated with the regression on x, and is calculated as:

$$SS(error) = SS(total) - SS(model) \qquad (8.107)$$

4. The model mean square (MS) is the model sum of squares divided by the number of degrees of freedom (d.f.) for the model. For linear regression, d.f.(model) is always 1.

$$MS(model) = SS(model)/d.f.(model) \qquad (8.108)$$

5. The error mean square is the error sum of squares divided by the error degrees of freedom (n-2).

$$MS(error) = SS(error)/d.f.(error) \qquad (8.109)$$

To test whether the slope is different from zero (Ho: b=0) we can use either an F-test or a t-test (Gaugush, 1986). For the F-test we first calculate the model and error mean squares. The test statistic is

$$F = MS(model)/MS(error) \qquad (8.110)$$

and is compared versus the critical F value with d.f.(numerator) equal to d.f.(model) and d.f.(denominator) equal to d.f.(error).

In our example we let SAS calculate the F value (Table VIII-31). The calculated F value of 28.410 was compared by SAS against F table values with the result that there is a probability of only 0.0001 that a greater F value could be obtained if Ho is true. In other words, we reject Ho with a significance of 99.99 percent (1-0.0001). The slope is not zero. (Perform this same test using the above equations to check the results from SAS.)

Gaugush (1986) notes that a t-test can also be used to test the null hypothesis of zero slope. The test statistic is

$$t = (b-0)/s_b = b/s_b \qquad (8.111)$$

where b is the calculated slope and s_b is the standard error of b. The standard error of b is calculated as (Gaugush, 1986, p. 64)

$$s_b = \sqrt{(MS_{error}/\Sigma(x_i - \bar{x})^2)} \qquad (8.112)$$

Again using the data from Table VIII-30, we calculate the sum of the deviations of x from the mean x (flow rate) to be 3,423.7373, so s_b is equal to 0.0022367. The t value is -0.01192/0.0022367, or -5.33. Using a two-sided test at α=0.05 with 14 degrees of freedom, the t value for comparison is (Appendix C) -2.1448. Clearly, the t-test results agree with the F-test results; the slope is not zero.

The same information calculated above was calculated by SAS and appears in Table VIII-31. The standard error of b is the standard error of FLOWRATE, the t value is the T FOR HO: PARAMETER=0 for FLOWRATE, and the PROB > |T| is the probability that a higher t value could be obtained if Ho is true. Hence, we reject Ho at 99.99 percent significance.

The null hypothesis of a zero intercept can also be tested with a t-test (Gaugush, 1986, p. 64). For this case the equation is

$$t = (a-0)/s_a = a/s_a \qquad (8.113)$$

where a is the calculated intercept and s_a is the standard error of a, calculated as

$$s_a = [MS_{error} \bullet (1/n + \bar{x}^2/\Sigma(x_i - \bar{x})^2)]^{1/2} \qquad (8.114)$$

From Table VIII-30 the mean of x (flow rate) is 28.8867. Using the above equation, s_a is calculated to be 0.0729142. Therefore, the calculated t value is 3.13172/0.0729142, or 42.951. We reject the null hypothesis of zero intercept.

The SAS output in Table VIII-31 shows s_a as the standard error for INTERCEP, the t value as T FOR HO: PARAMETER=0 for INTERCEP, and the significance as PROB > |T| for INTERCEP. Thus, SAS determined that we can reject Ho at 99.99 percent significance (1-0.0001).

SAS also has as an option under PROC REG a simultaneous test that (1) the intercept is equal to a specified value and (2) the slope is equal to another specified value (SAS Institute, Inc., 1985b). A good reason for using this test is that the level of significance determined above (using the same data) for a and b are incorrect due to the fact that the estimates for a and b are not independent (Freund, 1973). The following SAS steps are recommended for performing a test for a zero intercept and slope of one (Spooner et al., 1986):

```
PROC REG; MODEL Y = X;
TEST INTERCEPT = 0, X = 1;
```

We performed this test on the data in Table VIII-30, and the results are shown in Table VIII-31 in the area designated as TEST:. Clearly, from the F value (400605) and associated probability of a greater F if Ho is true (0.0001), we must reject the Ho that the intercept is zero and the slope is 1. To illustrate a case in which we cannot reject Ho, we ran the same test with intercept = 3.1 and slope = -0.01 (close to PARAMETER ESTIMATE values in Table VIII-4). The results of this test are an F value of 0.6172 and an associated probability of 0.5545 for a greater F value. Hence, there is a 55.45 percent chance that a greater F value could result if Ho is true. This joint Ho can not be rejected.

To determine the confidence intervals for a and b the following formulas are used (Freund, 1973, p. 415):

$$a \pm t_{\alpha/2} \cdot s_e \cdot [1/n + n \cdot x^2/(n(\Sigma x^2) - (\Sigma x)^2)]^{1/2} \qquad (8.115)$$

$$b \pm t_{\alpha/2} \cdot s_e / [(n(\Sigma x^2) - (\Sigma x)^2)/n]^{1/2} \qquad (8.116)$$

where s_e is the standard error of the estimate, equal to Freund, 1973, p. 414):

$$s_e = [(\Sigma y^2 - a(\Sigma y) - b(\Sigma xy))/n-2]^{1/2} \qquad (8.117)$$

As noted above, the confidence intervals for a and b are not independent, and it is best to not use the same data set for tests concerning both.

Using the data in Table VIII-30, we will calculate the confidence limits for the slope. The standard error of the estimate is 0.1309

and the resulting 95 percent confidence interval is -0.017 to -0.007 (verify this).

The coefficient of correlation, r, measures the strength of linear relationships (Freund, 1973). It ranges from -1 to 1, with the extreme values representing the strongest association and 0 representing no correlation. The statistic r is calculated as (Freund, 1973, p. 422):

$$r = \frac{n(\Sigma xy) - (\Sigma x) \bullet (\Sigma y)}{[n(\Sigma x^2) - (\Sigma x)^2]^{\frac{1}{2}} \bullet [n(\Sigma y^2) - (\Sigma y)^2]^{\frac{1}{2}}} \qquad (8.118)$$

The sign of r should be the same as the sign of the slope.

To explain the proportion of the variation in y explained by the model we calculate the coefficient of determination (the square of the coefficient of correlation) which is defined as (Anderson and McLean, 1974):

$$r^2 = \frac{SS(model)}{SS(total)} \qquad (8.119)$$

Values for the coefficient of determination range between 0 and 1, with 1 representing the case where all observed y values are on the regression line.

For the data in Table VIII-30, the coefficient of determination is:

$$r^2 = \frac{0.48662350}{0.70929333} = 0.686$$

We had SAS calculate this value also, and it can be seen in Table VIII-31 that the two results agree. Thus, the linear regression model, which has flow rate as the independent variable, explains 68.6 percent of the variability in y (split).

The correlation coefficient calculated from sample data is an estimate of the corresponding population parameter, p (rho), referred to as the population correlation coefficient (Freund, 1973). To establish a confidence interval for p we will perform normal correlation analysis which requires the assumption of x as a normally distributed random variable in addition to the assumptions of normal regression analysis (remember that we considered x to be a constant in linear regression).

Using Appendix K (Remington and Schork, 1970) we can determine graphically the 95 percent and 99 percent confidence limits for *p* knowing only n and r. For the data in Table VIII-30, n is 15 and r is -0.8283 (negative slope, square root of r-square). So, the 99 percent confidence limits from Appendix K are approximately -0.95 to -0.50.

A t-test can be used to test the null hypothesis that p is zero. The t statistic (n-2 degrees of freedom) for this test is (Freund, 1973, p. 427):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (8.120)$$

For the above data t would be -5.33. From Appendix C the two-sided t value for 95 percent significance (d.f.=13) is -2.1604. Therefore, we reject the null hypothesis (p = 0) and accept the Ha that p is not zero. This test will most likely seem trivial in many applications.

The Fisher Z transformation can be used to test null hypotheses of rho equal to values other than zero (Freund, 1973). For this test r is changed into a Z value using (Freund, 1973, p. 428):

$$Z = \frac{1}{2} \cdot \log_e [(1+r)/(1-r)] \qquad (8.121)$$

The test statistic is (Freund, 1973, p. 428):

$$z = \frac{Z-\mu_z}{1/\sqrt{n-3}} = (Z-\mu_z)\sqrt{n-3} \qquad (8.122)$$

where μ_z is the Z value for r equal to p. For illustration we test the Ho that p is -0.8 for the regression performed using the data from Table VIII-30. The z statistic becomes:

$$z = (-1.1827 + 1.0986)\sqrt{12} = -0.2913$$

The two-sided z statistic at 95 percent significance (Appendix B) is -1.960; therefore we cannot reject Ho.

A confidence interval for p can be determined by calculating such an interval for μ_z, and then converting that interval to r and p (Freund, 1973). The formula for the confidence interval for μz is (Freund, 1973, p. 429):

$$Z - \frac{z_{\alpha/2}}{\sqrt{n-3}} < \mu_z < Z + \frac{z_{\alpha/2}}{\sqrt{n-3}} \qquad (8.123)$$

Since μ_z equals Z when r equals p, the interval for μ_z gives an interval for Z in which p can be determined by substituting p for r in the above equation for Z. Again using our sample data, the 95 percent confidence interval for μ_z becomes:

$$-1.1827 - \frac{1.960}{\sqrt{12}} < \mu_z < -1.1827 + \frac{1.960}{\sqrt{12}}$$

or $$-1.7485 < \mu_z < -0.6169$$

Solving for p:

$$\text{Lower Limit of } p: \quad -1.7485 = 1/2 \ln \frac{(1 + p)}{(1 - p)}$$

$$\text{Upper Limit of } p: \quad -0.6169 = 1/2 \ln \frac{(1 + p)}{(1 - p)}$$

or $$-0.94 < p < -0.55$$

Freund (1973) provides a table of Z values to simplify this procedure. Note that the 95 percent confidence interval for p calculated using the Fisher Z transformation (-0.94 to -0.55) compares favorably with the 99 percent confidence interval (-0.95 to -0.50) determined graphically using Appendix K.

D.3.4.2.4 Using the regression line

The most obvious use of the regression line is for prediction of y values for selected values of x. For example, using the regression established above we can estimate the split for any flow rate (it is not good practice, however, to predict values beyond the range of test conditions). For a flowrate of 10 gpm the split is 3.01 percent, while for a flow rate of 50 gpm the split is 2.53 percent.

Since, in most cases, the regression line will not fit the data perfectly, we should quantify the uncertainty associated with our predicted values. We can use the regression line 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). This is due to the fact that single observations vary more than means.

The equation for the confidence interval for the population mean y at $x = x_o$ is (Remington and Schork, 1970, p. 268):

$$P\left[a + b(x_o - x) - t_{1-\alpha/2} s_{y \cdot x} \cdot [1/n + (x_o - x)^2/((n-1) \cdot s_x^2)]^{1/2} < \mu_{y \cdot x_o}\right.$$

$$\left. < a + b(x_o - x) + t_{1-\alpha/2} s_{y \cdot x} \cdot [1/n + (x_o - x)^2/((n-1) \cdot s_x^2)]^{1/2}\right] \quad (8.124)$$

This interval is most narrow at x and widens as x_o moves farther from the mean. By calculating the interval at each point along the

regression line a curve such as the one in Figure VIII-36 can be plotted.

The equation for the prediction interval for individual values of y at $x = x_o$ is (Remington and Schork, 1970, p. 269):

$$a + b(x_o - \bar{x}) \pm t_{1-\alpha/2} s_{y \bullet x} \bullet [1 + 1/n + (x_o - x)^2/(n-1) \bullet s_{x^2}]^{1/2} \quad (8.125)$$

Figure VIII-37 shows how this interval would appear as an overlay on the population mean confidence interval shown in Figure VIII-28.

Table VIII-31 contains both the mean confidence interval values and the individual prediction interval values calculated by SAS (SAS Institute, Inc., 1985b). A plot of these values is shown in Figure VIII-38. For the above predictions of split at 10 gpm and 50 gpm calculate the population mean confidence limits and prediction interval limits and check them against the plot in Figure VIII-38. They should fall on the appropriate curves.

One of the simplest (in theory) nonpoint source (NPS) 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 upon the parameter used. For NPS 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.) may be required before regression analysis can be successfully applied.

In many cases water quality parameters are regressed against flow. This is particularly relevant in NPS studies. In analysis of covariance regressions against flow are often performed prior to the analysis of variance (ANOVA) (Spooner et al., 1985). One of the implicit goals of NPS control is to change the relationship between flow and pollutant concentration or load. 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 NPS control efforts (Spooner et al., 1985). Again, this will be discussed in greater detail under analysis of covariance.

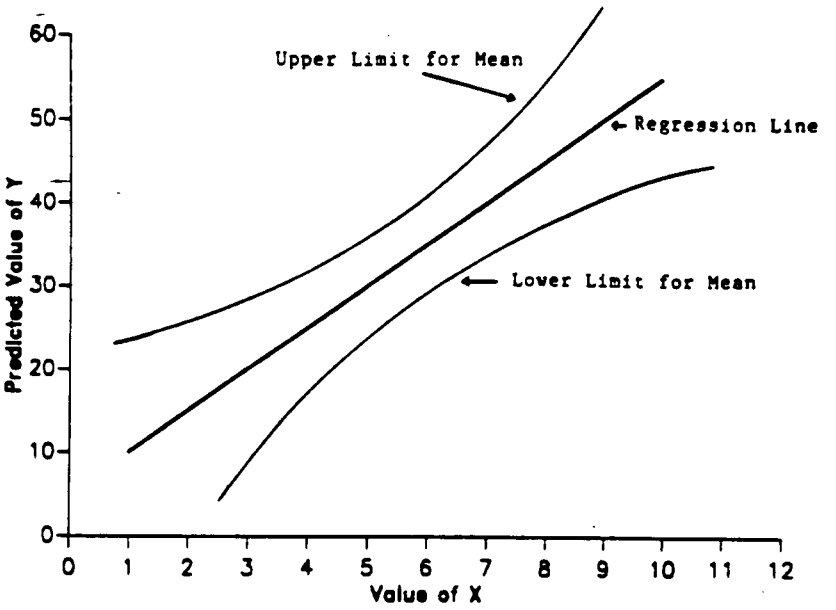


Figure VIII-36. Regression line with confidence intervals.

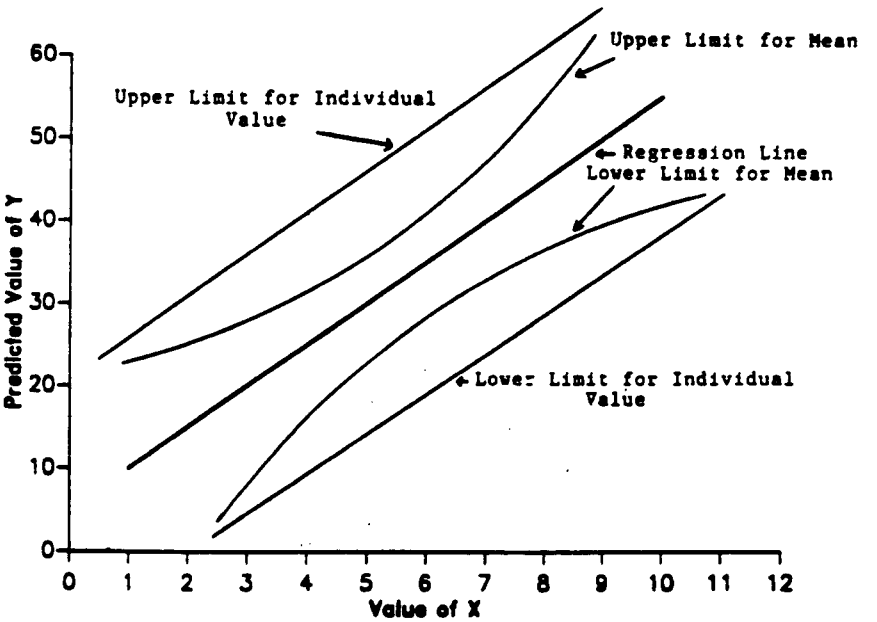


Figure VIII-37. Regression line with conficence intervals.

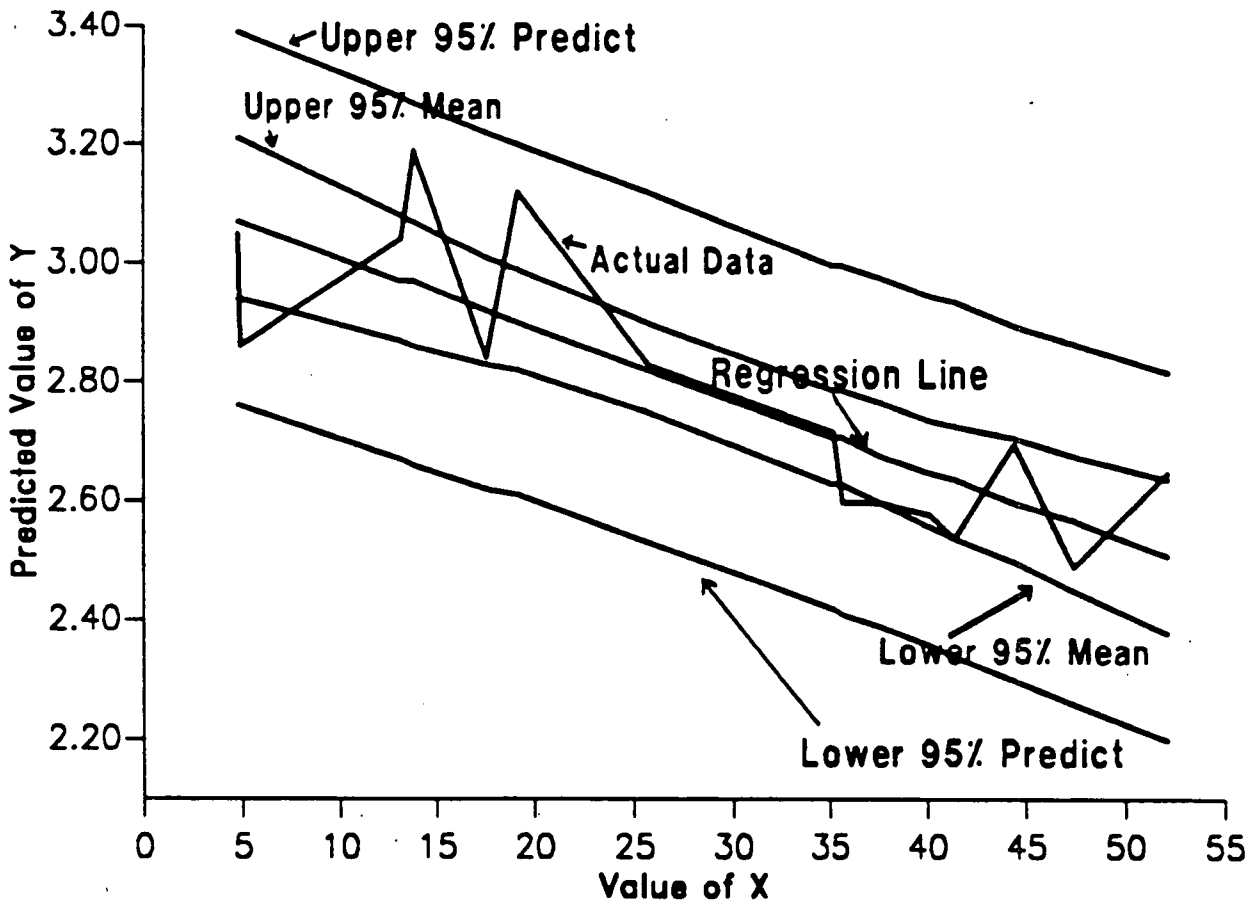


Figure VIII-38. Plot of mean and individual value confidence limits.

D.3.4.3 Non-linear regression

The discussion of non-linear, or curvilinear, regression is limited to cases where the non-linear 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).

Non-linear regression for curvilinear functions involves transformation to linear equations, followed by use of the methods described above for simple linear regression. Transformations for many functions are shown in Appendix L (Ponce, 1980a). Crawford et al. (1983) list numerous regression models most often applied by the USGS for flow adjusting concentrations.

Freund (1973, p. 400-402) illustrates application of non-linear regression techniques with an example for which an exponential curve of the form

$$y = a \bullet b^{x} \qquad (8.126)$$

best describes the relationship between y and x. The equation is converted to a linear form by taking the logarithms of both sides so that

$$\log y = \log a + x(\log b). \qquad (8.127)$$

By susbstituting log y for y, log a for a, and log b for b, the normal equations become

$$\Sigma \log y = n(\log a) + (\log b)(\Sigma x) \qquad (8.128)$$

$$\Sigma x(\log y) = (\log a)(\Sigma x) + (\log b)(\Sigma x^{2}). \qquad (8.129)$$

The equations can be solved for log a and log b.

D.3.4.4 Multiple regression

Multiple regression is applied to quantify a relationship between a dependent variable (Y) and more than one independent variables (X's) (Gaugush, 1986, p. 77). The assumptions made for linear regression also apply to multiple regression (Ponce, 1980a, p. 127). As for linear regression, the method of least squares is also used to determine the best multiple regression line. The general linear model to consider is (Ponce, 1980a, p. 128):

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n \qquad (8.130)$$

The corresponding normal equations are (Ponce, 1980a, p. 128):

$$(\Sigma x_1^2)b_1 + (\Sigma x_1 x_2)b_2) + (\Sigma x_1 x_3)b_3 + \ldots + (\Sigma x_1 x_n)b_n = \Sigma x_n y \quad (8.131)$$

$$(\Sigma x_1 x_2)b_1 + (\Sigma x_2^2)b_2 + (\Sigma x_2 x_3)b_3 + \ldots + (\Sigma x_2 x_n)b_n = \Sigma x_2 y \quad (8.132)$$

$$(\Sigma x_1 x_3)b_1 + (\Sigma x_2 x_3)b_2 + (\Sigma x_3^2)b_3 + \ldots + (\Sigma x_3 x_n)b_n = \Sigma x_3 y \quad (8.133)$$

.

.

$$(\Sigma x_1 x_n)b_1 + (\Sigma x_2 x_n)b_2 + (\Sigma x_3 x_n)b_3 + \ldots + (\Sigma xn^2)b_n = \Sigma x_n y \quad (8.134)$$

After solving for the b's, a can be calculated from (Ponce, 1980a, p. 128):

$$a = Y - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \quad (8.135)$$

Ponce (1980a, p. 128-133) presents an example of multiple regression using three independent variables (Appendix M). The reader is encouraged to follow through this 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 text book formulae, 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 researcher consults a statistician).

Gaugush (1986, p. 77-81) provides an example of multiple regression in which the SAS procedure GLM (SAS, Institute, Inc., 1985b) is used (Appendix N). 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:

1. An F test indicates the significance of the regression.

2. The coefficient of multiple determination (r-square), which is calculated as in simple linear regression, shows the proportion of variation in Y explained by the model.

3. 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, p. 711) 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. 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, p. 763). 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 (X) variables 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 if they are individually important to the regression on Y. The multiple correlation coefficient (R-square) is a measure of the fraction of variation in Y explained by the linear regression on X1, X2, ..., Xn 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).

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - F)^2}{\sum_{i=1}^{n} (Y_i - F)^2} = \frac{\text{SS Regression}}{\text{SS Total}} = 1 - \frac{\text{SS Error}}{\text{SS Total}} \qquad (8.136)$$

The null hypothesis that $R^2=0$ (i.e., $b_1 = b_2 = b_3 = \ldots = b_k = 0$) can be tested using the F-statistic to determine whether or not the regression model explains any of the variation in Y. The F-statistic is $(n-k-1)R^2/(k-1)(1-R^2)$ with (k-1) and (n-k-1) degrees of freedom. It should be noted that (k-1) is the degrees of freedom (df) for the regression model SS and (n-k-1) is the df for the error SS.

A small R^2 may be significantly different from zero if the sample size (n) is large. Conversely, a large R^2 may be insignificant if n is small compared to the number of X's (k) in the model.

If R^2 is small, then most of the variation in Y is unexplained by the linear regression model. This remaining "noise" may be random variation, or may 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 may change.

When new variables are added to the model, R^2 always increases. This explains why a large R^2 may 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:

$$R_a^{-2} = (1 - R^2)(n - 1)/(n-k-1) \qquad (8.137)$$

How does one test if 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, then the variable is important. The F-statistic is:

$$F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \qquad (8.138)$$

where df_R and df_F are the degrees of freedom for the reduced and full model SS, respectively. Appendix O is an example of using R^2 for analysis of agricultural nonpoint source project data (Spooner, 1984).

5. Multivariate regression

Multivariate regression can be a very useful technique in NPS M&E 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 well beyond the scope of this document as it involves considerable knowledge of multivariate normal theory and matrix theory. Readers are referred to statistics texts (e.g., Srivastava and Khatri, 1979) for more on multivariate regression.

Users of SAS (SAS Institute, Inc., 1985b, p. 655-710) 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, p. 666):

```
MODEL Y1 Y2 = X1 X2 X3                                    (8.139)
```

where

Y1 and Y2 are the dependent variables
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, p. 1-2):

1. Wilks' lambda
2. Pillai's trace
3. Hotelling-Lawley trace
4. Roy's maximum root

E. Extreme Events

The majority of NPS 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. One of the key characteristics that separate environmental, and in particularly NPS-influenced data is the presence of extreme events. In Section VIII.C, summary statistics were used to characterize the average conditions. Section VIII.D described procedures for determining whether conditions are changing. This section is appropriate for evaluating extreme conditions. This is important for evaluating standard violations or evaluating peak concentrations to determine if a BMP was effective.

E.1 Rainfall Analyses

E.1.1 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.
- Use Table VIII-32 using C_v to estimate values of K for different return periods.
- Compute the annual precipitation (X_c) for different return periods using Equation 8.140.

$$Xc = \bar{x}(1+ C_v K) \qquad (8.140)$$

The adequacy of the length of record n can be evaluated using Equation 8.141 (Mockus, 1960)

$$Y = (4.30t \log_{10} R)^2 + 6 \qquad (8.141)$$

where Y is the minimum record length in years, t is the student t quantile 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. Application of Equation 8.141 is iterative.

Table VIII-32. Theoretical Log-probability Frequency Factors (Chow, 1954).

	Return Period (years)					
	1.01	2	5	20	100	
	Probability in percentage equal to or greater than given variate					
C_s	99	50	20	5	1	C_v
0	-2.33	0	0.84	1.64	2.33	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

E.1.2 Storm Return Period

The method developed by Hershfield (1961) is the most usual 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 VIII-39 through VIII-42) and applying Equation 8.142 developed by Weiss (1962).

$$I = 0.0256(C-A)x + 0.000256[(D-C) - (B-A)]xy + 0.01(B-A)y + A \quad (8.142)$$

where I is the rainfall amount and A, B, C, and D are interpolated values from Figures VIII-39 through VIII-42, respectively. The return period, x, and duration, y, are taken from Table VIII-33 and VIII-34, respectively.

Table VIII-33. Linearized Rainfall Frequency Variate for Equation 8.142 (Weiss, 1962).

Return Period in Years	1	2	5	10	25	50	100
Linearize Variate, x	-6.93	0	9.2	16.1	25.3	32.1	39.1

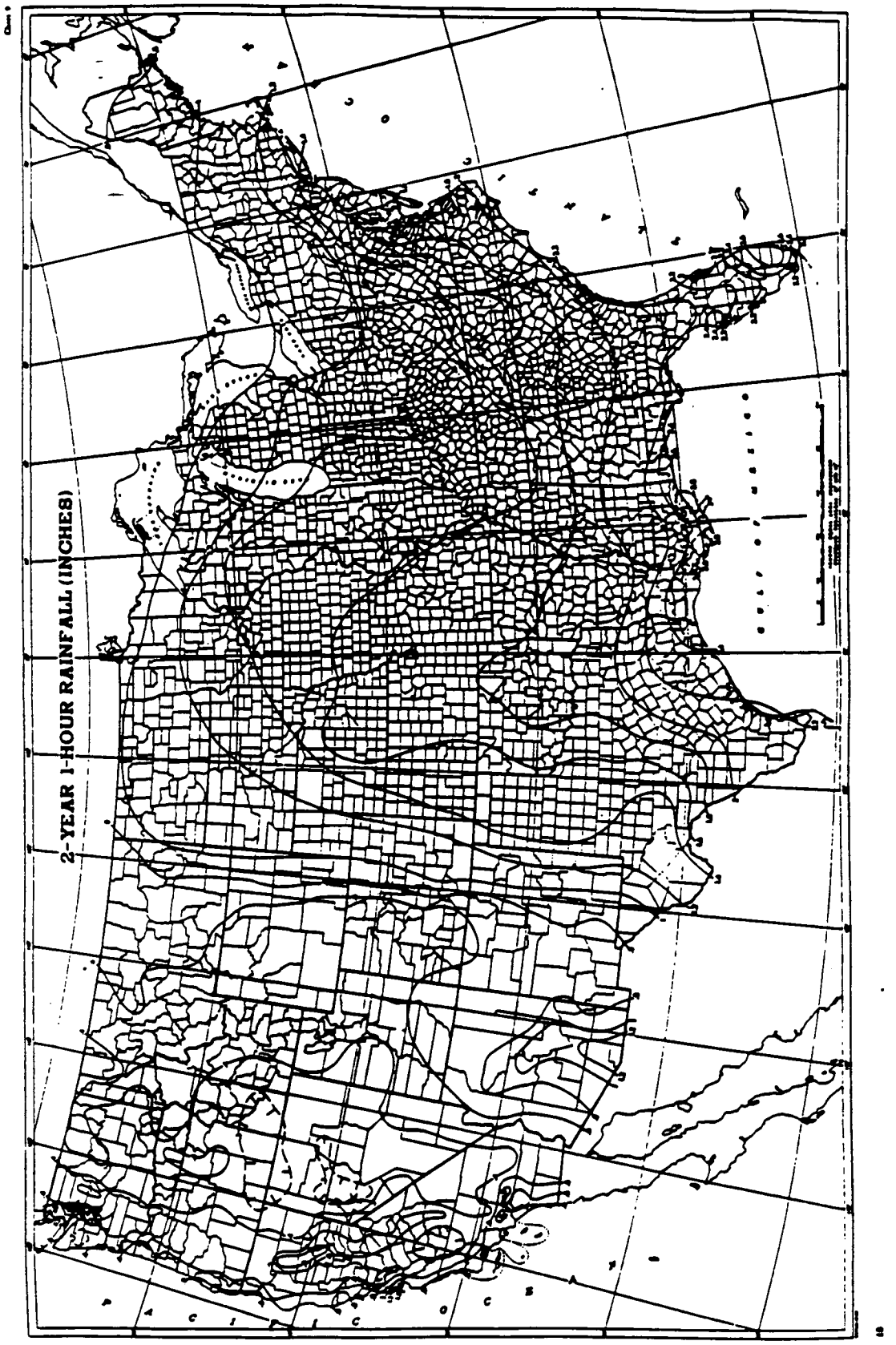


Figure VIII-39. One-hour rainfall to be expected at a return period of 2 years (Hershfield, 1961).

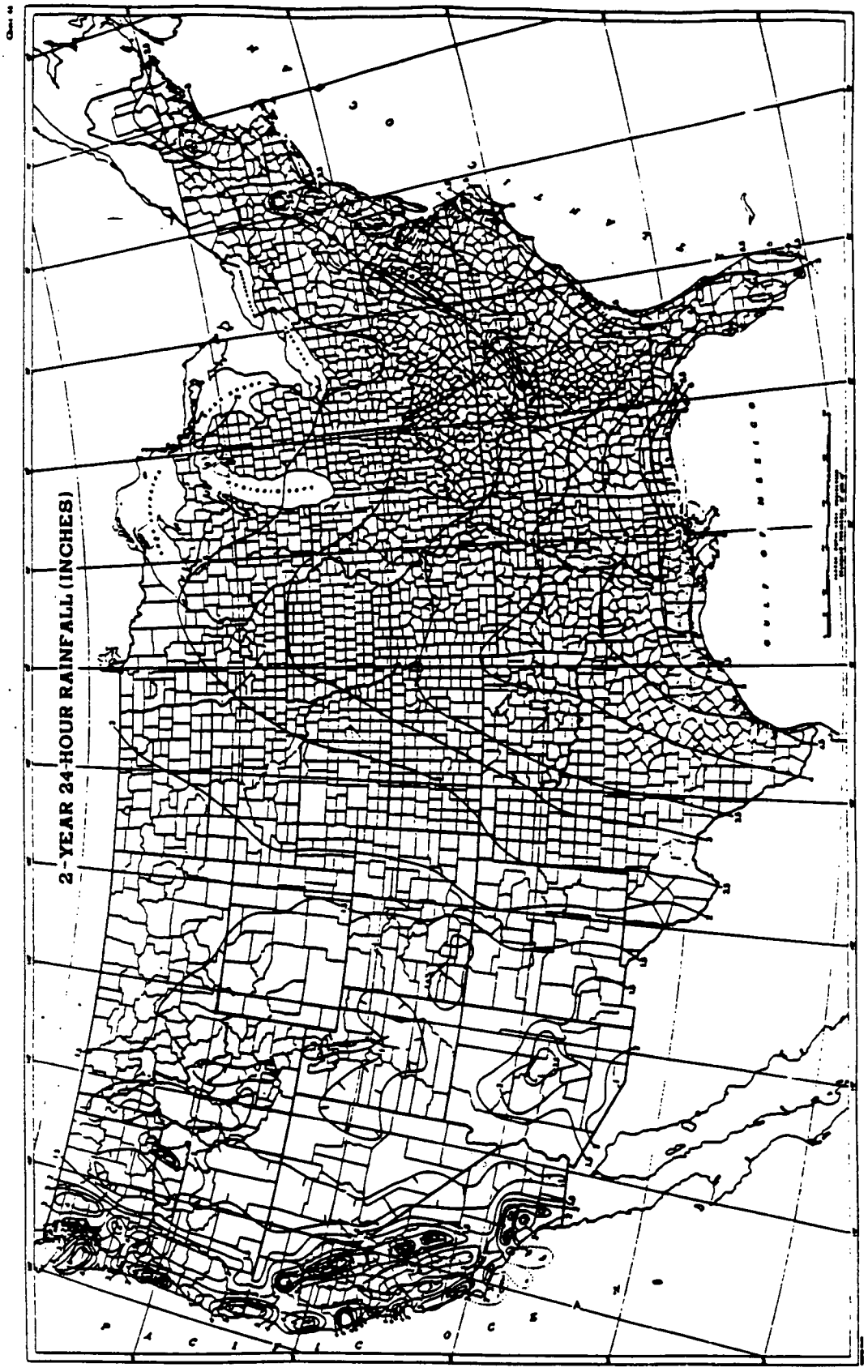


Figure VIII-40. 24-hour rainfall to be expected at a return period of 2 years (Hershfield, 1961).

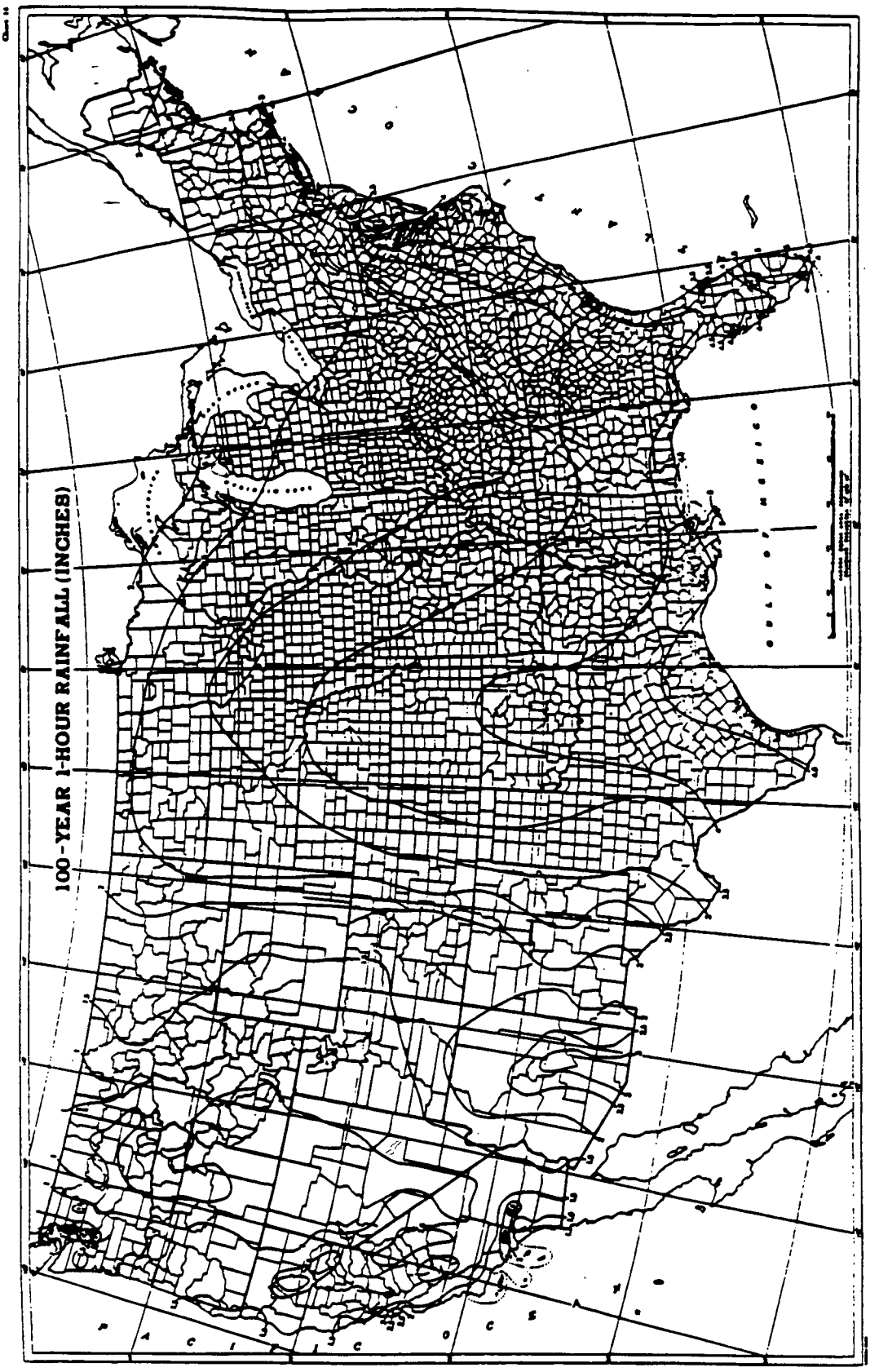


Figure VIII-41. One-hour rainfall to be expected at a return period of 100 years (Hershfield, 1961).

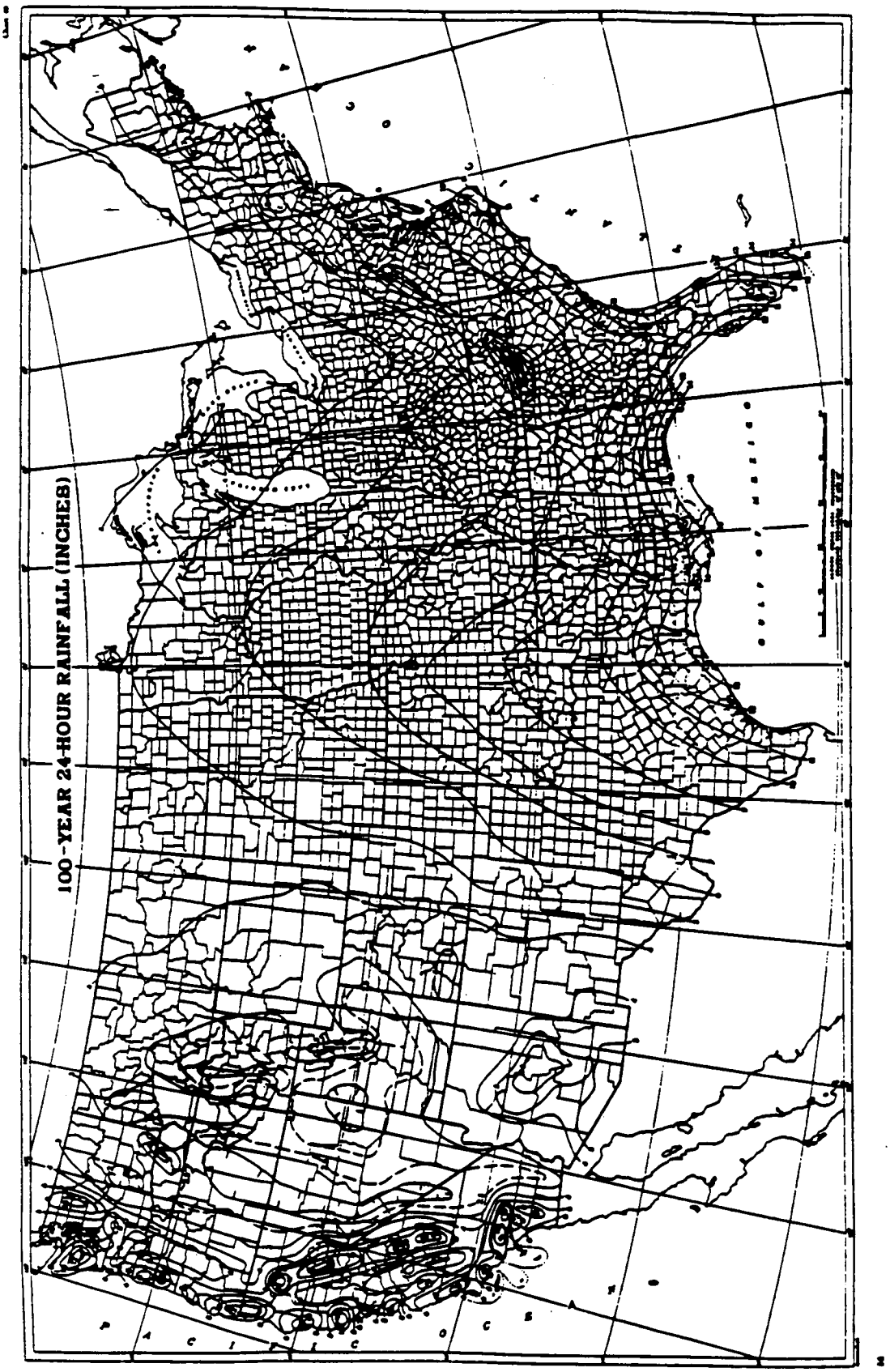


Figure VIII-42. 24-hour rainfall to be expected at a return period of 100 years (Hershfield, 1961).

Table VIII-34. Linearized Rainfall Duration Variate for Equation 8.142 (Weiss, 1962).

Duration in hours	0.17	0.33	0.5	0.67	1
Linearize Variate, $\bar{y}$	-37	-24	-15.6	-9.4	0
Duration in hours	2	3	6	12	24
Linearize Variate, y	17.6	28.8	49.9	73.4	100.0

E.2 Design Flows

This section describes the computational steps employed by DFLOW and DESON 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.

E.2.1 Extreme Value Design Flow

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 $1x10^{20}$).

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:

$$F0 = (NY - N)/NY \qquad (8.143)$$

Let P be the cumulative probability corresponding to the user-supplied return period of R years, adjusted for the presence of zero-flow years:

$$P = (1/R - F0)/(1 - F0) \qquad (8.144)$$

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):

$$Z = 4.91(P^{.14} - (1- P)^{.14}) \qquad (8.145)$$

or using Appendix B.

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):

$$K = (2/g)((1 + G*Z/6 - G^2/36)^3 - 1) \qquad (8.146)$$

STEP 5e. Compute the design flow as

$$\exp(U + K*S) \qquad (8.147)$$

E.2.2 Biologically-Based Design Flow

It is 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 $1x10^{20}$).

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)/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 = 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% of FL, then end with DFLOW = FU. If XL is within 0.005 of A, then end with DFLOW = FL. If XU is within 0.005 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:

$$FT = FL + (FU - FL)*(A - XL)/(XU - XL) \qquad (8.148)$$

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 <= 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 follows:

DAY	4-Day Avg. Flow, cfs	DAY	4-Day Avg. Flow, cfs
1	34	513	
2	65	to	< 100
3	25	545	
4-12	> 100	546	
13	57	to	> 100
14	34	end	
15	26		
16-512	> 100		

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 as follows:

CLUSTER	PERIOD	START DAY	LENGTH, DAYS	# EXCURSIONS In Period	# 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

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.

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.

E.2.3 Human Health Design Flow

The overall harmonic mean daily flow can serve as a design flow for human health water criteria that are based on lifetime exposures. (See Rossman (1990) for justifying the use of the harmonic mean.) Computation of the harmonic mean flow begins by reading daily flow values into a vector X. Then the following steps are followed:

STEP 1. Set NDAYS = 0, NZEROS = 0, DSUM = 0 and t = 1.

STEP 2. If X(t) equals UNKNOWN, then go to Step 5. Otherwise set NDAYS = NDAYS + 1.

STEP 3. If X(t) equals 0, then set NZEROS = NZEROS + 1 and go to Step 5.

STEP 4. Set DSUM = DSUM + 1/X(t).

STEP 5. Set t = t + 1. If the end of the record has not been reached then return to Step 2.

STEP 6. Compute the design flow HMEAN as
HMEAN = (NDAYS - NZEROS) / DSUM * DR
where DR = (NDAYS - NZEROS) / NDAYS.

Note that this procedure takes into account the possibility of days with zero flow. The final estimate of the harmonic mean is a weighted average of the harmonic mean of the non-zero flows and zero. The weight attached to the harmonic mean of the non-zero flows is simply the fraction of the total days of record that have non-zero flows.

E.3 Frequency of Extreme Events

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

$$p = u / n \qquad (8.149)$$

where u is the number of observations exceeding X_c and n is the number of observations. For $n \leq 30$, Appendix U can be used to develop nonparametric 90^{th} or 95^{th} percentile confidence limits. For n greater than 30, Equations 8.150 and 8.151 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.

$$\text{Lower limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{ (u-0.5) + \frac{Z_{1-\alpha/2}^2}{2} - Z_{1-\alpha/2}\left[(u-0.5) - \frac{(u-0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4} \right]^{1/2} \right\} \qquad (8.150)$$

$$\text{Upper limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{ (u+0.5) + \frac{Z_{1-\alpha/2}^2}{2} + Z_{1-\alpha/2}\left[(u+0.5) - \frac{(u+0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4} \right]^{1/2} \right\} \qquad (8.151)$$

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:

$$p \pm Z_{1-\alpha/2}\left[\frac{p(1-p)}{n}\right]^{1/2} \qquad (8.152)$$

The confidence intervals can be used to evaluate one-sample hypotheses such as

Ho: $p = 0.10$
Ha: $p \neq 0.10$

Very simply 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 Appendix V 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 set by 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

Ho: $p_1 = p_2$
Ha: $p_1 \neq p_2$

Moore and McCabe (1989) provide the test statistics as

$$z = \frac{p_1 - p_2}{s_p} \qquad (8.153)$$

where s_p and p are given by

$$s_p = \sqrt{p\,(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad (8.154)$$

$$p = \frac{u_1+u_2}{n_1+n_2} \qquad (8.155)$$

They 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 computed from Equation 8.153 is greater than the associated normal deviate (e.g., 1.96 for a two-sided test with α equal to 0.05), then the null hypothesis is rejected.

F. Multivariate analyses

There are several mulivariate procedures in addition to multivariate regression discussed above. 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 to be described briefly in this guidance include canoncial correlation, cluster analysis, principal components and factor analysis, and discriminant analysis. These procedures were selected for discussion based upon the work of Gaugush (1986, p. 147-188), which should be reviewed in addition to the detailed discussions provided in statistics texts for a better understanding of these multivariate analyses.

F.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, p. 139). 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, p. 139).

Gaugush (1986, p. 152) 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 Institue, Inc., 1985b, p. 140) 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, p. 140).

Gaugush (1986, p. 152) 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 which support hypothesis testing and/or prediction.

Gaugush (1986, p. 148) 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, p. 152). 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, p. 153).

To utilize canonical correlation in hypothesis testing it is important that the assumption of multivariate normality is satisfied (Gaugush, 1986, p. 153). Snedecor and Cochran (1980, p. 361) 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, p. 153) notes that the assumption of multivariate normality is often satisfied by "creating data distributions that are approximately univariate normal."

To satisfy the assumptions of canonical correlation Gaugush (1986, p. 153-154) recommends:

1. use transformations if needed to create roughly symmetric univariate data distributions

2. carefully examine the validity of outliers and run analyses with and without outliers to document their impact on the correlations

3. transform data if necessary to create linear relationships among the variables in each set of variables

Finally, Gaugush (1986, p. 154-161) gives an example application of canonical correlation using the SAS CANCORR procedure described above.

F.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, p. 45). The types of cluster analysis include (SAS Institute, Inc., 1985b, p. 45):

1. disjoint clusters which place each object in one and only one cluster

2. hierarchical clusters in which one cluster may be contained entirely within another cluster, but for which no other kind of overlap is allowed

3. overlapping clusters with or without constraints placed on the number of objects that belong to two clusters

4. fuzzy clusters which are defined by a probability of membership of each object in each cluster (these can be disjoint, hierarchical or overlapping)

SAS offers several clustering options under the CLUSTER procedure (SAS Institute, Inc., 1985b, p. 255-315).

Both Gaugush (1986, p. 163) and the SAS Institute (1985b, p. 48) acknowledge that Ward's minimum variance method and the average linkage method are two of the best approaches to hierarchical clustering. However, SAS also notes that evaluations of clustering methods have resulted in inconsistency and confusion (SAS Institute, Inc., 1985b, p. 48). As a result of this variability in assessments of clustering methods, it is recommended that the researcher become familiar with a particular method (Gaugush, 1986, p. 164) and/or use several methods on the same data base. By using several methods on the same data base it should be possible to interpret the result based upon agreement among the different methods. SAS recommends inclusion of at least one less biased method (e.g., density linkage) when trying several methods. Less biased methods are based on nonparametric density estimation (e.g.,

single linkage and density linkage), whereas more biased methods look for clusters based on size (e.g., k-means and Ward's minimum variance method), shape, or dispersion (e.g, average linkage) (SAS Institute, Inc., 1985b, p. 48).

Gaugush (1986, p. 165) recommends against performing formal statistical tests with the results of cluster analysis. With no such tests, no formal assumptions are imposed. However, Gaugush (1986, p. 165) advises that the data should be symmetrically distributed with no outliers. This could require data transformation prior to cluster analysis. As a further point, Gaugush (1986, p. 164-165) suggests that the data should be standardized (i.e., value replaced by its deviation from the mean divided by the standard deviation) if the variances of different variables are of different magnitudes. Otherwise, the raw (or tranformed) data are to be used.

Gaugush (1986, p. 165-169) illustrates the use of Ward's method of cluster analysis by grouping reservoirs on the basis of 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 on the basis of similar variances. Variables utilized in these cluster analyses were 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."

F.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, p. 621). 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, p. 169). The primary use of these procedures is exploratory analysis, i.e., hypothesis testing is not normally performed (SAS Institute, Inc., 1985b, p. 621).

Gaugush (1986, p. 172) notes that PCA is usually performed first. As an example, the author describes how PCA can be used to develop a trophic state index from biological, nutrient, and physical data. He goes on to illustrate how factor analysis can be used to enhance the scientific interpretation of the principal components thus developed.

In lay terms, PCA creates linear relationships (or factors) that account for a maximum of the variance contained in the original data set (Gaugush, 1986, p. 170). 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, p. 171) recommends that data distributions be approximately symmetric with no outliers. As in other cases, data transformations may be needed to meet these recommendations.

Due to problems of scale, Gaugush (1986, p. 171-172) 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, then 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, p. 169-178) goes several steps further in describing these procedures, including an illustrative example. SAS gives a fairly detailed mathematical description of PCA and factor analysis (SAS Institute, Inc., 1985b, p. 621-622 and p. 335-338, respectively), and offers procedures for performing both (PRINCOMP and FACTOR procedures).

F.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, p. 178). Some of the uses for discriminant analysis are (SAS Institute, Inc., 1985b, p. 41):

1. to find a mathematical rule (or "discriminant function") for predicting which class an observation belongs to given data for the independent quantitative variables

2. to find linear combinations of the independent quantitative variables that best reveal the differences among the classes

3. to find a subset of the independent quantitative variables that best shows the differences among the classes.

It should be understood that discriminant analysis requires prior knowledge of all classes (e.g., a sample), whereas cluster analysis has no such requirement (SAS Institute, Inc., 1985b, p. 42). In fact, cluster analysis is used to <u>define</u> the classes.

Gaugush (1986, p. 180) 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) which can be used for nonnormal data (SAS Institute, Inc., 1985b, p. 559-567).

Researchers are encouraged to follow the descriptions of discriminant analysis offered by SAS (SAS Institute, Inc., 1985b, p. 41-44) and Gaugush (1986, p. 178-181) before using the procedure. Gaugush (1986, p. 181-188) provides an example application of discriminant analysis in which pH and aluminum concentration are used to predict the presence or absence of brook trout.

G. Special Treatment of Environmental Data

Environmental data provide an interesting challenge to statisticians because of the natural variability and inherent difficulty in collecting environmental data that meet the typical assumptions. As a result, significant research over the last 20 years has been placed on developing new or modifying existing statistical procedures to handle these difficulties.

The previous sections of this document largely focused on the normality assumption in order to select a parametric or a nonparametric test. This section describes methods for addressing missing values, outliers, and censored data.

G.1 Missing data

Snedecor and Cochran (1980, pp. 274-279) present a few standard approaches to dealing with missing data. These standard approaches work from the assumption that the missing data fit the same mathematical model as the data one has. Often, this assumption might not be satisified in NPS monitoring efforts. For example, samplers frequently malfunction under icy conditions in northern climates. The data missing from these sampling periods might not at all resemble the data collected during other periods.

Snedecor and Cochran (1980, p. 274) point out that computerized versions of general analysis of variance, analysis of variance and covariance, and multiple linear regression usually handle missing data correctly. They caution, however, that one should check that such programs provide standard errors or instructions for computing them when comparisons on or among treatment means are made.

Snedecor and Cochran (1980, pp. 275-276) present equations for estimating a single missing observation in both a one-way and two-way classification, and for determining the standard error of the mean of the treatment with the missing value in comparisons of treatment means. They also provide a way to correct for bias in the calculation of the treatment's mean square in ANOVA.

Two iterative processes are described for estimating more than one missing value. Both procedures use initial guesses of the missing values. A third, noniterative method is also discussed by Snedecor and Cochran (1980, pp. 277-279).

Gilbert (1987) provides and illustrates methods for handling missing data in estimating the mean and variance (p. 177-185), in detecting and estimating trends (pp. 208-237), and in nonparametric tests (p. 241-244).

G.2 Outliers

The detection of outliers is often the most interesting component of the data analysis to the environmental manager. In environmental analyses, it is often very difficult to distinguish between erroneous observations (i.e., bad data) and extreme events. Although several authors have suggested methods for detecting outliers, the time to detect erroneous data is as soon after data collection as possible. Correcting old data is time consuming and the usual solution is to throw the questionable data out of the analysis.

Snedecor and Cochran (1980, p. 279) state that outliers should be examined initially to determine the cause of the extreme value. In some cases, the outlier can be traced to a mistake, and the correct value restored. In other cases, the extreme value is due to a mistake for which a correction cannot be made. For these situations the erroneous value should be dropped, and analyses should be performed with allowance made for the missing value. In still other cases the outlier cannot be traced to any known errors and must be presumed to be a "real" data point.

Snedecor and Cochran (1980, pp. 279-282) discuss methods for assessing single extreme values for samples from a normal distribution, two extreme values from a normal distribution, and suspected outliers in ANOVA in one-way and two-way classifications. Gilbert (1987, pp. 186-203) presents several methods for detecting and treating outliers.

G.3 Censored Data

Observations reported as less-than or non-detect are often quite troublesome for many statistical procedures (as well as databases). Quite simply, it is difficult to compute the mean (or any of a number of statistics) when one or more of the values is <0.5. Comparisons between censored data is also difficult. Is a <0.5 more or less than <0.9? Section VIII.C provides several options for estimating summary statistics when some of the data are censored. Helsel and Cohn (1988) provide definitive approaches for estimating summary statistics for when there are multiple censoring levels in the same data set although these methods are typically computationally intensive.

CHAPTER 9

QUALITY ASSURANCE

IX. Quality Assurance and Quality Control

A. 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 should be thought of as 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 the U.S. Environmental Protection Agency's (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.

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, 1994b).

Quality Control:

> A system of technical procedures and activities developed and implemented to produce measurements of requisite quality (Taylor, 1993; USEPA, 1994b).

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 IX-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.

2. Importance of QA/QC Programs

While 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

Table IX-1. Common QA and QC activities (adapted from Drouse et al., 1986, and Erickson et al., 1991).

QA Activities	QC 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	• 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

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, then 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 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 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 that are 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).

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 a 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 a 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.

2. 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 a NPS monitoring project to minimize or eliminate these problems (Erikson et al., 1991; USEPA, 1994a; USGS, 1992).

EPA's mandatory agency-wide Quality System policy requires each office or laboratory generating data to implement minimum procedures to assure that precision, accuracy, completeness, comparability, and representativeness of data are known and documented (USEPA, 1984; Erickson et al., 1991). 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, 1994a). In addition, each non-EPA organization must have a well-documented Quality Assurance Project Plan (QAPP) covering each monitoring or measurement activity associated with a project (USEPA, 1983, 1994b; Erickson et al., 1991).

The purpose of writing a QAPP prior to undertaking a 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 expected 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 a 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 these 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, 1994b).

B. Data Quality Objectives (DQOs)

Prior to collecting environmental data in support of a 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, 1994c).

1. The Data Quality Objectives Process

The Data Quality Objectives (DQO) Process 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, 1994c).

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, 1994c), 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 John Warren, Quality Assurance Management Staff, 202-260-9464).

Step 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.

Step 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," then 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.

Step 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.

Step 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.

Step 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.

Step 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 isn't [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.

Step 7: Optimize the design. Evaluate information from the previous steps and generate alternative data collection designs. The design 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.

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, 1994c). 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, 1994c). 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, 1994c).

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, 1994b, 1994c). 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, 1994c).

C. 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, 1994b). EPA requires that four types of elements be discussed in a Quality Assurance Project Plan (QAPP). These elements are listed in Table IX-2 and discussed briefly below. (For complete descriptions and requirements, be sure to see USEPA (1994b). 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 (1994b), "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.

Table IX-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

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 required, see Cross-Smiecinski and Stetzenback (1994) and USEPA (1994b).

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 IX-1). Using the diagram, explain the rationale for the organization (e.g., to maximize 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

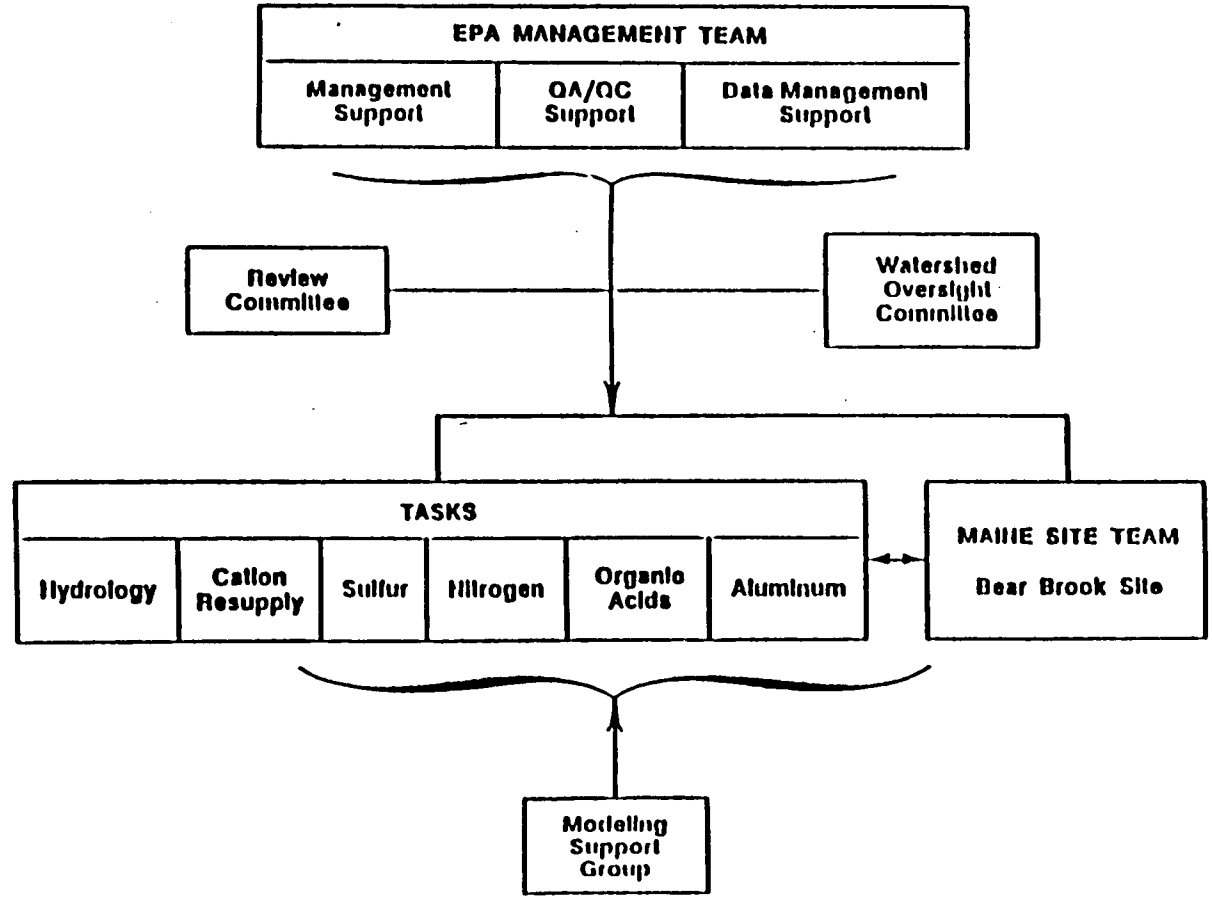


Figure IX-1. Sample organization chart for a quality assurance project plan (from Erickson et al., 1991).

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 IX-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).

$$CV = \left(\frac{s}{\bar{x}}\right) 100$$

s = sample standard deviation;
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):

$$RPD = \frac{2(x_1 - x_2)}{x_1 + x_2} (100)$$

x_1 = analyte concentration of first duplicate;
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).

Parameter	Units	Expected Range	Accuracy	Precision	Completeness
Particulate NO_3/SO_4					
SSI[1]	$\mu g/m^3$	10-1000	10%	20%	70%
47 mm TF/PC[2]	$\mu g/m^3$	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%

[1] PM_{10} Size Selective Inlet High Volume Sampler
[2] Teflon/Polycarbonate Filter

Figure IX-2. Sample quality assurance objectives (from 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):

$$\%R = \frac{A-B}{C}(100)$$

A = spiked sample result;
B = sample result;
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:

$$\%C = 100\left[\frac{V}{n}\right]$$

%C = percent complete;
V = number of measurements judged valid;
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 (1994b) 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.

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 (1994b) 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 IX-3). Copies of the form must be completed in the field, and signed by the field team when they transfer custody of the samples to the shipper. Upon receipt in the

CHAIN-OF-CUSTODY RECORD

PROJ NO	PROJECT NAME		NO OF CONTAINERS	TAG NUMBERS	PARAMETERS						REMARKS
SAMPLERS (Signature)										SAMPLE TAG VERIFICATION	
DATE	TIME	DESCRIPTION									

Relinquished by: (Signature)	Date/Time	Received by: (Signature)	Relinquished by (Signature)	Date/Time	Received by: (Signature)
Print Name:		Print Name	Print Name:		Print Name
Relinquished by: (Signature)	Date/Time	Received for Laboratory by: (Signature)	Date/Time	Remarks	
Print Name:		Print Name			

Figure IX-3. Sample custody chart (from Cross-Smiecinski and Stetzenback, 1994).

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 retainsthe 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.

In addition, a custody seal should be applied to the lid of each sample container in such a fashion that the container

cannot be opened without breaking the seal. Further, an additional custody seal should be placed across the lid of the shipping container in such a way that the container cannot be opened without breaking the seal (see Cross-Smiecinski and Stetzenback, 1994).

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) then 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 (USGS, 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, 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 (USGS, 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 data bases, 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).

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,

echnical 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.

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 may 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 data base 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 IX-14, 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.

D. Field Operations Program

Field operations are an important activity in a NPS monitoring program. Field operations involve organizing and designing 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 IX-3 summarizes many important items that should be considered in a field operations portion of a QA/QC program.

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

Table IX-3. Checklist of items that should be considered in the field operations section of a QA/QC program (from USEPA, 1990).

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		
Health and Safety	Field personnel properly trained	_____	
	Proper field gear and clothing issued to field personnel .	_____	

Table IX-3. (continued)

Field Operations			
Element	Specifics	Check-off	Responsibility
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	____	

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.

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 prior to the commencement of sampling. 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 permission being revoked at some point during the project.

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.

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 before 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. 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.

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).

E. 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.

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 IX-4). More detailed references on laboratory QA and QC should be consulted for further information.

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 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).

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 (USGS, 1992).

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 (USGS, 1992).

Table IX-4. Checklist of items that should be considered in the laboratory operations section of a QA/QC program (from USEPA, 1990).

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	_____ _____ _____	

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 beallowed 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 (USGS, 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. Testing of safety equipment should be conducted on a regular basis (USGS, 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 (USGS, 1992).

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.

F. Data and Reports

It is essential during the conduct of a 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.

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.

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 a 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.

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 problems that arose related to sampling 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.

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.

APPENDICES

APPENDICES

Appendix A.	Chain of Custody Procedures	A-1
Appendix B.	Normal Distribution	B-1
Appendix C.	t-Distribution	C-1
Appendix D.	Chi-square Distribution	D-1
Appendix E.	F Distribution	E-1
Appendix F.	Skewness Test	F-1
Appendix G.	Kurtosis Test	G-1
Appendix H.	Coefficients for W Test for Normality	H-1
Appendix I.	W Test Percentage Points	I-1
Appendix J.	Noncentral t-Distribution	J-1
Appendix K.	Population Correlation Coefficient	K-1
Appendix L.	Transformation of Linearization	L-1
Appendix M.	Multiple Regression Example - Hand Calculated	M-1
Appendix N.	Multiple Regression Example - Computerized	N-1
Appendix O.	Example Use of R-Square	O-1
Appendix P.	Confidence Limits for the Median of Any Continuous Distribution	P-1
Appendix Q.	Quantiles of the Spearman Test Statistic	Q-1
Appendix R.	Probabilities for the Mann-Kendall Nonparametric Test for Trend	R-1
Appendix S.	Quantiles of the Wilcoxon Signed-Rank Test Statistic	T-1
Appendix T.	Nonparametric 95 and 90 percentile confidence intervals on a proportions	U-1
Appendix U.	Sample sizes for one-sided non-parametric tolerance limits	V-1

APPENDIX A

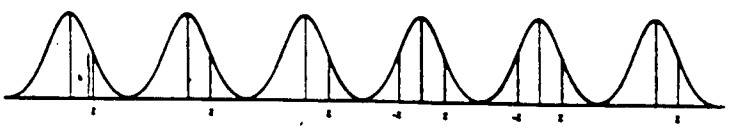
Chain of Custody Procedures

Water quality data to be used for legal purposes must be collected and processed through approved procedures, which are called "chain of custody" procedures. These procedures are described in existing documents (U.S.E.P.A., 1978b, 1982b; and Scalf, et al., 1981), and will, therefore, not be addressed in this M&E guidance.

APPENDIX B

The Normal Distribution
(SOURCE: Remington and Schork, 1970)

THE NORMAL DISTRIBUTION*



z	h	A†	B	C	D	E
0.00	0.3989	0.0000	0.5000	0.0000	1.0000	0.5000
0.01	0.3989	0.0040	0.4960	0.0080	0.9920	0.5040
0.02	0.3989	0.0080	0.4920	0.0160	0.9840	0.5080
0.0251	0.3988	0.01	0.49	0.02	0.98	0.51
0.03	0.3988	0.0120	0.4880	0.0239	0.9761	0.5120
0.04	0.3986	0.0160	0.4840	0.0319	0.9681	0.5160
0.05	0.3984	0.0199	0.4801	0.0399	0.9601	0.5199
0.0502	0.3984	0.02	0.48	0.04	0.96	0.52
0.06	0.3982	0.0239	0.4761	0.0478	0.9522	0.5239
0.07	0.3980	0.0279	0.4721	0.0558	0.9442	0.5279
0.0753	0.3978	0.03	0.47	0.06	0.94	0.53
0.08	0.3977	0.0319	0.4681	0.0638	0.9362	0.5319
0.09	0.3973	0.0359	0.4641	0.0717	0.9283	0.5359
0.10	0.3970	0.0398	0.4602	0.0797	0.9203	0.5398
0.1004	0.3969	0.04	0.46	0.08	0.92	0.54
0.11	0.3965	0.0438	0.4562	0.0876	0.9124	0.5438
0.12	0.3961	0.0478	0.4522	0.0955	0.9045	0.5478
0.1257	0.3958	0.05	0.45	0.10	0.9	0.55
0.13	0.3956	0.0517	0.4483	0.1034	0.8966	0.5517
0.14	0.3951	0.0557	0.4443	0.1113	0.8887	0.5557
0.15	0.3945	0.0596	0.4404	0.1192	0.8808	0.5596
0.1510	0.3944	0.06	0.44	0.12	0.88	0.56
0.16	0.3939	0.0636	0.4364	0.1271	0.8729	0.5636
0.17	0.3932	0.0675	0.4325	0.1350	0.8650	0.5675
0.1764	0.3928	0.07	0.43	0.14	0.86	0.57
0.18	0.3925	0.0714	0.4286	0.1429	0.8571	0.5714
0.19	0.3918	0.0753	0.4247	0.1507	0.8493	0.5753
0.20	0.3910	0.0793	0.4207	0.1585	0.8415	0.5793
0.2019	0.3909	0.08	0.42	0.16	0.84	0.58
0.21	0.3902	0.0832	0.4168	0.1663	0.8337	0.5832
0.22	0.3894	0.0871	0.4129	0.1741	0.8259	0.5871
0.2275	0.3888	0.09	0.41	0.18	0.82	0.59
0.23	0.3885	0.0910	0.4090	0.1819	0.8181	0.5910
0.24	0.3876	0.0948	0.4052	0.1897	0.8103	0.5948
0.25	0.3867	0.0987	0.4013	0.1974	0.8026	0.5987
0.2533	0.3863	0.10	0.40	0.20	0.80	0.60
0.26	0.3857	0.1026	0.3974	0.2051	0.7949	0.6026
0.27	0.3847	0.1064	0.3936	0.2128	0.7872	0.6064
0.2793	0.3837	0.11	0.39	0.22	0.78	0.61
0.28	0.3836	0.1103	0.3897	0.2205	0.7795	0.6103
0.29	0.3825	0.1141	0.3859	0.2282	0.7718	0.6141
0.30	0.3814	0.1179	0.3821	0.2358	0.7642	0.6179
0.3055	0.3808	0.12	0.38	0.24	0.76	0.62
0.31	0.3802	0.1217	0.3783	0.2434	0.7566	0.6217
0.32	0.3790	0.1255	0.3745	0.2510	0.7490	0.6255

* Abstracted from *National Bureau of Standards—Applied Mathematics Series*—23, U.S. Govt. Printing Office, Washington, D.C., 1953.

† Letters *A–E* refer to the shaded areas under the corresponding normal curves.

(*cont.*)

THE NORMAL DISTRIBUTION

z	h	A	B	C	D	E
0.33	0.3778	0.1293	0.3707	0.2586	0.7414	0.6293
0.3319	0.3776	0.13	0.37	0.26	0.74	0.63
0.34	0.3765	0.1331	0.3669	0.2661	0.7339	0.6331
0.35	0.3752	0.1368	0.3632	0.2737	0.7263	0.6368
0.3585	0.3741	0.14	0.36	0.28	0.72	0.64
0.36	0.3739	0.1406	0.3594	0.2812	0.7188	0.6406
0.37	0.3725	0.1443	0.3557	0.2886	0.7114	0.6443
0.38	0.3712	0.1480	0.3520	0.2961	0.7039	0.6480
0.3853	0.3704	0.15	0.35	0.30	0.70	0.65
0.39	0.3697	0.1517	0.3483	0.3035	0.6965	0.6517
0.40	0.3683	0.1554	0.3446	0.3108	0.6892	0.6554
0.41	0.3668	0.1591	0.3409	0.3182	0.6818	0.6591
0.4125	0.3664	0.16	0.34	0.32	0.68	0.66
0.42	0.3653	0.1628	0.3372	0.3255	0.6745	0.6628
0.43	0.3637	0.1664	0.3336	0.3328	0.6672	0.6664
0.4399	0.3622	0.17	0.33	0.34	0.66	0.67
0.44	0.3621	0.1700	0.3300	0.3401	0.6599	0.6700
0.45	0.3605	0.1736	0.3264	0.3473	0.6527	0.6736
0.46	0.3589	0.1772	0.3228	0.3545	0.6455	0.6772
0.4677	0.3576	0.18	0.32	0.36	0.64	0.68
0.47	0.3572	0.1808	0.3192	0.3616	0.6384	0.6808
0.48	0.3555	0.1844	0.3156	0.3688	0.6312	0.6844
0.49	0.3538	0.1879	0.3121	0.3759	0.6241	0.6879
0.4959	0.3528	0.19	0.31	0.38	0.62	0.69
0.50	0.3521	0.1915	0.3085	0.3829	0.6171	0.6915
0.51	0.3503	0.1950	0.3050	0.3899	0.6101	0.6950
0.52	0.3485	0.1985	0.3015	0.3969	0.6031	0.6985
0.5244	0.3477	0.2	0.3	0.40	0.6	0.70
0.53	0.3467	0.2019	0.2981	0.4039	0.5961	0.7019
0.54	0.3448	0.2054	0.2946	0.4108	0.5892	0.7054
0.55	0.3429	0.2088	0.2912	0.4177	0.5823	0.7088
0.5534	0.3423	0.21	0.29	0.42	0.58	0.71
0.56	0.3410	0.2123	0.2877	0.4245	0.5755	0.7123
0.57	0.3391	0.2157	0.2843	0.4313	0.5687	0.7157
0.58	0.3372	0.2190	0.2810	0.4381	0.5619	0.7190
0.5828	0.3366	0.22	0.28	0.44	0.56	0.72
0.59	0.3352	0.224	0.2776	0.4448	0.5552	0.7224
0.60	0.3332	0.2257	0.2743	0.4515	0.5485	0.7257
0.61	0.3312	0.2291	0.2709	0.4581	0.5419	0.7291
0.6128	0.3306	0.23	0.27	0.46	0.54	0.73
0.62	0.3292	0.2324	0.2676	0.4647	0.5353	0.7324
0.63	0.3271	0.2357	0.2643	0.4713	0.5287	0.7357
0.64	0.3251	0.2389	0.2611	0.4778	0.5222	0.7389
0.6433	0.3244	0.24	0.26	0.48	0.52	0.74
0.65	0.3230	0.2422	0.2578	0.4843	0.5157	0.7422
0.66	0.3209	0.2454	0.2546	0.4907	0.5093	0.7454
0.67	0.3187	0.2486	0.2514	0.4971	0.5029	0.7486
0.6745	0.3178	0.25	0.25	0.50	0.50	0.75
0.68	0.3166	0.2517	0.2483	0.5035	0.4965	0.7517
0.69	0.3144	0.2549	0.2451	0.5098	0.4902	0.7549
0.70	0.3123	0.2580	0.2420	0.5161	0.4839	0.7580
0.7063	0.3109	0.26	0.24	0.52	0.48	0.76
0.71	0.3101	0.2611	0.2389	0.5223	0.4777	0.7611
0.72	0.3079	0.2642	0.2358	0.5285	0.4715	0.7642
0.73	0.3056	0.2673	0.2327	0.5346	0.4654	0.7673
0.7388	0.3037	0.27	0.23	0.54	0.46	0.77
0.74	0.3034	0.2704	0.2296	0.5407	0.4593	0.7704
0.75	0.3011	0.2734	0.2266	0.5467	0.4533	0.7734
0.76	0.2989	0.2764	0.2236	0.5527	0.4473	0.7764
0.77	0.2966	0.2794	0.2206	0.5587	0.4413	0.7794

(cont.)

THE NORMAL DISTRIBUTION

z	h	A	B	C	D	E
0.7722	0.2961	0.28	0.22	0.56	0.44	0.78
0.78	0.2943	0.2823	0.2177	0.5646	0.4354	0.7823
0.79	0.2920	0.2852	0.2148	0.5705	0.4295	0.7852
0.80	0.2897	0.2881	0.2119	0.5763	0.4237	0.7881
0.8064	0.2882	0.29	0.21	0.58	0.42	0.79
0.81	0.2874	0.2910	0.2090	0.5821	0.4179	0.7910
0.82	0.2850	0.2939	0.2061	0.5878	0.4122	0.7939
0.83	0.2827	0.2967	0.2033	0.5935	0.4065	0.7967
0.84	0.2803	0.2995	0.2005	0.5991	0.4009	0.7995
0.8416	0.2800	0.30	0.20	0.60	0.40	0.80
0.85	0.2780	0.3023	0.1977	0.6047	0.3953	0.8023
0.86	0.2756	0.3051	0.1949	0.6102	0.3898	0.8051
0.87	0.2732	0.3078	0.1922	0.6157	0.3843	0.8078
0.8779	0.2714	0.31	0.19	0.62	0.38	0.81
0.88	0.2709	0.3106	0.1894	0.6211	0.3789	0.8106
0.89	0.2685	0.3133	0.1867	0.6265	0.3735	0.8133
0.90	0.2661	0.3159	0.1841	0.6319	0.3681	0.8159
0.91	0.2637	0.3186	0.1814	0.6372	0.3628	0.8186
0.9154	0.2624	0.32	0.18	0.64	0.36	0.82
0.92	0.2613	0.3212	0.1788	0.6424	0.3576	0.8212
0.93	0.2589	0.3238	0.1762	0.6476	0.3524	0.8238
0.94	0.2565	0.3264	0.1736	0.6528	0.3472	0.8264
0.95	0.2541	0.3289	0.1711	0.6579	0.3421	0.8289
0.9542	0.2531	0.33	0.17	0.66	0.34	0.83
0.96	0.2516	0.3315	0.1685	0.6629	0.3371	0.8315
0.97	0.2492	0.3340	0.1660	0.6680	0.3320	0.8340
0.98	0.2468	0.3365	0.1635	0.6729	0.3271	0.8365
0.99	0.2444	0.3389	0.1611	0.6778	0.3222	0.8389
0.9945	0.2433	0.34	0.16	0.68	0.32	0.84
1.00	0.2420	0.3413	0.1587	0.6827	0.3173	0.8413
1.01	0.2396	0.3438	0.1562	0.6875	0.3125	0.8438
1.02	0.2371	0.3461	0.1539	0.6923	0.3077	0.8461
1.03	0.2347	0.3485	0.1515	0.6970	0.3030	0.8485
1.036	0.2332	0.35	0.15	0.70	0.3	0.85
1.04	0.2323	0.3508	0.1492	0.7017	0.2983	0.8508
1.05	0.2299	0.3531	0.1469	0.7063	0.2937	0.8531
1.06	0.2275	0.3554	0.1446	0.7109	0.2891	0.8554
1.07	0.2251	0.3577	0.1423	0.7154	0.2846	0.8577
1.08	0.2227	0.3599	0.1401	0.7199	0.2801	0.8599
1.080	0.2226	0.36	0.14	0.72	0.28	0.86
1.09	0.2203	0.3621	0.1379	0.7243	0.2757	0.8621
1.10	0.2179	0.3643	0.1357	0.7287	0.2713	0.8643
1.11	0.2155	0.3665	0.1335	0.7330	0.2670	0.8665
1.12	0.2131	0.3686	0.1314	0.7373	0.2627	0.8686
1.1264	0.2115	0.37	0.13	0.74	0.26	0.87
1.13	0.2107	0.3708	0.1292	0.7415	0.2585	0.8708
1.14	0.2083	0.3729	0.1271	0.7457	0.2543	0.8729
1.15	0.2059	0.3749	0.1251	0.7499	0.2501	0.8749
1.16	0.2036	0.3770	0.1230	0.7540	0.2460	0.8770
1.17	0.2012	0.3790	0.1210	0.7580	0.2420	0.8790
1.175	0.2000	0.38	0.12	0.76	0.24	0.88
1.18	0.1989	0.3810	0.1190	0.7620	0.2380	0.8810
1.19	0.1965	0.3830	0.1170	0.7660	0.2340	0.8830
1.20	0.1942	0.3849	0.1151	0.7699	0.2301	0.8849
1.21	0.1919	0.3869	0.1131	0.7737	0.2263	0.8869
1.22	0.1895	0.3888	0.1112	0.7775	0.2225	0.8888
1.227	0.1880	0.39	0.11	0.78	0.22	0.89
1.23	0.1872	0.3907	0.1093	0.7813	0.2187	0.8907
1.24	0.1849	0.3925	0.1075	0.7850	0.2150	0.8925
1.25	0.1826	0.3944	0.1056	0.7887	0.2113	0.8944

(*cont.*)

THE NORMAL DISTRIBUTION

z	*h*	*A*	*B*	*C*	*D*	*E*
1.26	0.1804	0.3962	0.1038	0.7923	0.2077	0.8962
1.27	0.1781	0.3980	0.1020	0.7959	0.2041	0.8980
1.28	0.1758	0.3997	0.1003	0.7995	0.2005	0.8997
1.282	0.1755	0.40	0.10	0.80	0.20	0.90
1.29	0.1736	0.4015	0.0985	0.8029	0.1971	0.9015
1.30	0.1714	0.4032	0.0968	0.8064	0.1936	0.9032
1.31	0.1691	0.4049	0.0951	0.8098	0.1902	0.9049
1.32	0.1669	0.4066	0.0934	0.8132	0.1868	0.9066
1.33	0.1647	0.4082	0.0918	0.8165	0.1835	0.9082
1.34	0.1626	0.4099	0.0901	0.8198	0.1802	0.9099
1.341	0.1624	0.41	0.09	0.82	0.18	0.91
1.35	0.1604	0.4115	0.0885	0.8230	0.1770	0.9115
1.36	0.1582	0.4131	0.0869	0.8262	0.1738	0.9131
1.37	0.1561	0.4147	0.0853	0.8293	0.1707	0.9147
1.38	0.1539	0.4162	0.0838	0.8324	0.1676	0.9162
1.39	0.1518	0.4177	0.0823	0.8355	0.1645	0.9177
1.40	0.1497	0.4192	0.0808	0.8385	0.1615	0.9192
1.405	0.1487	0.42	0.08	0.84	0.16	0.92
1.41	0.1476	0.4207	0.0793	0.8415	0.1585	0.9207
1.42	0.1456	0.4222	0.0778	0.8444	0.1556	0.9222
1.43	0.1435	0.4236	0.0764	0.8473	0.1527	0.9236
1.44	0.1415	0.4251	0.0749	0.8501	0.1499	0.9251
1.45	0.1394	0.4265	0.0735	0.8529	0.1471	0.9265
1.46	0.1374	0.4279	0.0721	0.8557	0.1443	0.9279
1.47	0.1354	0.4292	0.0708	0.8584	0.1416	0.9292
1.476	0.1343	0.43	0.07	0.86	0.14	0.93
1.48	0.1334	0.4306	0.0694	0.8611	0.1389	0.9306
1.49	0.1315	0.4319	0.0681	0.8638	0.1362	0.9319
1.50	0.1295	0.4332	0.0668	0.8664	0.1336	0.9332
1.51	0.1276	0.4345	0.0655	0.8690	0.1310	0.9345
1.52	0.1257	0.4357	0.0643	0.8715	0.1285	0.9357
1.53	0.1238	0.4370	0.0630	0.8740	0.1260	0.9370
1.54	0.1219	0.4382	0.0618	0.8764	0.1236	0.9382
1.55	0.1200	0.4394	0.0606	0.8789	0.1211	0.9394
1.555	0.1191	0.44	0.06	0.88	0.12	0.94
1.56	0.1182	0.4406	0.0594	0.8812	0.1188	0.9406
1.57	0.1163	0.4418	0.0582	0.8836	0.1164	0.9418
1.58	0.1145	0.4429	0.0571	0.8859	0.1141	0.9429
1.59	0.1127	0.4441	0.0559	0.8882	0.1118	0.9441
1.60	0.1109	0.4452	0.0548	0.8904	0.1096	0.9452
1.61	0.1092	0.4463	0.0537	0.8926	0.1074	0.9463
1.62	0.1074	0.4474	0.0526	0.8948	0.1052	0.9474
1.63	0.1057	0.4484	0.0516	0.8969	0.1031	0.9484
1.64	0.1040	0.4495	0.0505	0.8990	0.1010	0.9495
1.645	0.1031	0.45	0.05	0.90	0.10	0.95
1.65	0.1023	0.4505	0.0495	0.9011	0.0989	0.9505
1.66	0.1006	0.4515	0.0485	0.9031	0.0969	0.9515
1.67	0.0989	0.4525	0.0475	0.9051	0.0949	0.9525
1.68	0.0973	0.4535	0.0465	0.9070	0.0930	0.9535
1.69	0.0957	0.4545	0.0455	0.9090	0.0910	0.9545
1.70	0.0940	0.4554	0.0446	0.9109	0.0891	0.9554
1.71	0.0925	0.4564	0.0436	0.9127	0.0873	0.9564
1.72	0.0909	0.4573	0.0427	0.9146	0.0854	0.9573
1.73	0.0893	0.4582	0.0418	0.9164	0.0836	0.9582
1.74	0.0878	0.4591	0.0409	0.9181	0.0819	0.9591
1.75	0.0863	0.4599	0.0401	0.9199	0.0801	0.9599
1.751	0.0862	0.46	0.04	0.92	0.08	0.96
1.76	0.0848	0.4608	0.0392	0.9216	0.0784	0.9608
1.77	0.0833	0.4616	0.0384	0.9233	0.0767	0.9616
1.78	0.0818	0.4625	0.0375	0.9249	0.0751	0.9625

(cont.)

THE NORMAL DISTRIBUTION

z	h	A	B	C	D	E
1.79	0.0804	0.4633	0.0367	0.9266	0.0734	0.9633
1.80	0.0790	0.4641	0.0359	0.9281	0.0719	0.9641
1.81	0.0775	0.4649	0.0352	0.9297	0.0703	0.9649
1.82	0.0761	0.4656	0.0344	0.9312	0.0688	0.9656
1.83	0.0748	0.4664	0.0336	0.9328	0.0672	0.9664
1.84	0.0734	0.4671	0.0329	0.9342	0.0658	0.9671
1.85	0.0721	0.4678	0.0322	0.9357	0.0643	0.9678
1.86	0.0707	0.4686	0.0314	0.9371	0.0629	0.9686
1.87	0.0694	0.4693	0.0307	0.9385	0.0615	0.9693
1.88	0.0681	0.4699	0.0301	0.9399	0.0601	0.9699
1.881	0.0680	0.47	0.03	0.94	0.06	0.97
1.89	0.0669	0.4706	0.0294	0.9412	0.0588	0.9706
1.90	0.0656	0.4713	0.0287	0.9426	0.0574	0.9713
1.91	0.0644	0.4719	0.0281	0.9439	0.0561	0.9719
1.92	0.0632	0.4726	0.0274	0.9451	0.0549	0.9726
1.93	0.0620	0.4732	0.0268	0.9464	0.0536	0.9732
1.94	0.0608	0.4738	0.0262	0.9476	0.0524	0.9738
1.95	0.0596	0.4744	0.0256	0.9488	0.0512	0.9744
1.960	0.0585	0.475	0.025	0.95	0.05	0.975
1.97	0.0573	0.4756	0.0244	0.9512	0.0488	0.9756
1.98	0.0562	0.4761	0.0239	0.9523	0.0477	0.9761
1.99	0.0551	0.4767	0.0233	0.9534	0.0466	0.9767
2.00	0.0540	0.4772	0.0228	0.9545	0.0455	0.9772
2.01	0.0529	0.4778	0.0222	0.9556	0.0444	0.9778
2.02	0.0519	0.4783	0.0217	0.9566	0.0434	0.9783
2.03	0.0508	0.4788	0.0212	0.9576	0.0424	0.9788
2.04	0.0498	0.4793	0.0207	0.9586	0.0414	0.9793
2.05	0.0488	0.4798	0.0202	0.9596	0.0404	0.9798
2.054	0.0484	0.48	0.02	0.96	0.04	0.98
2.06	0.0478	0.4803	0.0197	0.9606	0.0394	0.9803
2.07	0.0468	0.4808	0.0192	0.9615	0.0385	0.9808
2.08	0.0459	0.4812	0.0188	0.9625	0.0375	0.9812
2.09	0.0449	0.4817	0.0183	0.9634	0.0366	0.9817
2.10	0.0440	0.4821	0.0179	0.9643	0.0357	0.9821
2.11	0.0431	0.4826	0.0174	0.9651	0.0349	0.9826
2.12	0.0422	0.4830	0.0170	0.9660	0.0340	0.9830
2.13	0.0413	0.4834	0.0166	0.9668	0.0332	0.9834
2.14	0.0404	0.4838	0.0162	0.9676	0.0324	0.9838
2.15	0.0396	0.4842	0.0158	0.9684	0.0316	0.9842
2.16	0.0387	0.4846	0.0154	0.9692	0.0308	0.9846
2.17	0.0379	0.4850	0.0150	0.9700	0.0300	0.9850
2.18	0.0371	0.4854	0.0146	0.9707	0.0293	0.9854
2.19	0.0363	0.4857	0.0143	0.9715	0.0285	0.9857
2.20	0.0355	0.4861	0.0139	0.9722	0.0278	0.9861
2.21	0.0347	0.4864	0.0136	0.9729	0.0271	0.9864
2.22	0.0339	0.4868	0.0132	0.9736	0.0264	0.9868
2.23	0.0332	0.4871	0.0129	0.9743	0.0257	0.9871
2.24	0.0325	0.4875	0.0125	0.9749	0.0251	0.9875
2.25	0.0317	0.4878	0.0122	0.9756	0.0244	0.9878
2.26	0.0310	0.4881	0.0119	0.9762	0.0238	0.9881
2.27	0.0303	0.4884	0.0116	0.9768	0.0232	0.9884
2.28	0.0297	0.4887	0.0113	0.9774	0.0226	0.9887
2.29	0.0290	0.4890	0.0110	0.9780	0.0220	0.9890
2.30	0.0283	0.4893	0.0107	0.9786	0.0214	0.9893
2.31	0.0277	0.4896	0.0104	0.9791	0.0209	0.9896
2.32	0.0270	0.4898	0.0102	0.9797	0.0203	0.9898
2.326	0.0267	0.49	0.01	0.98	0.02	0.99
2.33	0.0264	0.4901	0.0099	0.9802	0.0198	0.9901
2.34	0.0258	0.4904	0.0096	0.9807	0.0193	0.9904
2.35	0.0252	0.4906	0.0094	0.9812	0.0188	0.9906

(*cont.*)

THE NORMAL DISTRIBUTION

z	h	A	B	C	D	E
2.36	0.0246	0.4909	0.0091	0.9817	0.0183	0.9909
2.37	0.0241	0.4911	0.0089	0.9822	0.0178	0.9911
2.38	0.0235	0.4913	0.0087	0.9827	0.0173	0.9913
2.39	0.0229	0.4916	0.0084	0.9832	0.0168	0.9916
2.40	0.0224	0.4918	0.0082	0.9836	0.0164	0.9918
2.41	0.0219	0.4920	0.0080	0.9840	0.0160	0.9920
2.42	0.0213	0.4922	0.0078	0.9845	0.0155	0.9922
2.43	0.0208	0.4925	0.0075	0.9849	0.0151	0.9925
2.44	0.0203	0.4927	0.0073	0.9853	0.0147	0.9927
2.45	0.0198	0.4929	0.0071	0.9857	0.0143	0.9929
2.46	0.0194	0.4931	0.0069	0.9861	0.0139	0.9931
2.47	0.0189	0.4932	0.0068	0.9865	0.0135	0.9932
2.48	0.0184	0.4934	0.0066	0.9869	0.0131	0.9934
2.49	0.0180	0.4936	0.0064	0.9872	0.0128	0.9936
2.50	0.0175	0.4938	0.0062	0.9876	0.0124	0.9938
2.51	0.0171	0.4940	0.0060	0.9879	0.0121	0.9940
2.52	0.0167	0.4941	0.0059	0.9883	0.0117	0.9941
2.53	0.0163	0.4943	0.0057	0.9886	0.0114	0.9943
2.54	0.0158	0.4945	0.0055	0.9889	0.0111	0.9945
2.55	0.0154	0.4946	0.0054	0.9892	0.0108	0.9946
2.56	0.0151	0.4948	0.0052	0.9895	0.0105	0.9948
2.57	0.0147	0.4949	0.0051	0.9898	0.0102	0.9949
2.576	0.0145	0.495	0.005	0.99	0.01	0.995
2.58	0.0143	0.4951	0.0049	0.9901	0.0099	0.9951
2.59	0.0139	0.4952	0.0048	0.9904	0.0096	0.9952
2.60	0.0136	0.4953	0.0047	0.9907	0.0093	0.9953
2.61	0.0132	0.4955	0.0045	0.9909	0.0091	0.9955
2.62	0.0129	0.4956	0.0044	0.9912	0.0088	0.9956
2.63	0.0126	0.4957	0.0043	0.9915	0.0085	0.9957
2.64	0.0122	0.4959	0.0041	0.9917	0.0083	0.9959
2.65	0.0119	0.4960	0.0040	0.9920	0.0080	0.9960
2.70	0.0104	0.4965	0.0035	0.9931	0.0069	0.9965
2.75	0.0091	0.4970	0.0030	0.9940	0.0060	0.9970
2.80	0.0079	0.4974	0.0026	0.9949	0.0051	0.9974
2.85	0.0069	0.4978	0.0022	0.9956	0.0044	0.9978
2.90	0.0060	0.4981	0.0019	0.9963	0.0037	0.9981
2.95	0.0051	0.4984	0.0016	0.9968	0.0032	0.9984
3.00	0.0044	0.4987	0.0013	0.9973	0.0027	0.9987
3.05	0.0038	0.4989	0.0011	0.9977	0.0023	0.9989
3.090	0.0034	0.499	0.001	0.998	0.002	0.999
3.10	0.0033	0.4990	0.0010	0.9981	0.0019	0.9990
3.15	0.0028	0.4992	0.0008	0.9984	0.0016	0.9992
3.20	0.0024	0.4993	0.0007	0.9986	0.0014	0.9993
3.25	0.0020	0.4994	0.0006	0.9988	0.0012	0.9994
3.291	0.0018	0.4995	0.0005	0.999	0.001	0.9995
3.30	0.0017	0.4995	0.0005	0.9990	0.0010	0.9995
3.35	0.0015	0.4996	0.0004	0.9992	0.0008	0.9996
3.40	0.0012	0.4997	0.0003	0.9993	0.0007	0.9997
3.45	0.0010	0.4997	0.0003	0.9994	0.0006	0.9997
3.50	0.0009	0.4998	0.0002	0.9995	0.0005	0.9998
3.55	0.0007	0.4998	0.0002	0.9996	0.0004	0.9998
3.60	0.0006	0.4998	0.0002	0.9997	0.0003	0.9998
3.65	0.0005	0.4999	0.0001	0.9997	0.0003	0.9999
3.70	0.0004	0.4999	0.0001	0.9998	0.0002	0.9999
3.75	0.0004	0.4999	0.0001	0.9998	0.0002	0.9999
3.80	0.0003	0.4999	0.0001	0.9999	0.0001	0.9999

APPENDIX C

The t-Distribution
(SOURCE: Remington and Schork, 1970)

PERCENTILES OF THE t DISTRIBUTION*

d.f.	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.9995}$
1	0.3250	0.7270	1.376	3.078	6.3138	12.706	31.821	63.657	636.619
2	0.2885	0.6172	1.061	1.886	2.9200	4.3027	6.965	9.9248	31.598
3	0.2766	0.5840	0.978	1.638	2.3534	3.1825	4.541	5.8409	12.924
4	0.2707	0.5692	0.941	1.533	2.1318	2.7764	3.747	4.6041	8.610
5	0.2672	0.5598	0.920	1.476	2.0150	2.5706	3.365	4.0321	6.869
6	0.2648	0.5536	0.906	1.440	1.9432	2.4469	3.143	3.7074	5.959
7	0.2632	0.5493	0.896	1.415	1.8946	2.3646	2.998	3.4995	5.408
8	0.2619	0.5461	0.889	1.397	1.8595	2.3060	2.896	3.3554	5.041
9	0.2610	0.5436	0.883	1.383	1.8331	2.2622	2.821	3.2498	4.781
10	0.2602	0.5416	0.879	1.372	1.8125	2.2281	2.764	3.1693	4.587
11	0.2596	0.5400	0.876	1.363	1.7939	2.2010	2.718	3.1058	4.437
12	0.2590	0.5387	0.873	1.356	1.7823	2.1788	2.681	3.0545	4.318
13	0.2586	0.5375	0.870	1.350	1.7709	2.1604	2.650	3.0123	4.221
14	0.2582	0.5366	0.868	1.345	1.7613	2.1448	2.624	2.9768	4.140
15	0.2579	0.5358	0.866	1.341	1.7530	2.1315	2.602	2.9467	4.073
16	0.2576	0.5358	0.865	1.337	1.7459	2.1199	2.583	2.9208	4.015
17	0.2574	0.5344	0.863	1.333	1.7396	2.1098	2.567	2.8982	3.965
18	0.2571	0.5338	0.862	1.330	1.7341	2.1009	2.552	2.8784	3.922
19	0.2569	0.5333	0.861	1.328	1.7291	2.0930	2.539	2.8609	3.883
20	0.2567	0.5329	0.860	1.325	1.7247	2.0860	2.528	2.8453	3.850
21	0.2566	0.5325	0.859	1.323	1.7207	2.0796	2.518	2.8314	3.819
22	0.2564	0.5321	0.858	1.321	1.7171	2.0739	2.508	2.8188	3.792
23	0.2563	0.5318	0.858	1.319	1.7139	2.0687	2.500	2.9073	3.767
24	0.2562	0.5315	0.857	1.318	1.7109	2.0639	2.492	2.7969	3.745
25	0.2561	0.5312	0.856	1.316	1.7081	2.0595	2.485	2.7874	3.725
26	0.2560	0.5309	0.856	1.315	1.7056	2.0555	2.479	2.7787	3.707
27	0.2559	0.5307	0.855	1.314	1.7033	2.0518	2.473	2.7707	3.690
28	0.2558	0.5304	0.855	1.313	1.7011	2.0484	2.467	2.7633	3.674
29	0.2557	0.5302	0.854	1.311	1.6991	2.0452	2.462	2.7564	3.659
30	0.2556	0.5300	0.854	1.310	1.6973	2.0423	2.457	2.7500	3.616
35	0.2553	0.5292	0.8521	1.3062	1.6896	2.0301	2.438	2.7239	3.5919
40	0.2550	0.5286	0.8507	1.3031	1.6839	2.0211	2.423	2.7045	3.5511
45	0.2549	0.5281	0.8497	1.3007	1.6794	2.0141	2.412	2.6896	3.5207
50	0.2547	0.5278	0.8489	1.2987	1.6759	2.0086	2.403	2.6778	3.4965
60	0.2545	0.5272	0.8477	1.2959	1.6707	2.0003	2.390	2.6603	3.4606
70	0.2543	0.5268	0.8468	1.2938	1.6669	1.9945	2.381	2.6480	3.4355
80	0.2542	0.5265	0.8462	1.2922	1.6641	1.9901	2.374	2.6388	3.4169
90	0.2541	0.5263	0.8457	1.2910	1.6620	1.9867	2.368	2.6316	3.4022
100	0.2540	0.5261	0.8452	1.2901	1.6602	1.9840	2.364	2.6260	3.3909
120	0.2539	0.5258	0.8446	1.2887	1.6577	1.9799	2.358	2.6175	3.3736
140	0.2538	0.5256	0.8442	1.2876	1.6558	1.9771	2.353	2.6114	3.3615
160	0.2538	0.5255	0.8439	1.2869	1.6545	1.9749	2.350	2.6070	3.3527
180	0.2537	0.5253	0.8436	1.2863	1.6534	1.9733	2.347	2.6035	3.3456
200	0.2537	0.5252	0.8434	1.2858	1.6525	1.9719	2.345	2.6006	3.3400
∞	0.2533	0.5244	0.8416	1.2816	1.6449	1.9600	2.326	2.5758	3.2905

* The data of this table are extracted with kind permission from *Documenta Geigy Scientific Tables*, 6th Ed., pp. 32-35, Geigy Pharmaceuticals, Division of Geigy Chemical Corporation, Ardsley, N.Y.

APPENDIX D

The Chi-Square Distribution
(SOURCE: Remington and Schork, 1970)

PERCENTILES OF THE CHI-SQUARE DISTRIBUTION[a]

d.f.	$\chi^2_{0.0000}$	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.99}$
1	0.000000393	0.0000393	0.000157	0.000982	0.00393	0.0158	0.0642	0.148	0.275
2	0.00100	0.0100	0.0201	0.0506	0.103	0.211	0.446	0.713	1.022
3	0.0153	0.0717	0.115	0.216	0.352	0.584	1.005	1.424	1.869
4	0.0639	0.207	0.297	0.484	0.711	1.004	1.649	2.195	2.753
5	0.158	0.412	0.554	0.831	1.145	1.610	2.343	3.000	3.655
6	0.299	0.676	0.872	1.237	1.635	2.204	3.070	3.828	4.570
7	0.485	0.989	1.239	1.690	2.167	2.833	3.822	4.671	5.493
8	0.710	1.344	1.646	2.180	2.733	3.490	4.594	5.527	6.423
9	0.972	1.735	2.088	2.700	3.325	4.168	5.380	6.393	7.357
10	1.265	2.156	2.558	3.247	3.940	4.865	6.179	7.267	8.295
11	1.587	2.603	3.053	3.816	4.575	5.578	6.989	8.148	9.237
12	1.934	3.074	3.571	4.404	5.226	6.304	7.807	9.034	10.182
13	2.305	3.565	4.107	5.009	5.892	7.042	8.634	9.926	11.129
14	2.697	4.075	4.660	5.629	6.571	7.790	9.467	10.821	12.079
15	3.108	4.601	5.229	6.262	7.261	8.547	10.307	11.721	13.030
16	3.536	5.142	5.812	6.908	7.962	9.312	11.152	12.624	13.983
17	3.980	5.697	6.408	7.564	8.672	10.085	12.002	13.531	14.937
18	4.439	6.265	7.015	8.231	9.390	10.865	12.857	14.440	15.893
19	4.912	6.844	7.633	8.907	10.117	11.651	13.716	15.352	16.850
20	5.398	7.434	8.260	9.591	10.851	12.443	14.578	16.266	17.809
21	5.896	8.034	8.897	10.283	11.591	13.240	15.445	17.182	18.768
22	6.405	8.643	9.542	10.982	12.338	14.041	16.314	18.101	19.729
23	6.924	9.260	10.196	11.688	13.091	14.848	17.187	19.021	20.690
24	7.453	9.886	10.856	12.401	13.848	15.659	18.062	19.943	21.652
25	7.991	10.520	11.524	13.120	14.611	16.473	18.940	20.867	22.616
26	8.538	11.160	12.198	13.844	15.379	17.292	19.820	21.792	23.579
27	9.093	11.808	12.879	14.573	16.151	18.114	20.703	22.719	24.544
28	9.656	12.461	13.565	15.308	16.928	18.939	21.588	23.647	25.509
29	10.227	13.121	14.256	16.047	17.708	19.768	22.475	24.577	26.475
30	10.804	13.787	14.953	16.791	18.493	20.599	23.364	25.508	27.442
35	13.788	17.192	18.509	20.569	22.465	24.797	27.836	30.178	32.282
40	16.906	20.707	22.164	24.433	26.509	29.051	32.345	34.872	37.134
45	20.136	24.311	25.901	28.366	30.612	33.350	36.884	39.585	41.995
50	23.461	27.991	29.707	32.357	34.764	37.689	41.449	44.313	46.864
60	30.340	35.535	37.485	40.482	43.188	46.459	50.641	53.809	56.620
70	37.467	43.275	45.442	48.758	51.739	55.329	59.898	63.346	66.396
80	44.791	51.172	53.540	57.153	60.391	64.278	69.207	72.915	76.188
90	52.276	59.196	61.754	65.647	69.126	73.291	78.558	82.511	85.993
100	59.897	67.328	70.065	74.222	77.930	82.358	87.945	92.129	95.808
120	75.468	83.852	86.924	91.573	95.705	100.624	106.806	111.419	115.465
140	91.393	100.655	104.035	109.137	113.659	119.029	125.758	130.766	135.149
160	107.598	117.680	121.346	126.870	131.756	137.546	144.783	150.158	154.856
180	124.033	134.885	138.821	144.741	149.969	156.153	163.868	169.588	174.580
200	140.661	152.241	156.432	162.728	168.279	174.835	183.003	189.049	194.319

(*cont.*)

PERCENTILES OF THE CHI-SQUARE DISTRIBUTION

d.f.	$\chi^2_{0.50}$	$\chi^2_{0.60}$	$\chi^2_{0.70}$	$\chi^2_{0.80}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$	$\chi^2_{0.9995}$
1	0.455	0.708	1.074	1.642	2.706	3.841	5.024	6.635	7.879	12.116
2	1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.210	10.597	15.202
3	2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.345	12.838	17.730
4	3.357	4.045	4.878	5.989	7.779	9.488	11.143	13.277	14.860	19.998
5	4.351	5.132	6.064	7.289	9.236	11.070	12.832	15.086	16.750	22.105
6	5.348	6.211	7.231	8.558	10.645	12.592	14.449	16.812	18.548	24.103
7	6.346	7.283	8.383	9.803	12.017	14.067	16.013	18.475	20.278	26.018
8	7.344	8.351	9.524	11.030	13.362	15.507	17.535	20.090	21.955	27.868
9	8.343	9.414	10.656	12.242	14.684	16.919	19.023	21.666	23.589	29.666
10	9.342	10.473	11.781	13.442	15.987	18.307	20.483	23.209	25.188	31.419
11	10.341	11.530	12.899	14.631	17.275	19.675	21.920	24.725	26.757	33.136
12	11.340	12.584	14.011	15.812	18.549	21.026	23.336	26.217	28.300	34.821
13	12.340	13.636	15.119	16.985	19.812	22.362	24.736	27.688	29.819	36.478
14	13.339	14.685	16.222	18.151	21.064	23.685	26.119	29.141	31.319	38.109
15	14.339	15.733	17.322	19.311	22.307	24.996	27.488	30.578	32.601	39.719
16	15.338	16.780	18.418	20.465	23.542	26.296	28.845	32.000	34.267	41.308
17	16.338	17.824	19.511	21.615	24.769	27.587	30.191	33.409	35.718	42.879
18	17.338	18.868	20.601	22.760	25.989	28.869	31.526	34.805	37.156	44.434
19	18.338	19.910	21.689	23.900	27.204	30.144	32.852	36.191	38.582	45.973
20	19.337	20.951	22.775	25.038	28.412	31.410	34.170	37.566	39.997	47.498
21	20.337	21.991	23.858	26.171	29.615	32.671	35.479	38.932	41.401	49.010
22	21.337	23.031	24.939	27.301	30.813	33.924	36.781	40.289	42.796	50.511
23	22.337	24.069	26.018	28.429	32.007	35.172	38.076	41.638	44.181	52.000
24	23.337	25.106	27.096	29.553	33.196	36.415	39.364	42.980	45.558	53.479
25	24.337	26.143	28.172	30.675	34.382	37.652	40.646	44.314	46.928	54.947
26	25.336	27.179	29.246	31.795	35.563	38.885	41.923	45.642	48.290	56.407
27	26.336	28.214	30.319	32.912	36.741	40.113	43.194	46.963	49.645	57.858
28	27.336	29.249	31.391	34.027	37.916	41.337	44.461	48.278	50.993	59.300
29	28.336	30.283	32.461	35.139	39.087	42.537	45.722	49.588	52.336	60.734
30	29.336	31.316	33.530	36.250	40.256	43.773	46.979	50.892	53.672	62.161
35	34.336	36.475	38.859	41.778	46.059	49.802	53.203	57.342	60.275	69.198
40	39.335	41.622	44.165	47.269	51.805	55.758	59.342	63.691	66.766	76.095
45	44.335	46.761	49.452	52.729	57.505	61.656	65.410	69.957	73.166	82.876
50	49.335	51.892	54.723	58.164	63.167	67.505	71.420	76.154	79.490	89.561
60	59.335	62.135	65.226	68.972	74.397	79.082	83.298	88.379	91.952	102.695
70	69.334	72.358	75.689	79.715	85.527	90.531	95.023	100.425	104.215	115.577
80	79.334	82.566	86.120	90.405	96.578	101.879	106.629	112.329	116.321	128.261
90	89.334	92.761	96.524	101.054	107.565	113.145	118.136	124.116	128.299	140.783
100	99.334	102.946	106.906	111.667	118.498	124.342	129.561	135.806	140.169	153.165
120	119.334	123.289	127.616	132.806	140.233	146.567	152.211	158.950	163.648	177.602
140	139.334	143.604	148.269	153.854	161.827	168.613	174.648	181.840	186.846	201.682
160	159.334	163.898	168.876	174.828	183.311	190.516	196.915	204.530	209.824	225.480
180	179.334	184.173	189.446	195.743	204.704	212.304	219.044	227.056	232.620	249.048
200	199.334	204.434	209.985	216.609	226.021	233.994	241.058	249.445	255.264	272.422

* The data of this table are extracted with kind permission from *Documenta Geigy Scientific Tables*, *6th Ed.*, pp. 36–39, Geigy Pharmaceuticals, Division of Geigy Chemical Corporation, Ardsley, N.Y. and Hald, A. and Sinkbaek, S. A., "A Table of Percentage Points of the χ^2-Distribution," *Skandinavisk Aktuarietidskrift*, 33, 168–175, 1950.

APPENDIX E

The F Distribution
(SOURCE: Remington and Schork, 1970)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.999}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9
4	74.14	61.25	56.18	53.44	61.71	50.53	49.66	49.00	48.47
5	47.18	37.12	33.20	31.09	29.75	28.82	28.16	27.64	27.24
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	15.08	10.16	8.28	7.26	6.62	6.18	5.85	5.59	5.39
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10

* Multiply these entries by 100.
f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

The relation $F_{\alpha/2}(f_1 f_2) = \dfrac{1}{F_{1-(\alpha/2)}(f_2 f_1)}$ gives lower percentiles of the F distribution. Note the reversal of degrees of freedom.

$F_{0.999}$

f_2 \ f_1	10	12	15	20	24	30	40	60	120	∞
1	6056*	6107*	6158*	6209*	6235*	6161*	6287*	6313*	6340*	6366*
2	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5	999.5
3	129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	123.5
4	48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	44.05
5	26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	23.79
6	18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	15.75
7	14.08	13.71	13/32	12.93	12.73	12.53	12.33	12.12	11.91	11.70
8	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	6.00
12	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
15	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
16	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
18	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
19	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
22	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
24	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
29	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	1.54
∞	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

$F_{0.005}$

f_2 \ f_1	10	12	15	20	24	30	40	60	120	∞
1	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.90	2.87
19	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
28	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
29	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
30	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

(*cont.*)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.99}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.79	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

$F_{0.99}$

f_2 \ f_1	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

(*cont.*)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.975}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.2	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.14	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

$F_{0.975}$

f_2 \ f_1	10	12	15	20	24	30	40	60	120	∞
1	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

(*cont.*)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.95}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

$F_{0.05}$

f_1 / f_2	10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

(*cont.*)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.90}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.39	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

$F_{0.05}$

f_1 / f_2	10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

(*cont.*)

PERCENTILES OF THE F DISTRIBUTION[1]

$F_{0.75}$

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.46	1.45
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27

f_1 is the number of degrees of freedom in the numerator.
f_2 is the number of degrees of freedom in the denominator.

[1] The data of this table are extracted with kind permission from *Biometrika Tables for Statisticians*, 3rd Ed., Vol. I, Table 18, 1966. London: Bentley House.

$F_{0.75}$

f_2 \ f_1	10	12	15	20	24	30	40	60	120	∞
1	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.25	1.24	1.22	1.19	1.10	1.16	1.14	1.12	1.08	1.00

APPENDIX F

Skewness Test
(SOURCE: Remington and Schork, 1990)

PERCENTILES OF THE SKEWNESS TEST STATISTIC a_3*

Size of sample n	$(a_3)_{0.05}$	$(a_3)_{0.01}$	*Size of sample* n	$(a_3)_{0.05}$	$(a_3)_{0.01}$
25	0.711	1.061	200	0.280	0.403
30	0.662	0.986	250	0.251	0.360
35	0.621	0.923	300	0.230	0.329
40	0.587	0.870	350	0.213	0.305
45	0.558	0.825	400	0.200	0.285
50	0.534	0.787	450	0.188	0.269
			500	0.179	0.255
60	0.492	0.723	550	0.171	0.243
70	0.459	0.673	600	0.163	0.233
80	0.432	0.631	650	0.157	0.224
90	0.409	0.596	700	0.151	0.215
100	0.389	0.567	750	0.146	0.208
			800	0.142	0.202
125	0.350	0.508	850	0.138	0.196
150	0.321	0.464	900	0.134	0.190
175	0.298	0.430	950	0.130	0.185
200	0.280	0.403	1000	0.127	0.180

For a full explanation see Sec. 8-5-1.

* The data of this table are extracted with kind permission from *Biometrika Tables for Statisticians*, 3rd Ed., Vol. I, Table 34, part B, 1966. London: Bentley House.

APPENDIX G

Kurtosis Test
(SOURCE: Remington and Schork, 1970)

PERCENTILES OF GEARY'S KURTOSIS TEST STATISTIC g^*

Size of sample n	$g_{.01}$	$g_{.05}$	$g_{.10}$	$g_{.90}$	$g_{.95}$	$g_{.99}$
11	0.6675	0.7153	0.7400	0.8899	0.9073	0.9359
16	0.6829	0.7236	0.7452	0.8733	0.8884	0.9137
21	0.6950	0.7304	0.7495	0.8631	0.8768	0.9001
26	0.7040	0.7360	0.7530	0.8570	0.8686	0.8901
31	0.7110	0.7404	0.7559	0.8511	0.8625	0.8827
36	0.7167	0.7440	0.7583	0.8468	0.8578	0.8769
41	0.7216	0.7470	0.7604	0.8436	0.8540	0.8722
46	0.7256	0.7496	0.7621	0.8409	0.8508	0.8682
51	0.7291	0.7518	0.7636	0.8385	0.8481	0.8648
61	0.7347	0.7554	0.7662	0.8349	0.8434	0.8592
71	0.7393	0.7583	0.7683	0.8321	0.8403	0.8549
81	0.7430	0.7607	0.7700	0.8298	0.8376	0.8515
91	0.7460	0.7626	0.7714	0.8279	0.8353	0.8484
101	0.7487	0.7644	0.7726	0.8264	0.8344	0.8460
201	0.7629	0.7738	0.7796	0.8178	0.8220	0.8322
301	0.7693	0.7781	0.7828	0.8140	0.8183	0.8260
401	0.7731	0.7807	0.7847	0.8118	0.8155	0.8223
501	0.7757	0.7825	0.7861	0.8103	0.8136	0.8198
601	0.7776	0.7838	0.7873	0.8092	0.8123	0.8179
701	0.7791	0.7848	0.7878	0.8084	0.8112	0.8164
801	0.7803	0.7857	0.7885	0.8077	0.8103	0.8152
901	0.7814	0.7864	0.7890	0.8071	0.8096	0.8142
1001	0.7822	0.7869	0.7894	0.8066	0.8090	0.8134

For a full explanation see Sec. 8-5-2.

* The data of this table are extracted with kind permission from *Biometrika Tables for Statisticians*, 3rd Ed., Vol. I, Table 34, part A, 1966. London: Bentley House.

APPENDIX H

Coefficients for W Test for Normality
(SOURCE: Anderson and Mclean, 1974)

Coefficients (a_{n-i+1}) for the W test for normality, for n = 2(1)50

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		.0000	.1677	.2413	.2806	.3031	.3164	.3244	.3291
3				.0000	.0875	.1401	.1743	.1976	.2141
4						.0000	.0561	.0947	.1224
5								.0000	.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	.3315	.3325	.3325	.3318	.3306	.3290	.3273	.3253	.3232	.3211
3	.2260	.2347	.2412	.2460	.2495	.2521	.2540	.2553	.2561	.2565
4	.1429	.1586	.1707	.1802	.1878	.1939	.1988	.2027	.2059	.2085
5	.0695	.0922	.1099	.1240	.1353	.1447	.1524	.1587	.1641	.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7			.0000	.0240	.0433	.0593	.0725	.0837	.0932	.1013
8					.0000	.0196	.0359	.0496	.0612	.0711
9							.0000	.0163	.0303	.0422
10									.0000	.0410

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	.3185	.3156	.3126	.3098	.3069	.3043	.3018	.2992	.2968	.2944
3	.2578	.2571	.2563	.2554	.2543	.2533	.2522	.2510	.2499	.2487
4	.2119	.2131	.2139	.2145	.2148	.2151	.2152	.2151	.2150	.2148
5	.1736	.1764	.1787	.1807	.1822	.1836	.1848	.1857	.1864	.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	.1092	.1150	.1201	.1245	.1283	.1316	.1346	.1372	.1395	.1415
8	.0804	.0878	.0941	.0997	.1046	.1089	.1128	.1162	.1192	.1219
9	.0530	.0618	.0696	.0764	.0823	.0876	.0923	.0965	.1002	.1036
10	.0263	.0368	.0459	.0539	.0610	.0672	.0728	.0778	.0822	.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12			.0000	.0107	.0200	.0284	.0358	.0424	.0483	.0537
13					.0000	.0094	.0178	.0253	.0320	.0381
14							.0000	.0084	.0159	.0227
15									.0000	.0076

Continued

Coefficients $\{a_{n-i+1}\}$ for the W test for normality,

for n = 2(1)50 (cont.)

i \ 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	.2921	.2898	.2876	.2854	.2834	.2813	.2794	.2774	.2755	.2737
3	.2475	.2463	.2451	.2439	.2427	.2415	.2403	.2391	.2380	.2368
4	.2145	.2141	.2137	.2132	.2127	.2121	.2116	.2110	.2104	.2098
5	.1874	.1878	.1880	.1882	.1883	.1883	.1883	.1881	.1880	.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	.1433	.1449	.1463	.1475	.1487	.1496	.1505	.1513	.1520	.1526
8	.1243	.1265	.1284	.1301	.1317	.1331	.1344	.1356	.1366	.1376
9	.1066	.1093	.1118	.1140	.1160	.1179	.1196	.1211	.1225	.1237
10	.0899	.0931	.0961	.0988	.1013	.1036	.1056	.1075	.1092	.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	.0585	.0629	.0669	.0706	.0739	.0770	.0798	.0824	.0848	.0870
13	.0435	.0485	.0530	.0572	.0610	.0645	.0677	.0706	.0733	.0759
14	.0289	.0344	.0395	.0441	.0484	.0523	.0559	.0592	.0622	.0651
15	.0144	.0206	.0262	.0314	.0361	.0404	.0444	.0481	.0515	.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17			.0000	.0062	.0119	.0172	.0220	.0264	.0305	.0343
18					.0000	.0057	.0110	.0158	.0203	.0244
19							.0000	.0053	.0101	.0146
20									.0000	.0049

i \ 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	.2719	.2701	.2684	.2667	.2651	.2635	.2620	.2604	.2589	.2574
3	.2357	.2345	.2334	.2323	.2313	.2302	.2291	.2281	.2271	.2260
4	.2091	.2085	.2078	.2072	.2065	.2058	.2052	.2045	.2038	.2032
5	.1876	.1874	.1871	.1868	.1865	.1862	.1859	.1855	.1851	.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	.1531	.1535	.1539	.1542	.1545	.1548	.1550	.1551	.1553	.1554
8	.1384	.1392	.1398	.1405	.1410	.1415	.1420	.1423	.1427	.1430
9	.1249	.1259	.1269	.1278	.1286	.1293	.1300	.1306	.1312	.1317
10	.1123	.1136	.1149	.1160	.1170	.1180	.1189	.1197	.1205	.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	.0891	.0909	.0927	.0943	.0959	.0972	.0986	.0998	.1010	.1020
13	.0782	.0804	.0824	.0842	.0860	.0876	.0892	.0906	.0919	.0932
14	.0677	.0701	.0724	.0745	.0765	.0783	.0801	.0817	.0832	.0846
15	.0575	.0602	.0628	.0651	.0673	.0694	.0713	.0731	.0748	.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	.0379	.0411	.0442	.0471	.0497	.0522	.0546	.0568	.0588	.0608
18	.0283	.0318	.0352	.0383	.0412	.0439	.0465	.0489	.0511	.0532
19	.0188	.0227	.0263	.0296	.0328	.0357	.0385	.0411	.0436	.0459
20	.0094	.0136	.0175	.0211	.0245	.0277	.0307	.0335	.0361	.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22			.0000	.0042	.0081	.0118	.0153	.0185	.0215	.0244
23					.0000	.0039	.0076	.0111	.0143	.0174
24							.0000	.0037	.0071	.0104
25									.0000	.0035

APPENDIX I

W Test Percentage Points
(SOURCE: Anderson and Mclean, 1974)

Percentage points of the W test for n = 3(1)50

	Level								
n	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	.687	.707	.748	.792	.935	.987	.992	.996	.997
5	.686	.715	.762	.806	.927	.979	.986	.991	.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	.730	.760	.803	.838	.928	.972	.979	.985	.988
8	.749	.778	.818	.851	.932	.972	.978	.984	.987
9	.764	.791	.829	.859	.935	.972	.978	.984	.986
10	.781	.806	.842	.869	.938	.972	.978	.983	.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	.805	.828	.859	.883	.943	.973	.979	.984	.986
13	.814	.837	.866	.889	.945	.974	.979	.984	.986
14	.825	.846	.874	.895	.947	.975	.980	.984	.986
15	.835	.855	.881	.901	.950	.975	.980	.984	.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	.851	.869	.892	.910	.954	.977	.981	.985	.987
18	.858	.874	.897	.914	.956	.978	.982	.986	.988
19	.863	.879	.901	.917	.957	.978	.982	.986	.988
20	.868	.884	.905	.920	.959	.979	.983	.986	.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	.878	.892	.911	.926	.961	.980	.984	.987	.989
23	.881	.895	.914	.928	.962	.981	.984	.987	.989
24	.884	.898	.916	.930	.963	.981	.984	.987	.989
25	.888	.901	.918	.931	.964	.981	.985	.988	.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	.894	.906	.923	.935	.965	.982	.985	.988	.990
28	.896	.908	.924	.936	.966	.982	.985	.988	.990
29	.898	.910	.926	.937	.966	.982	.985	.988	.990
30	.900	.912	.927	.939	.967	.983	.985	.988	.900
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	.904	.915	.930	.941	.968	.983	.986	.988	.990
33	.906	.917	.931	.942	.968	.983	.986	.989	.990
34	.908	.919	.933	.943	.969	.983	.986	.989	.990
35	.910	.920	.934	.944	.969	.984	.986	.989	.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	.914	.924	.936	.946	.970	.984	.987	.989	.990
38	.916	.925	.938	.947	.971	.984	.987	.989	.990
39	.917	.927	.939	.948	.971	.984	.987	.989	.991
40	.919	.928	.940	.949	.972	.985	.987	.989	.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	.922	.930	.942	.951	.972	.985	.987	.989	.991
43	.923	.932	.943	.951	.973	.985	.987	.990	.991
44	.924	.933	.944	.952	.973	.985	.987	.990	.991
45	.926	.934	.945	.953	.973	.985	.988	.990	.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	.928	.936	.946	.954	.974	.985	.988	.990	.991
48	.929	.937	.947	.954	.974	.985	.988	.990	.991
49	.929	.937	.947	.955	.974	.985	.988	.990	.991
50	.930	.938	.947	.955	.974	.985	.988	.990	.991

APPENDIX J

Noncentral t-Distribution
(SOURCE: Winer, 1971)

Noncentral t Distribution†

Values of the noncentrality parameter for a one-sided test with $\alpha = .050$

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

$$\Pr\{\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)\} = 1 - \beta.$$

† This table is reproduced from: The power of Student's t test. *Journal of the American Statistical Association*, 1965, 60, 320–333, with the permission of the author, D. B. Owen, and the editors.

(*Continued*)

Values of the noncentrality parameter for a one-sided test with $\alpha = .025$

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

$\Pr[\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)] = 1 - \beta.$

(*Continued*)

Values of the noncentrality parameter for a one-sided test with $\alpha = .010$

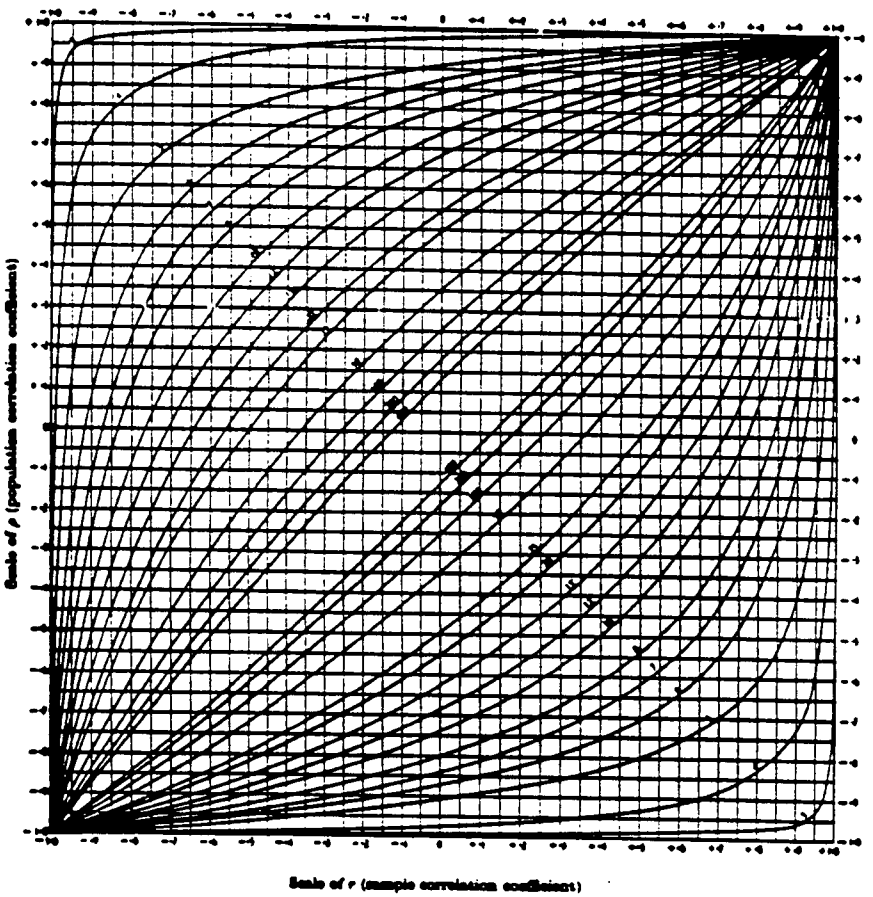
f	β = type 2 error										
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
1	82.00	62.40	52.37	40.80	33.00	26.79	21.47	16.69	12.27	8.07	4.00
2	15.22	12.26	10.74	8.96	7.73	6.73	5.83	4.98	4.12	3.20	2.08
3	9.34	7.71	6.86	5.87	5.17	4.59	4.07	3.56	3.03	2.44	1.66
4	7.52	6.28	5.64	4.88	4.34	3.88	3.47	3.06	2.63	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

$\Pr[\text{noncentral } t > t_{1-\alpha} \mid \delta = (\mu_1 - \mu_0)(\sqrt{n}/\sigma)] = 1 - \beta.$

APPENDIX K

Population Correlation Coefficient
(SOURCE: Remington and Schork, 1970)

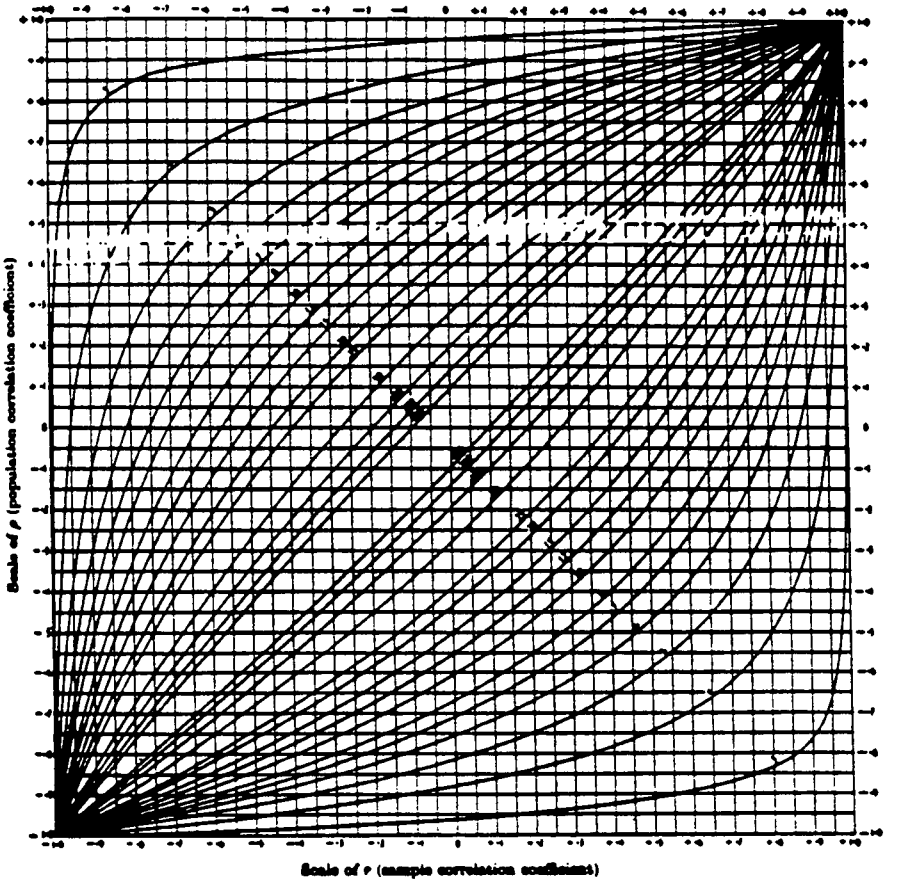
CHARTS FOR COMPUTING CONFIDENCE LIMITS FOR THE POPULATION CORRELATION COEFFICIENT ρ
CONFIDENCE LEVEL 95%*



* The data of this table are extracted with kind permission from *Biometrika Tables for Statisticians*, 3rd Ed., Vol. I, Table 15, 1966. London: Bentley House.

(*cont.*)

CHARTS FOR COMPUTING CONFIDENCE LIMITS FOR THE POPULATION CORRELATION COEFFICIENT ρ
CONFIDENCE LEVEL 99%*



* The data of this table are extracted with kind permission from *Biometrika Tables for Statisticians*, 3rd Ed., Vol. I, Table 15, 1966. London: Bentley House.

APPENDIX L

Transformation of Linearization
(SOURCE: Ponce, 1980a)

Table 39. Transformations of linearization of different functions (after Chow, 1967).

	Type of function	Straight-line coordinate		Equation in linear form
		Abscissa	Ordinate	
1	$y = a + bx$	x	y	$[y] = a + b[x]$
2	$y = bx^{cx}$	x	$\log y$	$[\log y] = \log b + (c \log e)[x]$
3	$y = ax^b$	$\log x$	$\log y$	$[\log y] = \log a + b[\log x]$
4	$y = a_0 + a_1 x + a_2 x^2$	$x - x_0$	$\frac{y - y_0}{x - x_0}$	$\left[\frac{y - y_0}{x - x_0}\right] = a_1 + 2a_2 x_0 + a_2[x - x_0]$
5	$y = a + b/x$	$1/x$	y	$[y] = a + b[1/x]$
6	$y = x/(a + bx)$	x	x/y	$[x/y] = a + b[x]$
7	$y = a/(b + cx)$	x	$1/y$	$[1/y] = (b/a) + (c/a)[x]$
8	$y = c + bc^{cx}$	x	$\log \frac{\Delta y}{\Delta x}$	$\left[\log \frac{dy}{dx}\right] = \log (cb) + (c \log e)[x]$
9	$y = c + ax^b$	$\log x$	$\log \frac{\Delta y}{\Delta x}$	$\left[\log \frac{dy}{dx}\right] = \log (ab) + (b - 1)[\log x]$
10	$y = c + \frac{b}{x - a}$	$x - x_0$	$\frac{x - x_0}{y - y_0}$	$\left[\frac{x - x_0}{y - y_0}\right] = -\frac{a - x_0}{c - y_0} + \frac{1}{c - y_0}[x - x_0]$
11	$y = c + \frac{x}{a + bx}$	x	$\frac{x - x_0}{y - y_0}$	$\left[\frac{x - x_0}{y - y_0}\right] = (a + bx_0) + \frac{b(a + bx_0)}{a}[x]$
12	$y = d + cx + bc^{cx}$	x	$\log \frac{\Delta^2 y}{\Delta x^2}$	$\left[\log \frac{d^2 y}{dx^2}\right] = \log (ab^2) + (c \log e)[x]$
13	$y = dc^{bx^2}$, where $m = a^x$	x	$\log \frac{\Delta^2(\log y)}{\Delta x^2}$	$[y - bc^{cx^2}] = d + e[x]$ $\left[\frac{\log d^2(\log y)}{dx^2}\right] = \log\left[\frac{(\log b)(\log a)^2}{(\log e)^2}\right] + (\log a)[x]$
14	$y = dx^a + bc^{cx}$	$\frac{y_{0+1}}{y_0}$	$\frac{y_{0+2}}{y_0}$	$[\log y - a^x \log b] = \log d + (\log e)[x]$ $\left[\frac{y_{0+2}}{y_0}\right] = -c^{2c} x_0 b c^{cx_0} + (c^{2cx_0} + c^{c\Delta x})\left[\frac{y_{0+1}}{y_0}\right]$
15	$y = e^{ax}(d \cos bx + e \sin bx)$	$\frac{y_{0+1}}{y_0}$	$\frac{y_{0+2}}{y_0}$	$[y e^{-ax}] = d + e[\sin bx]$ $\left[\frac{y_{0+2}}{y_0}\right] = -e^{2a\Delta x} - (2e^{a\Delta x} \cos b\Delta x)\left[\frac{y_{0+1}}{y_0}\right]$ $\left[\frac{y e^{-ax}}{\cos bx}\right] = d + e[\tan bx]$

Symbols: In types 14 and 15, y_0, y_{0+1}, and y_{0+2} are consecutive values for an increment Δx.

APPENDIX M

Multiple Regression Example - Hand Calculated
(SOURCE: Ponce, 1980a)

The mechanics of the least squares determination of the multiple regression line are outlined using an illustrated example. (The example presented here has been taken directly from Freese, 1967. It is included with only minor modification because I feel it presents the concept in a very clear and concise manner.) Consider the data presented in Table 37. With this data we would like to fit an equation of the form

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

Table 37. Data for the example illustrating the least squares determination of the multiple regression line (after Freese, 1967).

	Y	X_1	X_2	X_3
	65	41	79	75
	78	90	48	83
	85	53	67	74
	50	42	52	61
	55	57	52	59
	59	32	82	73
	82	71	80	72
	66	60	65	66
	113	98	96	99
	86	80	81	90
	104	101	78	86
	92	100	59	88
	96	84	84	93
	65	72	48	70
	81	55	93	85
	77	77	68	71
	83	98	51	84
	97	95	82	81
	90	90	70	78
	87	93	61	89
	74	45	96	81
	70	50	80	77
	75	60	76	70
	75	68	74	76
	93	75	96	85
	76	82	58	80
	71	72	58	68
	61	46	69	65
Sums	2,206	1,987	2,003	2,179
Means (n = 28)	78.7857	70.9643	71.5357	77.8214

According to the principle of least squares, the best estimates of the "b" coefficients can be obtained by solving the set of least squares normal equations.

b_1 equation: $(\Sigma x_1^2)b_1 + (\Sigma x_1x_2)b_2 + (\Sigma x_1x_3)b_3 = \Sigma x_1y$

b_2 equation: $(\Sigma x_1x_2)b_1 + (\Sigma x_2^2)b_2 + (\Sigma x_2x_3)b_3 = \Sigma x_2y$

b_3 equation: $(\Sigma x_1x_3)b_1 + (\Sigma x_2x_3)b_2 + (\Sigma x_3^2)b_3 = \Sigma x_3y$

where: $$\Sigma x_ix_j = \Sigma X_iX_j - \frac{(\Sigma X_i)(\Sigma X_j)}{n}$$

The corrected sums of squares and products are computed in the familiar manner:

$$\Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = (65^2 + \ldots + 61^2) - \frac{(2206)^2}{28} = 5{,}974.7143$$

$$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n} = (41^2 + \ldots + 46^2) - \frac{(1987)^2}{28} = 11{,}436.9643$$

$$\Sigma x_1y = \Sigma X_1Y - \frac{(\Sigma X_1)(\Sigma Y)}{n} = (41)(65) + \ldots + (46)(61) - \frac{(1{,}987)(2{,}206)}{28} = 6{,}428.7858$$

Similarly,

$\Sigma x_1x_2 = -1{,}171.4642$
$\Sigma x_1x_3 = 3{,}458.8215$
$\Sigma x_2^2 = 5{,}998.9643$
$\Sigma x_2x_3 = 1{,}789.6786$
$\Sigma x_2y = 2{,}632.2143$
$\Sigma x_3^2 = 2{,}606.1072$
$\Sigma x_3y = 3{,}327.9286$

Putting these values in the normal equations gives:

$$11{,}436.9643b_1 - 1{,}171.4642b_2 + 3{,}458.8215b_3 = 6{,}428.7858$$

$$-1{,}171.4642b_1 + 5{,}998.9643b_2 + 1{,}789.6786b_3 = 2{,}632.2143$$

$$3{,}458.8215b_1 + 1{,}789.6786b_2 + 2{,}606.1072b_3 = 3{,}327.9286$$

These equations can be solved by any of the standard procedures for simultaneous equations. One approach is as follows:

1. Divide through each equation by the numerical coefficient of b_1.

$b_1 - 0.102,427,897b_2 + 0.302,424,788b_3 = -0.562,105,960$

$b_1 - 5.120,911,334b_2 - 1,527,727,949b_3 = -2.246,943,867$

$b_1 + 0.517,424,389b_2 + 0.753,466,809b_3 = 0.962,156,792$

2. Subtract the second equation from the first and the third from the first so as to leave two equations in b_2 and b_3.

$5.018,483,437b_2 + 1.830,152,737b_3 = 2.809,049,827$

$-0.619,852,296b_2 - 0.451,042,021b_3 = -0.400,050,832$

3. Divide through each equation by the numerical coefficient of b_2.

$b_2 + 0.364,682,430b_3 = 0.559,740,779$

$b_2 + 0.727,660,494b_3 = 0.645,397,042$

4. Subtract the second of these equations from the first, leaving one equation in b_3.

$-0.362,978,064b_3 = -0.085,656,263$

5. Solve for b_3.

$$b_3 = \frac{-0.085,656,263}{-0.362,978,064} = 0.235,981,927$$

6. Substitute this value of b_3 in one of the equations (say the first) of step 3 and solve for b_2.

$b_2 + (0.364,682,430)(0.235,981,927) = 0.559,740,779$

$b = 0.473,682,316$

7. Substitute the solutions for b_2 and b_3 in one of the equations (say the first) of step 1, and solve for b_1.

$b_1 - (0.102,427,897)(0.473,682,316) + (0.302,424,788)(0.235,981,927)$

$= 0.562,105,960$

$b_1 = 0.539,257,459$

8. As a check, add up the original normal equations and substitute the solutions for b_1, b_2 and b_3.

$13,724.3216b_1 = 6,617,1787b_2 + 7,854.6073b_3 = 12,388.9287$

$12,388.92869 = 12,388.9287$, check.

Given the values of b_1, b_2, and b_3 we can now compute

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 - b_3\bar{X}_3 = -11.7320$$

Thus, after rounding of the coefficients, the regression equation is

$$Y = -11.732 + 0.539\ X_1 + 0.474\ X_2 + 0.236\ X_3$$

It should be noted that in solving the normal equations more digits have been carried than would be justified by the rules for number of significant digits. Unless this is done, the rounding errors may make it difficult to check the computations.

11.3.3 Testing the Significance of the Multiple Regression Line

At this point in our regression analysis we should ask, "How well does the regression line fit the data?" The analysis of variance procedure to be used here is similar to that outlined for the significance test of the simple linear regression line. However, in this case the degrees of freedom for the reduction are equal to the number of independent variables fitted. The reduction sum of squares for any least squares regression can be found using the general equation:

SS_{reg} = (est. coefficients)(right side of their respective normal equations) (53)

Therefore, for Freese's example

$$SS_{reg} = b_1(\Sigma x_1 y) + b_2(\Sigma x_2 y) + b_3(\Sigma x_3 y)$$

The anova table for the test of significance is as follows in Table 38:

Table 38. ANOVA results for the test of significance of the multiple regression developed from the data given in Table 36 (after Freese, 1967).

Source	df	SS	MS
Reduction due to X_1, X_2, and X_3	3	5,498.9389	1,832.9796
Residuals	24	475.7754	19.8240
Total	27	5,974.7143	

To test the significance of the regression we compute F_s where

$$F_s = \frac{\text{Reduction MS}}{\text{Residual MS}}$$

For the case at hand $F_s = 92.46$, which is significant at the 0.01 level.

In some instances we would like to know the contribution each independent variable makes in the prediction of the dependent variable. In other words, what portion of the total SS can be attributed to each individual independent variable. The statistical method to use for this type of analysis is stepwise regression. The procedures for stepwise regression are not outlined here, but can be found in any good statistics text.

11.3.4 Coefficient of Multiple Determination

The coefficient of multiple determination is calculated in the same manner as that for the simple linear regression.

$$r^2 = \frac{SS_{reg}}{SS_{tot}} \tag{54}$$

For our illustrative example, $r^2 = 0.92$, which means 92 percent of the variation in Y is associated with the regression.

APPENDIX N

Multiple Regression Example - Computerized
(SOURCE: Gaugush, 1986)

Multiple regression

173. The concepts applicable to simple linear regression extend to multiple regression. The purpose is still to quantify a relationship between a dependent response variable and some independent variable. However, in multiple regression there is more than one independent variable. Since there may exist interrelationships between the so-called independent variables, the regression coefficients for the dependent variables from a multiple regression are not the same as would be obtained if the dependent variable was regressed on each independent variable separately.

174. The formulas for a multiple regression will not be derived in this manual. For simple two-factor (i.e., two independent variables) multiple regression, formulas are available in some textbooks, but broad application of multiple regressions requires matrix algebra. We will presume that the calculations will be done on a computer and shall therefore concentrate on the interpretation of computer output. Computer table and graphic output presented in this chapter were obtained from SAS procedure GLM (SAS (1981) GRAPH User's Guide; SAS (1982) User's Guide).

175. An example of multiple regression data is given in Table 10. These data represent levels of a pollutant measured at various distances downstream from a source. The pollutant decreases with distance from the source of pollution, but additional factors are expected to influence the results of measurements taken on different occasions. Therefore, two other independent variables have been included in the multiple regression--temperature and discharge. The model is then

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i} + \epsilon_i$$

Table 10

Hypothetical Data for Multiple Regression Example

Observation	Pollutant	Distance, m	Temperature, °C	Discharge m^3/sec
1	15.5	1,000	24	0.8
2	12.9	1,000	20	1.0
3	14.8	1,000	19	1.3
4	10.3	1,000	25	1.8
5	10.7	1,000	15	2.0
6	14.9	2,000	17	0.5
7	6.6	2,000	20	1.0
8	9.5	3,000	21	1.0
9	5.1	3,000	15	2.0
10	7.4	4,000	15	0.5
11	11.9	4,000	24	0.8
12	5.4	4,000	21	1.0

where X_1 , X_2 , and X_3 refer to distance, temperature, and discharge, respectively.

176. Examination of the computer output (Table 11) will illustrate several concepts. The output first indicates that the analysis was performed on a dependent variable called POLLUTNT. The sources of variation are listed as MODEL and ERROR, which will sum to the CORRECTED TOTAL ($= \Sigma y^2$) as before. Note that the model now has 3 d.f., one for each of the independent variables. Since there were 12 observations, the CORRECTED TOTAL carries 11 d.f., one being lost when the variables are adjusted about their means. Three degrees of freedom were required for the three variables in the model, leaving eight as a measure of random error. The computer provides calculations of the sum of squares for each source, and the corresponding mean square. An F test is calculated for the mean square of the model, using the mean square error. For the example in Table 11, the F test indicates that the regression does account for a significant portion of the total variation. This F test is a joint test of all three of the independent variables and does not indicate which one(s) contribute significantly to the description of POLLUTNT.

Table 11

Computer Output for Multiple Regression Example

(General Linear Model Procedure)

DEPENDENT VARIABLE:	POLLUTANT			
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	3	97.73161461	32.57720487	4.84
ERROR	8	53.82505206	6.72813151	PR > F
CORRECTED TOTAL	11	151.55666667		0.0331

R-SQUARE	C.V.	ROOT MSE	POLLUTANT MEAN
0.644852	24.9011	2.59386420	10.41666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F
DISTANCE	1	54.19739726	8.06	0.0219
TEMP	1	7.19969359	1.07	0.3312
DISCHARG	1	36.33452376	5.40	0.0486

SOURCE	DF	TYPE III SS	F VALUE	PR > F
DISTANCE	1	77.32993941	11.9	0.0095
TEMP	1	1.69303058	0.25	0.6294
DISCHARG	1	36.33452376	5.40	0.0486

PARAMETER	ESTIMATE	R FOR HO: PARAMETER=0	PR > T	STD ERROR OF ESTIMATE
INTERCEPT	17.59409833	3.02	0.0166	5.82622109
DISTANCE	-0.00225161	-3.39	0.0095	0.00066415
TEMP	0.11241376	0.50	0.6294	0.22409610
DISCHARG	-3.78579922	-2.32	0.0486	1.62909000

177. That portion accounted for by the model is given by the r^2 value, which is the ratio of the MODEL sum of squares ($= \Sigma_j b_j \Sigma_i (X_j)^2$, the b_j being the estimate of the β_j) to the CORRECTED TOTAL sum of squares (Σy_i). Thus 65.49 percent of the variation was accounted for by the model. There are several important points to be made about these calculations. Usual practice is to adjust about the mean (i.e., to fit an intercept not necessarily equal to zero), but there are some exceptions. An option in the computer algorithm will provide results not adjusted about the means, but the r^2 value then does not have the same interpretation.

178. Following the synopsis of the model are the results of the individual independent variables for the regression. As before, two types of sums of squares are provided, but in the case of multiple regression they differ. They are the "sequential sums of squares" or, in the jargon of SAS GLM, Type I SS, and partial sums of squares or Type III SS.

179. The sequential sums of squares are commonly, but not necessarily, larger than the corresponding partial sums of squares; indeed, Table 11 provides an example of a case where one partial sum of squares is larger. The sequential sums of squares depend on the order in which they were entered into the computer program. Here distance was entered first, and on its own accounted for 54.1974 of the sum of squares carried by the model. This is exactly the same as would be obtained if a simple linear regression was done for POLLUTNT on DISTANCE, with no other variables in the model. TEMP was the second variable entered into the model, so the variable POLLUTNT was already adjusted for DISTANCE when TEMP was entered. A model containing DISTANCE and TEMP carries a sum of squares of 61.3971 (not given explicitly) so that contribution of TEMP, given that DISTANCE is already in the model, is 61.3971 - 54.1974 = 7.1997. DISCHARG was last to enter the model, so the effect of DISTANCE and TEMP jointly was already included. Accordingly, the sum of squares carried by DISCHARG, given that DISTANCE and TEMP are already in the model, is 97.7316 - 61.3971 = 36.3345.

180. The partial sums of squares for each variable are obtained as if all other variables were already included in the model; i.e., they give the contribution to the total sum of squares carried by each variable adjusted for the previous inclusion of all other variables jointly. Thus, the partial sum of squares for DISCHARG is here equal to the sequential sum of squares since this was the last variable entered. Clearly, the sequential sums of squares will differ according to the order that the variables are entered. The partial sums of squares are independent of this ordering.

181. The appropriate sums of squares for testing hypothesis depend on the circumstances. These tests are specific to the model so if one of the variables were to be omitted, or another variable added, the values would change.

182. For instance, in the example, the partial sums of squares show that there is virtually no merit in including TEMP in the model in addition to DISTANCE and DISCHARG (i.e., the hypothesis $\beta_2 = 0$ can be accepted). The sequential sums of squares also show that there would be little or no merit in including TEMP in addition to DISTANCE, but neither sequential nor partial sums of squares provide a test of adding DISCHARG to DISTANCE or vice versa. Nor do we see how a model that includes TEMP and DISCHARG would perform. For these we would have to fit two-variable models or enter the variables in a different order.

183. The final section of output from the computer contains the parameter estimates, or regression coefficients. These values, as with simple linear regression, can be used to estimate the mean pollution, given values of distance, temperature, and discharge. The equation for this example would be

POLLUTNT = 17.5941 - 0.00225*DISTANCE + 0.1124*TEMP - 3.7858*DISCHARG

184. Note, however, that the regression coefficients for a variable are always calculated with adjustments for all other variables in the model. The parameter would have to be reestimated if TEMP were excluded.

APPENDIX O

Example Use of R-Square
(SOURCE: Spooner, 1984)

AN EXAMPLE OF USING R^2 FOR ANALYSIS OF AG NPS PROJECT DATA.

From the above discussion it may be apparent that we can describe mean water quality concentration values for each monitoring station as a function of a series of independent variables. This regression model takes the form of:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + E$$

Where: X_1 may be precipitation or stream flow
X_2 may be soil type.
X_3 may be percentage of the drainage area with agricultural BMPs applied.

From our experiences we know that X_1 and Y are usually highly correlated (large R^2) at monitoring stations impacted primarily by NPS. X_2 and X_3 may or may not turn out to increase significantly the value of R^2. Obviously, we can add more X's (crop type, season, stream flow) to explain an increasing amount of the variation in Y.

Will such a model, however, be useful to us? Probably not if our purpose is to document in a limited time period whether the BMPs we are implementing are having a positive effect on in-stream water quality.

Thus, what we really want to know is whether Y is correlated (significant R^2) with X_3. Suppose we had a project area with 12 water sampling sites each with a corresponding drainage area. We will assume that precipitation patterns did not vary within the watershed and that all sites were sampled simultaneously (within a few hours). Although we implemented phosphorus-control BMPs throughout the project area, implementation had inevitably progressed further in some sub-basins of the project area than in others. Assuming that the land use percentages in each basin were approximately equal, this situation affords an excellent opportunity to use the R^2 from a simple regression model ($Y=B_0+B_3X + E$) to determine whether there was a significant association between the level of BMP application and the mean P concentration at the corre-

sponding station. Such a situation might have generated the following data:

Subbasin	Total P Conc(Y) (mg/l)(N=20)	% of Drainage Area with BMPs
A	0.52	19
B	0.41	41
C	0.15	29
D	0.31	11
E	0.17	70
F	0.60	22
G	0.40	20
H	0.39	12
I	0.46	14
J	0.10	47
K	0.48	31
L	0.22	58

The R^2 value for these data is only 0.36 which means that only a small fraction of the variation in Y is explained by the presence or absence of BMPs (Not a surprising result since we have neglected to account for rainfall or stream flow, which we know have a major effect). However, using the F test described above, we find that this R^2 value is highly significant (at the 95% confidence level). This means that the results probably did not occur by chance, and provide statistical evidence that the BMP implementation caused a decrease in pollutant concentrations.

(Final Suggestion: Look at your project area sub-basins and see if such paired watershed analysis might be appropriate!)

APPENDIX P

Confidence Limits for the Median of Any Continuous Distribution

n	α = 0.05		α = 0.01		n	α = 0.05		α = 0.01	
	l	u	l	u		l	u	l	u
5	-	-	-	-	41	14	28	12	30
6	1	6	-	-	42	15	28	13	30
7	1	7	-	-	43	15	29	13	31
8	1	8	1	8	44	16	29	14	31
9	2	8	1	9	45	16	30	14	32
10	2	9	1	10	46	16	31	14	33
11	2	10	1	11	47	17	31	15	33
12	3	10	2	11	48	17	32	15	34
13	3	11	2	12	49	18	32	16	34
14	3	12	2	13	50	18	33	16	35
15	4	12	3	13	51	19	33	16	36
16	4	13	3	14	52	19	34	17	36
17	5	13	3	15	53	19	35	17	37
18	5	14	4	15	54	20	35	18	37
19	5	15	4	16	55	20	36	18	38
20	6	15	4	17	56	21	36	18	39
21	6	16	5	17	57	21	37	19	39
22	6	17	5	18	58	22	37	19	40
23	7	17	5	19	59	22	38	20	40
24	7	18	6	19	60	22	39	20	41
25	8	18	6	20	61	23	39	21	41
26	8	19	7	20	62	23	40	21	42
27	8	20	7	21	63	24	40	21	43
28	9	20	7	22	64	24	41	22	43
29	9	21	8	22	65	25	41	22	44
30	10	21	8	23	66	25	42	23	44
31	10	22	8	24	67	26	42	23	45
32	10	23	9	24	68	26	43	23	46
33	11	23	9	25	69	26	44	24	46
34	11	24	10	25	70	27	44	24	47
35	12	24	10	26	71	27	45	25	47
36	12	25	10	27	72	28	45	25	48
37	13	25	11	27	73	28	46	26	48
38	13	26	11	28	74	29	46	26	49
39	13	27	12	28	75	29	47	26	50
40	14	27	12	29					

Source: After Geigy, 1982.

Given are the values l and u such that for the order statistics $x_{[l]}$ and $x_{[u]}$, Prob$[x_{[l]} <$ true median $< x_{[u]}] \geq 1 - \alpha$.

Appendix Q
(Source: Conover, 1980)

Quantiles of the Spearman test statistic.

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.

SOURCE. Adapted from Glasser and Winter (1961), with corrections, with permission from the *Biometrika* Trustees.

[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.

APPENDIX R

Probabilities for the Mann-Kendall Nonparametric Test for Trend

S	Values of n				S	Values of n		
	4	5	8	9		6	7	10
0	0.625	0.592	0.548	0.540	1	0.500	0.500	0.500
2	0.375	0.408	0.452	0.460	3	0.360	0.386	0.431
4	0.167	0.242	0.360	0.381	5	0.235	0.281	0.364
6	0.042	0.117	0.274	0.306	7	0.136	0.191	0.300
8		0.042	0.199	0.238	9	0.068	0.119	0.242
10		$0.0^{2}83$	0.138	0.179	11	0.028	0.068	0.190
12			0.089	0.130	13	$0.0^{2}83$	0.035	0.146
14			0.054	0.090	15	$0.0^{2}14$	0.015	0.108
16			0.031	0.060	17		$0.0^{2}54$	0.078
18			0.016	0.038	19		$0.0^{2}14$	0.054
20			$0.0^{2}71$	0.022	21		$0.0^{3}20$	0.036
22			$0.0^{2}28$	0.012	23			0.023
24			$0.0^{3}87$	$0.0^{2}63$	25			0.014
26			$0.0^{3}19$	$0.0^{2}29$	27			$0.0^{2}83$
28			$0.0^{4}25$	$0.0^{2}12$	29			$0.0^{2}46$
30				$0.0^{3}43$	31			$0.0^{2}23$
32				$0.0^{3}12$	33			$0.0^{2}11$
34				$0.0^{4}25$	35			$0.0^{3}47$
36				$0.0^{5}28$	37			$0.0^{3}18$
					39			$0.0^{4}58$
					41			$0.0^{4}15$
					43			$0.0^{5}28$
					45			$0.0^{6}28$

Source: From Kendall, 1975. Used by permission.

Repeated zeros are indicated by powers; for example, $0.0^{3}47$ stands for 0.00047.

Each table entry is the probability that the Mann-Kendall statistic *S* equals or exceeds the specified value of S when no trend is present.

This table is used in Section 16.4.1.

Appendix S
(Source: Conover, 1980)

Quantiles of the Wilcoxon signed ranks test statistic.

	$w_{0.005}$	$w_{0.01}$	$w_{0.025}$	$w_{0.05}$	$w_{0.10}$	$w_{0.20}$	$w_{0.30}$	$w_{0.40}$	$w_{0.50}$	$\frac{n(n+1)}{2}$
$n=4$	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.5	15
6	0	0	1	3	4	6	8	9	10.5	21
7	0	1	3	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.5	45
10	4	6	9	11	15	19	22	25	27.5	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.5	91
14	13	16	22	26	32	39	44	48	52.5	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.5	153
18	28	33	41	48	56	66	73	80	85.5	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	83	91	98	105	210
21	44	50	59	68	78	91	100	108	115.5	131
22	49	56	67	76	87	100	110	119	126.5	153
23	55	63	74	84	95	110	120	130	138	176
24	62	70	82	92	105	120	131	141	150	300
25	69	77	90	101	114	131	143	153	162.5	325
26	76	85	99	111	125	142	155	165	175.5	351
27	84	94	108	120	135	154	167	178	189	378
28	92	102	117	131	146	166	180	192	203	406
29	101	111	127	141	158	178	193	206	217.5	435
30	110	121	138	152	170	191	207	220	232.5	465
31	119	131	148	164	182	205	221	235	248	496
32	129	141	160	176	195	219	236	250	264	528
33	139	152	171	188	208	233	251	266	280.5	561
34	149	163	183	201	222	248	266	282	297.5	595
35	160	175	196	214	236	263	283	299	315	630
36	172	187	209	228	251	279	299	317	333	666
37	184	199	222	242	266	295	316	335	351.5	703
38	196	212	236	257	282	312	334	353	370.5	741
39	208	225	250	272	298	329	352	372	390	780
40	221	239	265	287	314	347	371	391	410	820
41	235	253	280	303	331	365	390	411	430.5	861
42	248	267	295	320	349	384	409	431	451.5	903
43	263	282	311	337	366	403	429	452	473	946
44	277	297	328	354	385	422	450	473	495	990
45	292	313	344	372	403	442	471	495	517.5	1035
46	308	329	362	390	423	463	492	517	540.5	1081
47	324	346	379	408	442	484	514	540	564	1128

	$w_{0.005}$	$w_{0.01}$	$w_{0.025}$	$w_{0.05}$	$w_{0.10}$	$w_{0.20}$	$w_{0.30}$	$w_{0.40}$	$w_{0.50}$	$\frac{n(n+1)}{2}$
48	340	363	397	428	463	505	536	563	588	1176
49	357	381	416	447	483	527	559	587	612.5	1225
50	374	398	435	467	504	550	583	611	637.5	1275

For n larger than 50, the pth quantile w_p of the Wilcoxon signed ranks test statistic may be approximated by $w_p = [n(n+1)/4] + x_p\sqrt{n(n+1)(2n+1)/24}$, where x_p is the pth quantile of a standard normal random variable, obtained from Table A1.

SOURCE. Adapted from Harter and Owen (1970), with permission from the Institute of Mathematical Statistics.

[a] The entries in this table are quantiles w_p of the Wilcoxon signed ranks test statistic T, given by Equation 5.7.6, for selected values of $p \leq .50$. Quantiles w_p for $p > .50$ may be computed from the equation

$$w_p = n(n+1)/2 - w_{1-p}$$

where $n(n+1)/2$ is given in the right hand column in the table. Note that $P(T < w_p) \leq p$ and $P(T > w_p) \leq 1-p$ if H_0 is true. Critical regions correspond to values of T less than (or greater than) but not including the appropriate quantile.

Appendix T
(Source: Gilbert, 1987)

Nonparametric 95% and 99% confidence intervals on a proportion.

u	n = 1				n = 2				n = 3				n = 4				n = 5				n = 6				u
0	0	0	.95	.99	0	0	.78	.90	0	0	.63	.78	0	0	.53	.68	0	0	.50	.60	0	0	.41	.54	0
1	.01	.05	1	1	.01	.03	.97	.99	.00	.02	.86	.94	.00	.01	.75	.86	.00	.01	.66	.78	.00	.01	.59	.71	1
2					.10	.22	1	1	.06	.14	.98	1	.04	.10	.90	.96	.03	.08	.81	.89	.03	.06	.73	.83	2
3									.22	.37	1	1	.14	.25	.99	1	.11	.19	.92	.97	.08	.15	.85	.92	3

u	n = 7				n = 8				n = 9				n = 10				n = 11				n = 12				u
0	0	0	.38	.50	0	0	.36	.45	0	0	.32	.43	0	0	.29	.38	0	0	.26	.36	0	0	.24	.25	0
1	.00	.01	.55	.64	.00	.01	.50	.59	.00	.01	.44	.57	.00	.01	.44	.51	.00	.00	.40	.50	.00	.00	.37	.45	1
2	.02	.05	.66	.76	.02	.05	.64	.71	.02	.04	.56	.66	.02	.04	.56	.62	.01	.03	.50	.59	.01	.03	.46	.55	2
3	.07	.13	.77	.86	.06	.11	.71	.80	.05	.10	.68	.75	.05	.09	.62	.70	.04	.08	.60	.66	.04	.07	.54	.65	3
4	.14	.23	.87	.93	.12	.19	.81	.88	.11	.17	.75	.83	.09	.15	.70	.78	.08	.14	.67	.74	.08	.12	.63	.70	4
5	.24	.34	.95	.98	.20	.29	.89	.94	.17	.25	.83	.89	.15	.22	.78	.85	.13	.20	.74	.81	.12	.18	.71	.77	5
6	.36	.45	.99	1	.29	.36	.95	.98	.25	.32	.90	.95	.22	.29	.85	.91	.19	.26	.80	.87	.17	.24	.76	.83	6

u	n = 13				n = 14				n = 15				n = 16				n = 17				n = 18				u
0	0	0	.23	.32	0	0	.23	.30	0	0	.22	.28	0	0	.20	.26	0	0	.19	.26	0	0	.18	.25	0
1	.00	.00	.34	.43	.00	.00	.32	.42	.00	.00	.30	.39	.00	.00	.30	.36	.00	.00	.28	.35	.00	.00	.27	.34	1
2	.01	.03	.43	.52	.01	.03	.42	.50	.01	.02	.39	.46	.01	.02	.37	.45	.01	.02	.35	.43	.01	.02	.33	.41	2
3	.04	.07	.52	.59	.03	.06	.50	.58	.03	.06	.47	.54	.03	.05	.44	.52	.03	.05	.42	.50	.03	.05	.41	.47	3
4	.07	.11	.59	.68	.06	.10	.58	.64	.06	.10	.53	.61	.06	.09	.50	.58	.05	.08	.49	.57	.05	.08	.47	.53	4
5	.11	.17	.66	.73	.10	.15	.63	.70	.09	.14	.61	.67	.09	.13	.56	.64	.08	.12	.54	.62	.08	.12	.53	.59	5
6	.16	.22	.74	.79	.15	.21	.68	.75	.13	.19	.67	.72	.13	.18	.63	.70	.12	.17	.59	.66	.11	.16	.59	.66	6
7	.21	.26	.78	.84	.19	.24	.76	.81	.18	.22	.71	.77	.17	.20	.70	.74	.16	.19	.65	.73	.15	.18	.63	.69	7
8	.27	.34	.83	.89	.25	.32	.79	.85	.23	.29	.78	.82	.21	.27	.73	.79	.20	.25	.72	.76	.18	.24	.67	.75	8
9	.32	.41	.89	.93	.30	.37	.85	.90	.28	.33	.81	.87	.26	.30	.80	.83	.24	.28	.75	.80	.23	.27	.73	.77	9

u	n = 19				n = 20				n = 21				n = 22				n = 23				n = 24				u
0	0	0	.17	.24	0	0	.16	.22	0	0	.15	.21	0	0	.15	.20	0	0	.14	.19	0	0	.13	.19	0
1	.00	.00	.25	.32	.00	.00	.24	.31	.00	.00	.23	.29	.00	.00	.22	.28	.00	.00	.21	.27	.00	.00	.20	.26	1
2	.01	.02	.32	.39	.01	.02	.32	.37	.01	.02	.30	.37	.01	.02	.29	.35	.01	.02	.27	.33	.01	.02	.26	.32	2
3	.02	.04	.39	.46	.02	.04	.37	.44	.02	.04	.35	.42	.02	.04	.34	.40	.02	.04	.32	.39	.02	.03	.31	.39	3
4	.05	.08	.45	.52	.04	.07	.42	.50	.04	.07	.40	.47	.04	.06	.39	.45	.04	.06	.39	.45	.04	.06	.37	.43	4
5	.07	.11	.50	.56	.07	.10	.47	.56	.07	.10	.46	.53	.06	.09	.45	.50	.06	.09	.43	.50	.06	.09	.41	.48	5
6	.10	.15	.55	.61	.10	.14	.53	.60	.09	.13	.51	.58	.09	.13	.50	.55	.08	.12	.48	.55	.08	.11	.46	.52	6
7	.14	.17	.61	.68	.13	.16	.58	.64	.12	.15	.55	.63	.12	.15	.55	.60	.11	.14	.52	.58	.11	.13	.50	.57	7
8	.17	.22	.66	.71	.16	.21	.63	.69	.15	.20	.60	.66	.15	.19	.58	.65	.14	.18	.57	.62	.13	.17	.54	.61	8
9	.21	.25	.69	.76	.20	.24	.68	.73	.19	.23	.65	.71	.18	.22	.62	.68	.17	.21	.61	.67	.16	.20	.59	.64	9
10	.24	.31	.75	.79	.22	.29	.71	.78	.21	.28	.70	.74	.20	.26	.66	.72	.19	.25	.64	.70	.19	.23	.63	.68	10
11	.29	.34	.78	.83	.27	.32	.76	.80	.26	.30	.72	.79	.24	.29	.71	.76	.23	.27	.68	.73	.22	.26	.66	.72	11
12	.32	.39	.83	.86	.31	.37	.79	.84	.29	.35	.77	.81	.28	.34	.74	.80	.27	.32	.73	.77	.26	.31	.69	.74	12

u	n = 25				n = 26				n = 27				n = 28				n = 29				n = 30				u
0	0	0	.13	.18	0	0	.12	.17	0	0	.12	.17	0	0	.12	.16	0	0	.11	.16	0	0	.11	.16	0
1	.00	.00	.19	.26	.00	.00	.19	.25	.00	.00	.18	.24	.00	.00	.17	.23	.00	.00	.17	.22	.00	.00	.16	.22	1
2	.01	.01	.25	.31	.01	.01	.24	.30	.01	.01	.23	.30	.01	.01	.23	.29	.01	.01	.22	.28	.01	.01	.21	.27	2
3	.02	.03	.30	.37	.02	.03	.30	.36	.02	.03	.29	.34	.02	.03	.28	.33	.02	.03	.27	.32	.01	.03	.26	.31	3
4	.03	.06	.36	.41	.03	.05	.34	.40	.03	.05	.33	.38	.03	.05	.32	.36	.03	.05	.31	.37	.03	.05	.30	.36	4
5	.05	.08	.40	.46	.05	.08	.38	.44	.05	.08	.37	.44	.05	.07	.36	.42	.05	.07	.36	.41	.04	.07	.35	.39	5
6	.08	.11	.44	.50	.07	.11	.42	.49	.07	.10	.41	.48	.07	.10	.41	.46	.07	.09	.39	.44	.06	.09	.38	.43	6
7	.10	.13	.48	.54	.10	.12	.47	.53	.09	.12	.46	.52	.09	.12	.44	.50	.09	.11	.43	.48	.08	.11	.41	.47	7
8	.13	.16	.52	.59	.12	.15	.51	.56	.12	.15	.50	.56	.11	.14	.48	.54	.11	.14	.46	.52	.10	.13	.45	.51	8
9	.16	.19	.56	.63	.15	.19	.54	.60	.14	.18	.54	.59	.14	.17	.52	.58	.13	.17	.50	.56	.13	.16	.48	.54	9
10	.18	.22	.60	.66	.17	.21	.58	.64	.17	.20	.57	.62	.16	.19	.56	.62	.16	.18	.54	.59	.15	.18	.52	.57	10
11	.21	.25	.64	.69	.19	.24	.62	.68	.18	.23	.60	.66	.18	.23	.59	.64	.17	.22	.57	.63	.16	.21	.55	.61	11
12	.25	.30	.68	.74	.23	.28	.66	.70	.22	.27	.63	.70	.21	.26	.62	.67	.21	.25	.61	.65	.20	.24	.59	.64	12
13	.26	.32	.70	.75	.25	.30	.70	.75	.24	.29	.67	.72	.23	.28	.65	.71	.22	.27	.64	.68	.22	.26	.62	.67	13
14	.31	.36	.75	.79	.30	.34	.72	.77	.28	.33	.71	.76	.27	.32	.68	.73	.26	.31	.66	.72	.25	.30	.65	.69	14
15	.34	.40	.78	.82	.32	.38	.76	.81	.30	.37	.73	.78	.29	.35	.72	.77	.28	.34	.69	.74	.27	.32	.68	.73	15

Source: After Blyth and Still, 1983.

Inner entries give the 95% interval, and outer entries the 99% interval. For example, for $n = 13$, $u = 3$, the 95% interval is (0.07, 0.52) and the 99% interval is (0.04, 0.59). n = number of observations. u = number of those that exceed some specified value x_c.

This table is used in Section 11.11.

Appendix U
(Source: Conover, 1980)

Sample sizes for one-sided nonparametric tolerance limits[a].

$1-\alpha$	$q=.500$	.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

$$P(q \text{ of the population} \leq X^{(n)}) \geq 1-a$$

REFERENCES

REFERENCES

Air National Guard. 1993. *Draft final Remedial Investigation/Feasibility 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.

APHA-AWWA-WPCF. 1975. *Standard methods for the examination of water and wastewater.* 14th ed. American Public Health Association, Washington, DC.

ASIWPCA. 1986. *America's clean water: the states' nonpoint source assessment 1985. Appendix.* Association of State and Interstate Water Pollution Control Administrators, Washington, DC.

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.

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., 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.

Barnett, V. and T. Lewis. 1979. *Outliers in statistical data.* John Wiley & Sons, Inc., New York, New York.

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.

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.

Brown, M.P. 1981. *Quality assurance plan for research and monitoring of the West Branch Delaware River model implementation program.* New York State Dept. of environmental Conservation, Albany, New York.

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.

Carsel, R.F., C.N. Smith, L.A. Mulkey, J.D. Dean, and P. Jowise. 1984. *Users manual for the pesticide root zone model (przm), release 1.* EPA/600/3-84/109. U.S. Environmental Protection Agency, Washington, DC.

Cassell, E.A.,and D.W. Meals. 1981. *Comprehensive water quality monitoring and evaluation program for the St. Albans Bay rural clean water program watershed project.* Vermont Water Resources Research Center, Burlington, Vermont.

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

Chow, V.T. 1951. *The log-probability law and its engineering applications.* Separate No. 536, 80. American Society of Civil Engineering. November.

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., 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.

Clean Water Partnership. 1989. *Water quality monitoring for the Clean Water Partnership. A guidance document*. Minnesota Pollution Control Agency, Division of Water Quality. March.

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.

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.

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.

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.

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.

Cox, D.R., and Stuart, A. 1955. Some quick tests for trend in location and dispersion. *Biometrika* 42:80-95.

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.

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. 1984. *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. 1984. *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.

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., and F.J. Humenik. 1986a. *Rural clean water program status report on the cm&e projects 1985.* 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.

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., R.P. Maas, M.D. Smolen, and F.J. Humenik, eds. 1984. *Proceedings of the rural clean water program.* Biological and Agricultural Engineering Dept., North Carolina State University, Raleigh, North Carolina.

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.

Erlebach, W.E., 1979. A systematic approach to monitoring trends in the quality of surface waters. In Everett, L.G. and K.D. Schmidt, Eds., *Establishment of water quality monitoring programs.* Proceedings of a symposium held in San Francisco, CA June 12-14, 1978, AWRA, Minneapolis, Minnesota.

Everett, L.G. 1980. *Groundwater monitoring.* General Electric Company, Schenectady, New York.

Field, R., A.N. Tafuri, and H.E. Masters. 1977. *Urban runoff pollution control technology overview.* EPA-600/2-77-047. U.S. Environmental Protection Agency, Washington, DC.

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.

Freund, J.E. 1973. *Modern elementary statistics.* Prentice-Hall, Englewood Cliffs, New Jersey.

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.

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.

Gray, J.S. 1989. Effects of environmental stress on species-rich assemblages. *Biol. J. Linn. Soc.* 37:19-32.

Grayman, W.M. 1984. *Application of the routing and graphical display system to the streams of the United States.* U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Green, R.H. 1993. Application of repeated measures designs in environmental impact and monitoring studies. *Austr. J. Ecol.* 18:81-98.

Gulland, J.A. 1983. *Fish stock assessment: A manual of basic methods.* FAO/Wiley Series, Vol. 1. Wiley & Sons, New York.

Guy, H.P., and V.W. Norman. 1970. *Field methods for measurement of fluvial sediment.* U.S. Geological Survey, TWRI Book 3, Chapter 2, Washington, D.C.

Hall, D. C., and W.R. Berkas. 1988. Comparison of instream and laboratory methods of measuring sediment oxygen demand. *Water Resour. Bull.* 24:533-544.

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.

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.

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.

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.

Hawkins, R.H., R.G. Werner, and J.O. Crevelling. 1974. *Summary report-runoff quality in a suburban stream.* New England Interstate Water Pollution Control Commission.

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.

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.

Hershfield, D.N. 1961. *Rainfall Frequency Atlas of the United States.* U.S. Weather Bureau Technical Paper 40. May.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomologist* ?:31-39.

Hipel, K.W. 1988. Nonparametric approaches to environmental impact assessment. *Water Resour. Bull.* 24(3).

Hipel, K.W., A.I. McLeod and R.R. Weiler. 1988. Data analysis of water quality time series in Lake Erie. *Water Resour. Bull.* 24:533-544.

Hirsch, R.M., J.R. Slack and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* 18:107-121.

Hirsch, R.M., and J.R. Slack. 1984. *A nonparametric trend test for seasonal data with serial dependence.* Water Resources Research 20:727-732.

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.

Hollander, M., and D.A. Wolfe. 1973. *Nonparametric statistical methods.* John Wiley and Sons, Inc., New York, New York.

Horn, C.R. and W.M. Grayman. 1985. *National and local modeling of water quality using the EPA reach file and related data.* Draft. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Horn, C.R. 1986. Draft Memorandum of Understanding. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

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.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54(2):187-211.

I.E.P.A. 1979. *Water quality management plan, Volume III.* Illinois Environmental Protection Agency, Springfield, Illinois.

IHD-WHO Working Group on Quality of Water. 1978. *Water quality surveys: a guide for the collection and interpretation of water quality data.* Unesco/WHO, Sydenhams Printers, Poole, Dorset, United Kingdom.

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.

ITFM. 1994. *Water-quality monitoring in the United States.* Intergovernmental Task Force on Monitoring Water Quality, Interagency Advisory Committee on Water Data, Water Information Coordination Program. Washington, DC. January.

ITFM. 1994. *Water-quality monitoring in the United States. Technical appendixes.* Intergovernmental Task Force on Monitoring Water Quality, Interagency Advisory Committee on Water Data, Water Information Coordination Program. Washington, DC. January.

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.

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.

Keith, L.H., ed. 1988. *Principles of environmental sampling.* American Chemical Society, Washington, D.C.

Keith, L.H., W. Crummett, J. Deegan Jr., R.A. Libby, J.K. Taylor and G. Wentler. 1983. Principles of environmental analysis. *Analytical Chemistry* (55):2210-2218.

Kendall, M.G. 1975. *Rank correlation methods, 4th edition.* Charles Griffin, London.

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 Research.* 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.

Knisel, W.G., ed. 1980. *CREAMS: a field-scale model for chemicals, runoff, and erosion from agricultural management systems.* Conservation Research Report No. 26. U.S.Department of Agriculture, Washington, DC.

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.

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.

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.

Maas, R.P. 1989. *Designing effective nonpoint source water quality monitoring programs.* Prepared for the Tennessee Valley Authority. November.

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.

MACS. 1993. *Standard operating procedures and technical basis. Macroinvertebrate collection and habitat assessment for low gradient non-tidal streams.* Draft. Prepared by The Mid-Atlantic Coastal Streams Workgroup. August 24.

Mandel, J. 1964. *The statistical analysis of experimental data.* Interscience Publishers division of John Wiley and Sons, New York, New York.

Mann, H.B. 1945. Non-Parametric tests against trend. *Econometrica* 13:245-259.

Martin, D.M. 1984. *Annual report, Rock Creek rural clean water program comprehensive monitoring and evaluation.* Idaho Department of Health and Welfare, Division of Environment, Boise, Idaho.

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.

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.

Mendenhall, W. 1971. *Introduction to Probability and Statistics.* Duxbury Press, Belmont, California. Cited in MacDonald et al., 1991.

Metcalf and Eddy, Inc. 1979. Wastewater engineering: treatment, disposal, reuse. Second Edition. McGraw-Hill, New York, New York.

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. *Amer. Soc. Engin. Trans.* 3:48-51,54.

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.

National Research Council. 1986. *Ecological Knowledge and Environmental Problem-Solving. Concepts and Case Studies.* National Academy Press, Washington, DC.

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.

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.

Ponce, S.L. 1980b. *Water quality monitoring programs.* WSDG Technical Paper WSDG-TP-00002. U.S. Department of Agriculture, U.S. Forest Service.

Puri, L.M., and P.K. Sen. 1971. *Nonparametric methods in multivariate analysis.* John Wiley & Sons, Inc., New York, New York.

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.

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.

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.

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.

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.

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.

Sen, P.K. 1968. *Estimates of the regression coefficient based on Kendalls tau.* Journal of the American Statistical Association 63:1379-1389.

Shelley, P.E. 1979. *Monitoring requirments, methods, and costs for the nationwide urban runoff program.* U.S. Environmental Protection Agency, Water Planning Division.

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.

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.

Simonson, T., and J. Lyons. 1992a. *Evaluation Monitoring of Stream Habitat During Priority Watershed Projects.* Wisconsin Department of Natural Resources. May.

Simonson, T., and J. Lyons. 1992b. *Evaluation Monitoring of Stream Fish Communities During Priority Watershed Projects.* Wisconsin Department of Natural Resources. May.

Skalski, J.R., and D.H. McKenzie. 1982. A design for aquatic monitoring programs. *J. Environ. Manage.* 14:237-251.

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 Dept., 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. 1986b. *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.

Snedecor, G.W. and W.G. Cochran. 1967. *Statistical methods.* 6th 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. 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., 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.

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.

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.

Stein, J., ed. 1980. *The Random House college dictionary*. Random House, Inc.

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. February.

The Volunteer Monitor. Published by the Coastal Resources Center, The University of Rhode Island, Rhode Island Sea Grant Program, Narragansett, Rhode Island.

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.

USDA. 1981. *Forest management for water quality*. Forest Service, Area Planning and Development.

USDA-FS. 1994. *Evaluating the effectiveness of forestry best management practices in meeting water quality goals or standards*. Miscellaneous Publication 1520. G.E. Dissmeyer. USDA Forest Service, Southern Region, Atlanta, Georgia. June.

USDA-SCS. 1993. *Water Quality Monitoring*. Draft. U.S. Department of Agriculture, Soil Conservation Service, Washington, DC. June.

USEPA. 1973. *Biological field and laboratory methods for measuring the quality of surface waters and effluents*. EPA-670/4-73-001. Program Element 1BA027. C.I. Weber, ed. U.S. Environmental Protection Agency, National Environmental Research Center, Office of Research and Development, Cincinnati, Ohio. July.

USEPA. 1975. *Model state water monitoring program*. National Water Monitoring Panel. EPA-440/9-74-002. 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. 1979. *Methods for chemical analysis of water and wastes.* EPA-600/4-79-020. 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. 1982a. *Chesapeake bay: introduction to an ecosystem.* Chesapeake Bay Program, Annapolis, Maryland.

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 and Appendices. U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1983. *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. 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. 1983. *Water Quality Standards Handbook.* U.S. Environmental Protection Agency, Office of Water Regulations and Standards. Washington, DC. December.

USEPA. 1983a. *Chesapeake Bay program: findings and recommendations.* U.S.Environmental Protection Agency, Region 3, Philadelphia, Pennsylvania.

USEPA. 1983b. *Results of the nationwide urban runoff program, volume I - final report.* U.S. Environmental Protection Agency, Water Planning Division, Washington, DC.

USEPA. 1984. *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. 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.

USEPA. 1984a. *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. 1984b. *Directory of environmental data bases.* PB86-214434. U.S. Environmental Protection Agency, Water Division, Region 5, Chicago, Illinois.

USEPA. 1984c. *Draft ground-water monitoring issue paper.* U.S. Environmental Protection Agency, Office of Ground Water Protection, Washington, DC.

USEPA. 1984d. *Ocean monitoring strategy.* U.S. Environmental Protection Agency, Monitoring and Data Support Division, Washington, DC.

USEPA. 1984e. *Report to congress: nonpoint source pollution in the U.S.* U.S. Environmental Protection Agency, Office of Water Program Operations, Washington, DC.

USEPA. 1985. *Guidance for State Water Monitoring and Wasteload Allocation Programs.* EPA 440/4-85-031. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC. October.

USEPA. 1985. *Summary of U.S. EPA-approved methods, standard methods, and other guidance for 301(h) monitoring variables.* EPA 503/4-90-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. September.

USEPA. 1985a. *Fiscal year 1985 monitoring strategy update for NPDES compliance.* Draft. U.S. Environmental Protection Agency, Office of Water Enforcement and Permits, Washington, DC.

USEPA. 1985b. *Planning workshop to develop recommendations for a ground-water monitoring strategy.* U.S. Environmental Protection Agency, Office of Ground Water Protection, Washington, DC.

USEPA. 1985c. *Resource document for the ground-water monitoring strategy workshop.* U.S. Environmental Protection Agency, Office of Ground Water Protection, Washington, DC.

USEPA. 1986. *Handbook: Stream Sampling for Waste Load Allocation Applications.* EPA/625/6-86/013. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. September.

USEPA. 1986. *RCRA Ground-Water Monitoring Technical Enforcement Guidance Document.* OSWER-9950.1. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC. September.

USEPA. 1987. *Bioaccumulation Monitoring Guidance: Selection of Target Species and Review of Available Bioaccumulation Data.* EPA 430/9-86-005. Volume I. U.S. Environmental Protection Agency, Office of Marine and Estuarine Protection, Washington, DC. March.

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. 1987a. *Handbook, ground water.* EPA/625/6-87/016. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1987b. *Nonpoint source guidance.* U.S. Environmental Protection Agency, Office of Water, Washington, DC.

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. EPA/600/3-89/013. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. March.

USEPA. 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. May.

USEPA. 1989. *Statistical analysis of ground-water monitoring data at RCRA facilities - Interim final guidance.* EPA/530-SW-89-026. NTIS PB89-151047. U.S. Environmental Protection Agency, Office of Solid Waste, Waste Management Division, Washington, DC. February.

USEPA. 1990. *Environmental Monitoring and Assessment Program: Ecological Indicators.* EPA/600/3-90/060. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. September.

USEPA. 1990. *Monitoring Lake and Reservoir Restoration.* EPA 440/4-90-007. U.S. Environmental Protection Agency, Office of Water, Washington, DC. August.

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. 1990. *Volunteer Water Monitoring: A Guide for State Managers.* EPA 440/4-90-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1991. *A Review of Methods for Assessing Nonpoint Source Contaminated Ground-Water Discharge to Surface Water.* EPA 570/9-91-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC. April.

USEPA. 1991. *Monitoring Guidance for the National Estuary Program.* EPA 503/8-91-002. U.S. Environmental Protection Agency, Office of Water. August.

USEPA. 1991. *Volunteer Lake Monitoring: A Methods Manual.* EPA 440/4-91-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. December.

USEPA. 1991. *Watershed Monitoring and Reporting for Section 319 National Monitoring Program Projects.* U.S. Environmental Protection Agency, Office of Water, Washington, DC. August.

USEPA. 1992. *A Guide to Selected National Environmental Statistics in the U.S. Government.* EPA-230-R-92-003. U.S. Environmental Protection Agency, Washington, DC. April.

USEPA. 1992. *Inventory of Exposure-Related Data Systems Sponsored by Federal Agencies.* EPA-600-R-92-078. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1992. *NPDES Storm Water Sampling Guidance Document.* EPA 833-B-92-001. U.S. Environmental Protection Agency, Office of Water. Washington, DC. July.

USEPA. 1992. *Office of Water Environmental and Program Information Systems Compendium.* EPA-800-B-92-001. U.S. Environmental Protection Agency, Washington, DC.