San Francisco Bay Regional Water Enhancement Program IRWM Round 2 Implementation Proposal

Attachment 6 References

Project 1. Bay Area Regional Conservation and Education Program

N/A

Project 2. East Bayshore Recycled Water Project Phase 1A (Emeryville)

N/A

Project 3. Lagunitas Creek Watershed Sediment Reduction and Management Project

1. MMWD. 2011. *Lagunitas Creek Stewardship Plan*.

   See Attachment 3, Project 3 References.

2. O'Connor Environmental. 2012. *Sediment and Streambed Monitoring Plan for Lagunitas Creek.* Prepared for MMWD.

   See Attachment 3, Project 3 References

Project 4. Marin/Sonoma Conserving Our Watersheds: Agricultural BMP Projects

1. California Department of Fish and Wildlife (CDFW). 2010. California Salmonid Stream Habitat Restoration Manual. 4th Edition.
   Available: http://www.dfg.ca.gov/fish/Resources/HabitatManual.asp

   See Attachment 3, Project 4 References

2. Lennox, M., N. Scolari, and D. Lewis. 2010. Riparian Zone Monitoring Plan. Prepared by University of California Cooperative Extension for Marin Resource Conservation District, Point Reyes Station CA. 75 p. http://cesonoma.ucdavis.edu/files/76316.pdf

   See Attachment 3, Project 4 References

3. Tomales Bay Watershed Council. 2003. Tomales Bay Watershed Water Quality Monitoring Plan.
   Available: http://www.tomalesbaywatershed.org/waterqualitymonitoring.pdf

4. USDA Natural Resources Conservation Service. 2013. Field Office Technical Guide.
   Available: http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/fotg/

5. USEPA. 1997. Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls. EPA 841-B-96-004.
Available: http://water.epa.gov/polwaste/nps/monitoringguidance.cfm

Project 5. Napa Milliken Creek Flood Damage Reduction and Fish Passage Barrier Removal

N/A

Project 6. North Bay Water Reuse Program—Sonoma Valley CSD 5th Street East/McGill Road Recycled Water Project

N/A

Project 7. Oakland Sausal Creek Restoration Project

1. California Department of Fish and Wildlife (CDFW). 2010. California Salmonid Stream Habitat Restoration Manual. 4th Edition.
Available: http://www.dfg.ca.gov/fish/Resources/HabitatManual.asp

See Attachment 3, Project 4 References

2. State Water Resources Control Board. 2007. Surface Water Ambient Monitoring Program (SWAMP) Standard Operating Procedures.
Available: http://swamp.mpsl.mlml.calstate.edu/resources-and-downloads/standard-operating-procedures#FieldMeasures

Project 8. Pescadero Water Supply and Sustainability Project

1. HydroScience Engineers, Inc. 2013. Technical Memorandum #1: Water Supply Reliability. March 19, 2013.

See Attachment 3, Project 8 References

Project 9. Petaluma Flood Reduction, Water & Habitat Quality, and Recreation Project for Capri Creek

N/A

Project 10. Redwood City Bayfront Canal and Atherton Channel Flood Improvement and Habitat Restoration Project

N/A

Project 11. Regional Groundwater Storage and Recovery Project Phase 1A - South Westside Basin, Northern San Mateo County

1. SFPUC. 2012. *South Westside Basin Groundwater Management Plan*. July.

See Attachment 3, Project 11 References

Project 12. Richmond Breuner Marsh Restoration Project

N/A

Project 13. Roseview Heights Infrastructure Upgrades for Water Supply and Quality Improvement, Santa Clara County

N/A

Project 14. San Francisco Bay Climate Change Pilot Projects Combining Ecosystem Adaptation, Flood Risk Management and Wastewater Effluent Polishing

N/A

Project 15. San Francisco International Airport Reclaimed Water Facility

N/A

Project 16. San José Green Streets & Alleys Demonstration Projects

N/A

Project 17. San Pablo Rheem Creek Wetlands Restoration Project

1. USACE and SFRWQCB Mitigation Monitoring and Reporting Plan (MMRP)

Project 18. St. Helena Upper York Creek Dam Removal and Ecosystem Restoration Project

1. City of St. Helena and U.S. Army Corps of Engineers. 2009. *Administrative Draft Mitigation, Monitoring, and Reporting Plan (MMRP) Upper York Creek Ecosystem Restoration Project*. November.

Project 19. Students and Teachers Restoring a Watershed (STRAW) Project—North and East Bay Watersheds

1. Lennox, M., N. Scolari, and D. Lewis. 2010. *Riparian Zone Monitoring Plan* (RZMP). Prepared by University of California Cooperative Extension for Marin Resource Conservation District, Point Reyes Station CA. 75 p. http://cesonoma.ucanr.edu/files/76316.pdf.

   See Attachment 3, Project 19 References

2. PRBO Science Education Evaluation Methodology, supported and approved by Marin Community Foundation Evaluation Resources.

3. Ralph, C.J., G.R. Geupel, P. Pyle, T.E. Martin and D.F. DeSante. 1993. *Handbook of field methods for monitoring landbirds*. USDA Forest Service General Technical Report PSW-GTR-144**.**

   See Attachment 3, Project 19 References

4. Seavy, N. E., and T. Gardali. 2012. “Developing a riparian bird index to communicate restoration success in Marin County, California.” *Ecological Restoration* 30:157-160.
5. State Water Resources Control Board. 2001. Standard Operating Procedure (SOP) 4.2.1.4. Stream Photo Documentation Procedure.

Administrative Draft

Mitigation, Monitoring, and Reporting Plan

Upper York Creek Ecosystem Restoration Project

Napa County, California

(State Clearinghouse No. 2006092096)

CEQA Lead Agency

City of St. Helena
1480 Main Street
St. Helena, CA 94574

Contact: John Ferons, Director of Public Works/City Engineer
707-968-2658

A Joint Project of the

City of St. Helena, Napa County, California
U.S. Army Corps of Engineers, San Francisco District, South Pacific Division

November 2009

This page is intentionally blank.

TABLE OF CONTENTS

Page

1 INTRODUCTION 1-1

2 MITIGATION, MONITORING, AND REPORTING PLAN 2-1

2.1 Aesthetics 2-1

2.2 Air Quality 2-2

2.3 Biological Resources 2-4

2.4 Cultural Resources 2-10

2.5 Geology, Soils, Landslides, and Seismic Activity 2-10

2.6 Global Climate Change 2-10

2.7 Hazardous Materials 2-11

2.8 Hydrology and Hydraulics 2-13

2.9 Noise 2-13

2.10 Traffic and Transportation 2-14

2.11 Water Quality 2-15

This page is intentionally blank.

1 Introduction

A Mitigation, Monitoring, and Reporting Plan (MMRP) is required when a lead agency adopts findings regarding significant effects analyzed in an Environmental Impact Report (EIR) (PRC §21081.6(a); CCR §§15091(d), 15097). The Draft EIR (DEIR) for the *Upper York Creek Ecosystem Restoration Project* identified six potentially significant and unavoidable impacts after mitigation for some or all of the project alternatives. These include:

- Aesthetics (AES-NP4): Placement of materials along the shore of the lower reservoir may result in a small reduction of water surface area as viewed from above; impact may occur in all project alternatives.

- Biological Resources (BIO-NP1): Upper St. Helena Dam would continue to act as an unnatural migration barrier to fish and other aquatic species; impact would occur only in the No Project alternative.

- Cultural Resources (CUL-PA1): Removal of the UYCD and associated structures would have a significant adverse impact on the historic feeling, design, and association of a resource considered eligible for listing in the National Register of Historic Places and the California Register of Historic Resources; impact would occur in all action alternatives.

- Global Climate Change (GCC-NP1): On-going actions to remove sediment and debris would increase carbon dioxide emissions by committing the City to long-term, repeated maintenance activities; impact would occur only in the No Project alternative.

- Hydrology & Hydraulics: (H&H-PA1): Sediment aggradation in lower York Creek may increase frequency and severity of flooding; impact may occur in all action alternatives. The City will prepare a monitoring and maintenance plan (Mitigation H&H-PA1) as described in DEIR Section 3.8.4.

- Noise (NOI-NP2): Maintenance or construction activities would temporarily increase noise at the dam and disposal sites; impact would occur in all project alternatives.

The MMRP must include all measures incorporated into the project that will mitigate or avoid significant impacts on the environment (CCR §15097). The program must ensure compliance with the mitigation and avoidance measures by providing for appropriate monitoring and reporting prior to, during, and subsequent to project implementation. The plan identifies who is responsible for compliance and when. In the case of the *Upper York Creek Ecosystem Restoration Project*, the City is not only the lead agency for compliance with the California Environmental Quality Act (CEQA), it is also the project sponsor. Thus, all long-term monitoring activities and report preparation are the responsibility of the City.

In addition to the measures contained in Section 2 of the MMRP, two other documents are referenced: a Storm Water Pollution Prevention Plan (SWPPP) and DEIR Appendix 2, the *Lower York Creek Reservoir Revegetation and Mitigation Plan*. A SWPPP is developed and implemented for projects that will impact an acre or more; it specifies Best Management Practices (BMPs) that are designed to prevent construction pollutants from contacting storm water. Adherence to the SWPPP will ensure that all products of erosion from earthwork are prevented from moving into receiving waters.

DEIR Appendix 2, the *Lower York Creek Reservoir Revegetation and Mitigation Plan,* was developed to provide guidance to minimize impacts on aesthetics and biological resources through planting disturbed areas with appropriate native species and stabilizing surface soils. The plan requires that the City be responsible to ensure that the revegetation and monitoring actions described in the DEIR and the MMRP are taken to meet the objectives and stated success criteria by employing professionals with experience in native revegetation, wildlife habitat restoration, and erosion control. The revegetation plan provides for both wetland and upland restoration; details include storage of transplants from the fill area, native plant lists, and methods for successful plantings, monitoring, and maintenance, including remedial actions if established success criteria are not met.

2 Mitigation, Monitoring, and Reporting Plan

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.1 Aesthetics				
AES-NP1 Aesthetic impacts may occur during project activities at the upper and lower reservoirs through removal of vegetation and use of heavy equipment.	• Establish a buffer, defined by a qualified biologist, to protect established vegetation along Spring Mountain Road. • Use BMPs to prevent "track-out" of material from the construction area to the paved public road, including washing down equipment before moving from the property onto a paved public road and cleaning visible track-out on the paved public road using wet sweeping or a HEPA filter-equipped vacuum device within 24 hours. • Revegetate with native species in disturbed areas to provide erosion control and aesthetically pleasing wildlife habitat. • Work during daylight hours only to prevent impacts from light/glare.	Contractor: - During construction City: - Post-construction vegetation success monitoring	Dust control standards: • 90% or more of dust track-out checks required by the SWPPP are clean. Wetland plantings: • Cover at the end of one year is ≥70% of the cover in nearby reference wetlands. If ≥70% cover is not present, a certified professional will determine if additional planting is required or if an additional growing season for existing plants will allow growth to meet the cover criterion. If after two years criterion is not met, require additional plantings until area has ≥70% of reference cover. Upland plantings: • 80% survival of the woody plants after three years; require replacement planting if success criterion is met.	Dust control: BMPs recorded in on-site SWPPP binder. Vegetation success: Post-construction and annual reports to ecological regulators, as required, until plants are successfully established.
AES-NP2 Placement of materials removed from the upper reservoir may result in aesthetic impacts at the lower reservoir.	• Limit disturbance to existing vegetation and revegetate with native species that will cover the surface of the sediment and blend in with existing vegetation. • Have all work conducted under the supervision of a licensed and/or certified professional with experience in native revegetation, wildlife habitat restoration, and erosion control. • Follow all additional revegetation requirements in Appendix 2.	City and Contractor: - During construction	See wetland and upland plantings success criteria in AES-NP1 above.	City to submit post-construction and annual monitoring reports to ecological regulators, as required, until plants are successfully established.

Impact	Mitigation	Who & When	Success Criteria	Reporting
AES-NP4 Placement of materials along the shore of the lower reservoir may result in impacts to views of the water surface area.	Install native vegetation in accordance with Appendix 2.	City and Contractor: - During construction	See wetland and upland plantings success criteria in AES-NP1 above.	See AES-NP2 above.
2.2 Air Quality				
AIR-PA1 Asbestos dust could be released into the air during removal of sediment and/or dam materials at the upper reservoir and during placement at reuse/ disposition sites.	• Use BMPs to avoid releasing dust from construction activities at the project sites and along public and private roads. Dam material will be used as fill only in locations where it can be capped with clean fill to limit the release of asbestos into the environment. • Secure BAAQMD Asbestos Dust Mitigation Plan approval and implement specific measures to limit asbestos dust from becoming airborne. • Secure materials in place at LYCR with both 3 inches of nonasbestos-containing soil and vegetative planting above the expected average water level; see, also, Appendix 2. • Do not transport dam material to Spring Mountain Vineyard or Clover Flat landfill.	Contractor: - During construction	BAAQMD asbestos dust mitigation standards: • No visible dust track-out on road. • No visible dust blowing across property line. • Implement all BMPs required by the Asbestos Dust Mitigation Plan.	BMPs recorded in on-site SWPPP binder
AIR-NP2 Hydrogen sulfide (H_2S) gas may pose a threat to air quality if it is released during movement of sediment in the reservoir.	• Equip at least one worker in each construction area with personal H_2S monitoring device. • Follow OSHA protocols for safety of construction workers; if H_2S levels exceed safe levels, halt work until H_2S dissipates.	Contractor: - During construction	Complete construction without any worker health incidents.	Contractor to furnish report to City within 48 hours documenting any health incident or post-construction checklist documenting lack of incidents.

Impact	Mitigation	Who & When	Success Criteria	Reporting
AIR-NP3 Project activities could result in an increase in small particulate matter (PM10) emissions during sediment removal.	Utilize BAAQMD CEQA Guidelines for PM10 emission BMPs for construction in the SWPPP, including: • Water all active construction areas at least twice daily. • Cover all trucks hauling soil, sand, and other loose materials or require all trucks to maintain at least two feet of freeboard. • Pave, apply water three times daily, or apply nontoxic soil stabilizers on all unpaved access roads, parking areas, and staging areas at construction sites. • Sweep all paved access roads, parking areas, and staging areas at construction sites daily with water sweepers. • Sweep streets daily with water sweepers if visible soil material is carried onto adjacent public streets. • See, also, HAZ-PA1 below.	Contractor: - During construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the SWPPP monitoring program.	BMPs recorded in on-site SWPPP binder.
AIR-PA1 Asbestos dust could be released into the air during removal of sediment and/or dam materials at the upper reservoir and during placement at reuse/disposition sites.	• Use BMPs to avoid releasing dust from construction activities at the project sites and along public and private roads. Dam material will be used as fill only in locations where it can be capped with clean fill to limit the release of asbestos into the environment. • Secure BAAQMD Asbestos Dust Mitigation Plan approval and implement specific measures to limit asbestos dust from becoming airborne. • Materials placed off site at the lower reservoir will be secured in place with both 3 inches of nonasbestos-containing soil and vegetative planting above the expected average water level of the reservoir. • Do not transport dam material to Spring Mountain Vineyard or Clover Flat landfill.	Contractor: - During construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the SWPPP monitoring program.	BMPs recorded in on-site SWPPP binder.

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.3 Biological Resources				
BIO-PA1 Construction activities could result in temporary, short-term downstream sediment releases, kills of fish and other aquatic species, and destruction of habitat downstream of the dam.	• Complete work during summer low flow conditions between June 15 and October 15. • Develop detailed dewatering and species protection plan and ensure implementation by a NMFS and CDFG-approved biologist. • Perform work in isolation from the flowing stream by utilizing cofferdams or other approved water diversion structures constructed of a nonerodible material that does not contain soil or fine sediment. • Utilize erosion control measures throughout all phases of construction when potential sediment runoff from exposed slopes may enter water; do not allow silt-laden runoff to enter the stream or be directed to where it may enter the stream. • Dispose of dredged material at an alterative site with proper erosion control measures in place. • Provide monitoring by a qualified biologist throughout construction to ensure water quality standards are being met and sediment is not entering the watercourse. • Provide a preconstruction training session for construction crew members by the qualified biologist, including discussion of the sensitive biological resources within the project area, the potential impacts of accidental sediment releases, and project boundaries.	Dewatering: Contractor under supervision of project biologist. Preconstruction surveys, and training, and construction monitoring: Project biologist - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.

Impact	Mitigation	Who & When	Success Criteria	Reporting
BIO-PA2 Construction activities could result in direct disturbance, displacement, and/or mortality to aquatic species.	• Develop detailed dewatering and species protection plan and ensure implementation by a NMFS and CDFG-approved biologist. • Consult with CDFG and NMFS to obtain authorization to relocate aquatic species. • Perform work in isolation from the flowing stream by utilizing cofferdams or other approved water diversion structures constructed of a nonerodible material that does not contain soil or fine sediment; have qualified biologist on site during construction and decommissioning of water diversion structures. • Use approved screen pump intake with 3/32-inch screen mesh to divert water; see *Juvenile Fish Screen Criteria for Pump Intakes* (NMFS 1996). • Have qualified biologist relocate aquatic species prior to commencing project construction. • Provide preconstruction training session for construction crew members by the qualified biologist, including discussion of the sensitive aquatic resources, potential impacts of accidental sediment releases, and project boundaries. • Sweep site periodically during construction to ensure no aquatic species have moved into the area. • Provide monitoring by a qualified biologist throughout construction to ensure water quality standards are being met and sediment is not entering the watercourse. • Utilize erosion control measures throughout all phases of construction when potential sediment runoff from exposed slopes may enter water; do not allow silt-laden runoff to enter the stream or be directed to where it may enter the stream. • Dispose of dredged material at an alterative site with proper erosion control measures in place.	Dewatering: Contractor under supervision of project biologist. Preconstruction surveys, and training, and construction monitoring: Project biologist - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit post-construction report to ecological regulators, as required.

Impact	Mitigation	Who & When	Success Criteria	Reporting
BIO-PA3 Construction activities could result in incidental take of California red-legged frog (CRLF) and foothill yellow-legged frog (FYLF).	• Consult with USFWS and CDFG prior to project activities and obtain authorization to relocate frogs. Provide preconstruction survey re CRLF and FYLF prior to beginning work; limit work to areas that have been surveyed. • Provide crew members with training by a qualified biologist on the status, life history characteristics, and avoidance measures for CRLF and FYLF. • Complete earth work during the dry season, June 15 to October 15. • If CRLF and/or FYLF are encountered during construction, contact USFWS and CDFG for guidance, and/or relocate by a permitted biologist. • Qualified biologist will visit project area frequently during construction to ensure no CRLF, FYLF, or other species have entered the work area and are being impacted by construction activities.	Project biologist: - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.
BIO-PA4 Construction activities could result in direct disturbance, displacement, and/or mortality to special-status and common bat species.	• Perform survey of the site by qualified biologist for bat roosts prior to commencing work • Do not allow removal of occupied roost trees until the roost is unoccupied. • Provide crew members with training by a qualified biologist on the status, life history characteristics, and avoidance measures for bats. • Limit construction to daylight hours to avoid interference with the foraging abilities of bats.	Project biologist: - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.

Impact	Mitigation	Who & When	Success Criteria	Reporting
BIO-PA5 Construction activities could result in both direct and indirect impacts to nesting birds as a result of nest destruction, mortality, and/or disturbance.	• To the extent feasible, perform construction activities outside of the critical breeding period, mid-March to mid-August in the St. Helena area. • If activities must occur during the normal breeding season, qualified biologist to perform survey of work areas prior to commencing. • If active nests or behavior indicative of nesting birds are encountered, avoid those areas and have biologist designate a 50-foot buffer area for small songbirds and 250-foot buffer for larger birds (e.g., owls, raptors) to be avoided until the nests have been vacated. • Perform on-going monitoring for nesting activity within the project area. If state and/or federally listed species (e.g., northern spotted owl) are found breeding within the project site, halt activities and consult with USFWS and CDFG; follow conditions of agreements with regulators.	Project biologist: - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.
BIO-PA6 Construction activities could result in direct disturbance, displacement, and/or mortality to common terrestrial wildlife species (e.g., reptiles, amphibians, and mammals).	• Perform preconstruction survey by a qualified biologist (on the day preceding work and/or ahead of the construction crew) to ensure that no terrestrial species are occupying the site. • If terrestrial species are observed within the project site or immediate surroundings, avoid these areas until the animal(s) has (have) vacated the area, and/or the animal(s) have been relocated out of the project area by a qualified biologist. • Survey periodically during construction to ensure that no terrestrial species are being impacted by construction activities.	Project biologist: - Before and during construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.

Impact	Mitigation	Who & When	Success Criteria	Reporting
BIO-PA7 Disposal of materials at Lower York Creek Reservoir and alteration of Upper York Creek Reservoir could result in the disturbance/loss of jurisdictional wetlands and/or other waters of the U.S.	• If feasible, avoid filling or altering jurisdictional wetlands and/or other waters of the U.S. during construction. • If impacts on jurisdictional wetlands and/or other waters of the U.S. are unavoidable, obtain appropriate state and federal permits from the Corps Regulator Branch, RWQCB, and CDFG; adhere to conditions of these permits/agreements. • Mitigate for loss of jurisdictional wetlands and/or other waters of the U.S. at a ratio of 1:1 (or as agreed upon by the permitting agencies) within the project site or at a ratio of 2:1 (or as agreed upon by the permitting agencies) off site or for out-of-kind compensation (i.e., mitigation for impacts on one habitat type by creating, restoring, or enhancing another habitat type). • Implement 5-year monitoring program with applicable performance standards.	City: -- Before, during, and after construction	• ≥80% survival for restoration plantings. • Absence of invasive plant species. • Absence of erosion features. • Presence of a functioning, self-sustainable wetland system.	City or representative will submit post-construction reports to ecological regulators, as required.
BIO-PA8 Construction activities could result in damage and/or mortality to special-status plant species if present within the work area.	Spring 2007 plant surveys were performed by the project botanist. No special-status species were observed. See FEIR Appendix 8 for details.	N/A	No special status plants will be affected. No further mitigation is required.	N/A

Impact	Mitigation	Who & When	Success Criteria	Reporting
BIO-PA10 Native upland vegetation may be impacted during construction that may result in temporary loss of plant and wildlife habitat, increased erosion, and establishment of invasive nonnative plants, which could inhibit native vegetation establishment and result in a permanent loss of biotic functions and values.	• Minimize vegetation removal during all work activities to the maximum extent practicable. • Clearly flag grading limits to minimize disturbance from construction equipment. • Comply with the City's Tree Ordinance for any tree removal associated with project activities. • Replace upland native trees greater than 12 inches diameter at breast height that are removed as a result of project activities at a minimum 1:1 ratio with equivalent native species. • Obtain all propagules from local nursery stock, if available. • Revegetate all disturbed areas with native plantings and/or a native seed mix as soon as practicable to minimize erosion and recruitment of invasive nonnative plant species. • Implement BMPs to avoid dispersal of invasive, nonnative plants, including using only certified, weed-free materials dominated by native species for erosion control and revegetation.	Contractor: - During construction	Upland plantings will be considered successful if ≥80% of the woody plants installed are alive and healthy and have been in the ground for 3 years; replacement plantings will be required until success criterion is met.	City or representative will prepare and submit post-construction report and yearly monitoring reports to ecological regulators, as required, for five years.
BIO-PA11 Construction activities could result in direct disturbance, displacement, and/or mortality to northwestern pond turtle.	• Perform preconstruction survey prior to disturbance of the site to ensure that no pond turtles are present. . • If observed, avoid area until the animal(s) has (have) vacated the area, and/or the animal(s) have been relocated out of the project area by a qualified biologist. • If nests are encountered, avoid those areas plus a 25-foot buffer area until the nests have been vacated. • Survey site periodically during construction to ensure that no turtles are being impacted by construction activities.	Contractor: - During construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the biologist's post-construction report.	Project biologist will submit a post-construction report to ecological regulators, as required.

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.4 Cultural Resources				
CUL-PA1 Removal of Upper St. Helena Dam would adversely impact the historic feeling, design, and association of a resource considered eligible for listing in the National Register of Historic Places and the California Register of Historic Resources.	• Create a record of the dam using photo documentation, drawings, and written data. • Maintain documentation depicting dam prior to alteration. • Develop interpretive materials to be incorporated into the City's website within one year of the date of the dam modification. • Create an exhibit of photographs and graphics of the dam for installation in a public facility in St. Helena within one year of the date of the dam alteration.	City: - Information to be gathered prior to project construction. Public exhibit and website to be developed over the course of the following year.	The website and exhibit will be completed within one year of the dam modification.	N/A
2.5 Geology, Soils, Landslides, and Seismic Activity				
GEO-PA1 The project may expose people, property, or sensitive natural resources to potential substantial adverse effects involving slope failure due to removal of the dam at the toe of a landslide.	• Stabilize the landslide area by leaving existing spillway in place and buttressing it with fill from the project site. • Install two rows of soil anchors and subdrains. • Monitoring existing landslide area with equipment such as slope inclinometers and piezometers. • During construction, review slopes within the project site to evaluate slope stability, identify/evaluate possible slope defects, and determine if supplemental stabilization measures should be implemented.	Project engineer: - During construction	Mitigation measure is intrinsic to construction.	N/A
2.6 Global Climate Change				
GCC-PA1 Use of equipment will generate CO_2 during construction activities.	Plant approximately 564 trees in the riparian areas to sequester ±65 metric tons CO_2 equivalent by 2020.	Contractor: - During construction	See success criteria for wetland and upland plantings in AES-NP1 above.	City or representative will prepare and submit post-construction report and yearly monitoring reports to ecological regulators, as required, for five years.

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.7 Hazardous Materials				
HAZ-NP1 Potential exposure of construction workers to unsafe levels of hydrogen sulfide.	• Equip at least one worker in each construction area with personal H_2S monitoring device. • Follow OSHA protocols for safety of construction workers; if H_2S levels exceed safe levels, halt work until H_2S dissipates. • See, also, AIR-NP2 above.	Contractor: - During construction	Complete construction without any worker health incidents.	Contractor to furnish report to City within 48 hours documenting any health incident or post-construction checklist documenting lack of incidents.
HAZ-NP2 Potential for standard toxics associated with construction vehicle operation to be released into the environment.	BMPs for construction equipment and vehicles include: • Adhere to the SWPPP for the project site. • Maintain vehicles to prevent oil or other leaks; use off-site fueling stations and repair shops; keep vehicles and equipment clean; do not allow excessive build-up of oil and grease; place stockpiled spill cleanup materials where they are readily accessible; check incoming vehicles and equipment (including delivery trucks and employee and subcontractor vehicles) for leaking oil and fluids; do not allow leaking vehicles or equipment on-site; segregate and recycle wastes, such as greases, used oil or oil filters, antifreeze, cleaning solutions, automotive batteries, and hydraulic and transmission fluids; avoid mobile fueling of construction equipment; establish vehicle and equipment storage, cleaning, and maintenance areas in designated, confined areas, away from significant drainage courses and out of the riparian corridor. • If fueling must occur on site, designate areas away from drainage; locate fuel storage tanks over retention area designed to hold the total tank volume; do not "top-off" fuel tanks. • Direct concentrated storm water run-on/runoff around storage and service areas. Minimize contact of storm water and run-on/runoff with stored equipment by raising equipment on pallets or other similar devices.	Contractor: - During construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the SWPPP monitoring program.	BMPs recorded in on-site SWPPP binder.

Impact	Mitigation	Who & When	Success Criteria	Reporting
HAZ-PA1 Asbestos may become airborne through earthmoving activities during dam removal creating a health hazard for construction workers.	• Designate a "competent person" as defined in 29 CFR §1926.1101(b) to develop and oversee a monitoring program; if monitoring shows levels that exceed the Permissible Exposure Limits (PELS) or excursion limits (1.0 fiber/cm^3 over 30 minutes), competent person will require respiration equipment or other measures to protect workers' health. • Designate and demark a "regulated area" around dam removal site following recommendations of the competent person. Only authorized personnel may enter the regulated area. • Train construction workers regarding asbestos hazards, self-protection, and appropriate hygienic practices within the regulated area, including not eating, drinking, or smoking and wearing disposable coveralls. • Use water and other BMPs to prevent asbestos from becoming airborne; see Mitigation AIR-NP3 above for details. • Secure BAAQMD Asbestos Dust Mitigation Plan approval and implement specific measures to limit asbestos dust from becoming airborne. • Secure materials in place at LYCR with both 3 inches of nonasbestos-containing soil and vegetative planting above the expected average water level. • Do not transport dam material to Spring Mountain Vineyard or Clover Flat landfill.	Contractor: - During construction	Mitigation measures are intrinsic to construction. Documentation of measures will be included in the SWPPP monitoring program.	BMPs recorded in on-site SWPPP binder.

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.8 Hydrology and Hydraulics				
H&H-PA1 Sediment aggradation in lower York Creek may increase the frequency and severity of flooding.	Implement a channel monitoring and maintenance program that: • Surveys and documents changes in channel cross-sectional area on a regular, predetermined schedule. • Tracks sediment accumulation at bridges and other critical locations. • Provides guidance on acceptable levels of aggradation. • Establishes a mitigation plan should channel aggradation reach a point that it significantly impacts channel capacity and flooding by defining and prioritizing options for sediment management (e.g., sediment retention basins or in-stream sediment removal). • Defines and prioritizes options for sediment management. • Obtains permits required for sediment management practices.	City: - Channel monitoring program will be implemented within one year of dam removal.	To be developed as part of channel monitoring program.	N/A
H&H-PA2 Construction activities may lead to streambank erosion within the project site.	• Include soil anchors, rock bank protection to the 100-year flow elevation, and extensive riparian revegetation in restoration design. • See, also, BIO-NP4 above.	Contractor: - During construction	Mitigation measure is intrinsic to construction.	
2.9 Noise				
NOI-NP2 Construction activities would temporarily increase noise at the dam and disposal sites.	• Work during daylight hours. • Schedule construction activities efficiently to limit impacts on neighbors.	Contractor: - During construction	Mitigation measure is intrinsic to construction.	N/A

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.10 Traffic and Transportation				
TRA-NP1 Trucks entering and leaving Spring Mountain Road may cause traffic delays.	• Prepare a traffic control plan using standard Caltrans protocols that includes one or more flaggers for trucks entering and leaving the project sites. • Give emergency vehicles priority so that even when traffic is stopped, emergency vehicles will not be detained.	Contractor: - During construction	Mitigation measure is intrinsic to construction.	N/A
TRA-PA2 Project could generate peak hour traffic to roads that exceed acceptable level of service.	• Limit project truck traffic on Hwy. 29 through the City to between 9 am and 3 pm.	Contractor: - During construction	Mitigation measure is intrinsic to construction.	N/A

Impact	Mitigation	Who & When	Success Criteria	Reporting
2.11 Water Quality				
WQ-NP1 Water quality in York Creek could be degraded during project maintenance or construction activities and result in adverse effects to downstream aquatic life.	• Implement dewatering plan; see BIO-PA2 above. • Implement SWPPP that outlines erosion control and basic construction BMPs to limit the potential for sediment release. • Monitor upstream and downstream of the project area at least once daily during activities that could impact water quality. • Use Basin Plan standards guidelines unless more stringent permit conditions are imposed.	Contractor: - During construction	Water temperature standards: • Do not allow increase in water temperature of more than 5° F; if temperature increases more than 5° F, require changes to the dewatering pipe to allow cool water to move through the project site; keep pipe underground, out of direct sunlight, to keep temperatures cool. pH standards: • Maintain pH levels between 6.5 and 8.5 with no change greater than 0.5; if pH level change is greater than 0.5, require careful inspection of dewatering system to ensure water from the project area does not contaminate water in the pipe. Dissolved oxygen (DO) standards: • Maintain minimum level of 7.0 mg/L with measured levels of DO within 5% of the upstream level and no lower than 7.0 mg/L before leaving project area; if DO level in the bypass is lower than upstream conditions or 7.0 mg/L, agitate water to provide oxygenation prior to release downstream. Note that agitation could be inherent in the design of the pipe (corrugated pipe would naturally agitate water) or with the installation of an energy dissipater downstream.	BMPs recorded in on-site SWPPP binder.

Impact	Mitigation	Who & When	Success Criteria	Reporting
WQ-NP2 Project activities could result in increased turbidity that adversely affects water quality downstream of the dam to the Napa River.	• Implement SWPPP; see WQ-NP1 above. • Implement dewatering plan; see BIO-PA2 above. • Monitor turbidity upstream and downstream of the project site at least daily during all stages of project activities that may affect water quality to meet requirements set forth by the Basin Plan.	Contractor: - During construction	Basin Plan turbidity standards: • Turbidity not to exceed 10% of natural if background is over 50 NTU, or at a level that would cause a nuisance or adversely affect beneficial uses; if turbidity levels exceed 10%, halt work until turbidity is below 10%.	BMPs recorded in on-site SWPPP binder. Success of measure will be reported to RWQCB in post-construction report.
WQ-NP4 Hydrogen sulfide (H_2S) could be released into York Creek from an accidental spill during sediment removal.	• Monitor H_2S levels upstream and downstream of the construction area at least once daily during activities that may affect water quality to ensure levels do not exceed background; halt work if safe levels are exceeded and allow to return to background level. • See, also, AIR-NP2 and HAZ-NP1 above.	Contractor: - During construction	H_2S standards: • Maintain background levels upstream and downstream of project area.	Success of measure will be reported to RWQCB in post-construction report.

PRBO Conservation Science

Publication Brief

Release:
2 August 2012
www.prbo.org

Contact:
Nat Seavy

Phone:
(415) 868-0655 x311

E-mail:
nseavy@prbo.org

Developing a Riparian Bird Index to Communicate Restoration Success

There is a need to develop monitoring programs that clearly define restoration success and provide pathways to improve restoration and ecosystem performance from investments in restoration.

To address these challenges in the context of riparian restoration in coastal California, we used historical bird monitoring data from reference and restoration sites in Marin County to develop a Riparian Bird Index.

This index is essentially a species richness score for a given area that is weighted by the degree to which each species detected is associated with target riparian vegetation and can be converted into a simple "poor," "fair," "good," or "excellent" rating to communicate restoration success to a diverse audience.

The Riparian Bird Index is a simple yet biologically meaningful way to evaluate restoration performance and to communicate to a wide range of stakeholders. It can be used to initiate discussions among agency staff, biologists, restoration practitioners, and individual landowners on how to improve restoration performance.

This index provides a simple model for integrating bird monitoring with restoration that can be replicated in other regions.

Next steps.
For such a tool to reach its full potential, it will need to be easily accessible and widely available. One approach would be to develop the informatics tools that facilitate data entry, tracking, and data visualizations for this format. This strategy would take our rating system from a format that is assembled by hand, to a format where the same information is entered and assembled online to provide near real-time feedback to landowners and managers through an online progress report.

We are working to extend this concept to our work on private lands on the banks of the San Joaquin River in California's Great Valley.

Main Points

- Historical monitoring data can be used to develop indices that describe wildlife communities.
- The Riparian Bird Index helps multiple stakeholders agree on restoration goals.
- Stakeholders can use the Riparian Bird Index to measure progress toward those goals, and then use the information in an adaptive management framework if the goals are not being met.
- This index provides a simple model for integrating bird monitoring with restoration that can be replicated in other regions.

Paper citation:
N. E. Seavy, and T. Gardali. 2012. Developing a Riparian Bird Index to Communicate Restoration Success in Marin County, California. Ecological Restoration 30: 157-160. PRBO publication #1865.

Standard Operating Procedure (SOP) 4.2.1.4

Stream Photo Documentation Procedure

(CARCD 2001, Written by TAC Visual Assessments work group)

Introduction:

Photographs provide a qualitative, and potentially semi-quantitative, record of conditions in a watershed or on a water body. Photographs can be used to document general conditions on a reach of a stream during a stream walk, pollution events or other impacts, assess resource conditions over time, or can be used to document temporal progress for restoration efforts or other projects designed to benefit water quality. Photographic technology is available to anyone and it does not require a large degree of training or expensive equipment. Photos can be used in reports, presentations, or uploaded onto a computer website or GIS program. This approach is useful in providing a visual portrait of water resources to those who may never have the opportunity to actually visit a monitoring site.

Equipment:

Use the same camera to the extent possible for each photo throughout the duration of the project. Either 35 mm color or digital color cameras are recommended, accompanied by a telephoto lens. If you must change cameras during the program, replace the original camera with a similar one comparable in terms of media (digital vs. 35 mm) and other characteristics. A complete equipment list is suggested as follows:

Required:

- Camera and backup camera
- Folder with copies of previous photos (do not carry original photos in the field)
- Topographic and/or road map
- Aerial photos if available
- Compass
- Timepiece
- Extra film or digital disk capacity (whichever is applicable)
- Extra batteries for camera (if applicable)
- Photo-log data sheets or, alternatively, a bound notebook dedicated to the project
- Yellow photo sign form and black marker, or, alternatively, a small black board and chalk

Optional:

- GPS unit
- Stadia rod (for scale on landscape shots)
- Ruler (for scale on close up views of streams and vegetation)
- Steel fence posts for dedicating fixed photo points in the absence of available fixed landmarks

How to Access Aerial Photographs:

Aerial Photos can be obtained from the following federal agencies:

USGS Earth Science Information Center
507 National Center
12201 Sunrise Valley Drive
Reston, VA 22092
800-USA-MAPS

USDA Consolidated Farm Service Agencies
Aerial Photography Field Office
222 West 2300 South
P.O. Box 30010
Salt Lake City, UT 84103-0010
801-524-5856

Cartographic and Architectural Branch
National Archives and Records Administration
8601 Adelphi Road
College park, MD 20740-6001
301-713-7040

Roles and Duties of Team:

The team should be comprised of a minimum of two people, and preferably three people for restoration or other water quality improvement projects, as follows:

1. Primary Photographer
2. Subject, target for centering the photo and providing scale
3. Person responsible for determining geographic position and holding the photo sign forms or blackboard.

One of these people is also responsible for taking field notes to describe and record photos and photo points.

Safety Concerns:

Persons involved in photo monitoring should **ALWAYS** put safety first. For safety reasons, always have at least two 2 volunteers for the survey. Make sure that the area(s) you are surveying either are accessible to the public or that you have obtained permission from the landowner prior to the survey.

Some safety concerns that may be encountered during the survey include, but are not limited to:

- Inclement weather
- Flood conditions, fast flowing water, or very cold water
- Poisonous plants (e.g.: poison oak)
- Dangerous insects and animals (e.g.: bees, rattlesnakes, range animals such as cattle, etc.)
- Harmful or hazardous trash (e.g.: broken glass, hypodermic needles, human feces)

We recommend that the volunteer coordinator or leader discuss the potential hazards with all volunteers prior to any fieldwork.

General Instructions:

From the inception of any photo documentation project until it is completed, always take each photo from the same position (photo point), and at the same bearing and vertical angle at that photo point. Photo point positions should be thoroughly documented, including photographs taken of the photo point. Refer to copies of previous photos when arriving at the photo point. Try to maintain a level (horizontal) camera view unless the terrain is sloped. (If the photo can not be horizontal due to the slope, then record the angle for that photo.) When photo points are first being selected, consider the type of project (meadow or stream restoration, vegetation management for fire control, ambient or event monitoring as part of a stream walk, etc.) and refer to the guidance listed on *Suggestions for Photo Points by Type of Project*.

When taking photographs, try to include landscape features that are unlikely to change over several years (buildings, other structures, and landscape features such as peaks, rock outcrops, large trees, etc.) so that repeat photos will be easy to position. Lighting is, of course, a key ingredient so give consideration to the angle of light, cloud cover, background, shadows, and contrasts. Close view photographs taken from the north (i.e., facing south) will minimize shadows. Medium and long view photos are best shot with the sun at the photographer's back. Some artistic expression is encouraged as some photos may be used on websites and in slide shows (early morning and late evening shots may be useful for this purpose). Seasonal changes can be used to advantage as foliage, stream flow, cloud cover, and site access fluctuate. It is often important to include a ruler, stadia rod, person, farm animal, or automobile in photos to convey the scale of the image. Of particular concern is the angle from which the photo is taken. Oftentimes an overhead or elevated shot from a bridge, cliff, peak, tree, etc. will be instrumental in conveying the full dimensions of the project. Of most importance overall, however, is being aware of the goal(s) of the project and capturing images that clearly demonstrate progress towards achieving those goal(s). Again, reference to *Suggestions for Photo Points by Type of Project* may be helpful.

If possible, try to include a black board or yellow photo sign in the view, marked at a minimum with the location, subject, time and date of the photograph. A blank photo sign form is included in this document.

Recording Information:

Use a systematic method of recording information about each project, photo point, and photo. The following information should be entered on the photo-log forms (blank form included in this document) or in a dedicated notebook:

- Project or group name, and contract number (if applicable, e.g., for funded restoration projects)
- General location (stream, beach, city, etc.), and short narrative description of project's habitat type, goals, etc.
- Photographer and other team members
- Photo number
- Date
- Time (for each photograph)
- Photo point information, including:
    - Name or other unique identifier (abbreviated name and/or ID number)
    - Narrative description of location including proximity to and direction from notable landscape features like roads, fence lines, creeks, rock outcrops, large trees, buildings, previous photo points, etc. – sufficient for future photographers who have never visited the project to locate the photo point
    - Latitude, longitude, and altitude from map or GPS unit
- Magnetic compass bearing from the photo point to the subject
- Specific information about the subject of the photo
- Optional additional information: a true compass bearing (corrected for declination) from photo point to subject, time of sunrise and sunset (check newspaper or almanac), and cloud cover.

For ambient monitoring, the stream and shore walk form should be attached or referenced in the photo-log.

When monitoring the implementation of restoration, fuel reduction, or Best Management Practices (BMP) projects, include or attach to the photo-log a narrative description of observable progress in achieving the goals of the project. Provide supplementary information along with the photo, such as noticeable changes in habitat, wildlife, and water quality and quantity.

Archive all photos, along with the associated photo-log information, in a protected environment.

The Photo Point: Establishing Position of Photographer:

1. Have available a variety of methods for establishing position: maps, aerial photos, GPS, permanent markers and landmarks, etc. If the primary method fails (e.g., a GPS or lost

marker post) then have an alternate method (map, aerial photo, copy of an original photograph of the photo-point, etc).

2. Select an existing structure or landmark (mailbox, telephone pole, benchmark, large rock, etc.), identify its latitude and longitude, and choose (and record for future use) the permanent position of the photographer relative to that landmark. Alternatively, choose the procedure described in *Monitoring California's Annual Rangeland Vegetation* (UC/DANR Leaflet 21486, Dec. 1990). This procedure involves placing a permanently marked steel fence post to establish the position of the photographer.

3. For restoration, fuel reduction, and BMP projects, photograph the photo-points and carry copies of those photographs on subsequent field visits.

Determining the Compass Bearing:

1. Select and record the permanent magnetic bearing of the photo center view. You can also record the true compass bearing (corrected for declination) but do not substitute this for the magnetic bearing. Include a prominent landmark in a set position within the view. If possible, have an assistant stand at a fixed distance from both the photographer and the center of the view, holding a stadia rod if available, within the view of the camera; preferably position the stadia rod on one established, consistent side of the view for each photo (right or left side).
2. Alternatively, use the procedure described in *Monitoring California's Annual Rangeland Vegetation* (UC/DANR Leaflet 21486, Dec. 1990). This procedure involves placing a permanently marked steel fence post to establish the position of the focal point (photo center).
3. When performing ambient or event photo monitoring, and when a compass is not available, then refer to a map and record the approximate bearing as north, south, east or west.

Suggestions for Photo Points by Type of Project:

Ambient or Event Monitoring, Including Photography Associated with Narrative Visual Assessments:

1. When first beginning an ambient monitoring program take representative long and/or medium view photos of stream reaches and segments of shoreline being monitored. Show the positions of these photos on a map, preferably on the stream/shore walk form. Subjects to be photographed include a representative view of the stream or shore condition at the beginning and ending positions of the segment being monitored, storm drain outfalls, confluence of tributaries, structures (e.g., bridges, dams, pipelines, etc.).

2. If possible, take a close view photograph of the substrate (streambed), algae, or submerged aquatic vegetation.

3. Time series: Photographs of these subjects at the same photo points should be repeated annually during the same season or month if possible.

4. Event monitoring refers to any unusual or sporadic conditions encountered during a stream or shore walk, such as trash dumps, turbidity events, oil spills, etc. Photograph and record information on your photo-log and on your Stream and Shore Walk Visual Assessment form. Report pollution events to the Regional Board. Report trash dumps to local authorities.

All Restoration and Fuel Reduction Projects – Time Series:

Take photos immediately before and after construction, planting, or vegetation removal. Long term monitoring should allow for at least annual photography for a minimum of three years after the project, and thereafter at 5 years and ten years.

Meadow Restoration:

1. Aerial view (satellite or airplane photography) if available.

2. In the absence of an aerial view, a landscape, long view showing an overlapping sequence of photos illustrating a long reach of stream and meadow (satellite photos, or hill close by, fly-over, etc.)

3. Long view up or down the longitudinal dimension of the creek showing riparian vegetation growth bounded on each side by grasses, sedges, or whatever that is lower in height

4. Long view of conversion of sage and other upland species back to meadow vegetation

5. Long view and medium view of streambed changes (straightened back to meandering, sediment back to gravel, etc.)

6. Medium and close views of structures, plantings, etc. intended to induce these changes

Stream Restoration/stabilization:

1. Aerial view (satellite or airplane photography) if available.

2. In the absence of an aerial view, a landscape, long-view showing all or representative sections of the project (bluff, bridge, etc.)

3. Long view up or down the stream (from stream level) showing changes in the stream bank, vegetation, etc.

4. Long view and medium view of streambed changes (thalweg, gravel, meanders, etc.)

5. Medium and close views of structures, plantings, etc. intended to induce these changes.

6. Optional: Use a tape set perpendicular across the stream channel at fixed points and include this tape in your photos described in 3 and 4 above. For specific procedures refer to Harrelson, Cheryl C., C.L. Rawlins, and John P. Potyondy, *Stream Channel Reference Sites: An Illustrated Guide to Field Techniques*, United States Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RM-245.

Vegetation Management for Fire Prevention ("fuel reduction"):

1. Aerial view (satellite or airplane photography) if available.

2. In the absence of an aerial view, a landscape, long view showing all or representative sections of the project (bluff, bridge, etc.)

3. Long view (wide angle if possible) showing the project area or areas. Preferably these long views should be from an elevated vantage point.

4. Medium view photos showing examples of vegetation changes, and plantings if included in the project. It is recommended that a person (preferably holding a stadia rod) be included in the view for scale

5. To the extent possible include medium and long view photos that include adjacent stream channels.

Stream Sediment Load or Erosion Monitoring:

1. Long views from bridge or other elevated position.

2. Medium views of bars and banks, with a person (preferably holding a stadia rod) in view for scale.

3. Close views of streambed with ruler or other common object in the view for scale.

4. Time series: Photograph during the dry season (low flow) once per year or after a significant flood event when streambed is visible. The flood events may be episodic in the south and seasonal in the north.

5. Optional: Use a tape set perpendicular across the stream channel at fixed points and include this tape in your photos described in 1 and 2 above. For specific procedures refer to Harrelson, Cheryl C., C.L. Rawlins, and John P. Potyondy, *Stream Channel Reference Sites: An Illustrated Guide to Field Techniques*, United States Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RM-245.

PHOTO- LOG FORM

Project:

Location:

Date:

Photographer:

Team members:

Photo #	Time	Photo Point ID	Photo Pt. Description & Location	Bearing to Subject	Subject Description

General Notes or Comments (weather, cloud cover, time of sunrise and sunset, other pertinent information):

PHOTO SIGN FORM: Print this form on yellow paper. Complete the following information in black marker for each photograph. Include in the photographic view so that it will be legible in the finished photo.

Location:

Subject Description:

Date:

Time:

Tomales Bay Watershed Water Quality Monitoring Plan

Program goal: Based on the information gathered through this monitoring program, the Council will work to identify water quality problems, to develop solutions to these problems, and to provide support to realize these solutions by working with partners and landowners in the watershed to improve and protect water quality.

I. Introduction

One of three goals established in the *Tomales Bay Watershed Stewardship Plan: A Framework For Action* (TBWC, 2003) is to ensure water quality in Tomales Bay and tributary streams sufficient to support natural resources and beneficial uses. The Tomales Bay Watershed Council identified several steps to achieve this goal including the initial action of developing a water quality monitoring plan for Tomales Bay and tributary streams that will allow us to develop long-term trends and identify source areas of concern for water quality management in the bay and its tributary streams.

The water quality monitoring program will be implemented by the Tomales Bay Watershed Council (Council) and its watershed partners. To meet the technical needs associated with defining the design, monitoring and data management aspects of this program, the Council has created a Water Quality Technical Advisory Committee (WQ TAC). The WQ TAC will continue to oversee the development and implementation of this program to ensure that data is collected in efficient, effective and appropriate ways in order to meet the program objectives.

Implementation of this program by the Council members will be in accordance to their respective statutory mandates, charters, and resource availability. To that end, and in order to foster commitment between partners and the spirit of collaboration and coordination that will be necessary, the Tomales Bay Watershed Council's Water Quality Committee will develop and ratify an agreement that describes the role of key local organizations, agencies and the Council in the development and implementation of this water quality monitoring program (see Section III, Task 4). Core partners and committee members include: California Department of Health Services; S.F. Bay Regional Water Quality Control Board; Marin County Environmental Health Services; U.C. Cooperative Extension; Salmon Protection and Watershed Network; Tomales Bay Agricultural Group; Tomales Bay Shellfish Technical Advisory Committee; Gulf of the Farallones National Marine Sanctuary; California Department of Fish and Game; Pacific Coast Science and Learning Center; Point Reyes National Seashore; Point Reyes National Seashore Association; California State Parks; Marin Municipal Water District; Inverness Public Utility District; ranchers; private landowners; shellfish growers; and local non-governmental organizations.

In this plan, water quality is defined as the chemical, physical, and biological characteristics of surface waters (or these characteristics within the water column). Water quality is often described with respect to a specific activity or a legally recognized "beneficial use" like water contact recreation, water supply for residential or agricultural purposes, mariculture, estuarine and wildlife habitats, wetlands, and others. In providing this definition, it is important to remember that the focus on surface water is a starting point with the intended hope that additional monitoring interests can be coordinated to expand the program in the future.

It is the desire of the Council to provide needed water quality information that will assist individuals, organizations and agencies that are responsible for and/or advocating for water quality protection and improvement within the Tomales Bay watershed. The information collected through this program will ultimately be used to increase our collective understanding about the benefits of specific efforts to improve water quality, and our ability to effectively and adaptively manage human impacts on water quality. Data sensitivity is a significant concern amongst both public agencies and various stakeholder groups, and the appropriate use of data, data limitations, etc. will be defined prior to the collection and/or dissemination of any program data. Private property rights will be recognized, statutory responsibilities will be maintained, and voluntary cooperation will be encouraged and protected with data sensitivity considerations.

II. Plan Objectives

This plan provides direction for a water quality monitoring program with an initial 10-year timeframe. It is envisioned, however, that the design will include monitoring parameters and a sampling regime that can be carried out indefinitely. The plan and program objectives are to:

1) Provide the watershed community with the required data and analysis to determine improving, constant, or declining trends in bay and tributary water quality;
2) Form and maintain a clearinghouse of water quality data and monitoring activities that facilitates effective and efficient use of limited resources;
3) Serve as source of information that will direct and promote actions to improve water quality; and
4) Provide an understanding of source areas and categories for constituents of concern in the bay and on a sub-watershed and/or tributary scale.

III. Questions to be addressed by this monitoring program:

1) What are the natural ranges and the storm, seasonal and annual variability in water quality parameters in the Bay and its tributaries?
2) At what locations do parameters fall outside the natural range and to what duration and extent?
3) What are the pollutant loadings from controllable and uncontrollable sources and in the watershed, and how do the Bay and tributaries relate in this regard?
4) What are the trends in the levels, fate and transport of pollutants in the watershed and the Bay, and how do the Bay and tributaries relate in these regards?
5) How effective are actions to reduce pollutant loads?

IV. Program Components

A. Long Term Trend Monitoring

Trend monitoring will generate water quality data of sufficient duration and representation to assess long-term shifts in water quality within Tomales Bay and it tributaries. There are numerous stakeholder efforts to manage sources of pollution for which feedback is needed to assess impacts and the effectiveness of restoration efforts. There are also regulatory and statutory needs for long-term trend water quality monitoring; these include the Pathogen TMDL (SFBRWQCB, 2002), and the Shellfish Lease monitoring by California Department of Health Services. Similarly, water quality monitoring results have created the basis for regulatory attention as illustrated by the County of Marin's health advisories at popular beach and swimming areas during 2002-03. This component of the monitoring program will give the watershed community the needed benchmarks to determine the success of management efforts and efficacy of regulatory policies.

1. Current efforts

There currently is no long-term comprehensive water quality monitoring at a watershed level. Numerous stakeholders and regulatory agencies have conducted comprehensive monitoring for short duration or for a limited focus on water quality parameters or geographic boundaries. Examples of this sort of water quality monitoring would be the National Park Service's continuing monitoring of Olema Creek, and the Shellfish Technical Advisory Committee's two-year pathogen study on Tomales Bay and its tributaries.

The recent but now terminated efforts of California Department of Fish and Game to monitor ammonia concentrations in Stemple Creek and Tomales Bay watersheds is a relevant example of the type of effort and intended use of the data to be generated for this program. In that case, what was initially perceived as a regulatory threat, quickly evolved into a management tool, and eventually became documentation that the management of pollution sources was having a beneficial impact on water quality.

2. Lead

The lead organization for long-term trend monitoring will be the Tomales Bay Watershed Council. To the greatest extent possible, the Council's efforts will be coordinated with the on-going monitoring programs of its partners.

3. Subtasks

i. Parameters

Water quality samples collected in Tomales Bay and tributary creeks will be analyzed for fecal coliform, transparency, turbidity, conductivity/salinity, pH, dissolved oxygen, ammonia, and temperature. In addition to these water quality parameters or "response variables", descriptive or "explanatory variables" will be collected. These will include tidal stage, discharge, cumulative precipitation, and possibly others. Analytical methods will follow accepted procedures such as those outlined in the Standard Methods for the Examination of Water and Wastewater (Clesceri et al., 1998). If there is sufficient funding, samples will also be analyzed

for additional parameters including suspended sediment concentration, E.coli and Enteroccocus.

ii. Frequency and duration

Trend sampling shall be conducted on a weekly basis, and in the future the data will be evaluated to determine if seasonally based monitoring can be used to reduce the sampling frequency. This level of sample collection will afford documentation of seasonal and annual changes in water quality. It will also allow for the development of geometric means as required under the San Francisco Bay Basin Plan, which includes Tomales Bay (CRWQCB, 1995). It will not necessarily provide documentation of variability due to storm conditions. This program component will be carried for an initial ten years with the anticipation that it will be continued for an additional 20 years if not longer.

iii. Sampling locations

a. **Tomales Bay sampling locations:** Sampling locations will include four bay sites to represent the tidal conditions and separation of the Bay (Fisher et al., 1996). These sites will be re-evaluated to ensure that variability within the Bay, from east to west and north to south, is captured. These sites will be coordinated with the existing monitoring sites in the bay that are used by the Department of Health Services, Regional Water Quality Control Board, and others. The WQ TAC will consider the forthcoming UC Berkeley hydrodynamic model of the bay, and opportunities to identify long-term monitoring sites that may be useful boundary conditions for the model development and operation. [T. Hollibough sites to be considered here]

b. **Watershed sampling locations:** Nine permanent tributary sampling locations will be established and will include one per sub-watershed. These sites will be coordinated with the four existing gauging stations (mainstem Lagunitas, Olema, Walker, and San Geronimo); Chileno Creek; 1 site on the east shore of Tomales Bay (Millerton Creek); and 1 site on the west shore (1st or 2nd Valley Creek). East and west shore sites will include at least one reference stream (milepost 36.17 on the east shore of the Bay and another on the west to be determined) that flows through sub-watersheds with minimal human land use activities to represent the two dominant geologic formations that comprise the majority of the Tomales Bay watershed.

iv. Statistical Analysis

The data generated from this effort will have a high level of variability. The descriptive variables will be used to normalize concentration results or to calculate flux and load for a given parameter. These steps will allow for comparison of results across the different locations. Admittedly, the comparison of tributary and bay locations requires additional normalization because of the simultaneous influences of discharge and tides. These data will also be valuable as boundary conditions to calibrate and test the UC Berkeley Tomales Bay hydrodynamic model.

Analysis of trends will be conducted graphically and through time series analysis. Graphical analysis will include the representation of concentration, flux, and load values as a function of time. These graphics will provide anecdotal indications of

water quality trends including seasonal and annual fluctuations. Time series analysis for upward or downward trends in concentration, flux, and load will be conducted according to Helsel and Hirsch (1995) or other suitable and accepted methods (Hahn, ___; Hirsch et al., 1991; and Helsel, 1987). This will include nonparametric statistical methods including data transformation to account for lack of normal distribution in the data.

B. Source Area Monitoring

Source area monitoring efforts will be focused on identifying sources and quantities of water pollutants to Tomales Bay and its freshwater tributaries. While trend monitoring is dependant on long-term sampling at a suite of permanent sampling sites, source area monitoring is both flexible and responsive based on the data collected. The intent of source area monitoring will be to support and prioritize future watershed or sub-watershed water quality improvement efforts, and to document conditions in order to evaluate the effectiveness of past efforts to improve water quality on private and public lands.

Sampling sites will be determined based upon the results of previous sampling. Source area monitoring will be initially concentrated at the sub-watershed scale, and as needed in the Bay. Additional sampling will be conducted where sources are identified, while sampling will be curtailed in areas where conditions were of less concern. These monitoring activities will compliment the long-term monitoring program to document general trends within the watershed and the Bay. Differentiating source areas and land use issues would be valuable for the long-term management of water quality in the Bay and tributaries.

1. Current Efforts

Point Reyes National Seashore currently conducts a source area monitoring program within the boundaries of the Seashore. The intent is to document loading from the various park watersheds, in order to concentrate long-term management and restoration efforts into areas where the watershed will benefit the most. Through this program, the Seashore has identified source areas and management or structural practices intended to reduce pollutant loading to aquatic systems.

The Salmon Protection and Watershed Network (SPAWN) is initiating a water quality monitoring program that will document conditions and source areas within the San Geronimo watershed. Their efforts, funded through the State 319(h) grant program will investigate additional water quality parameters, and may be used to determine if other parameters should be investigated in the long-run.

The community of Marshall has recently conducted a voluntary septic monitoring program via a partnership between the East Shore Planning Group (ESPG) and the County of Marin. Based on the monitoring results, the ESPG and the County have subsequently developed substantial grant funds for septic improvements in Marshall. Currently, the community and an engineering team are exploring options for septic improvements, feasibility and cost benefits. These studies will allow Marshall and Marin County to collaborate further on

decisions regarding the most cost effective use of grants and homeowner finances for septic system improvements and local community septic management.

2. Lead(s)

The lead organization for the source area monitoring program would be the Tomales Bay Watershed Council. These efforts would be coordinated with ongoing monitoring efforts in the watershed including the Tomales Bay Agricultural Group, SPAWN, Point Reyes National Seashore, Regional Water Quality Control Board, and others.

3. Subtasks

i. Identify monitoring locations

The intent of this program is to conduct monitoring at publicly accessible locations, for the purpose of understanding pollutant loading to the bay from different geographic areas. Various organizations through multiple monitoring efforts have conducted water quality sampling at more than 50 locations within the Tomales Bay watershed. Initial sites for source area monitoring will be determined from the results of the 1995-96 and 2000-2001 Shellfish TAC sampling programs. After reviewing the existing data, priority areas will be identified for the initial source area monitoring efforts, and these sites will be grouped according to the primary watersheds (see list below in *iv. Statistical data analysis*). Site sampling will be conducted by sub-watershed to ensure that similar conditions are sampled within each sub-watershed, and that the results are comparable.

ii. Parameters

The monitoring program includes collection of field sampled parameters and laboratory analyzed parameters. The water quality parameters to be sampled will be coordinated with the long-term monitoring program. In the initial years of this study, the water quality monitoring constituents will include field collected parameters (e.g. temperature, DO, conductivity, pH, etc.) as well as analytical lab analysis (e.g. fecal coliform, ammonia, etc.). Additional parameters may be added as we learn more about the system and the issues in the Tomales Bay watershed.

The source area monitoring program will target winter and spring runoff events when nonpoint source pollution loading is most prevalent. This would include two to three storm events per winter season at all sites to capture nonpoint source loading information, and additional sampling upstream of sites that show high levels of loading. In addition, a summer sampling event will also be conducted to document seasonal differences in the parameters. Stream discharge should also be measured in conjunction with water quality sampling. In this manner, loading may be estimated for different watersheds and subwatersheds. Because discharge measurements are often time consuming, installation of staff plates and estimates of flow may be deemed appropriate.

iii. Site reconnaissance and documentation

Once sites are selected, they will be spatially documented using GPS, photographs, and a narrative description of site access. Field sampling protocols are well established and will vary based upon the actual equipment used in the monitoring effort. Laboratory analysis is standardized by analytical method, but program sampling protocols and handling procedures will be similar to those identified in the long-term monitoring plan.

iv. Statistical data analysis and interpretation

Data analysis related to the source area monitoring program should provide information that may be used to identify priority watersheds. Comparative sample concentrations and loading estimates for the various sampling location should be evaluated to determine watersheds and sub-watersheds where further sampling will be implemented. TBWC will coordinate data analysis and interpretation with the intent of using information to focus watershed outreach and management efforts.

a. Primary sub- watershed and Bay groupings (which will be used to compare data and to focus management efforts):

Lagunitas Creek
San Geronimo Creek
Olema Creek
West Side Bay Tributaries
East Side Bay Tributaries
Walker Creek
Chileno Creek
Keyes Creek
Sites within the Bay to be determined

v. Watershed outreach and coordination

Results of the source area sampling program will be used to advise outreach and prioritization of water quality management measures supported by partners that are working on the ground to address water quality issues on public and private lands (e.g. the Marin RCD, SPAWN, PRNS, NRCS, Marin County, and others).

C. Quality Assurance (QA) and Quality Control (QC) for All Water Quality Data Collected

All QA/QC procedures will be performed pursuant to the State Water Resources Control Board's Quality Assurance Project Plan (QAPP) for the Surface Water Ambient Monitoring Program (SWAMP). These procedures are outlined in the California State Water Resources Control Board, Division of Water Quality, December 2002: *Quality Assurance Management Plan for the State of California's Surface Water Ambient Monitoring Program, 1st Version*. This includes sample collection and handling, as well as sample analysis. The QA/QC plan for implementation of this program will be included in the implementation protocol.

D. Documentation of Baseline Conditions and Formation of a Database

Research, collect and compile reliable baseline data describing the concentrations of contaminants in the waters of Tomales Bay and tributary streams. Develop a database on water quality that is compatible with data being generated in ongoing studies of Tomales Bay and tributary streams; which will include past and current water quality information, and the identification of gaps that need to be filled. The construction of a water quality monitoring database for the entire watershed, and the capacity to analyze data and to

develop trends, will benefit the agencies and organizations that are currently collecting data, and those responsible for tracking and protecting water quality.

1. Current Efforts

More than a dozen agencies and/or organizations are currently collecting or planning to collect water quality data on Tomales Bay and within the watershed (see attached table). These efforts to characterize baseline water quality are hampered by the lack of a single database and analysis protocol for all WQ data currently collected on the bay, its watersheds and sub-watersheds. Thus, a central database is needed for data storage and comprehensive analysis. The TBWC has already contacted and gained agreement from most of the groups collecting water quality. Each has agreed to contribute electronic copies of their data to a central database coordinated by the TBWC.

2. Lead(s)

The Pacific Coast Science and Learning Center (PCSLC) at Point Reyes National Seashore (PRNS) will provide WQ database development, data acquisition, and cursory spatial and temporal analysis of trends. The Point Reyes National Seashore Association (or another group identified by TBWC) will act as the fiscal agent.

3. Subtasks

Develop database, collect data from existing programs, summarize baseline data, and produce reports. This series of events should take 6 months. The final report will include graphs and cursory statistical analysis of trends (over time and grouped by season) of all WQ variables by site. Data will also be presented to aid analysis of whether additional data collection sites and frequencies are required to fully describe the water quality patterns in the watershed. Metadata including QA/QC will be maintained on all data sources as well as the database as a whole.

The following steps will be taken to develop a database fully compatible with historic and existing water quality data collection and analyses:

1. Database manager hired to work with PCSLC database staff.
2. Data obtained from WQ data collecting groups.
3. NPS WQ database template modified to incorporate types of data from donors.
4. Donor data imported into database.
5. Publication of a summary report of available water quality data including graphs, tables, etc.

E. Database Maintenance and Management

Maintain and regularly update the water quality database. Provide a clearinghouse for use by landowners, stakeholders, regulatory agencies, watershed managers and the general public through such media as the Internet (web page), published reports, and readily available, computer searchable database. Inclusion and dissemination of water quality data will respect data sensitivity defined by the partners and respective monitoring programs that have developed this data.

1. Current Efforts

No unified database exists for all of Tomales Bay watershed. This section addresses the maintenance of the database that will be constructed as an initial step in this program.

2. Lead(s)

The Pacific Coast Science and Learning Center (PCSLC) at Point Reyes National Seashore (PRNS) will provide WQ database development, data acquisition, and cursory spatial and temporal analysis of trends. PRNSA (or another group identified by TBWC) will act as the fiscal agent.

3. Subtasks

Data from partner groups collecting data will be transferred to the Tomales Bay WQ database quarterly with summary reports and trends for all variable reported on the web and electronic (PDF) copies to all interested parties. Analyses and reports will be as in Task 1 above. Individual requests for particular analyses will be performed by the database manager or, if more complex, data will be provided to a WQ specialist of the Council's choosing.

The WQ database will be maintained in Microsoft Access and reside at the PCSLC with weekly backups to local server and CD-ROMs stored off site. The WQ database will be linked to a GIS of Tomales Bay's habitats and species (in development at PCSLC) to investigate spatial relationships between WQ, habitats, species, and land use.

F. Information dissemination and outreach

Information that is gathered through the monitoring program will be used to increase the community's understanding of water quality problems in the watershed including the impacts of septic systems, recreation, agriculture, sewage ponds, the landfill, mines, and other human activities. Additionally, this program will support coordination and cooperation amongst the various partners that are collecting water quality data to ensure timely and effective information dissemination about water quality monitoring results to agencies, organizations, bay users and communities in the Tomales Bay watershed via electronic posting of results, regularly published reports, press releases, etc.

1. Current efforts

Current efforts to disseminate information about water quality in the Tomales Bay watershed include:

- Written reports compiled independently by agencies and organizations to summarize their data and findings on an irregular basis (e.g. as projects end, as annual reports are due, as funding cycles require, etc.);
- Posting of Marin County recreational water quality monitoring at http://california.earth911.org/usa/WaterQuality/default.asp?beach_id=888&cluster=6041 on a regular basis;
- Water contact advisories that are posted by Marin County or State Parks when State Water Quality Standards are exceeded at locally popular swimming holes and beaches in the watershed;

- Press releases/media coverage issued by agencies and organizations; and
- Water quality summaries/brief mention in newsletters that are produced and distributed locally.

2. Lead

The Tomales Bay Watershed Council will be the lead organization to provide a clearinghouse for coordinating and disseminating information from the Tomales Bay Watershed Water Quality Monitoring Program via its Water Quality and Outreach committees. The Council's partners will provide outreach support by bringing this information to their constituencies, staff and/or members.

G. Estimated Annual Program Budget

The estimated annual program budget is $154,000 per year, not including the in-kind contributions of agencies, local organizations and volunteers.

<u>To date, in-kind/cash contributions provided during 2003:</u>

1. Volunteer time ($25-35/hr. for program planning, meetings, etc.) $4,500
2. Technical support ($75-100/hr. for plan development, meetings, etc.) $4,000
3. Coordinator time ($40/hr. plan development, meetings, etc.) $10,000

<u>Future in-kind contributions on an annual basis are estimated at:</u>

1. Database: $4,000 for use of existing program and assistance to add water quality data
2. Use of the Pacific Coast Learning Center office ($500/mo.): $6,000 /year
3. Oversight by data manager: $10,000 during the first 6-12 months of the project.
   The water quality data manager will be supervised and receive technical assistance from the Tomales Bay Biodiversity Inventory Data Manager, the Director of the PCSLC, and NPS WQ monitoring coordinator. NPS will also provide a WQ database template used for the park that will be easily modified for Tomales Bay watershed data.
4. Technical support: Staff from National Parks, Regional Water Quality Control Board, University of California Cooperative Extension, and possibly others will provide technical support through the Water Quality Technical Advisory Committee on an as-needed basis during the start-up phase of this program (years 1-2).
5. Volunteer time: volunteers will participate in this program on an as-needed basis.
6. Data analysis and trend development: Researchers from the University of California and potentially other academic institutions will be asked to assist with data analysis and trend development.

<u>Start-up costs</u>

Certain capital expenditures will be necessary to initiate the water quality monitoring program. These include:

1. Database intern (40% time) to create the database during first 6 months: $20,000
2. 2 sets of field sampling equipment: $5,000

1. Program manager

One full-time employee will be necessary to oversee and implementation of this water quality program. This individual will:

- manage the water quality program,
- collect field samples,
- enter new data into the database,
- write water quality reports on a regular basis (annual summaries, with larger reports analyzing program data and developing management recommendations every 3 years),
- post or link water quality data electronically to the TBWC website, and
- facilitate outreach and education to disseminate water quality information about the Tomales Bay watershed.

This person will collect both long-term trend and source area samples, as well as remotely accessible data that is available for gauging stations, etc. During winter months when sampling frequency will be increased to capture storm events, a part-time assistant will be hired and volunteers will potentially be used to assist the program manager with sample collection. During the summer when sampling frequency is decreased, the program manager will work on the data compilation, report writing, and other related tasks. Intensive statistical analyses would require additional funding and expertise.

Annual Program Budget Detail

Description	*No. units*	*Per unit cost*	*Total*
Personnel 1. Program manager/field sampler (1 FTE) 2. Seasonal sampling assistant (wet season) 3. Publications	 1 1 1	 $50,000/yr. $15,000/yr. $2,500/yr.	 $50,000 $15,000 $2,500
Sampling and equipment 1. *Long term trend monitoring* at 9 stations, weekly sampling (52/year) for: a. Bacteria, ammonia, nitrate, etc. b. Total suspended solids 2. *Source area monitoring* a. Analysis for source area samples b. Equipment maintenance	 470 470 100 1	 $80/sample $5/sample $50-100/sample $1,00/year	 $37,600 $2,350 $15,000 $1,000
Mileage/travel for field sampling	13,800	$0.36/mile	$5,000
Subtotal			$128,450
Project coordination and administrative overhead (20%)			$25,690
Total			***$154,140***

References

San Francisco Bay Regional Water Quality Control Board (SFBRWQCB). 2002. *Total Maximum Daily Load for Pathogens in Tomales Bay: Preliminary Project Report.*

Clesceri, L.S., A.E. Greenberg, and A.D. Eaton. Editors. 1998. *Standard methods for the examination of water and wastewater.* 20th Edition. APHA, AWWA, and WEF.

CRWQCB (California Regional Water Quality Control Board). 1995. San Francisco Bay Basin (Region 2) Water Quality Control Plan.

Fisher, D.T.; Smith, S.V.; Churchill, R.R. 1996. *Simulation of a century of runoff across the Tomales Watershed, Marin County, California.* Journal of Hydrology. 186:253-73.

Hahn, _____

Helsel, D.R. 1987. *Advantages of nonparametric procedures for analysis of water quality data.* Hydrologic Sciences Journal, V32(N2):179-190.

Helsel, D.R. and R.M. Hirsch. 1995. *Statistical methods in water resources.* Elsevier, Amsterdam, Neatherlands. 529 pgs.

Hirsch, R.M., R.B. Alexander, and R.A. Smith. 1991. *Selection of methods for detection and estimation of trends in water quality.* Water Resource Research, V27(N5):803-813.

I. A Coordinated Resource Management Plan For The Tomales Bay Watershed Water Quality Monitoring Program

The following memorandum of understanding specifies the role of each participating organization and agency and will be ratified by all participants in the program to foster collaboration and support for the long-term implementation of this monitoring program. This MOU is incorporated into a coordinate resource management plan below.

1. Purpose and Need

The undersigned are stakeholders in the Tomales Bay Watershed and are also signatories to the *Tomales Bay Watershed Stewardship Plan: A Framework for Action* (the "Plan"). The Plan envisions as its most urgent action the development of a coordinated and comprehensive water quality monitoring plan for Tomales Bay and its tributary streams, which will document baseline conditions of water quality and provide information needed to evaluate the efficacy of projects and programs to improve water quality in the Watershed. This document is intended to focus on the program by which the water quality measurements will be made, the data recorded, maintained and utilized to evaluate and design programs for the improvement of the water quality of Tomales Bay and the restoration of the environmental integrity of the entire watershed. To accomplish this we hereby enter into this coordinated resource management plan ("CRMP") to work together to develop a long-term water quality monitoring program ("Program"), and the tools for the analysis of sources and trends in water quality.

2. Objectives

1. A summary of current monitoring activities in the Tomales Bay Watershed is presented in Table A attached hereto. This summary includes an indication of what agencies are currently doing to determine water quality in the CRMP area, and future monitoring plans. The participants herein agree to transmit to the Tomales Bay Watershed Council (the "Council") all such data collected by the participants to the date hereof and all data collected after the date of this agreement.
2. The participants agree to collaborate to identify gaps in the available data with respect to the type of information obtained, the location at which the data is obtained and the time and date of such measurement in order to develop a comprehensive baseline view of the water quality of the Tomales Bay Watershed.
3. The participants will assist in the creation and maintenance of a database where the measurement results will be recorded and stored.
4. The participants will cooperate in the development of analytical tools for the evaluation of water quality trends in the Watershed.

3. Date of Adoption

[To be added]

4. Location and Description of Planning Area:
Tomales Bay and its watershed in Marin County, California

Please see the *Tomales Bay Watershed Stewardship Plan: A Framework for Action* (2003) for additional background information.

5. Natural Resources and Resource Uses in the Planning Area:
Please see the *Tomales Bay Watershed Stewardship Plan: A Framework for Action* (2003) for additional background information.

6. Resource Management Issues:
The CRMP area is not in a natural state with large portions having been modified by human activities. Sedimentation from tributary streams has reduced the area of the bay over the last 200 years. Salmon habitat has been diminished to less than half of the original range, and remaining habitat has been compromised by human activities over the last century. During recent years, water quality monitoring has resulted in the posting of human health advisories for the bay and tributaries for water contact, and the consumption of seven species of sport fishes regularly caught in Tomales Bay. Human activities that have affected the watershed include: residential development, agriculture, forestry, impoundment of water and the creation of large-scale reservoirs, mining, recreation, road construction, septic and waste disposal, and shellfish harvesting.

Please see the *Tomales Bay Watershed Stewardship Plan: A Framework for Action* (2003) for additional background information.

7. CRMP Strategy:
The Tomales Bay Watershed Council and its partners in this CRMP will implement the Water Quality Monitoring Program. Implementation by CRMP participants will be in accordance with their respective statutory mandates, charters and resource availability. The participants by execution of this agreement agree to collaborate in the Program and to support its development and implementation. As part of that collaboration the participants agree to consider the compatibility of their current and planned monitoring activities, and to cooperate in the coordination of those activities as part of a comprehensive monitoring program.

Additionally, the participants agree to coordinate water quality monitoring and data collection amongst the other partners in this program. The participants also agree to provide data to update the database on a regular basis.

The participants agree to work together to define the financial needs of the water quality program and to secure financial commitments (e.g. annual financial support, laboratory analysis, technical support, etc.)

… more detail to come as roles are defined.

8. Participants:

We, the undersigned, have participated in the development of the Tomales Bay Watershed Water Quality Monitoring Program Coordinated Resource Management Plan, concur with the Plan, and will act as outlined herein to implement the Plan to the best of our ability.

Meetings for the review of this Program by the participants will be scheduled and organized. The purpose of the meetings will be to discuss problems, make revisions and adjustments for implementation of planned activities.

Marin County
By: _____
Title: _____

National Park Service – Point Reyes National Seashore
By: _____
Title: _____

California State Parks
By: _____
Title: _____

California Department of Health Services
By: _____
Title: _____

Marin County Department Of Health Services
By: _____
Title: _____

San Francisco Bay Area Regional Water Quality Control Board
By: _____
Title: _____

Marin Municipal Water District
By: _____
Title: _____

North Marin Water District
By: _____
Title: _____

Inverness Public Utilities District
By: _____
Title: _____

California Department of Fish and Game
By: _____
Title: _____

University of California Cooperative Extension
By: _____
Title: _____

SPAWN
By: _____
Title: _____

Appendix A: Summary of water monitoring in the Tomales Bay watershed

Technical Resources

- Conservation Planning
- Ecological Science
- Natural Resources Assessment
- Data, Maps & Analysis
- Tools & Applications
- Technical References
- Field Office Technical Guide (FOTG)
- Engineering
- Economics
- Alphabetical Listing

Go to Your State's FOTG

Field Office Technical Guide (FOTG)

What is FOTG?

Technical guides are the primary scientific references for NRCS. They contain technical information about the conservation of soil, water, air, and related plant and animal resources.

Technical guides used in each field office are localized so that they apply specifically to the geographic area for which they are prepared. These documents are referred to as Field Office Technical Guides (FOTGs).

Appropriate parts of the Field Office Technical Guides are automated as data bases, computer programs, and other electronic-based materials such as those included in these web based pages.

What information is located in FOTG

Section I — General References
In this section you will find general state maps, descriptions of Major Land Resource Areas, watershed information, and links to NRCS reference manuals and handbooks. Section I contains links to researchers, universities, and agencies we work. Section I also contains conservation practice costs, agricultural laws and regulations, cultural resources, and information about protected plant and animal species.

Section II — Soil and Site Information
In this section you will find detailed information about soil, water, air, plant, and animal resources. NRCS Soil Surveys, Hydric Soils Interpretations, Ecological Site Descriptions, Forage Suitability Groups, Cropland Production Tables, Wildlife Habitat Evaluation Guides, Water Quality Guides, and other related information can be found here as it becomes available.

Section III — Conservation Management Systems
In this section you will find information on NRCS Quality Criteria, which establish standards for resource conditions that help provide sustained use.

Section IV — Practice Standards and Specifications
In this section you will find the NRCS Conservation Practices. Practice Standards define the practice and where it applies. Practice specifications are detailed requirements for installing the practice in the state.

Section V — Conservation Effects
In this section you will find background information on how Conservation Practices affect each identified resource concerns in the state.

Go to Your State's FOTG

United States
Environmental Protection
Agency

Office Of Water
(4503F)

EPA 841-B-96-004
September 1997

Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls

EPA/841-B-96-004
September 1997

MONITORING GUIDANCE FOR DETERMINING THE EFFECTIVENESS OF NONPOINT SOURCE CONTROLS

FINAL

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

September 1997

CONTENTS

PAGE

List of Tables v

List of Figures vii

Credits ix

Glossary xi

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

2. Developing a Monitoring Plan
   2.1 Introduction 2-1
   2.2 Monitoring Objectives 2-4
      2.2.1 Monitoring Objective Category: Problem Definition 2-5
      2.2.2 Monitoring Objective Category: Model Development 2-6
      2.2.3 Monitoring Objective Category: Evaluation 2-6
      2.2.4 Monitoring Objective Category: Conduct Research 2-6
   2.3 Data Analysis and Presentation Plans 2-7
   2.4 Variable Selection 2-7
      2.4.1 Physical and Chemical Water Quality Data 2-10
      2.4.2 Biological Data 2-10
      2.4.3 Precipitation Data 2-10
      2.4.4 Land Use Data 2-11
      2.4.5 Topographic Data 2-12

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

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

4. Data Analysis
4.1 Introduction 4-1
4.1.1 Estimation and Hypothesis Testing 4-1
4.1.2 Characteristics of Environmental Data 4-4
4.1.3 Recommendations for Selecting Statistical Methods 4-4
4.1.4 Data Stratification 4-8
4.1.5 Recommended Reading List and Available Software 4-9
4.2 Summary (Descriptive) Statistics 4-10
4.2.1 Point Estimation 4-10
4.2.2 Interval Estimation 4-18
4.3 Graphical Data Display 4-20
4.4 Evaluation of Test Assumptions 4-25
4.4.1 Tests of Normality 4-26
4.4.2 Tests of Equal Variance 4-31
4.4.3 Tests of Randomness 4-33

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

5. Quality Assurance and Quality Control
5.1 Introduction 5-1
5.1.1 Definitions of Quality Assurance and Quality Control 5-1
5.1.2 Importance of QA/QC Programs 5-1
5.1.3 EPA Quality Policy 5-2
5.2 Data Quality Objectives 5-3
5.2.1 The Data Quality Objectives Process 5-4
5.2.2 Data Quality Objectives and the QA/QC Program 5-7
5.3 Elements of a Quality Assurance Project Plan 5-8
5.3.1 Group A: Project Management 5-8
5.3.2 Group B: Measurements and Acquisition 5-14
5.3.3 Group C: Assessment/Oversight 5-17
5.3.4 Group D: Data Validation and Usability 5-18
5.4 Field Operations 5-19
5.4.1 Field Design 5-19
5.4.2 Sampling Site Selection 5-21
5.4.3 Sampling Equipment 5-22
5.4.4 Sample Collection 5-22

5.4.5 Sample Handling and Transport 5-23
5.4.6 Safety and Training 5-23
5.5 Laboratory Operations 5-24
5.5.1 General Laboratory QA and QC 5-24
5.5.2 Instrumentation and Materials for Laboratory Operations 5-24
5.5.3 Analytical Methods 5-26
5.5.4 Method Validation 5-26
5.5.5 Training and Safety 5-26
5.5.6 Procedural Checks and Audits 5-26
5.6 Data and Reports 5-26
5.6.1 Generation of New Data 5-27
5.6.2 Uses of Historical Data 5-27
5.6.3 Documentation and Record Keeping 5-28
5.6.4 Report Preparation 5-28

References R-1

Index I-1

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

LIST OF TABLES

TABLE	TITLE	PAGE
1-1	Sources of nonpoint source pollution and their contribution to the impairment of water quality in the United States	1-5
2-1	General characteristics of monitoring types	2-4
2-2	Applications of six sampling designs to estimate means and totals	2-21
3-1	General strengths and limitations of biological monitoring and assessment	3-5
3-2	Five tiers of the rapid bioassessment protocols	3-18
3-3	Scoring criteria for the core metrics as determined by the 25th percentile of the metric values from the Middle Rockies-Central Ecoregion, Wyoming	3-23
3-4	Scoring criteria for the metrics as determined by the 25th percentile of the metric values for the two aggregated subecoregions for Florida streams	3-24
3-5	Comparison of probabilistic and targeted monitoring designs	3-28
3-6	Waterbody stratification hierarchy	3-30
3-7	Summary of the primary technical issues related to biological monitoring for nonpoint source evaluations	3-39
3-8	Selected biomonitoring program components, Delaware DNREC	3-40
3-9	Selected biomonitoring program components, Florida DEP	3-41
3-10	Selected biomonitoring program components, Montana DHES	3-42
3-11	Selected biomonitoring program components, North Dakota DEH	3-43
3-12	Selected biomonitoring program components, Vermont DEC	3-44
3-13	Fish IBI metrics used in various regions of North America	3-46
3-14	Examples of metric suites used for analysis of macroinvertebrate assemblages	3-49
4-1	Errors in hypothesis testing	4-2
4-2	Methods for characterizing data	4-5
4-3	Methods for routine data analysis	4-6
4-4	Total nitrogen runoff concentrations for a single storm event in Florida	4-13
4-5	Total nitrogen runoff concentrations for a single storm event in Florida and example calculations for the EMC	4-14
4-6	Raw data by time period	4-15
4-7	Loadings rate data	4-15
4-8	Calculation of plotting position for the sulfate data from Station 16 in Figure 4-8	4-27
4-9	Table of skewness test for normality for sample sizes less than 150	4-29
4-10	Selected summary statistics for the sulfate data from Station 16 in Figure 4-8...	4-29
4-11	Values of kurtosis test for normality for small sample sizes	4-30
4-12	Example analysis of the Sharpiro-Wilk *W* test using the sulfate data from Station 16 in Figure 4-8	4-32
4-13	Highland Silver Lake TSS data for site 1	4-38
4-14	Evaluation of power using the postimplementation TSS data	4-43
4-15	Nonparametric evaluation of postimplementation data using the Wilcoxon Signed Ranks test	4-45
4-16	Sign test for comparing paired BOD_5 concentrations	4-46
4-17	Summary of parametric tests used to evaluate difference between means	4-49

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

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

LIST OF FIGURES

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

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

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

2. DEVELOPING A MONITORING PLAN

2.1 INTRODUCTION

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

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

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

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

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

Once the monitoring goals have been established, existing data and constraints should be considered. A thorough review of literature pertaining to water quality studies previously conducted in the geographic region of interest should be completed before starting a new study. The review should help determine whether existing data provide sufficient information to address the monitoring goals and what data gaps exist.

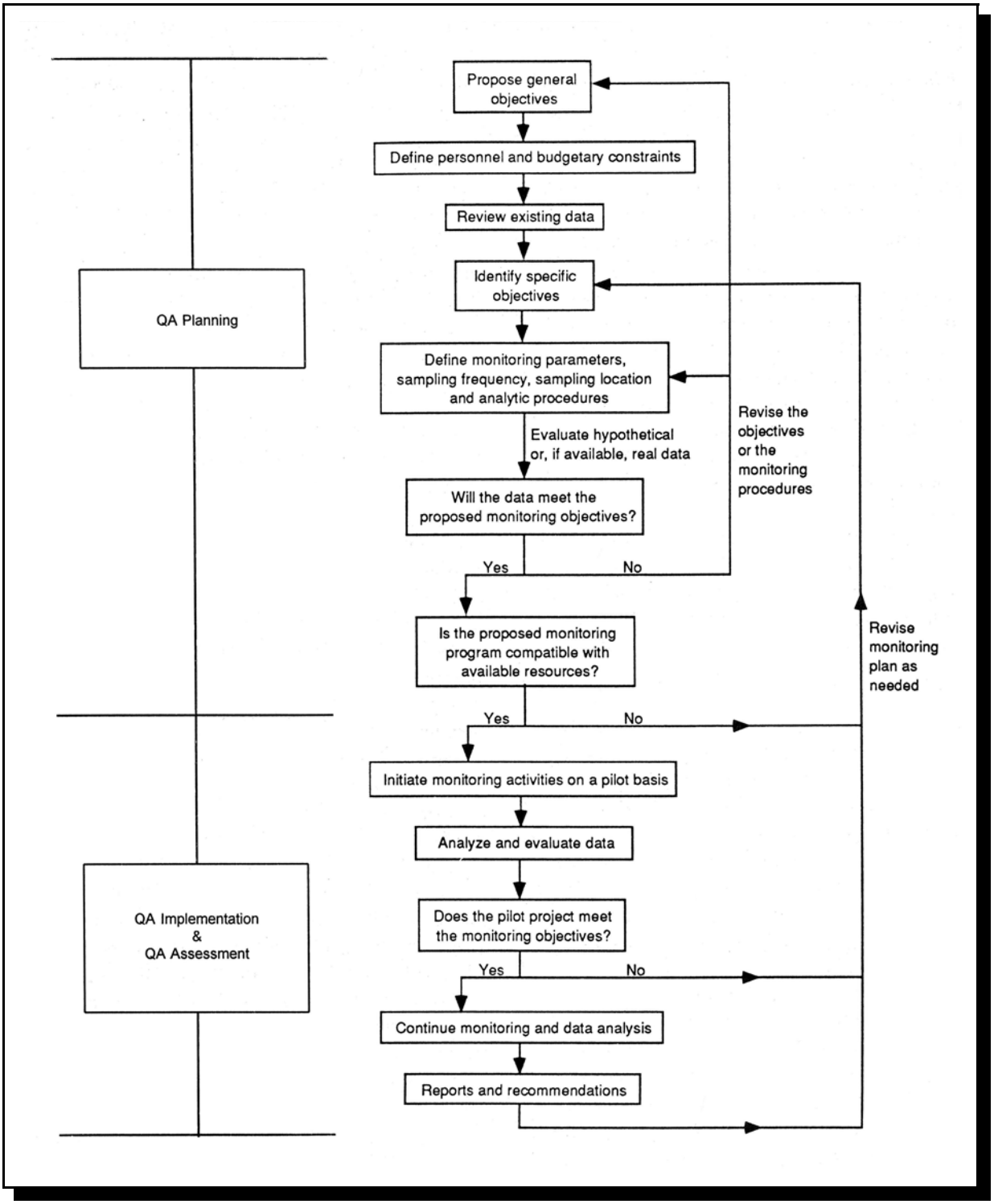


Figure 2-1. Development of a monitoring project (after MacDonald et al., 1991).

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

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

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

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

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

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

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

Table 2-1. General characteristics of monitoring types .

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

Source: MacDonald et al., 1991.

2.2 MONITORING OBJECTIVES

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

Monitoring programs can be implemented for one or many reasons. The more common types of monitoring program objectives are summarized below. The emphasis of this guidance is on evaluation monitoring, but information contained herein might also be used to address other types of monitoring. The reader is cautioned that even though two different monitoring programs might share some objective listed below, their designs can be radically different.

2.2.1 Monitoring Objective Category: Problem Definition

(1) Determine whether an impairment exists

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

(2) Determine the extent of the impairment

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

(3) Determine the causes and sources of impairment

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

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

2.2.2 Monitoring Objective Category: Model Development

(1) Calibrate models

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

(2) Validate models

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

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

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

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

(2) Analyze trends

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

2.2.4 Monitoring Objective Category: Conduct Research

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

2.3 Data Analysis and Presentation Plans

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

- Perform a thorough review of the legal basis for the management effort and define the resulting “implications” for monitoring.

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

- Review the ability of the monitoring program to supply information.

- Formulate an information expectations report for the monitoring system.

- Present the information expectations report to all users of the information.

- Develop consensus as to an agreeable formulation of information expectations and related monitoring system design criteria.

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

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

2.4 VARIABLE SELECTION

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

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

Expectations Report Outline

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

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

- Ability of Monitoring Systems to Produce Water Quality Information
  - Narrative Information
  - Numerical Information-Data
  - Graphical Information
  - Statistical Information
  - Average Conditions
  - Changing Conditions
  - Extreme Conditions
  - Water Quality Indices

- Suggested Information Expectations for Monitoring System
  - Management Information Goal(s)
  - Definition of Water Quality
  - Monitoring System Goal(s)
  - Information Product of Monitoring System
  - Narrative
  - Graphical
  - Statistical

- Resulting Monitoring Network Design Criteria
  - Variable Selection
  - Site Selection
  - Sampling Frequency Determination

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

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

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

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

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

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

2.4.1 Physical and Chemical Water Quality Data

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

2.4.2 Biological Data

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

2.4.3 Precipitation Data

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

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

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

2.4.4 Land Use Data

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

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

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

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

2.4.5 Topographic Data

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

2.4.6 Soil Characteristics Data

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

2.5 PROGRAM DESIGN

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

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

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

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

Simply identifying the site location and sampling frequency is not sufficient to describe the where and when of sampling programs. Additional considerations include the depth of sampling, the origins of the aliquot(s) taken in each sample bottle, the time frame over which measurements are made, and others. For example, if a stream is well mixed, a single grab sample from the center of the stream might be sufficient, whereas it might be more appropriate to take an integrated sample from a wider stream. In deeper estuaries, it is a common practice to collect samples near the top and bottom of the waterbody as well as just above and just below the pycnocline. Frequency of sampling should be based on several factors (Sherwani and Moreau, 1975):

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

- Variability and types of the inputs
- Magnitude of response

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

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

Other types of sampling uncertainty include random sampling errors and gross errors. Random sampling errors arise from the variability of population units (Gilbert, 1987) and explain why the sample means from two surveys are never equal. Gross mistakes can occur at any point in the process beginning with sample collection and ending with the reporting of study results. Adherence to accepted sampling and laboratory protocols combined with thorough quality control and data screening procedures and experience, dedication, and care will minimize the chances for gross errors.

2.5.1 Probabilistic Designs

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

Simple random sampling

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

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

where

n = number of samples,
t = Student's t value,
s = sample standard deviation,
d = absolute margin of error,
N = number of population units, and
α = confidence interval.

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

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

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

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

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

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

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

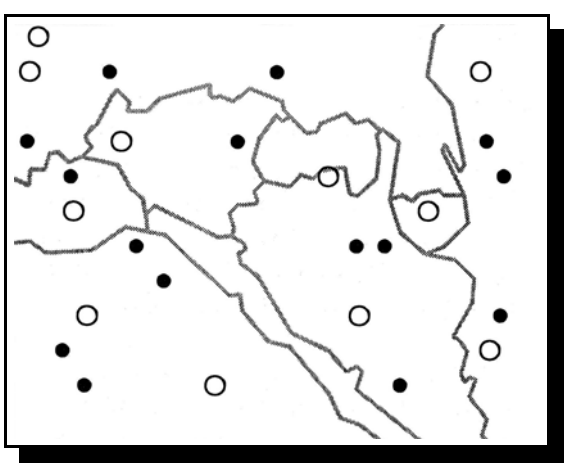


Figure 2-3. Simple random sampling for silviculture. Dots represent harvest sites. All harvest sites of interest are represented on the map, and the sites to be sampled (open dots—○) were selected randomly from all harvest sites on the map. The shaded lines on the map could represent county, watershed,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

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

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

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

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

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

Gilbert (1987) and Cochran (1977) address additional aspects of simple random sampling. Included in these texts are estimation of the mean and total for sampling with and without replacement, equations for determining the number of samples required for both independent and correlated data, and the impact of measure-ment errors. In most cases, environmental sampling is done without replacement (e.g., aliquots of stream water are not placed back into the stream), *N* is relatively large, the samples are assumed to be independent, and measurement error is ignored, thus making many of these specialized cases less critical. However, the reader should be aware that these issues might become paramount depending on the monitoring objectives and sampling design.

Stratified Random Sampling

In stratified random sampling, the target population is divided into groups called strata for the purpose of obtaining a better estimate of the mean or total for the entire population (Figure 2-4). Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables (e.g., season, flow condition) to group observations into more units that reduce

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

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

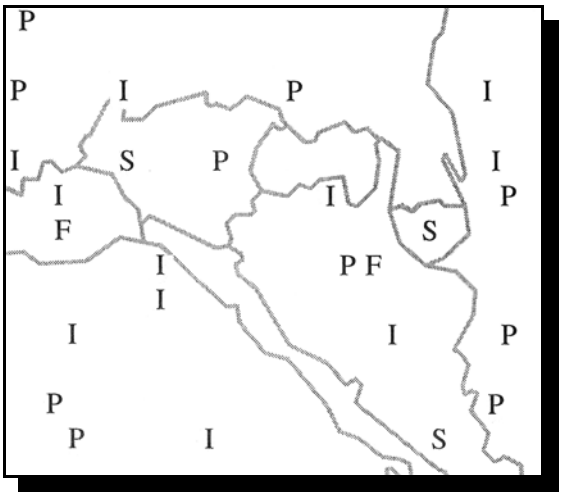


Figure 2-4. Stratified random sampling for silviculture. Letters represent harvest sites, subdivided by type of ownership (P = private nonindustrial, I = industrial, F = federal, S = state). All harvest sites of interest are represented on the map. From all of the sites in one ownership category, sites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

the variability of observations within each unit. As an example, stratified random sampling can be used to evaluate chemical concentrations in waterbodies when evaluating nonpoint source loadings. One approach would be to stratify stream flow into base and various storm flow periods to account for the energy relationship between precipitation and pollutant generation. Random sampling would then be performed in each stratum.

Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a linear trend, followed in order by systematic sampling (discussed later) and simple random sampling. He also states that stratification normally results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample.
In a stratified random sampling program when N, the number of population units, is large, the optimum number of samples can be estimated with (Cochran, 1977)

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

where

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

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

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

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

cost per population unit, c_h, is different among the strata:

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

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

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

$$\overline{x_{st}} = \sum_{h=1}^{L} W_h \overline{x_h} \qquad (2\text{-}7)$$

Systematic Sampling

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

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

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

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

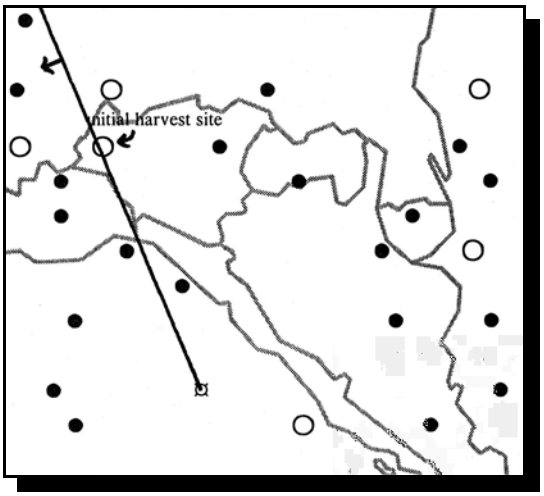


Figure 2-5. Systematic sampling for silviculture. Dots (● and ○) represent harvest sites of interest. A single point on the map (¤) and one of the harvest sites were randomly selected. A line was stretched outward from the point to (and beyond) the selected harvest site. The line was then rotated about the map and every fifth dot that it touched was selected for sampling (open dots—○). The direction of rotation was determined prior to selection of the point of the line's origin and the beginning harvest site. The shaded lines on the map could represent county

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

Cluster Sampling

Cluster sampling is applied in cases where it is more practical to measure randomly selected groups of individual units than to measure randomly selected individual units (Gilbert, 1987). In cluster sampling, the total population is divided into a number of relatively small subdivisions, or clusters, and then some of these subdivisions are randomly selected for sampling (Figure 2-6). For one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is then performed within each cluster (Gaugush, 1986). An example of one-stage cluster sampling is the collection of all macroinvertebrates on randomly selected rocks within a specified sampling area. The stream bottom might contain hundreds of rocks with thousands of organisms attached to them, thus making it difficult to sample the organisms as individual units. However, it is often possible to randomly select rocks and then inspect every organism on each selected rock.

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

Two-stage Sampling

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

Two-stage sampling might also include systematic sampling within a randomly selected subset of the population primary units. For example, if the target population is the average annual pollutant concentration in a stream, the primary units could be daily average concentrations (n = 365). A subset of these daily concentrations (e.g., n = 24) could be selected at random for further systematic sampling of hourly concentrations. For example if four systematic, hourly samples could be taken on each of 24 different days, with the hour for the first sample determined randomly, followed by three more hourly samples taken every sixth hour, 96 hourly composite samples would be available for the calculation of the population mean and variance.

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

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

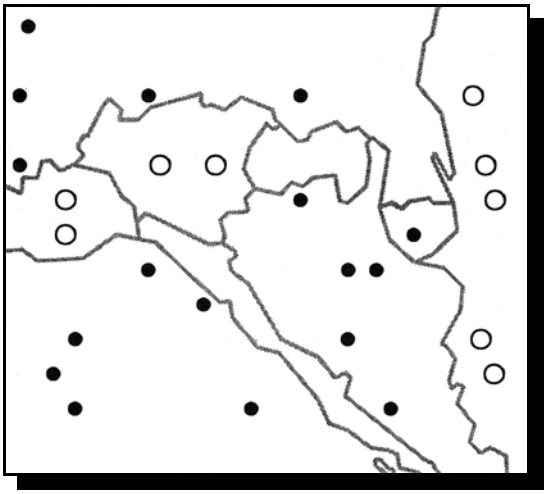


Figure 2-6. Cluster sampling for silviculture. All harvest sites in the area of interest are represented on the map (closed {●} and open {○} dots). The shaded lines on the map represent county boundaries. Some of the counties were randomly selected, and all harvest sites within those counties (open dots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

Double Sampling

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

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

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

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

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

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

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

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

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

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

2.5.2 Targeted Site Location Study Designs

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

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

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

Source: After Gilbert, 1987.

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

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

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

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

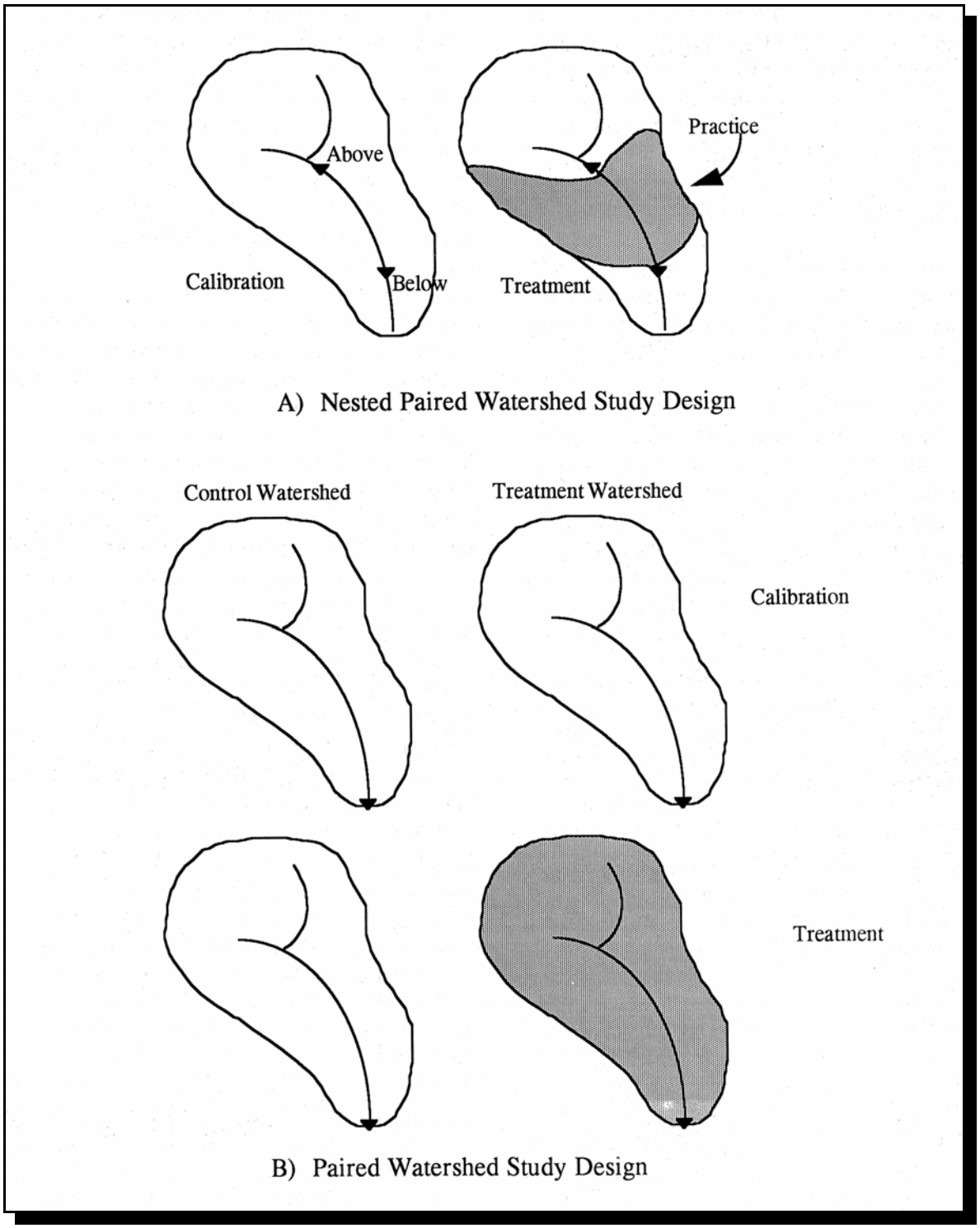


Figure 2-7. Nested paired and paired watershed study designs.

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

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

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

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

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

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

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

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

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

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

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

2.6 Example Program Design

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

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

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

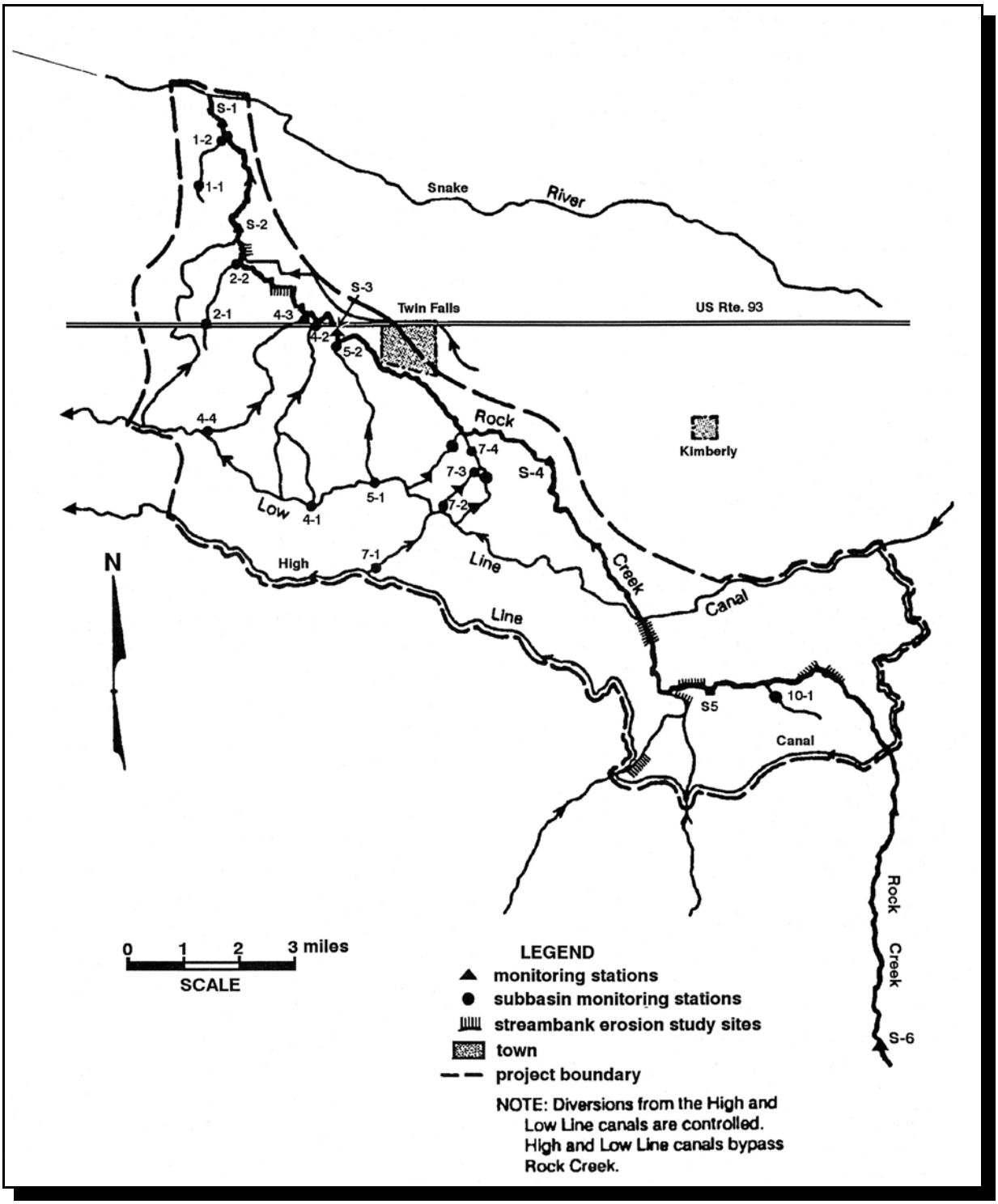


Figure 2-8. Map of the Rock Creek Rural Clean Water Program study area, Twin Falls County, Idaho. (Source: Clark, 1986)

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

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

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

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

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

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

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

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

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

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

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

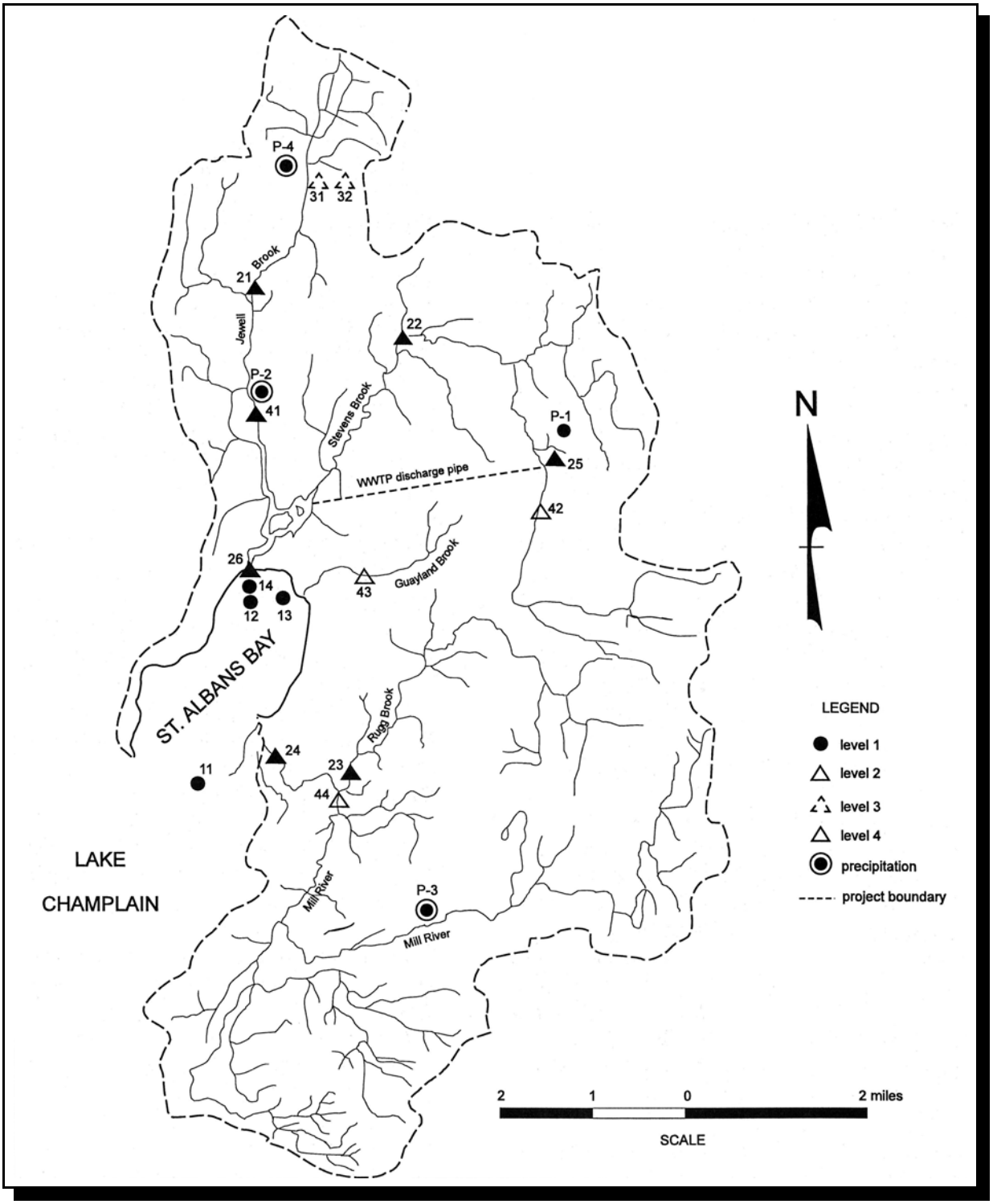


Figure 2-9. St. Albans Bay watershed, Franklin County, Vermont, sampling locations. (Source: Vermont RCWP Coordinating Committee, 1986)

2.7 Roles and Responsibilities

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

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

2.8 Quality Assurance Project Planning

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

2.9 Chemical and Physical Monitoring

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

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

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

  Typically, a grab sample is a sample taken at one place a single time. Care should be taken to make sure that a grab sample is representative. If there is spatial variability (e.g., across a stream, at different depths in a lake) or

temporal variability (e.g., during a storm event) it might be more appropriate to take a composite or time-integrated sample rather than a grab sample.

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

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

Continuous sampling requires electronic measuring devices and is therefore limited to variables that are amenable to this type of sampling, such as dissolved oxygen, conductivity, pH, and salinity. It is generally not suitable for measurements of metals, organics, or pesticides. Continuous sampling is typically used for research and fate and transport studies.

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

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

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

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

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

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

2.10 RECOMMENDED REFERENCES

Important monitoring references that should be consulted include the following:

American Public Health Administration. 1995. *Standard methods for the examination of water*

and wastewater. 19th ed. American Public Health Association, Washington, DC.

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

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

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

Clark, W.H. 1990. *Coordinated nonpoint source water quality monitoring program for Idaho.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, ID. January.

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

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

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

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

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

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

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

4. DATA ANALYSIS

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

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

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

4.1 INTRODUCTION

4.1.1 Estimation and Hypothesis Testing

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

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

Evaluating the effectiveness of controls and changing environmental conditions is one of the key monitoring program objectives described in Chapter 2. In addition to summarizing key statistics that describe the central tendency and spread of water quality variables and biological

metrics, statistical analysis usually involves hypothesis testing. Two common types of hypothesis testing done in environmental monitoring are step changes and monotonic trends. Step changes are typically evaluated when comparing at least two different sample populations such as an impacted site and a reference site or when comparing one sample population to an action level. Step changes can also be evaluated when comparing samples collected during different time periods. Monotonic trends (e.g., consistently increasing or decreasing concentrations) are typically evaluated when the analyst is investigating long-term gradual changes over time.

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

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

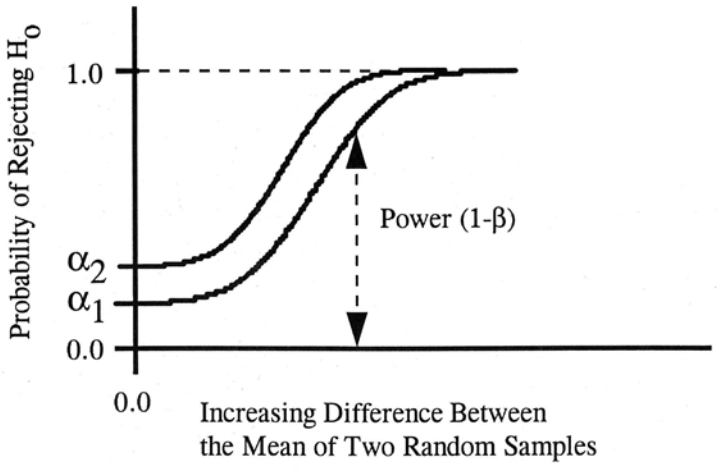
Type II: The null hypothesis (H_o) is accepted when H_o is really false.

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

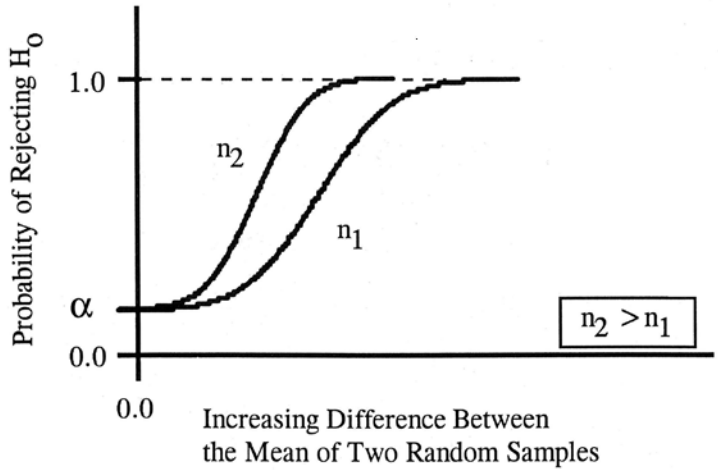
Type II error depends on the significance level, sample size, and variability, and which alternative hypothesis is true. The power of a test (1-β) is defined as the probability of correctly rejecting H_o when H_o is false. In general, for a fixed sample size, α and β vary inversely. For a fixed value of α, β can be reduced by increasing the sample size (Remington and Schork, 1970). Figure 4-1 illustrates this relationship. Suppose this interest is in testing whether there is a significant difference between the means from two independent random samples. As the difference in the two sample means increases (as indicated on the x-axis), the probability of rejecting H_o, the power, increases. If the real difference between the two sample means is zero, the probability of rejecting H_o is equal to the significance level, α. Figure 4-1A shows the general relationship between α and β if α is changed. Figure 4-1B shows the relationship between α and β if the sample size is increased.

Table 4-1. Errors in hypothesis testing.

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



A) Increasing Significance Level from α_1 to α_2



B) Increasing Sample Size from n_1 to n_2

Figure 4-1. Comparison of α and β.

4.1.2 Characteristics of Environmental Data

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

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

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

4.1.3 Recommendations for Selecting Statistical Methods

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

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

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

Table 4-2. Methods for characterizing data.

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

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

Adapted from Ward et al., 1990.

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

The remainder of this section provides recommendations for selecting statistical methods that can be applied on a routine basis for evaluating the average, changing, and extreme conditions of environmental variables (Table 4-3, adapted from Ward et al., 1990). In some instances, more appropriate methods might be available depending on the specific information needs. For routine analyses, both parametric and nonparametric methods are recommended. Nonparametric procedures are recommended together with parametric procedures since nonparametric procedures tend to be resilient to characteristics commonly found in nonpoint source monitoring data (Berryman et al., 1988; Gilliom and Helsel, 1986; Harcum et al., 1992; Harris et al., 1987; Helsel and Hirsch, 1995; Hirsch et al.,

Table 4-3. Methods for routine data analysis.

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

Adapted from Ward et al., 1990.

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

1982; van Belle and Hughes, 1984; Lettenmaier, 1988). However, the data analyst must be aware that violating assumptions associated with parametric or nonparametric tests can lead to incorrect conclusions about the collected data.

Average conditions

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

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

Changing conditions

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

The Mann-Whitney test is recommended for comparing two random samples when the distribution of the data is unknown or sufficiently nonnormal. The Student's *t* test can be used when the data are normally distributed. It has been demonstrated that the Student's *t* test can be successfully applied when the data are not normally distributed and might be more powerful under selected circumstances (Montgomery and Loftis, 1987), but that approach is not recommended here. The Kruskal-Wallis test (an extension of the Mann-Whitney test) is recommended for when there are three or more random samples. For example, numerous biological surveys are initiated by collecting data during the spring, summer, and fall. The hypothesis might be to determine whether there is a significant difference in key biological indices between the different seasons (index periods). An analysis of variance could be used if the data were normally distributed. Applying the Mann-Whitney or Student's *t* test to each pair of random samples is not appropriate.

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

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

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

Extreme values

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

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

4.1.4 Data Stratification

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

In urban areas the time periods should be set to correspond to the pollutant of concern and urban activities. Depending on the monitoring objectives, it might be necessary to consider periods of activity and nonactivity. If phosphorus is the pollutant of concern, periods that correspond to lawn maintenance activities and spring flush should be considered. If sediment is the problem, periods that correspond to the construction season should be considered. For irrigated agriculture, two periods should be established to correspond to irrigation and nonirrigation time.

In nonirrigated agricultural settings the periods selected should conform to the normal agricultural management pattern of the watershed. These periods should be based on amount of surface covered, precipitation patterns, and the timing of land and/or water management activities. By

defining time periods, the analyst can evaluate a hypothesis regarding whether significant differences in nitrogen and phosphorus losses occur during different agricultural seasons. Alberts et al. (1978) used this concept to examine seasonal losses of nitrogen and phosphorus in Missouri during three periods:

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

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

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

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

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

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

4.1.5 Recommended Reading List and Available Software

Recommended reading list

Over the last 20 years, considerable effort by researchers and practitioners has gone into the development of improved statistical methods for analyzing environmental data. Nonetheless, there is probably no single reference that fully covers all of the issues that the data analyst must consider when selecting methods for analyzing environmental data. The following list provides a summary of selected references that provide more details about a wider variety of issues. These references are strongly recommended for those who need a more in-depth discussion than that provided in this chapter.

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

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

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

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

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

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

Available software

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

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

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

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

4.2 Summary (Descriptive) Statistics

4.2.1 Point Estimation

Central tendency

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

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

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

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

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

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

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

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

$$HM_x = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \tag{4-4}$$

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

$$\textit{Weighted mean} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{4-5}$$

Summarizing storm event data

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

$$FWMC = \frac{\sum_{i=1}^{n} C_i T_i Q_i}{\sum_{i=1}^{n} T_i Q_i} \tag{4-6}$$

$$TWMC = \frac{\sum_{i=1}^{n} C_i T_i}{\sum_{i=1}^{n} T_i} \tag{4-7}$$

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

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

Figure 4-2 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 representative curves in Figure 4-2 or directly from the data. The nitrogen concentrations are typical of a "first flush" in which the concentrations are higher during the early part of the runoff. Tables 4-4 and 4-5 present the raw nitrogen values from Figure 4-2 together with the example calculations for computing the FWMC and EMC, respectively.

The first column in Table 4-4 is the time since the beginning of the storm. The fourth column is the time interval, T_i, represented by each sample. For example, the first entry, T_1, of 540 seconds is computed as (0.24 hours - 0.09 hours) times 3600 seconds/hour. The value of 0.24 is halfway between 0.20 and 0.28 hours. Selecting the halfway point between 0.20 and 0.28 hours centers the water quality observation in the time period being evaluated. The second entry, T_2, of 306 seconds is computed as (0.325 hours - 0.24 hours) times 3,600 seconds/hour. The value of 0.325 is halfway between 0.28 and 0.37 hours. The value of 0.24 is halfway between 0.20 and 0.28 hours. The fifth column is equal to flow (column 2) multiplied by the time interval (column 4). For

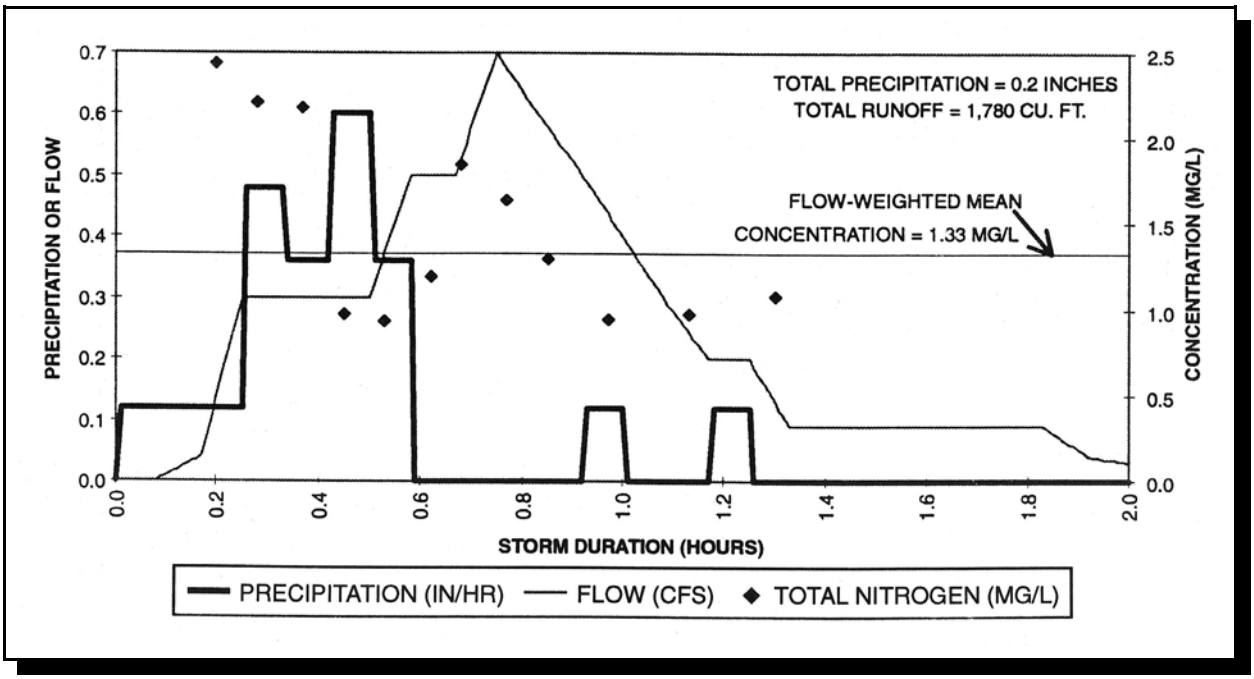


Figure 4-2. Precipitation, runoff, total nitrogen, and total phosphorus from a single storm event in Florida.

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

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

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

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

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

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

Time (hr)	Flow (cfs)	TN (mg/L)	T_{i+1} $-T_i$	$C_i \bullet Q_i$	$C_{i+1} \bullet Q_{i+1}$	$Q_i + Q_{i+1}$	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 (EMC) = 0.6722 / 0.4981 = 1.35 mg/L

Loading rates

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

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

Summarizing data with censored observations

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

Table 4-6. Raw data by time period.

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

Table 4-7. Loadings rate data.

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

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

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

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

Dispersion

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

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

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

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

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

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

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

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

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

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

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

For n equal to 10, the lower and upper quartile are equal to the 2.75th (0.25 x 11) and 8.25th (0.75 x 11) ordered observation. Using the data from above, Q_1 is equal to 0.11 + 0.75 x (0.16-0.11) or 0.1475 and Q_3 is equal to 0.85 + 0.25 x (0.98-0.85) or 0.8825. Similar to the *CV*, the coefficient of

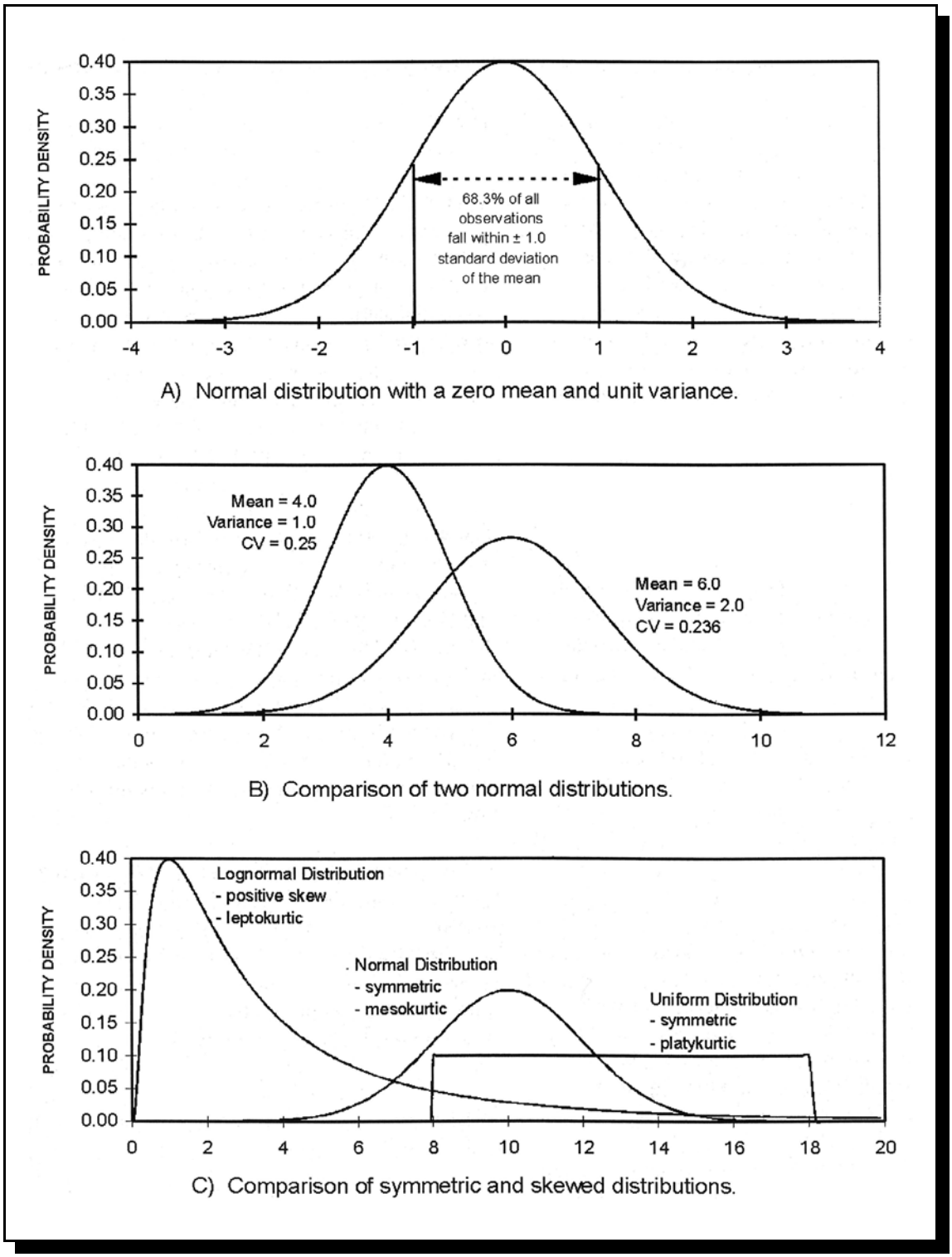


Figure 4-3. Comparison of several theoretical distributions.

quartile variation (V) can be used to compare different data sets:

$$V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \tag{4-11}$$

Skewness and Kurtosis

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

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

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

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

4.2.2 Interval Estimation

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

Mean

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

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

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

<u>Problem:</u>

Fifty-four samples were collected to determine the fraction of water collected (i.e., the split) by a water and sediment sampler for plot and field studies (Dressing et al., 1987). The data were tested and found to be normally distributed with a mean split of 0.0265 and a standard deviation of 0.0040. Determine the 95 and 99 percent confidence intervals for the population mean, μ.

Solution:

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

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

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

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

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

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

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

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

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

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

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

Standard deviation

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

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

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

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

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

Median and Quartiles

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

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

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

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

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

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

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

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

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

Problem:

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

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

Solution:

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

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

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

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

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

4.3 Graphical Data Display

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

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

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

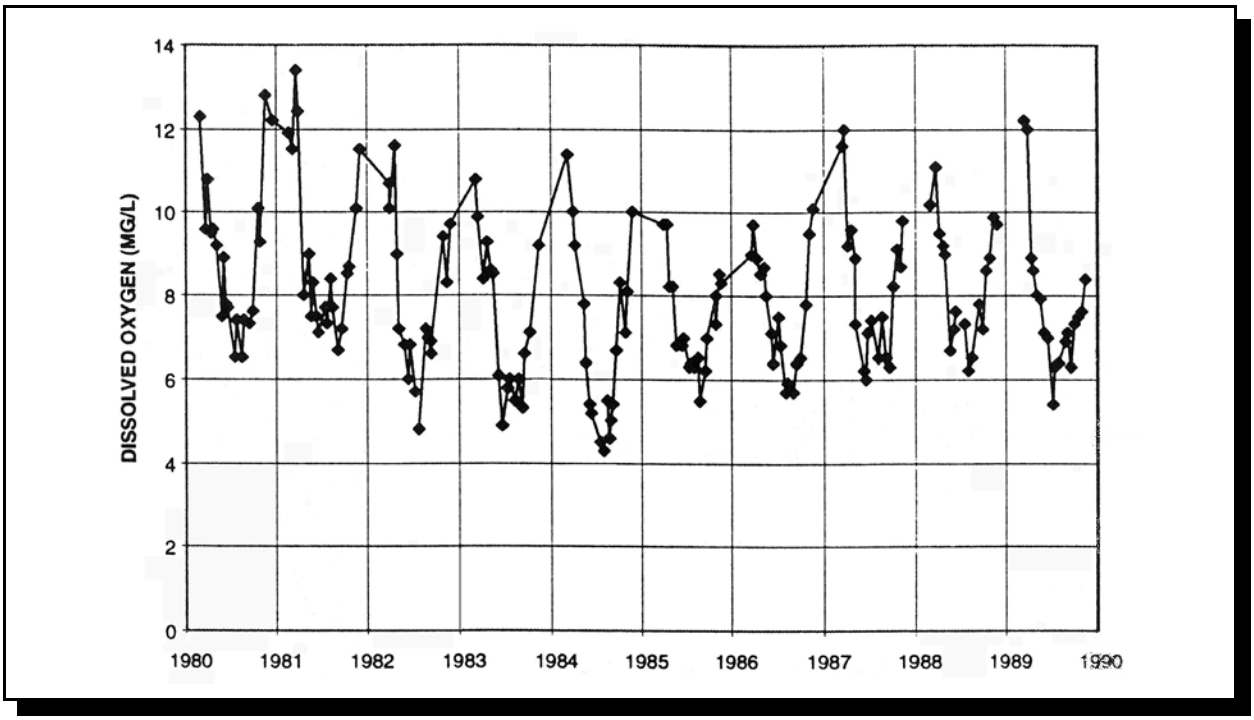


Figure 4-4. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a time series plot.

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

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

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

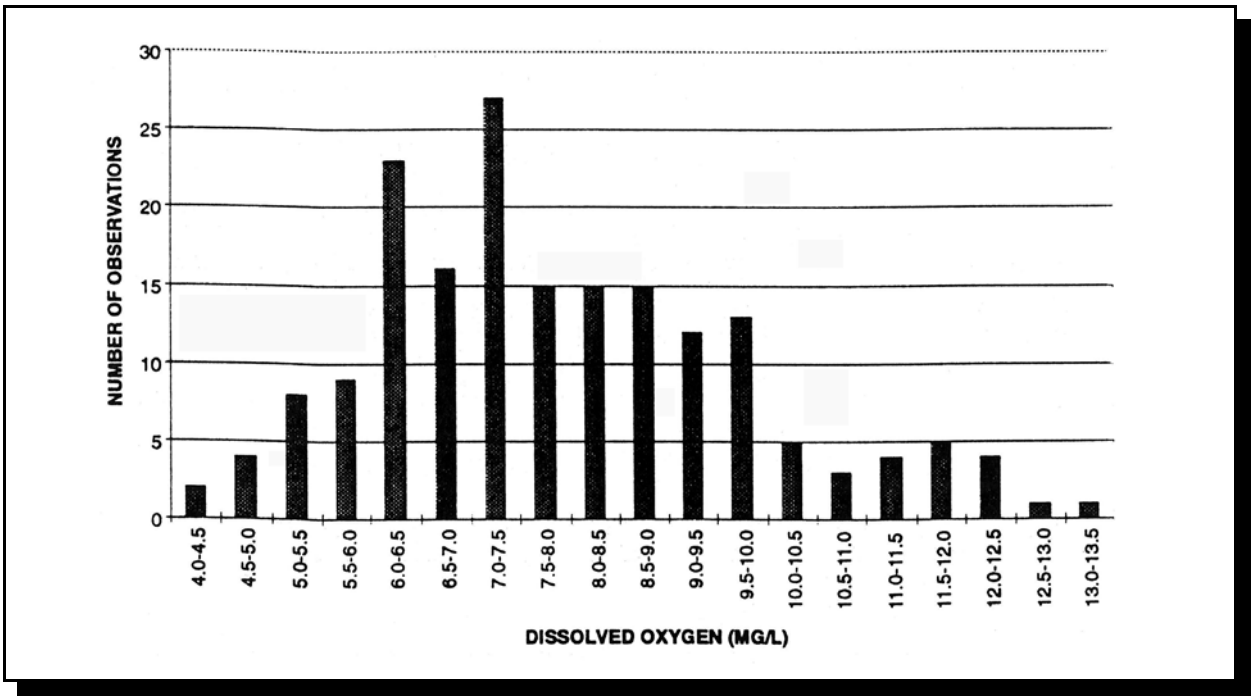


Figure 4-5. Dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware, using a histogram.

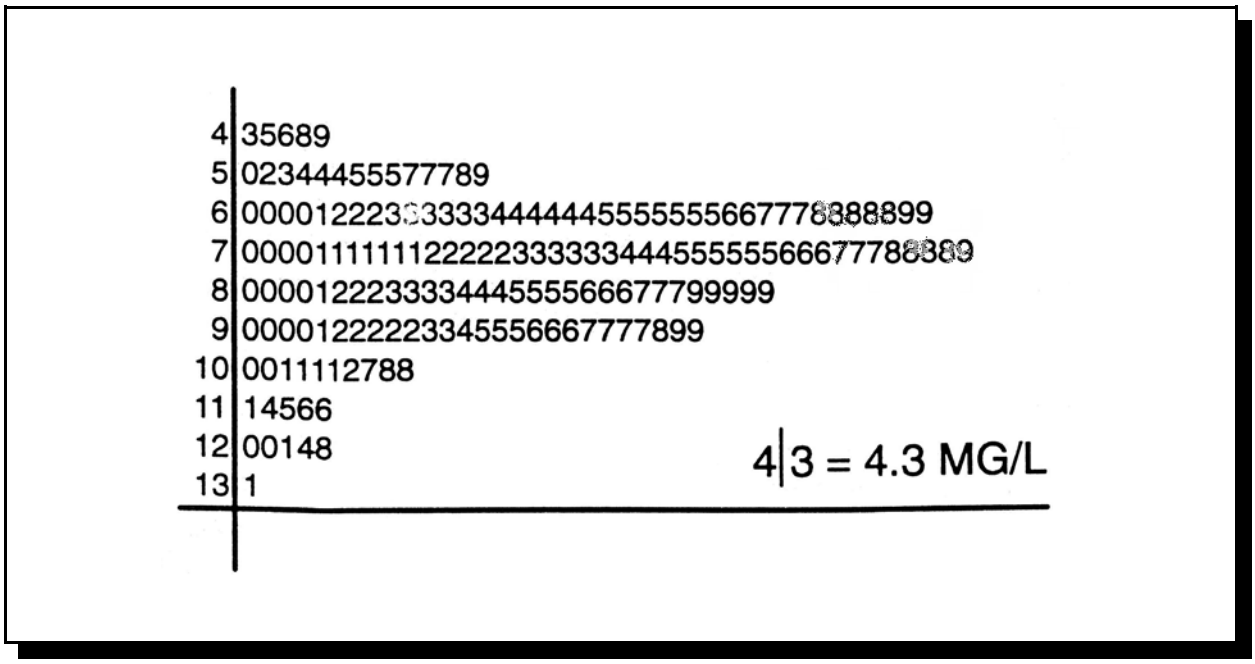


Figure 4-6. Stem-and-leaf plot of dissolved oxygen concentrations from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

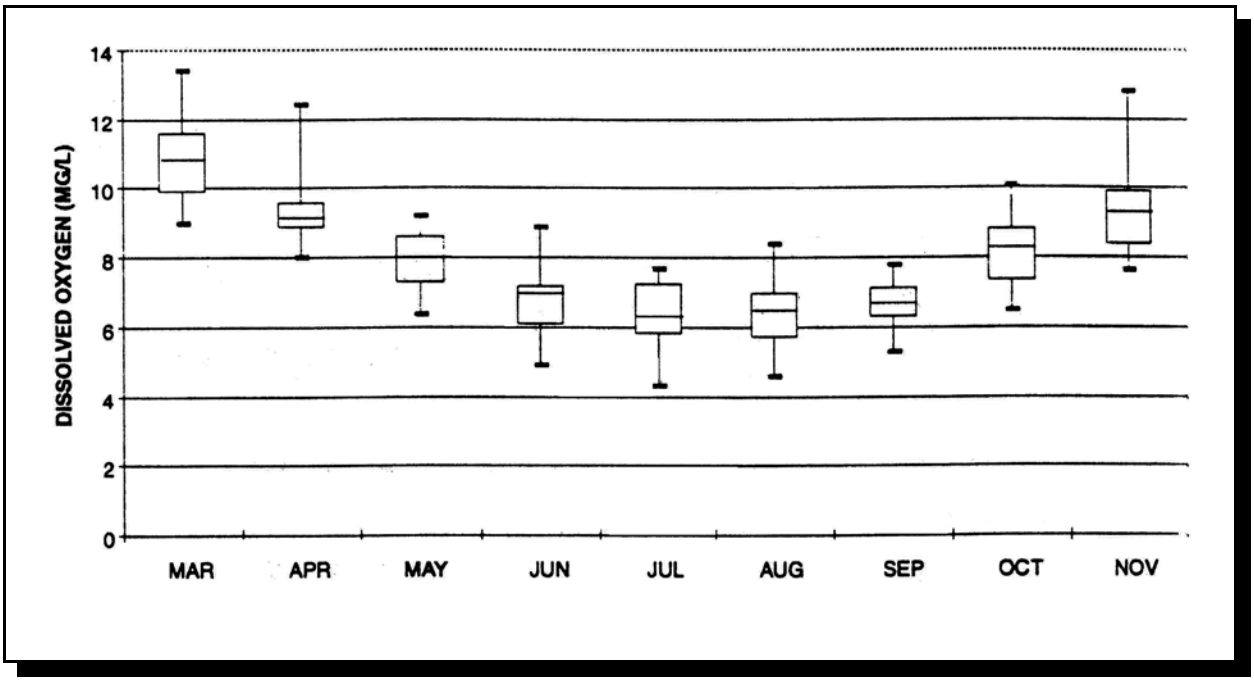


Figure 4-7. Boxplots of dissolved oxygen concentrations by month from 1980 through 1989 for the Delaware River at Reedy Island, Delaware.

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

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

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

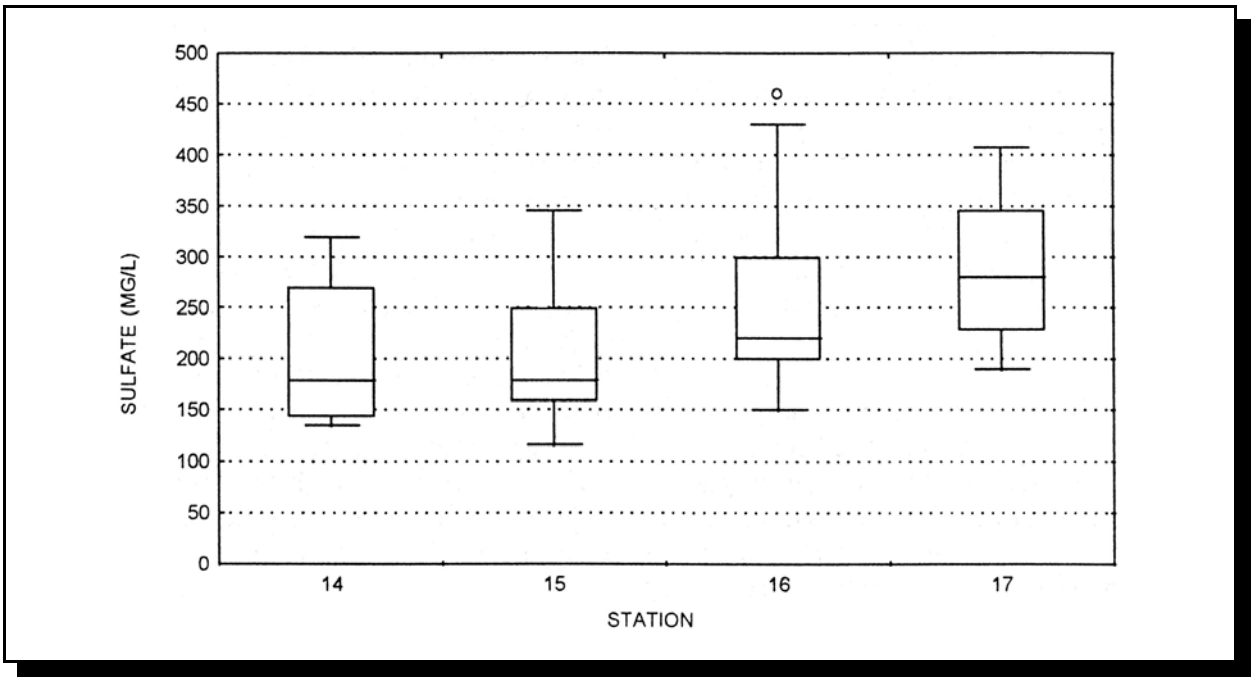


Figure 4-8. Boxplot of sulfate concentrations from 1993 and 1994 for the Rio Grande near El Paso, Texas.

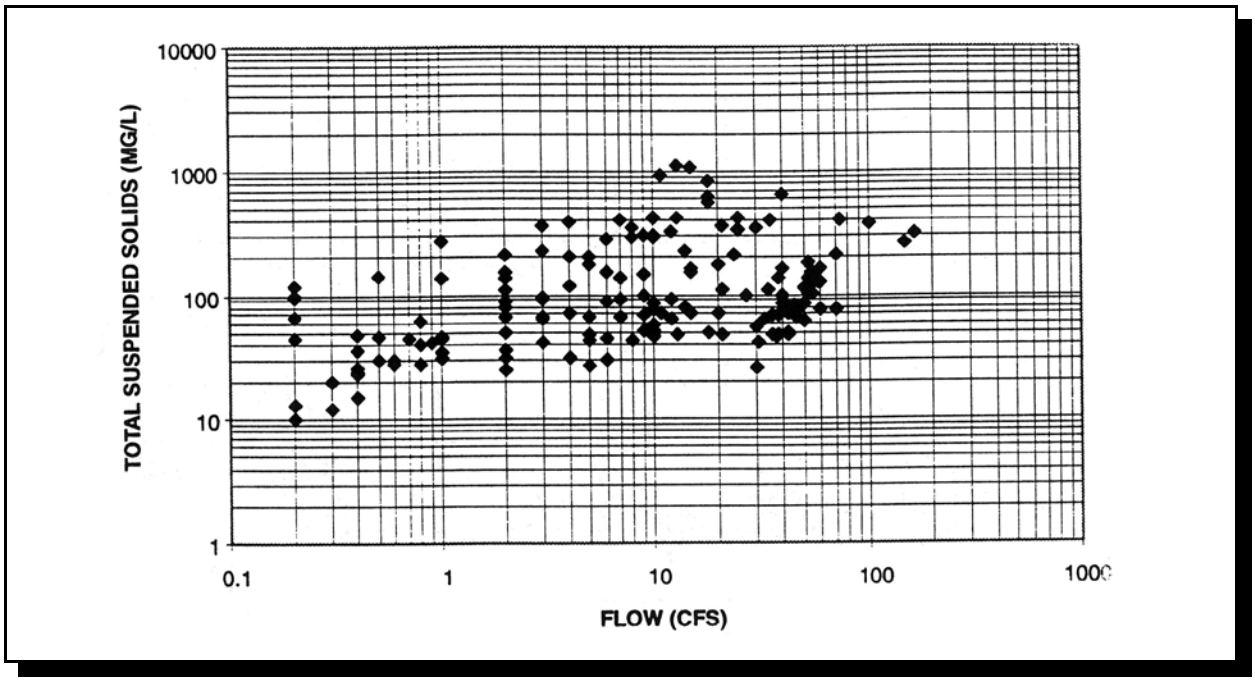


Figure 4-9. Bivariate scatter plot of total suspended solids and flow at 36th Street storm sewer in Denver, Colorado.

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

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

4.4 Evaluation of Test Assumptions

One of the basic criteria for selecting between parametric tests is whether the data being analyzed have a specific distribution (usually normal). For data with unknown distributions, nonparametric methods should be used since these methods do not

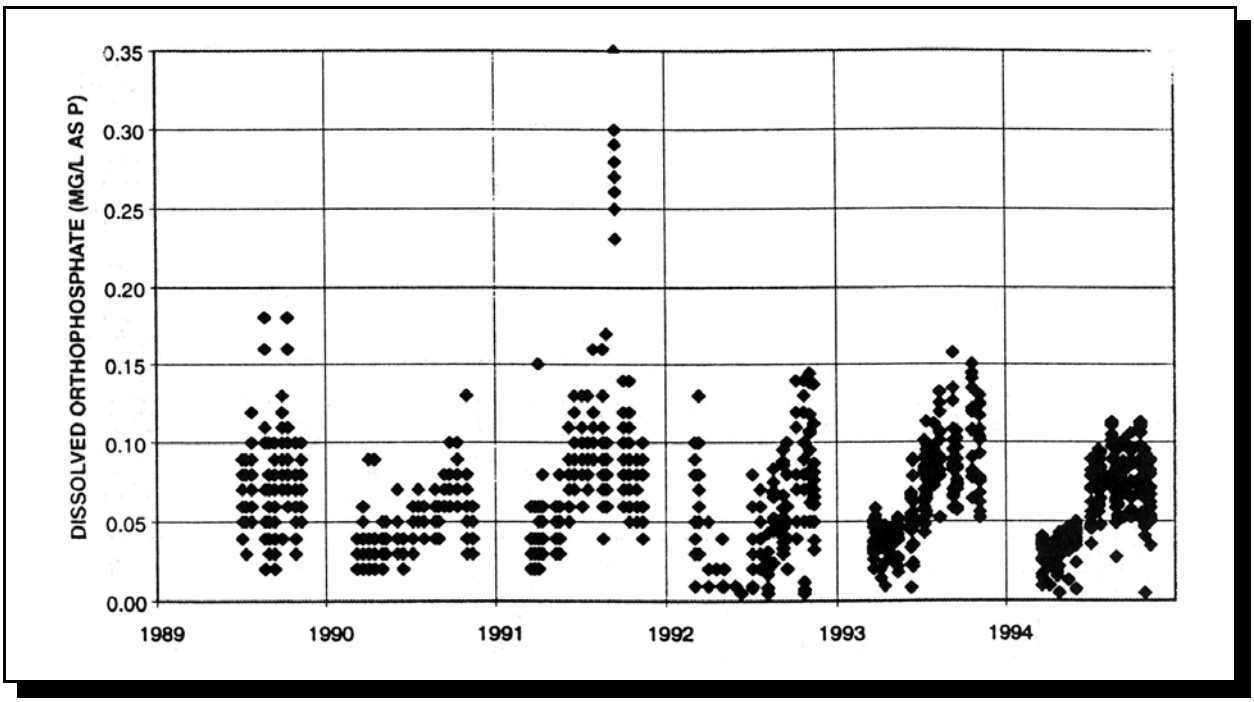


Figure 4-10. Time series plot of dissolved orthophosphate from 1989 through 1994 for portions of the Delaware River.

require that the data have a defined distribution. In addition, numerous tests require that the observations be independent (that is, randomly collected) and that the variances of the populations being compared be equal or of known ratio (Ponce, 1980a).

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

4.4.1 Tests of Normality

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

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

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

Graphical Methods

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

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

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

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

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

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

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

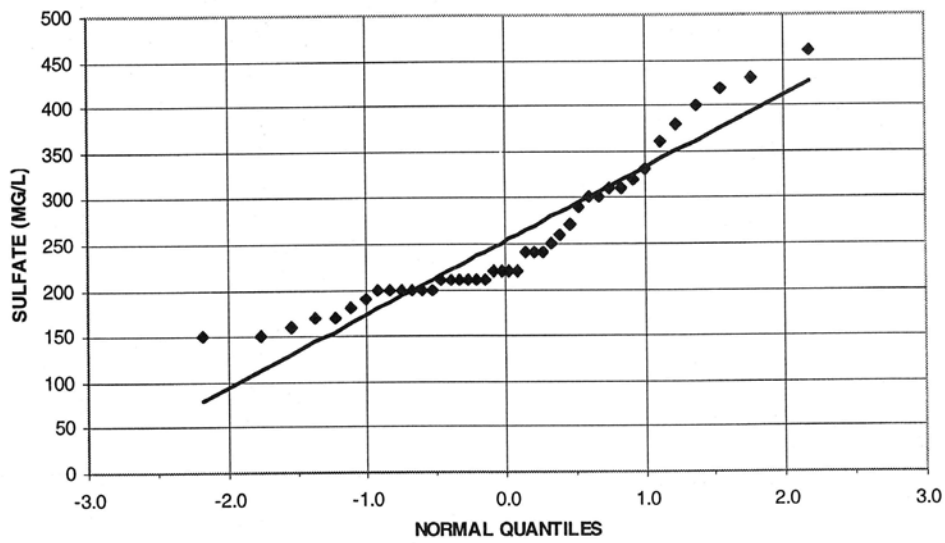
fall along the straight line. Clearly, the data do not fit a normal distribution, but are more typical of a positively skewed data set. As an alternative, the data can be log-transformed and the same analysis performed. In this case, the log-transformed data are less skewed (Figure 4-11B). The conclusion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

Skewness

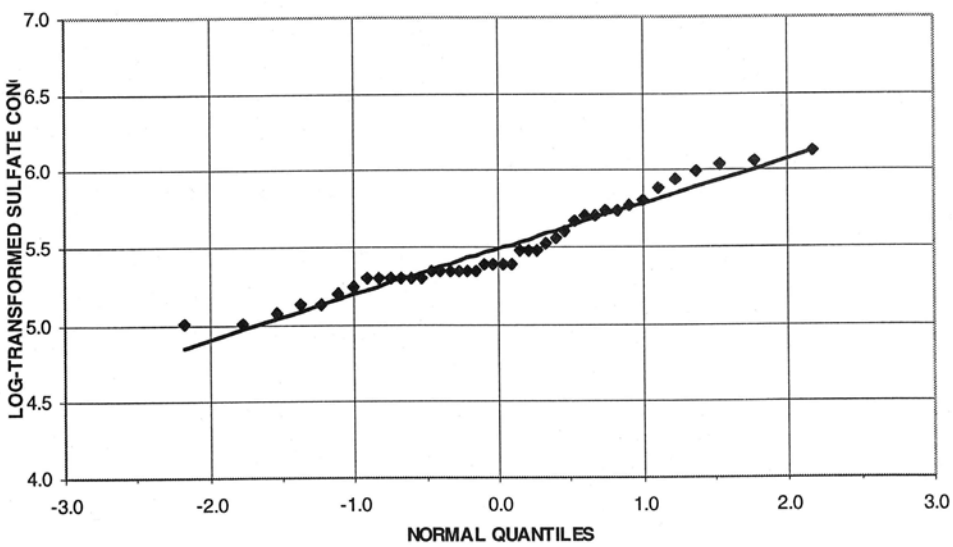
The approach used in testing for normality using skewness (Equation 4-12) is that a nonnormal distribution may be skewed, whereas a normal distribution is not skewed. If there are more than 150 observations and the data are normally distributed, the confidence limits on skewness from a normal distribution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

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

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



A) Probability Plot of Sulfate Concentration Data.



B) Probability Plot of Log-transformed Sulfate Data.

Figure 4-11. Probability plot of sulfate data from Station 16 in Figure 4-8.

Table 4-9. Table of skewness test for normality for sample sizes less than 150.

	α			α	
n	**0.02**	**0.10**	***n***	**0.02**	**0.10**
25	1.061	0.711	70	0.673	0.459
30	0.986	0.662	80	0.631	0.432
35	0.923	0.621	90	0.596	0.409
40	0.870	0.587	100	0.567	0.389
45	0.825	0.558	125	0.508	0.350
50	0.787	0.534	150	0.464	0.321
60	0.723	0.492	175	0.430	0.298

After Snedecor and Cochran, 1967.

Using the sulfate data from the previous example, selected statistics were computed and are summarized in Table 4-10. Selected statistics were also calculated for the log-transformed data. Using Equation 4-12, γ is equal to $(42/(41x40))x(2.1E+07/79.64^3)$ or 1.05. Using an estimated critical value from Table 4-9 of 0.575 for α equal to 0.10, the null hypothesis is rejected. The sulfate data do not come from a normal distribution. The log-transformed data (last column of Table 4-10) have a skewness equal to 0.54. The value is less than 0.575, and the null hypothesis is accepted. The reader should compare these results to those obtained using the graphical method presented in Figure 4-11.

Both Remington and Schork (1970) and the SAS Institute (1985a) caution that the test for skewness is only a partial indicator of normality. With small samples (less than 25), the test is particularly

Table 4-10. Selected summary statistics for the sulfate data from Station 16 in Figure 4-8.

	Sulfate	**log(sulfate)**
Number of observations (n)	42	42
Sum	10,620.00	230.55
Mean ($\bar{x}$)	252.86	5.49
Variance (s^2)	6,342.86	0.09
St. Dev. (s)	79.64	0.29
Skewness (γ)	1.05	0.54
Kurtosis (k)	0.32	-0.46
$\Sigma\lvert x_i-\bar{x}\rvert$	2,694.29	10.05
$\Sigma(x_i-\bar{x})^2$	2.6E+05	3.50
$\Sigma(x_i-\bar{x})^3$	2.1E+07	0.53
$\Sigma(x_i-\bar{x})^4$	5.1E+09	0.72

unreliable. That is, because of the small sample size, very large departures from normality are required before statistical tests will reject the null hypothesis of normality. Cochran (1977) proposed a general rule for determining how large *n* must be (i.e., *n* in the equation below) to allow safe use of the normal approximation in computing confidence limits for the mean. This rule is used most effectively for distributions with positive skewness, which are most common for environmental data.

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

where γ_1

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

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

Kurtosis

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

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

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

If the calculated value of k_1 falls outside the values in Table 4-11 for the selected level of confidence, there is evidence of non-normal kurtosis. Using the same example data, k_1 can be computed as 0.80 and 0.82 for the raw and log-transformed data, respectively. From this analysis, it is concluded that the raw and the log-transformed data have a

Table 4-11. Values of kurtosis test for normality for small sample sizes.

	α = 0.02		α = 0.10	
n	Lower	Upper	Lower	Upper
11	0.6675	0.9359	0.7153	0.9073
21	0.6950	0.9001	0.7304	0.8768
31	0.7110	0.8827	0.7404	0.8625
41	0.7216	0.8722	0.7470	0.8540
51	0.7291	0.8648	0.7518	0.8481
61	0.7347	0.8592	0.7554	0.8434
71	0.7393	0.8549	0.7583	0.8403

After Remington and Schork, 1970.

kurtosis that is consistent with a normal distribution since k_1 is between the range of 0.7470 to 0.8540 for α equal to 0.10.

Shapiro-Wilk *W* test

The Shapiro-Wilk *W* test can be used to test the distribution of a data set for sample sizes of less than 2,000 (SAS Institute, Inc., 1990). This test uses the *W* statistic, which is "the ratio of the best estimator of the variance to the usual corrected sum of squares estimator of the variance" (SAS Institute, Inc., 1990). The null hypothesis for this test is that the data set is a random sample from a normal distribution. Values of *W* are greater than zero and less than or equal to one. The null hypothesis is rejected with small values. For sample sizes greater than 2,000, the Kolmogorov *D* statistic may be used (SAS Institute, Inc., 1990).

Anderson and McLean (1974) recommend the Shapiro-Wilk *W* test for normality and note that it is superior to the Kolmogorov-Smirnov and chi-squared tests in detecting non-normality over sample sizes ranging from 10 to 50. The following procedure for using the test is adapted from Anderson and McLean (1974) and Gilbert (1987):

1. Order the *n* observations as $x_1 \le x_2 \le ... \le x_n$.

2. Compute $d = (n\text{-}1)s^2$.

3. Compute *k*. If *n* is even, $k = n/2$. If *n* is odd, $k = (n\text{-}1)/2$.

4. Compute

$$W = \frac{1}{d}\left[\sum_{i=1}^{k} a_i(x_{n-i+1} - x_i)\right]^2 \qquad (4\text{-}22)$$

where the values of a_i appear in Table D4. The value x_{n-i+1} is equal to x_n when *i* is equal to 1 and x_{n-k+1} when *i* is equal to *k*.

5. Reject H_o (of normality) at the α significance level if *W* is less than the quantile given in Table D5.

Table 4-12 presents the sulfate data from Station 16 in Figure 4-8 in a format ready for analysis. The results for step 2 can be computed from the statistics in Table 4-10. Since there are 42 observations, *k* is equal to 21. The first column in Table 4-12 indicates the value of *i* for each row in the table. The second column corresponds to the values of a_i from Table D4. (Note that the values in Table D4 are for $a_{n\text{-}i+1}$ and are exactly the same as a_i.) The third and fourth column, x_i and $x_{n\text{-}i+1}$, represent the raw sulfate data. The third column represents the first half of the observations, and the fourth column represents the last half of the data in reverse order (e.g., 460 is the largest sample observation). The fifth and sixth columns correspond to the log-transformed data from columns 3 and 4. For example, log(150) is equal to 5.01. The last two columns provide intermediate calculations associated with Equation 4-22 (i.e., $a_i(x_{n\text{-}i+1} - x_i)$) for the raw and log-transformed data, respectively.

Summing the last two columns results in completing the summation specified in Equation 4-22. The *W* statistic may now be computed using Equation 4-22 to yield 0.88 and 0.89 for the raw and log-transformed data, respectively. From Table D5, the quantile for 42 observations (95 percent confidence level) is 0.942. As a result, it can be concluded that the raw data and the log-transformed data are normally distributed.

4.4.2 Tests of Equal Variance

When performing hypothesis tests of two samples using parametric procedures, it is typically necessary to make sure that the two data sets have

Table 4-12. Example analysis of the Shapiro-Wilk *W* test using the sulfate data from Station 16 in Figure 4-8.

i	a_i	x_i	x_{n-i+1}	log x_i	log x_{n-i+1}	Intermediate Calculations	
1	0.3917	150	460	5.01	6.13	121.43	0.44
2	0.2701	150	430	5.01	6.06	75.63	0.28
3	0.2345	160	420	5.08	6.04	60.97	0.23
4	0.2085	170	400	5.14	5.99	47.96	0.18
5	0.1874	170	380	5.14	5.94	39.35	0.15
6	0.1694	180	360	5.19	5.89	30.49	0.12
7	0.1535	190	330	5.25	5.80	21.49	0.08
8	0.1392	200	320	5.30	5.77	16.70	0.07
9	0.1259	200	310	5.30	5.74	13.85	0.06
10	0.1136	200	310	5.30	5.74	12.50	0.05
11	0.1020	200	300	5.30	5.70	10.20	0.04
12	0.0909	200	300	5.30	5.70	9.09	0.04
13	0.0804	200	290	5.30	5.67	7.24	0.03
14	0.0701	210	270	5.35	5.60	4.21	0.02
15	0.0602	210	260	5.35	5.56	3.01	0.01
16	0.0506	210	250	5.35	5.52	2.02	0.01
17	0.0411	210	240	5.35	5.48	1.23	0.01
18	0.0318	210	240	5.35	5.48	0.95	0.00
19	0.0227	210	240	5.35	5.48	0.68	0.00
20	0.0136	220	220	5.39	5.39	0.00	0.00
21	0.0045	220	220	5.39	5.39	0.00	0.00
					SUM	479.00	1.81

the same variance. Testing for equal variances between two populations can be done by evaluating the ratio of the two sample variances (F_1) with the following equation:

$$F_1 = \frac{s_a^2}{s_b^2} \qquad (4\text{-}23)$$

where

$$s_a^2 \geq s_b^2$$

The null hypothesis in this test is that the variance ratio is equal to 1, and the alternative hypothesis is that the ratio is not equal to 1. The ratio is compared to a critical value from the *F* distribution (Table D6) that is based on the sample sizes (n_a and n_b) and the selected level of significance (α). Since the numerator is selected to be the variance with the larger value, it is necessary to look at only one critical value even though a two-sided test is being used.

For the sulfate data from Stations 16 and 17 in Figure 4-8, F_1 can be computed as 6,342.9/5,536.3 or 1.15 with 41 (42-1) and 10 (11-1) degrees of freedom. Using Table D6, the critical *F* value (for a two-sided 95 percent confidence level test where α/2 is equal to 0.025) is approximately 3.25. Therefore, the null hypothesis is accepted and it is concluded that the variances of the sulfate data from Stations 16 and 17 are the same.

4.4.3 Tests of Randomness

Another type of hypothesis testing involves time series at a single station. The DO data plotted in Figure 4-4 are one example. An approach to evaluate randomness is to compute the total number of runs (u) above and below the median (Freund, 1973). A run is a string of values all above or all below the median. A string of one value is acceptable. In this test, the median is determined, all values are placed in chronological order, and each value is assigned an "a" if it is above the median and a "b" if it is below the median. For example, the following is a set of data in chronological order:

5, 5, 6, 9, 13, 12, 2, 3, 2, 8, 14, 13, 11, 20, 4, 6, 9, 1, 7, 11, 12.

The median for this set of values (n=21) is 8. The series of values in terms of "a" and "b" is

b, b, b, a, a, a, b, b, b, omit, a, a, a, a, b, b, a, b, b, a, a

The number of runs (u) in the example data set is 8. Note that in this test all values equal to the median are omitted. Also, the number of values above (n_1) and below (n_2) must each be 10 or more to allow use of the following statistics. For n_1 and n_2 less than 10, special tables are required (Freund, 1973). The test statistic (derived from the normal distribution) is:

$$Z = \frac{u - \mu_u}{\sigma_u} \tag{4-24}$$

where

$$\mu_u = \frac{2n_1n_2}{n_1+n_2} + 1 \tag{4-25}$$

and

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2\ (n_1+n_2-1)}} \tag{4-26}$$

Applying these equations to the above example data,

$$\mu_u = \frac{2(10)(10)}{10+10} + 1 = 11$$

$$\sigma_u = \sqrt{\frac{2(10)(10)\,[2(10)(10) - (10) - (10)]}{(10+10)^2\ (10+10-1)}} = 2.1764$$

$$Z = \frac{8 - 11}{2.1764} = -1.38$$

With α equal to 0.05 in a two-tailed test, the Z values (for α/2) are 1.96 and -1.96 (Table D1). Since -1.38 falls within this range, the null hypothesis that the sample is random is accepted.

4.5 Evaluation of One or Two Independent Random Samples

The data collected for evaluating changes will typically come as (1) two or more sets of random samples or (2) a time series at a single station. In the first case, the analyst will test for a shift or step change (e.g., a significant difference between conditions before and after treatment). This might be typical for data collected from two stations along a stream segment. Or, when performing a biological assessment, for example, the goal might be to determine whether there is a significant difference (i.e., a step change) between biological metrics for data collected at randomly selected reference and test (targeted) sites. It is also possible to compare a single random sample to a particular value. This might be the case when comparing data to a standard or reference

condition. This section describes common approaches for comparing one or two independent random samples. Comparing more than two independent random samples or time series is discussed later.

Depending on the objective, it is appropriate to select a one- or two-sided test. For example, if the analyst knows that TSS would only decrease as a result of BMP implementation or is interested only if the TSS decreases, a one-sided test can be formulated. Alternatively, if the analyst does not know whether TSS will go up or down, a two-sided test is necessary. If the analyst simply wants to compare two random samples to decide if they are significantly different, a two-sided test can be used. Appropriate uses of a one-sided test include testing for decreased sediment or nutrient loads after implementing a flood control dam or best management practice, or comparing a suspected contaminated site to an upstream or control site. Typical null hypotheses (H_o) and alternative hypotheses (H_a) for one- and two-sided tests are provided below:

<u>One-sided test</u>
H_o: TSS (postimplementation) ≥ TSS (pre-implementation)
H_a: TSS (postimplementation) < TSS (pre-implementation)

<u>Two-sided test</u>
H_o: TSS (postimplementation) = TSS (pre-implementation)
H_a: TSS (postimplementation) ≠ TSS (pre-implementation)

Selecting a one-sided test instead of a two-sided test results in an increased power for the same significance level (Winer, 1971). That is, if the conditions are appropriate, a corresponding one-sided test is more desirable than a two-sided test given the same level of significance (α) and sample size. The manager and analyst should take great care in selecting one- or two-sided tests.

4.5.1 Tests for One Sample or Paired Data

Suppose the analyst is interested in evaluating compliance with a water quality standard or reference condition, e.g., a target determined from a load allocation or a percent substrate embeddedness less than the amount that hinders fisheries. In these situations the analyst might collect a random sample and compare it to a reference value. The Student's *t* and the Wilcoxon Signed Ranks tests are the two most appropriate tests when evaluating one *independent* random sample. The sign test can also be used, but it is generally limited to random samples that cannot be transformed into a symmetric distribution.

In addition, the analyst might be interested in determining whether a water quality variable increased between two sites located along a stream. In this situation the analyst might collect two random samples with matched or paired observations. Paired observations are a series of data collected as pairs at a given time or location. For example, if BOD_5 is sampled at two stream locations at a regular time interval, the result is a pair of BOD_5 observations for each time period. The same statistical tests used for one independent sample can be used to compare paired observations. The tests are adjusted by computing and analyzing the difference between the paired observations. The associated *t* test is referred to as the paired *t* test.

Tests for One Sample or Paired Data

Test[a]	Additional Assumptions
Student's *t* (paired *t*)	Normal distribution
Wilcoxon Signed Ranks	Symmetric distribution
Sign	None

[a] The standard forms of these tests require independent random samples.

Student's *t* test

The participants in the Highland Silver Lake RCWP project (Jamieson, 1986) formulated a null hypothesis that a BMP would not reduce the post-implementation mean TSS concentrations to less than 25 mg/L. (Presumably, the participants hoped that the mean TSS concentration would be less than 25 mg/L so that H_0 could be rejected.) A formalized statement of the null and alternative hypotheses using a one-sided test would be:

H_o: $\mu \geq 25$ mg/L
H_a: $\mu < 25$ mg/L

In this case it is assumed that the mean TSS concentration is a good measure of central tendency and is the best measure for evaluation. It is also assumed that any change in TSS mean concentration is due to the BMP alone. H_a is stated such that a one-sided test can be applied because there is concern specifically about whether the postimplementation mean TSS concentration is lower than 25 mg/L since this might have been the target in a load allocation.

The Student's *t* test statistic (*t*) with *n*-1 degrees of freedom (df) can be used if the data are independent and normally distributed:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad (4\text{-}27)$$

where μ would be equal to the hypothesized value, 25 mg/L in this case. Assuming a one-sided test is used, the critical value for *t* would be obtained from Table D2 with *n*-1 degrees of freedom and a significance level of α. If a two-sided test were used (H_o: $\mu = 25$ mg/L; H_a: $\mu \neq 25$ mg/L), a value corresponding to a significance level of $\alpha/2$ would be obtained from Table D2.

The TSS data from the Highland Silver Lake RCWP project (Table 4-13) are from May 21, 1981, through October 31, 1984. The period after April 1, 1983, is the postimplementation period. Before testing H_o with a statistical test, the data must be inspected and the assumptions of randomness and normality must be tested. These tests are performed on the preimplementation and postimplementation data sets although only the postimplementation data in the current example are used. Using the SAS Univariate procedure (SAS Institute, Inc., 1985a), summary statistics and graphical presentations can be generated for the two data sets (Figures 4-12 and 4-13).

The values for skewness (0.82) and kurtosis (- 0.42) indicate positive skew and low kurtosis in the pre-BMP sample distribution. The Shapiro-Wilk *W* statistic (0.893) and associated probability (0.063) show that the null hypothesis (that the sample is normally distributed) can be rejected with 93.7 percent confidence. In other words, there is only a 6.3 percent chance that a lower *W* value could be obtained if the sample were indeed taken from a normal distribution. Hence, the assumption of a normal distribution is rejected and the alternative hypothesis that the distribution is non-normal is accepted.

Modification for Paired *t* Test

The sample mean, $\bar{x}$, and standard deviation, s, in Equation 4-27 refer to the mean and standard deviation of the differenced data (i.e., $d_i = x_i - y_i$).

The differenced data must be normally distributed.

The number of degrees of freedom is equal to the number of paired observations minus one.

```
                                   PRE-BMP IMPLEMENTATION DATA SET

                                          Univariate Procedure

Variable=PRETSS

                Moments                                  Quantiles(Def=4)                                  Extremes

 N                16  Sum Wgts         16      100% Max        60       99%        60         Lowest   Obs   Highest Obs
 Mean         29.375  Sum             470       75% Q3      40.75       95%        60             11(   1)       37(   6)
 Std Dev    16.14879  Variance   260.7833       50% Med        23       90%        60             12(   4)       42(  16)
 Skewness   0.817104  Kurtosis   -0.41702       25% Q1         17       10%      11.7             13(   3)       48(  13)
 USS           17718  CSS         3911.75        0% Min        11        5%        11             16(  12)       60(   8)
 CV          54.9746  Std Mean   4.037197                              1%        11             20(  11)       60(   9)
 T:Mean=0   7.276088  Pr>|T|       0.0001      Range           49
 Num ~= 0         16  Num > 0          16      Q3-Q1        23.75
 M(Sign)           8  Pr>=|M|      0.0001      Mode            20
 Sgn Rank         68  Pr>=|S|      0.0001
 W:Normal   0.893033  Pr<W         0.0630

      Stem Leaf                     #             Boxplot                     Normal Probability Plot
         6 00                       2                |              65+                                    *    *+++++++
         5                                           |                |                                     ++++++
         4 28                       2             +-----+            |                              +*+*++
         3 057                      3             |     |            |                         +*+*+*
         2 00015                    5             *--+--*            |                    *+*+**+*
         1 1236                     4             +-----+          15+       *    * +*+*++
           ----+----+----+----+                                        +----+----+----+----+----+----+----+----+----+----+
       Multiply Stem.Leaf by 10**+1                                        -2        -1         0        +1        +2

                                                  Frequency Table

                          Percents                        Percents                        Percents                        Percents
            Value Count  Cell   Cum        Value Count  Cell   Cum      Value Count  Cell   Cum      Value Count  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 4-12. Preimplementation data set.

```
                                   POST-BMP IMPLEMENTATION DATA SET

                                          Univariate Procedure

Variable=POSTTSS

                Moments                                  Quantiles(Def=4)                                  Extremes

 N                15  Sum Wgts         15      100% Max        42       99%        42         Lowest    Obs     Highest    Obs
 Mean       19.86667  Sum             298       75% Q3         32       95%        42              6(       2)       30(       3)
 Std Dev    12.17061  Variance   148.1238       50% Med        14       90%      40.8              7(       4)       32(       5)
 Skewness   0.697811  Kurtosis   -0.99379       25% Q1         10       10%       6.6             10(      13)       32(       7)
 USS            7994  CSS        2073.733        0% Min         6        5%         6             10(       8)       40(      11)
 CV         61.26147  Std Mean   3.142439                                1%         6             11(      15)       42(       9)
 T:Mean=0   6.322054  Pr>|T|       0.0001      Range          36
 Num ~= 0         15  Num > 0          15      Q3-Q1          22
 M(Sign)         7.5  Pr>=|M|      0.0001      Mode           10
 Sgn Rank         60  Pr>=|S|      0.0001
 W:Normal   0.877487  Pr<W         0.0438

      Stem Leaf                     #             Boxplot                                  Normal Probability Plot
         4 02                       2                |                     42.5+                                    *    *+++
         3                                           |                         |                                     ++++
         3 022                      3             +-----+                     |                              * * ++++
         2                                        |     |                     |                              ++++
         2 2                        1             |  +  |                     |                          ++*+
         1 6                        1             |     |                     |                      ++++ *
         1 001244                   6             *-----*                     |                 *+*+*+*+**
         0 67                       2                |                      7.5+        *    *+++
           ----+----+----+----+                                                +----+----+----+----+----+----+----+----+----+----+
       Multiply Stem.Leaf by 10**+1                                                -2        -1         0        +1        +2

                                                   Frequency Table

                          Percents                          Percents                          Percents                          Percents
            Value Count  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 4-13. Postimplementation data set.

Table 4-13. Highland Silver Lake TSS data for site 1.

Preimplementation		Postimplementation	
Date	**TSS (mg/L)**	**Date**	**TSS (mg/L)**
5/21/81	11	4/12/83	14
6/18/81	21	5/10/83	6
7/30/81	13	6/7/83	30
9/3/81	12	7/27/83	7
10/6/81	30	10/5/83	32
11/5/81	37	11/18/83	14
12/8/81	20	12/29/83	32
4/8/82	60	1/25/84	10
4/27/82	60	2/20/84	42
5/25/82	20	4/11/84	22
6/22/82	20	5/15/84	40
7/20/82	16	7/17/84	12
9/20/82	48	8/21/84	10
10/26/82	35	9/26/84	16
11/23/82	25	10/31/84	11
12/2/82	42		

Overall:	*n* = 31 mean = 24.77	s = 14.93 median = 20
PreImplementation:	*n* = 16 mean = 29.38	s = 16.15 median = 23
PostImplementation:	*n* = 15 mean = 19.87	s = 12.17 median = 14

In the post-BMP sample distribution, the values for skewness (0.70) and kurtosis (-0.99) again indicate positive skew and low kurtosis. The Shapiro-Wilk *W* statistic (0.88) and associated probability (0.044) show that the null hypothesis (that the sample is from a normal distribution) can be rejected with 95.6 percent confidence. Also rejected is the assumption of a normal distribution for the post-BMP data set.

Taking the logarithm (base 10) of each data point for the pre-BMP and post-BMP data sets, the SAS Univariate procedure is run to see if the assumption of normality would be appropriate for the log-transformed data set. The output plots and statistics are shown in Figures 4-14 and 4-15. Note that the skewness (0.10) is much less pronounced, but the kurtosis (-1.09) is more negative for the transformed pre-BMP data set. The higher *W* statistic (0.951) and associated probability (0.493) indicate that the null hypothesis that the transformed data are normally distributed should be accepted.

For the log-transformed post-BMP data, the skewness (0.072) is also reduced and the kurtosis (-1.23) is more negative than for the raw data set. The *W* statistic (0.939) and associated probability (0.367) indicate that the null hypothesis that the transformed data are normally distributed should be accepted. In fact, there is a 63.3 percent probability that a lower *W* statistic could be obtained if the sample is from a normal distribution.

```
                                   PRE-BMP IMPLEMENTATION DATA SET (LOG TRANSFORMED)

                                                Univariate Procedure

Variable=PRELTSS

                Moments                                  Quantiles(Def=4)                                  Extremes

 N                16  Sum Wgts         16      100% Max  1.778151       99%  1.778151         Lowest    Obs     Highest    Obs
 Mean       1.407004  Sum        22.51207       75% Q3   1.609487       95%  1.778151       1.041393(       1) 1.568202(       6)
 Std Dev    0.238929  Variance   0.057087       50% Med   1.36008       90%  1.778151       1.079181(       4) 1.623249(      16)
 Skewness   0.102176  Kurtosis   -1.08679       25% Q1   1.228347       10%  1.067845       1.113943(       3) 1.681241(      13)
 USS        32.53089  CSS        0.856304        0% Min  1.041393        5%  1.041393        1.20412(      12) 1.778151(       8)
 CV         16.98138  Std Mean   0.059732                                1%  1.041393        1.30103(      11) 1.778151(       9)
 T:Mean=0   23.55521  Pr>|T|       0.0001      Range     0.736759
 Num ~= 0         16  Num > 0          16      Q3-Q1      0.38114
 M(Sign)           8  Pr>=|M|      0.0001      Mode       1.30103
 Sgn Rank         68  Pr>=|S|      0.0001
 W:Normal   0.951026  Pr<W         0.4929

      Stem Leaf                     #             Boxplot                                  Normal Probability Plot
        17 88                       2                |                     1.75+                                  *+++ *
        16 28                       2             +-----+                      |                                *+*++
        15 47                       2             |     |                      |                             *+*++
        14 08                       2             |  +  |                      |                        ++*+
        13 0002                     4             *-----*                      |                   * *+** *
        12 0                        1             +-----+                      |                 *+++
        11 1                        1                |                         |              ++*+
        10 48                       2                |                     1.05+       *++++*
         ----+----+----+----+                                                     +----+----+----+----+----+----+----+----+----+----+
      Multiply Stem.Leaf by 10**-1                                                   -2        -1         0        +1        +2

                                                Frequency Table

                          Percents                        Percents                        Percents                        Percents
            Value Count  Cell   Cum        Value Count  Cell   Cum        Value Count  Cell   Cum        Value Count  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 4-14. Log-transformed preimplementation data set.

```
                                POST-BMP IMPLEMENTATION DATA SET (LOG TRANSFORMED)

                                                Univariate Procedure

Variable=POSTLTSS

                Moments                                  Quantiles(Def=4)                                  Extremes

 N                15  Sum Wgts         15      100% Max  1.623249       99%  1.623249         Lowest    Obs     Highest    Obs
 Mean        1.21969  Sum         18.29535       75% Q3    1.50515       95%  1.623249       0.778151(       2) 1.477121(       3)
 Std Dev    0.273571  Variance   0.074841       50% Med  1.146128       90%  1.610536       0.845098(       4)  1.50515(       5)
 Skewness   0.072331  Kurtosis   -1.22633       25% Q1          1       10%  0.818319              1(      13)  1.50515(       7)
 USS        23.36244  CSS        1.047777        0% Min  0.778151        5%  0.778151              1(       8)  1.60206(      11)
 CV         22.42956  Std Mean   0.070636                                1%  0.778151       1.041393(      15) 1.623249(       9)
 T:Mean=0   17.26731  Pr>|T|       0.0001      Range     0.845098
 Num ~= 0         15  Num > 0          15      Q3-Q1      0.50515
 M(Sign)         7.5  Pr>=|M|      0.0001      Mode             1
 Sgn Rank         60  Pr>=|S|      0.0001
 W:Normal   0.939469  Pr<W         0.3661

      Stem Leaf                     #             Boxplot                                  Normal Probability Plot
        16 02                       2                |                    1.65+                                  * +++*
        15 11                       2             +-----+                    |                                 * * +++
        14 8                        1             |     |                    |                                * ++++
        13 4                        1             |     |                 1.35+                            *+++
        12 0                        1             |  +  |                    |                         ++*
        11 55                       2             *-----*                    |                      ++*+*
        10 0048                     4             +-----+                 1.05+                  *+*+* *
         9                                           |                       |                ++*
         8 5                        1                |                       |             +++*
         7 8                        1                |                    0.75+          +*+
         ----+----+----+----+                                                +----+----+----+----+----+----+----+----+----+----+
      Multiply Stem.Leaf by 10**-1                                                -2        -1         0         +1        +2

                                                Frequency Table

                          Percents                        Percents                        Percents                        Percents
          Value Count  Cell   Cum        Value Count  Cell   Cum        Value Count  Cell   Cum        Value Count  Cell   Cum
       0.778151     1   6.7   6.7     1.041393     1   6.7  33.3      1.20412     1   6.7  60.0      1.50515     2  13.3  86.7
       0.845098     1   6.7  13.3     1.079181     1   6.7  40.0     1.342423     1   6.7  66.7      1.60206     1   6.7  93.3
              1     2  13.3  26.7     1.146128     2  13.3  53.3     1.477121     1   6.7  73.3     1.623249     1   6.7 100.0
```

Figure 4-15. Log-transformed postimplementation data set.

To test the randomness of the data sets, the test described in Section 4.4.3 can be used. Since the test requires only the number of runs and the number of values above and below the median, it does not matter whether the raw data or transformed data are used. Using the raw data in Table 4-13, the number of runs for the pre-implementation data set is 6 while the number for the postimplementation data set is 9. The resulting *z* statistics (from Equation 4-24) for the preimplementation and postimplementation data sets are 1.5526 and 0.8971, respectively. These values are compared to a critical value of 1.96 (using $\alpha/2 = 0.025$) from Table D1 and the null hypothesis is accepted. Both samples are random.

Once the data sets are randomly sampled and normally distributed (after log-transformations), the one-sample hypothesis test using the log-transformed post-BMP data set can be performed. As shown in Figure 4-15, the mean of the log-transformed post-BMP data set is 1.21969 and the standard deviation is 0.273571. The log of the hypothesized value (25 mg/L) is 1.3979. Note that it is recommended that these values be rounded to the correct number of significant digits when reporting the results. The *t* statistic (Equation 4-27) is used to determine whether the post-BMP mean TSS concentration is less than 25 mg/L.

$$t = \frac{1.21929 - 1.3979}{0.273571 / \sqrt{15}} = -2.53$$

The schematic representation of this test is shown in Figure 4-16A, where the critical *t* value (-1.761) for the one-sided test (df = 14, $\alpha = 0.05$) is taken from Table D2. The computed *t* statistic falls to the left of the critical value, so the null hypothesis is rejected. In turn, the alternative hypothesis that the post-BMP mean TSS concentration is less than 25 mg/L is accepted.

Alternatively, had the participants in the Highland Silver Lake RCWP project selected a two-sided test where H_o and H_a are given as

H_o: $\mu = 25$ mg/L
H_a: $\mu \neq 25$ mg/L

a two-sided *t* test would be appropriate. The critical *t* value for the two-sided test from Table D2 (df = 14, $\alpha/2 = 0.025$) would be ± 2.145. In this case, the computed *t* statistic (-2.52) still falls outside this range and it is concluded that the post-BMP mean TSS concentration is less than 25 mg/L. Notice how the rejection region (shaded portion) in Figure 4-16B differs from Figure 4-16A. The total shaded area in the two curves is the same (i.e., 5 percent); however, it is in one piece in Figure 4-16A and is split into two parts in Figure 4-16B.

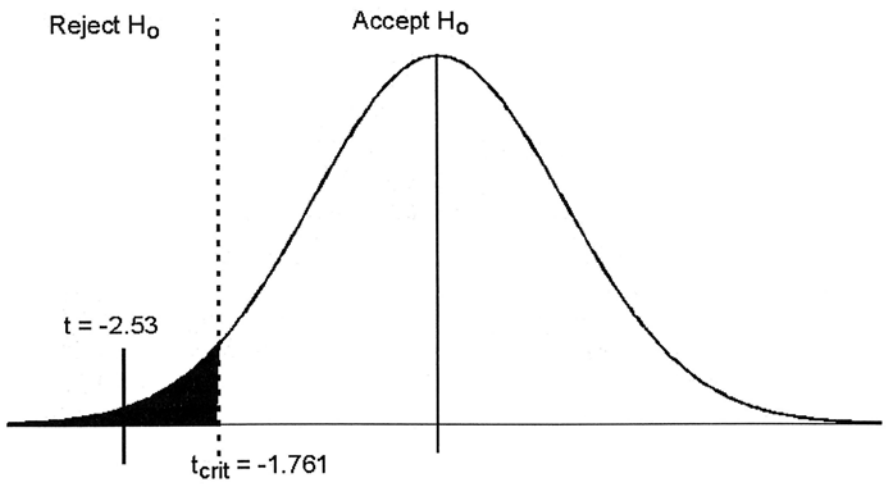
The power of this test can be evaluated using the noncentral *t* distribution with respect to various alternative hypotheses. The noncentral *T* statistic with *n*-1 degrees of freedom is given by

$$T_\Delta = \frac{\bar{x} - \mu_1 + \Delta}{s/\sqrt{n}} = \frac{\bar{x} - u_0}{s/\sqrt{n}} \qquad (4\text{-}28)$$

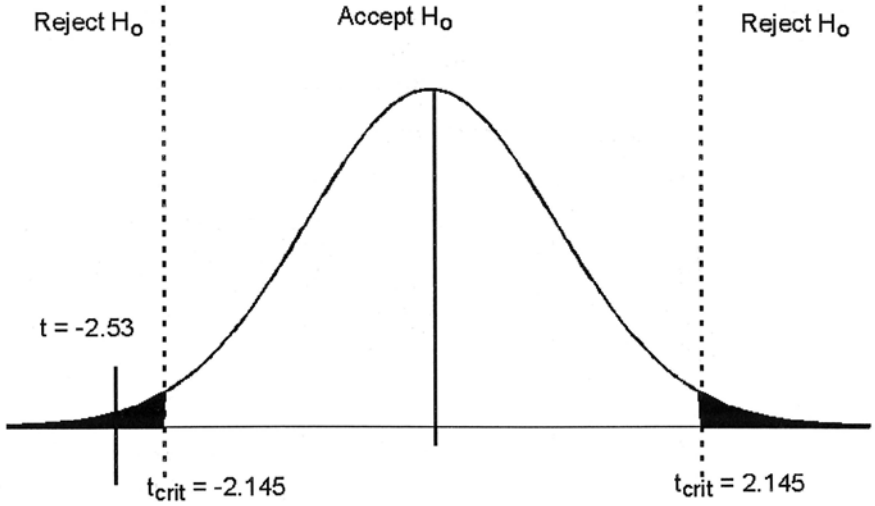
where $\Delta = \mu_1 - \mu_0$, the difference between the real and hypothesized mean. The noncentrality parameter (δ) is given by

$$\delta = \frac{\sqrt{n}\Delta}{s} \qquad (4\text{-}29)$$

Values of δ are given in Table D7 for a one-sided noncentral *t* distribution. Continuing with the current example, it is possible to develop a power curve that indicates the trade-offs between Type I and II errors. (Background discussion on power curves is provided in Section 4.1.1.) From Table D7 (df = 14, $\alpha = 0.05$), one value of δ is obtained



A) One-sided *t*-test with 14 degrees of freedom and α equal to 0.5.



B) Two-sided *t*-test with 14 degrees of freedom and α equal to 0.5.

Figure 4-16. One- and two-sided *t* test for post-BMP mean TSS concentration.

for each level of β (Table 4-14). In Table 4-14, power is computed as 1-β and Δ is obtained by rearranging Equation 4-29 and using *s* equal to 0.273571 and *n* equal to 15. Note that Δ, referred to as the minimum detectable difference, is in log-transformed units.

Power can be plotted as a function of the minimum detectable difference (see Figure 4-17). The dotted line indicates an approximate extrapolation back to α when the minimum detectable difference is equal to zero. Using the log-transformed postimplementation data, Δ is equal to 0.178 (= 1.3979 - 1.21969). Interpolating from Table 4-14 or Figure 4-17 yields that there is a 77 percent probability (i.e., power = 0.77) that a significant difference would be detected (i.e., reject H_o) if the difference between the estimated mean and true mean using log-transformed data were 0.178. For Δ less than 0.027, there is only a 10 percent chance of detecting a significant difference, whereas for Δ greater than 0.3 there is almost a 100 percent chance of detecting a significant difference.

Table 4-14. Evaluation of power using the post-implementation TSS data.

Power (1-β)	β	δ	Δ
0.10	0.90	0.38	0.027
0.20	0.80	0.84	0.059
0.30	0.70	1.18	0.083
0.40	0.60	1.46	0.103
0.50	0.50	1.73	0.122
0.60	0.40	2.00	0.141
0.70	0.30	2.28	0.161
0.80	0.20	2.62	0.185
0.90	0.10	3.08	0.218
0.95	0.05	3.46	0.244
0.99	0.01	4.18	0.295

Wilcoxon Signed Ranks test

Alternatively, if the log (or some other) transformation did not result in normally distributed data, the analyst could consider the Wilcoxon Signed Ranks test. Although less restrictive than the *t* test, this test requires that the

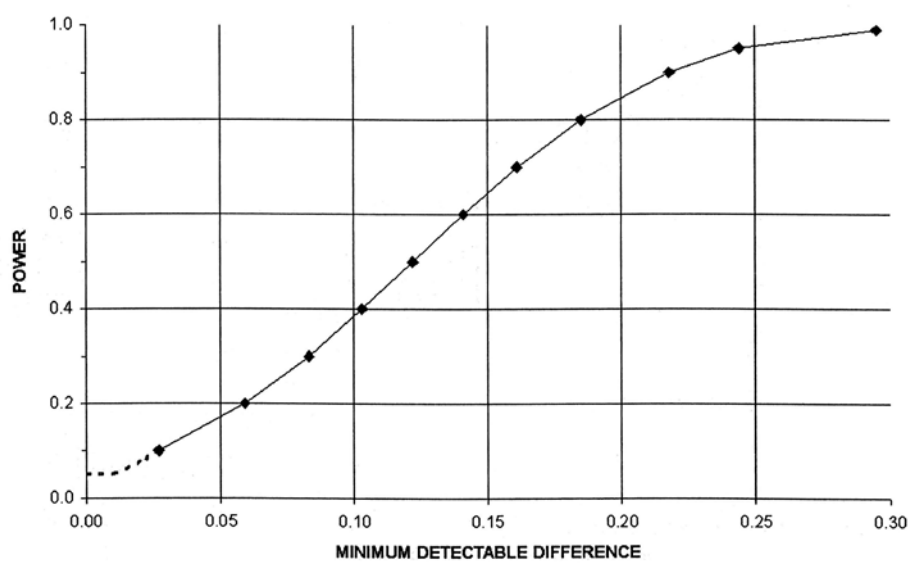


Figure 4-17. Evaluation of power using the log-transformed postimplementation TSS data.

data are independent and come from a symmetric distribution. As the name implies, a symmetric distribution is one in which the distribution of data above the midpoint is a mirror image of the data distribution below the midpoint. (The normal distribution is a special case of a sym-metric distribution.) When the data distribution is symmetric, the mean and median coincide and therefore inferences about the median are also valid for the mean (Conover, 1980). For this presentation, the median concentration is evaluated rather than the mean using the following hypotheses:

H_o: $P_{.50} \geq 25$ or H_o: $25\text{-}P_{.50} \leq 0$
H_a: $P_{.50} < 25$ H_a: $25\text{-}P_{.50} > 0$

The test statistic, T, is normally distributed and is given by Conover (1980) as

$$T = \frac{\sum_{i=1}^{n} Rank\,|d_i|}{\left[\sum_{i=1}^{n} Rank\,|d_i|^2\right]^{0.5}} \qquad (4\text{-}30)$$

where d_i is equal to the difference between the hypothesized value (25 mg/L) and the actual data and the rank is assigned a negative value if d_i is negative. El-Shaarawi and Damsleth (1988) provide a modified version of the Wilcoxon Signed Ranks test for use with serially correlated data.

Modification for Paired Data

d_i corresponds to the difference between paired observations (e.g., $d_i = x_i - y_i$).

From the previous example, it is already known that the raw postimplementation data are lognormal and thus not symmetric. Therefore, the log-transformed data are analyzed since it has already been determined that the log-transformed observations are symmetric as well as indepen-dent. Table 4-15 shows the calculations used to evaluate the log-transformed post-implementation data set. For convenience the data are sorted from smallest to largest observation. The dif-ference, d_i, is computed as log(25) - log(TSS_i). For example, the first entry is equal to log(25) - log(6) or 0.620. Since the log-transformed data were symmetric, d_i will also be symmetric. The fourth column is the absolute value of the dif-ference, $|d_i|$. The last two columns are the rank and rank-squared of $|d_i|$ where the rank is assigned a negative value if d_i is negative. T is equal to $76/(1238.5)^{0.5}$ or 2.16. Since 2.16 is greater than 1.645 (which is obtained from Table D1 using $\alpha = 0.05$), the null hypothesis is rejected and it is concluded that the median concentration is less than 25 mg/L. Had the raw data that are not symmetric been incorrectly used, T would have been equal to 1.54 and the null hypothesis would have been incorrectly accepted.

Sign test

Suppose that the postimplementation data could not be transformed into a symmetric distribution. By using the sign test, the symmetric distribution assumption can be relaxed (i.e., it is not required). In this case, the appropriate hypotheses for a one-sided test are

H_o: $P(+) \geq P(-)$
H_a: $P(+) < P(-)$

where P(+) is defined as the probability of an observation's being greater than the hypothesized value (in this case 25 mg/L). As stated, H_o implies that 50 percent or more of the population is greater than or equal to the hypothesized value.

Table 4-15. Nonparametric evaluation of postimplementation data using the Wilcoxon Signed Ranks test.

TSS (mg/L)	Log(TSS)	d_i=Log(25) -Log(TSS_i)	$\lvert d_i \rvert$	rank $\lvert d_i \rvert$ [a]	rank $\lvert d_i \rvert$ [2]
6	0.778	0.620	0.620	15	225
7	0.845	0.553	0.553	14	196
10	1.000	0.398	0.398	12.5	156.25
10	1.000	0.398	0.398	12.5	156.25
11	1.041	0.357	0.357	11	121
12	1.079	0.319	0.319	10	100
14	1.146	0.252	0.252	8.5	72.25
14	1.146	0.252	0.252	8.5	72.25
16	1.204	0.194	0.194	5	25
22	1.342	0.056	0.056	1	1
30	1.477	-0.079	0.079	-2	4
32	1.505	-0.107	0.107	-3.5	12.25
32	1.505	-0.107	0.107	-3.5	12.25
40	1.602	-0.204	0.204	-6	36
42	1.623	-0.225	0.225	-7	49
			SUM =	76	1238.5

[a] Assign the negative of the rank if d_i is negative.

Modification for Paired Data

The comparison is made between the paired observations rather than with a hypothesized value.

By comparing each observation from the random sample to the hypothesized value, the data set is converted into a series of "+," "-," and ties. The test statistic, *T*, is equal to the number of "+." The more "+" that result from the comparisons, the more H_o is supported.

Using the raw postimplementation data, *T* is equal to 5 and *n* is equal to 15. There are no ties. In this one-sided test, small values of *T* indicate that "-" are more probable. For sample sizes less than 20, use Table D8 with *p* equal to 0.5 and *n* equal to the number of "+" and "-" (ties are excluded). Find the table entry, *y*, that approximately equals α, rejecting H_o if $T \le y$. If *n* is greater than 20, *y* can be computed as

$$y = \frac{1}{2}(n + Z_\alpha \sqrt{n}) \qquad (4\text{-}31)$$

were Z_α is obtained from Table D1. For example, if α is equal to 0.05 in a one-sided test, $Z_{0.05}$ is equal to -1.645. Using the example data, a *y* equal to 4 (α=0.0592) is obtained from Table D7. *T* is greater than 4, so H_o is accepted.

Had the hypotheses been stated in the other direction (i.e., H_o: $P(+) \le P(-)$; H_a: $P(+) > P(-)$), H_o would be rejected if $T \ge n - y$. Had this been a two-sided test, the rejection region would be for $T \le y$ or $T \ge n - y$ where *y* is obtained from Table D8 or Equation 4-31 using α/2.

Table 4-16 presents paired observations for BOD_5 collected at two locations from the same stream. In this case, the hypothesis that there is no difference in BOD_5 concentrations between the two locations with a $\alpha = 0.10$ is being tested:

Hypotheses	Description
H_o: P(+) = P(-)	BOD_5 concentrations at the two locations are the same.
H_a: P(+) ≠ P(-)	BOD_5 concentrations at location 1 tends to be larger or smaller than the BOD_5 concentration at location 2.

In this case, a two-sided test is appropriate where P(+) indicates the probability that an observation from location 1 is greater than an observation from location 2. The fourth column indicates whether the BOD_5 concentration at location 1 is larger (+), smaller (-), or equal to (tie) the BOD_5 concentration at location 2. In this analysis there are 8 "+" and a total of 13 observation pairs without ties. From Table D8 with $\alpha/2 = 0.05$ and $n = 13$, $y = 3$ ($\alpha = 0.0461$) is obtained. H_o is accepted since $3 \leq 8 \leq (13-3)$.

Comparison of example results

In this case, the Student's *t* test and Wilcoxon Signed Ranks test give the same conclusion. It is proposed that the results from the *t* test are more appropriate for this example since all of the assumptions of the parametric test were met. Had the assumptions not been met, the results from the Wilcoxon Signed Ranks test would have been more appropriate. That is, if all assumptions are met, parametric procedures are more powerful than their nonparametric alternative. The sign test, while not incorrect, was not a good choice for the example data because the distributional assumptions were met and more powerful tests could be applied. Applying the Wilcoxon Signed Ranks test to data that are not symmetric results in a level of significance (α) that is somewhat lower than what is specified, whereas applying the *t* test to data that are not normally distributed results in an α that is much larger than specified (Helsel and Hirsch, 1995).

Table 4-16. Sign test for comparing paired BOD_5 concentrations.

Day	Conc. at Location 1 (mg/L)	Conc. at Location 2 (mg/L)	Sign of Difference
1	29	19	+
2	22	20	+
3	10	5	+
4	26	24	+
5	12	15	-
6	32	24	+
7	23	25	-
8	11	23	-
9	32	32	tie
10	27	30	-
11	28	20	+
12	23	16	+
13	18	33	-
14	35	25	+
15	20	20	tie

4.5.2 Two-sample Tests

In many instances, paired observations are not a practical or appropriate sampling methodology. Instead, two random samples are collected. The pre- and postimplementation data in Table 4-13 from the Highland Silver Lake RCWP are one example. The Student's *t* test for two samples and the Mann-Whitney test are the most appropriate tests for these types of data.

Two-sample *t* test

Suppose that a comparison of the pre- and post-implementation TSS data sets is desired to see if the BMPs have had an effect on TSS levels in Highland Silver Lake. Remembering the assumptions made earlier about using the mean TSS concentration as a good measure of central tendency and assuming that any change in TSS mean concentration is due to the BMP alone, the pre- and postimplementation data sets can be used in a one-sided hypothesis:

H_o: TSS (Post) $\geq$ TSS (Pre) or
H_o: TSS (Post) - TSS (Pre) ≥ 0

H_a: TSS (Post) < TSS (Pre) or
H_a: TSS (Post) - TSS (Pre) < 0

Note that in this case the H_o that the post-implementation TSS is greater than or equal to the preimplementation TSS concentration is tested with an H_a that postimplementation TSS is lower. The results from this analysis will be interpreted as simply indicating whether the BMPs worked. This could also have been set up as a two-sided test where H_o and H_a would be

H_o: TSS (Post) = TSS (Pre) or
H_o: TSS (Post) - TSS (Pre) = 0

H_a: TSS (Post) $\neq$ TSS (Pre) or
H_a: TSS (Post) - TSS (Pre) $\neq 0$

With confidence that the BMP would have only an effect of reducing TSS concentrations, H_o is tested using a one-sided *t* test. Both the pre-implementation and postimplementation data sets are random samples and normal when log-transformed. However, the two-sample *t* test also requires that the variances of the two populations be equal (Gaugush, 1986). Since a major effect of many nonpoint source control practices is to reduce the occurrence of large loading events, it is very likely that these practices will have an effect on the variance of nonpoint source loads. Thus, an *F* test is performed to evaluate variance homogeneity before proceeding with the *t* test even though the *t* test is robust with respect to moderate departures from homogeneous variance (Winer, 1971).

Since the log-transformed data (Figures 4-14 and 4-15) are being used, the variance of the transformed data must also be used in the *F* test. The resulting *F* statistic is computed from Equation 4-23:

$F_1 = 0.075/0.057 = 1.32$

The variances are substituted into Equation 4-23 so that the *F* statistic is greater than unity to account for the organization of Table D6. The critical *F* value from Table D6 ($f_n = 14$, $f_d = 15$, $\alpha/2 = 0.025$) is 2.89. The value 1.32 is compared to 2.89, and the null hypothesis of equal variance is accepted.

Tests for Two Independent Random Samples

Test[a]	Key Assumptions
Two-Sample *t*	• Both data sets must be normally distributed • Data sets should equal variances[b]
Mann-Whitney	• None

[a] The standard form of these tests requires independent random samples.
[b] The variance homogeneity assumption can be relaxed (see Table 4-17).

Satisfied that the data meet all of the assumptions required of the two-sample hypothesis test, H_o (TSS (Post) ≥ TSS (Pre)) is now tested. The two-sample *t* statistic with n_1+n_2-2 degrees of freedom is (Remington and Schork, 1970)

$$t = \frac{(\overline{x_1} - \overline{x_2}) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (4\text{-}32)$$

where s_p is the pooled standard deviation, which is defined by

$$s_p = \left[\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \right]^{0.5} \qquad (4\text{-}33)$$

The difference quantity (Δ_o) can be any value, but in this case it is set to zero. Δ_o can be set to a non-zero value to test whether the difference between the two data sets is greater than a selected value. Using the transformed data for preimplementation (n_1 = 16, s_1^2 = 0.057087, $\overline{x}_1$ = 1.407) and postimplementation conditions (n_2 = 15, s_2^2 = 0.074812, $\overline{x}_2$ = 1.21969), s_p is calculated as

$$s_p = \left[\frac{0.057087(16-1) + 0.074812(15-1)}{16 + 15 - 2} \right]^{0.5} = 0.2562$$

and the *t* statistic is calculated as

$$t = \frac{(1.407 - 1.21969) - 0}{0.2562 \sqrt{\frac{1}{16} + \frac{1}{15}}} = 2.034$$

Comparing this *t* statistic in a one-tailed test to the *t* value from Table D2 (α = 0.05, df = $n_1 + n_2 - 2$ = 29), it is found that the 2.034 exceeds the table value of 1.6991. Therefore, the null hypothesis is rejected and it is concluded that the postimplementation mean log-transformed TSS concentration is lower than the preimplementation level (i.e., the BMPs worked given earlier assumptions). Note that if a two-tailed test had been used, the null hypothesis would have been accepted since the corresponding *t* value from Table D2 is 2.0452. Remington and Schork (1970) give test statistics for other cases in which the difference between means is being tested. These cases and corresponding equations are given in Table 4-17. In particular, note Case #3, which allows for unequal variances.

The power of this test can be estimated using the noncentrality parameter (Larsen and Marx, 1981):

$$\delta = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{2}\ \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (4\text{-}34)$$

where σ is approximated with the pooled standard deviation. Using the data in this example,

$$\delta = \frac{1.407 - 1.21969}{\sqrt{2}\ 0.2562 \sqrt{\frac{1}{16} + \frac{1}{15}}} = 1.438$$

From Table D7 (df = 29, α = 0.05), a β approximately equal to 0.60 is obtained, so the power is equal to 0.40. Had the difference in

Table 4-17. Summary of parametric tests used to evaluate difference between means (Remington and Schork, 1970).

Case 1: Difference between means when variances are known (test statistic is standard normal distribution)		
Null Hypothesis	**Test Statistic**	**Assumptions**
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{(\sigma_1^2/n_1 + \sigma_2^2/n_2)^{0.5}}$	Independent, random samples of size n_1 and n_2 from two normally distributed populations.
Case 2: Difference between means when variances are unknown but equal (test statistic is Student's *t* distribution with n_1+n_2-2 degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{s_p(1/n_1 + 1/n_2)^{0.5}}$	Independent, random samples of size n_1 and n_2 from two normally distributed populations with equal variances
Case 3: Difference between means when variances are known and unequal (test statistic is approximately Student's *t*; see below for degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_1 - \mu_2 = \Delta_o$	$t = \dfrac{(\overline{x_1} - \overline{x_2}) - \Delta_o}{(s_1^2/n_1 + s_2^2/n_2)^{0.5}}$ $df^* = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}}$ $df = df^* - 2$	Independent, random samples of size n_1 and n_2 from two normally distributed populations with unknown and presumably unequal variances
Case 4: Pairing—the mean difference (test statistic is Student's *t* distribution with *n*-1 degrees of freedom)		
H_o	**Test Statistic**	**Assumptions**
$\mu_d = \Delta_o$ $\mu_d = \mu_2 - \mu_1$	$t = \dfrac{\overline{d} - \Delta_o}{s_d/n^{0.5}}$	Random sample of size *n* paired differences from a normally distributed populations of differences

means (of the log-transformed data) been larger (e.g., 0.30), δ would be 2.31 and the power would be equal to 73 percent.

Mann-Whitney (Wilcoxon's rank sum) test

The Mann-Whitney test can also be used to compare two independent random samples. This test is very flexible since there are no assumptions about the distribution of either sample or whether the distributions have to be the same (Helsel and Hirsch, 1995). Wilcoxon (1945) first introduced this test for equal-sized samples. Mann and Whitney (1947) modified the original Wilcoxon's test to apply it to different sample sizes. This test tests whether one data set tends to have larger observations than the other. Example two- and one-sided hypotheses are as follows:

Two-sided

H_o: Prob [TSS (Post) > TSS (Pre)] = 0.5

Description: The probability that the post-implementation TSS is larger than the pre-implementation TSS is equal to 50 percent.

H_a: Prob [TSS (Post) > TSS (Pre)] ≠ 0.5

Description: The postimplementation TSS is larger or smaller than the preimplementation TSS.

One-sided

H_o: Prob [TSS (Post) > TSS (Pre)] ≥ 0.5

Description: The probability that the post-implementation TSS is larger than the pre-implementation TSS is equal to or greater than 50 percent.

H_a: Prob [TSS (Post) > TSS (Pre)] < 0.5

Description: The postimplementation TSS is smaller than the preimplementation TSS.

If the distributions of the two samples are similar except for location (i.e., similar spread and skew), H_a can be refined to imply that the median concentration from one sample is "greater than," "less than," or "not equal to" the median concentration from the second sample. To achieve this greater detail in H_a, transformations such as logs can be used.

Table 4-18 shows the intermediate calculations using the same TSS data presented earlier. First, all observations from the pre- and post-implementation are sorted together and ranks assigned. Note that ties are assigned the average rank. The test statistic is equal to the sum of the ranks for the group with the smaller number of observations—in this case, the postimplementation data set.

Tables of Mann-Whitney test statistics (e.g., Conover, 1980) may be consulted to determine whether to reject H_o for small sample sizes. If n_1 and n_2 are greater than or equal to 10 observations, the test statistic can be computed from the following equation (Conover, 1980):

$$T_1 = \frac{T - n_1 \frac{N+1}{2}}{\sqrt{\frac{n_1 n_2}{N(N-1)} \sum_{i=1}^{N} R_i^2 - \frac{n_1 n_2 (N+1)^2}{4(N-1)}}} \tag{4-35}$$

where

n_1 = number of observations in sample with fewer observations (e.g., post-implementation);
n_2 = number of observations in sample with more observations (e.g., pre-implementation);
N = $n_1 + n_2$;
T = sum of ranks for sample with fewer observations; and
R_i = rank for the ith ordered observation used in both samples.

Table 4-18. Nonparametric evaluation of post-implementation data using the Mann-Whitney test.

Rank	Pre-Implement. TSS (mg/L)	Post-Implement. TSS (mg/L)	Rank	Pre-Implement. TSS (mg/L)	Post-Implement. TSS (mg/L)
1	-	6	17	21	-
2	-	7	18	-	22
3.5	-	10	19	25	-
3.5	-	10	20.5	30	-
5.5	11	-	20.5	-	30
5.5	-	11	22.5	-	32
7.5	12	-	22.5	-	32
7.5	-	12	24	35	-
9	13	-	25	37	-
10.5	-	14	26	-	40
10.5	-	14	27.5	42	-
12.5	16	-	27.5	-	42
12.5	-	16	29	48	-
15	20	-	30.5	60	-
15	20	-	30.5	60	-
15	20	-			
Sum of ranks for post-implementation, T = 193.5 Sum of all ranks squared, ΣR_i^2 = 10,409.5					

This equation is appropriate for situations when there are many ties. Applying this equation yields

$$T_1 = \frac{193.5 - 15\dfrac{31+1}{2}}{\sqrt{\dfrac{15 \bullet 16}{31(31-1)}\,10409.5 - \dfrac{15 \bullet 16(31+1)^2}{4(31-1)}}} = -1.84$$

T_1 is normally distributed, and Table D1 can be used to determine the appropriate quantile. Since the test was one-sided and α is equal to 0.05, the appropriate quantile from Table D1 is -1.645. T_1 is less than -1.645, and therefore the null hypothesis is rejected. The post-implementation TSS concentrations are significantly less than the pre-implementation TSS concentrations. Had a two-sided test been used, the appropriate quantile from Table D1 would have been -1.96 and the H_o would have been accepted. In this case, the two-sample *t* test and the Mann-Whitney test result in the same conclusion.

4.5.3 Magnitude of Differences

So far, Section 4.5 has described statistical tests for comparing one and two random samples for significant differences. A question remains: How big is the difference? For data that are normally distributed, the difference can be computed as the difference between the two sample means. The confidence interval (CI) for the differences can be computed under the equal variance scenario as (Winer, 1971):

$$CI = (\overline{x_1} - \overline{x_2}) \pm t_{(\alpha/2, n_1+n_2-2)}\, s_p \sqrt{1/n_1 + 1/n_2} \qquad (4\text{-}36)$$

If the standard deviations were not similar, the CI would be

$$CI = (\overline{x_1} - \overline{x_2}) \pm t_{(\alpha/2, df)} \sqrt{s_1^2/n_1 + s_2^2/n_2} \qquad (4\text{-}37)$$

where df is from Table 4-17 (Case 3).

Helsel and Hirsch (1995) recommend that a Hodges-Lehmann estimator ($\hat{\Delta}$) be used if the data have been transformed for testing or if the data are not normally distributed. The Hodges-Lehmann estimator (Hodges and Lehmann, 1963) can be used as a nonparametric estimator of the difference between the two samples. To compute the Hodges-Lehman estimate, the analyst computes the difference between all n_1 and n_2 observations. Using the TSS data used earlier, there are 16•15 or 240 differences to compute. The Hodges-Lehmann estimator is the median of these differences or 8 mg/L. This estimator is preferred to the difference between the medians of the random samples (Helsel and Hirsch, 1995). For sample sizes larger than 10, the upper and lower confidence intervals for $\hat{\Delta}$ can be estimated:

$$R_l = \frac{n_1 n_2 - z_{\alpha/2}\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{3}}}{2} \qquad (4\text{-}38)$$

$$R_u = \frac{n_1 n_2 + z_{\alpha/2}\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{3}}}{2} + 1 = n_1 n_2 - R_l + 1 \qquad (4\text{-}39)$$

where R_l and R_u correspond to the lth and u^{th} ranked difference. The 95 percent confidence interval for the difference between the pre- and post-implementation data would be computed as

$$R_l = \frac{16 \cdot 15 - 1.96\sqrt{\dfrac{16 \cdot 15(16+15+1)}{3}}}{2} = 70.4 \approx 70$$

$$R_u = 16 \cdot 15 - 70.4 + 1 = 170.6 \approx 171$$

Therefore, the confidence interval on the median difference is equal to the 70th and 171st ranked difference or $-1 \le \hat{\Delta} \le 19$.

4.6 Comparison of More Than Two Independent Random Samples

The analysis of variance (ANOVA) and Kruskal-Wallis are extensions of the two-sample *t* and Mann-Whitney tests, respectively, and can be used for analyzing more than two independent random samples. Unlike the *t* test described earlier, the ANOVA can have more than one factor or explanatory variable. In the Highland Silver Lake RCWP project example used in Section 4.5, one factor described whether the data were collected before or after implementation of a BMP. In the example that will be analyzed in this section, trout population, there are two factors. One factor is based on the stream from which the trout were collected; the other factor is based on the region from which the trout were collected. The Kruskal-Wallis test accommodates only one factor, whereas the Friedman test can be used for two factors. In addition to applying one of the above tests to determine whether one of the samples is significantly different from the others, it is also necessary to do postevaluations to determine which of the samples is different. This section recommends Tukey's method to analyze the raw or rank-transformed data only if one of the previous tests (ANOVA, rank-transformed ANOVA, Kruskal-Wallis, or Friedman) indicates a significant difference between groups. The reader is cautioned that when performing an ANOVA using standard software, the ANOVA test used must match the data.

4.6.1 One-Factor Comparisons

ANOVA

The ANOVA for one factor is a procedure for comparing the mean value from each group with the overall mean. H_o is typically stated that there are no differences between the group means, whereas H_a states that at least one group's mean is significantly different from the overall mean or

H_o: $\mu_1 = \mu_2 = ... = \mu_k$.
H_a: At least one group mean is different.

The basic assumptions made in using an ANOVA are as follows (Remington and Schork, 1970):

- Each sample is a random sample from the corresponding population, and observations from different populations are independent.
- The measurement variable is normally distributed in each of the *k* groups.
- The groups have the same variance (homoscedasticity).

The variation (or total noise) in the data can be split into the treatment sum of squares (SST) and the errors sum of squares (SSE) (see Equation 4-40) (Helsel and Hirsch, 1995) where

k = number of groups,
n_j = number of observations in the jth group,
$\bar{x}_j$ = mean of the jth group.
x_{ij} = i^{th} observation in the jth group,
$\bar{x}$ = overall mean, and

This notation is also used in Table 4-19, which indicates each observation, group sample size, group sample mean, and group true mean. Note that sample sizes for the different groups need not be the same. The reader should compare the notation in Table 4-19 to that used in Equation 4-40.

The observations (x_{ij}) within each group are assumed normally distributed about the mean, μ_j and variance, σ^2. The variance is the same for all classes, but the mean can vary among classes. The overall mean is denoted as μ, and the corresponding linear model is expressed as (Snedecor and Cochran, 1980)

$$x_{ij} = \mu + \alpha_j + \varepsilon_{ij} \qquad (4\text{-}41)$$
$$[\, i = 1,...n_j;\ j = 1,...k;\ \varepsilon_{ij} = N(0,\sigma^2)]$$

This fixed effects model shows that each observed value is the sum of an overall mean (μ), a treatment or class deviation (α_j), and a random element (ε_{ij}) from a normally distributed population with a zero mean and a standard deviation equal to σ. The model is referred to as "fixed" because the α_j, while unknown, are constant for a group. The random element represents variations due to such factors as unit-to-unit variation in treatment effect, measurement errors, or individual characteristics of the unit (Snedecor and Cochran, 1980). To detect a significant difference, the variation within the group (i.e., ε_{ij}) must be sufficiently smaller than the variation between groups.

$$\textit{Total sum of squares} = SST + SSE$$
$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 = \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{x})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2 \qquad (4\text{-}40)$$

Table 4-19. ANOVA notation.

	Factor Level			
	1	2	...	k
	x_{11}	x_{12}	...	x_{1k}
	x_{21}	x_{22}		x_{2k}
	.	.		.
	.	.	...	.
	.	.		.
	$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_k k}$
Sample size:	n_1	n_2	...	n_k
Sample mean:	$\overline{x}_1$	$\overline{x}_2$	...	$\overline{x}_k$
True mean:	μ_1	μ_2	...	μ_k

The ANOVA test statistic, *F*, is based on a ratio of the treatment mean squares (MST) and error mean squares (MSE):

$$F = \frac{MST}{MSE} \tag{4-42}$$

where

$$MST = \frac{\sum_{j=1}^{k} n_j(\overline{x}_j - \overline{x})^2}{k-1} = \frac{SST}{k-1} \tag{4-43}$$

$$MSE = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2}{N-k} = \frac{SSE}{N-k} \tag{4-44}$$

and *N* is the total number of observations.

An *F* value of 1 represents the condition where H_o is true, and large *F* values indicate differences among the μ_j. Snedecor and Cochran (1980) note that the *F* test is more affected by nonnormality and heterogeneity of variances when sample sizes, n_j, are not equal.

Table 4-20 presents a common format for the results from a one-factor ANOVA analysis generated by typical software. The first column identifies which portion of the linear model is being displayed and corresponds to the top portion of Equation 4-40. The second column presents the sum of squares for each source of variation, the third column presents the degrees of freedom, and the fourth column presents the treatment and error mean squares (Equations 4-43 and 4-44). *F* is calculated using Equation 4-42. The *p* value corresponds to the significance level associated with the computed *F*. The "*F* crit" corresponds to the critical value from Table D6 using *k*-1 and *N*-*k* degrees of freedom and a selected α. Note that some software packages do not present "*F* crit." If the *p* value is less than the selected α, H_o is rejected because at least one of the groups has a different mean.

As an example one-factor ANOVA, consider the situation where the trout populations of three streams are measured by the multiple-step Zippin

Table 4-20. Common one-way ANOVA output format.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Between Groups (Treatment)	SST	k-1	MST = SST/(k-1)	MST/MSE	p	F value for selected α
Within Groups (Error)	SSE	N-k	MSE = SSE/(N-k)			
Total	SST + SSE	N-1				

approach for electrofishing at five randomly selected sites in the Coastal Plain region (Platts et al., 1983). The data from this monitoring effort are shown in Table 4-21.

Using the one-factor ANOVA procedure from a standard spreadsheet, trout population as a function of stream was modeled to test the null hypothesis that stream has no effect on trout population (i.e., the treatment effect is zero). The results of this test are shown in Table 4-22. Note that the *F* value of 6.332 is equal to MST (92.867) divided by MSE (14.667). The *p* value is 0.013. The critical value from Table D6 with 2 and 12 degrees of freedom and $\alpha = 0.05$ is 3.885. H_o is rejected since at least one of the stream's trout populations has a different mean. Since H_o is rejected, it is appropriate to continue with postevaluations to determine which group has a different mean. Had H_o not been rejected, postevaluations would be meaningless and inappropriate.

One approach (Least Significant Difference) to determining which of the means is different is to compare each pair of means. To do a pairwise

Table 4-21. Trout population from streams in the coastal plain region.

	Stream		
	Black Creek	Blue Creek	Red Creek
Site	Trout Population (Pounds/Acre/Year - Year Class 2)		
1	60	49	50
2	65	60	56
3	64	54	51
4	63	58	60
5	58	57	52
$\bar{x}_j$	62.0	55.6	53.8

Table 4-22. One-way ANOVA of stream trout data from the coastal plain region using stream as the treatment.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Between Groups (Treatment)	185.733	2	92.867	6.332	0.013	3.885
Within Groups (Error)	176.000	12	14.667			
Total	361.733	14				

comparison, the standard error of the difference between two means is calculated as (Snedecor and Cochran, 1980)

$$s_{\overline{D}} = \sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \qquad (4\text{-}45)$$

with *N-k* degrees of freedom. Using the data from Table 4-22, $s_{\overline{D}}$ is equal to 2.422 with 12 degrees of freedom. For α/2 equal to 0.025 (and df = 12), the value of *t* from Table D2 is 2.1788. Therefore, if any pair of means exceeds a difference of 2.422 x 2.1788 or 5.3, the difference is significant. The mean trout populations for Black, Blue, and Red Creeks are 62, 55.6, and 53.8 pounds/acre/year, respectively. The trout population in Black Creek is significantly higher than the trout population in Blue Creek or Red Creek. Note that a pairwise comparison was made between the three groups (i.e., three pairwise comparisons) with α = 0.05; therefore, the overall error rate is $1-(1-0.05)^3$, or about 14 percent. Other approaches for multiple comparisons are discussed in Section 4.6.4.

Kruskal-Wallis test

The Kruskal-Wallis test is an extension of the Mann-Whitney test described earlier. This test can be used when there are several independent samples that do not have the same distribution. In this case, H_o and H_a are as follows:

H_o: All *k* groups have identical distributions.
H_a: At least one of the groups tends to yield larger observations than at least one other group.

If the distributions of all groups are similar except for location (i.e., similar spread and skew), H_a can be refined to imply that the median concentration from one group is different from the median concentration from at least one other group. To achieve this greater detail in H_a, transformations such as logs can be used.

Again consider the notation used in Table 4-19 where there are *k* groups and each group has n_j observations. *N* is the total number of observations. To compute the Kruskal-Wallis statistic, the following steps (Conover, 1980) can be used:

- Rank all of the data from lowest to highest, assigning the average of ranks to ties. The rank of observation x_{ij} is denoted as $R(x_{ij})$.

- Compute R_j for all k random samples using

$$R_j = \sum_{i=1}^{n_j} R(x_{ij}) \qquad for\ j = 1,2,...,k \tag{4-46}$$

- Compute the test statistic, T:

$$T = \frac{1}{S^2}\left[\sum_{j=1}^{k} \frac{R_j^2}{n_j} - \frac{N(N+1)^2}{4}\right] \tag{4-47}$$

where

$$S^2 = \frac{1}{N-1}\left[\sum_{j=1}^{k}\sum_{i=1}^{n_j} R(x_{ij})^2 - N\frac{(N+1)^2}{4}\right] \tag{4-48}$$

For $k = 3$, all n_j are 5 or less, and there are no ties, special tables should be used to determine the rejection region for T (see Conover, 1980). If these criteria do not apply, Table D3 with $p = 1-\alpha$ and k-1 degrees of freedom should be used. If the computed T statistic from Equation 4-47 is greater than the value obtained from the table, H_o is rejected.

Table 4-23 presents the rank of the trout population data used in the previous example; R_j for each group has already been computed. Applying Equation 4-48 with the individual ranks from Table 4-23 and $N = 15$, S^2 is equal to 19.82. Substituting S^2, $N = 15$, $n_j = 5$ (for all j) into Equation 4-47 along with the R_j summarized in Table 4-23, T is equal to 7.21. From Table D3 with $\alpha = 0.05$ and 2 degrees of freedom, the critical value is 5.991. H_o is rejected. Had there been no ties, the exact critical value would be 5.66 (Conover, 1980).

Since H_o has been rejected, it is acceptable to do a multiple comparisons evaluation. One approach is to compare the ranks from each pairwise group. The groups i and j are different if the following inequality is satisfied (Conover, 1980):

$$\left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right| > t_{1-(\alpha/2)}\left(S^2\frac{N-1-T}{N-k}\right)^{0.5}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{0.5} \tag{4-49}$$

In this example, all n_j are equal to 5 and the above equation can be reduced to

Table 4-23. Rank of trout population from streams in the coastal plain region.

	Stream		
	Black Creek	Blue Creek	Red Creek
Site	Rank		
1	11	1	2
2	15	11	6
3	14	5	3
4	13	8.5	11
5	8.5	7	4
R_j	61.5	32.5	26

$$\left|\frac{R_i}{5}-\frac{R_j}{5}\right| > 2.1788\left(19.82\frac{15-1-7.21}{15-3}\right)^{0.5}\left(\frac{1}{5}+\frac{1}{5}\right)^{0.5}$$

or

$$\left| R_i - R_j \right| > 23.08$$

where t is obtained from Table D2 with 15-3 degrees of freedom. By comparing the above result with R_j in Table 4-23, it can be concluded that the trout population in Black Creek is significantly greater than the trout population in Blue Creek or Red Creek.

4.6.2 Two-Factor Comparisons

ANOVA

In a two-way ANOVA the variation due to two factors is quantified. One factor cannot be a subset of the other factor. Subsetted factors are referred to as nested factors, a subject that is not considered here. The reader is referred to Gaugush (1986) and Snedecor and Cochran (1980) for more thorough discussions regarding factorial experiments and hierarchical arrange-ments for fixed effects models. In this section, Equation 4-41 is extended to include a second factor (Helsel and Hirsch, 1995; Snedecor and Cochran, 1980),

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad (4\text{-}50)$$

where $i = 1, ..., a$; $j = 1, ..., b$; and $k = 1, ..., n$. The number of levels in factors A and B are represented by a and b, respectively. There are axb treatment groups. The number of replicates is equal to n and is constant across all treatment levels i and j. That is, there are the same number of observations for each unique combination of factors A and B. In this case, each observed value is the sum of an overall mean (μ), the influence of the ith category of factor A (α_i), the influence of the jth category of factor B (β_j), the interaction effect between factors A and B ($(\alpha\beta)_{ij}$), and a residual error (ε_{ijk}). If $(\alpha\beta)_{ij}$ is equal to zero, there is no interaction. No interaction means that a change in factor B has the same impact on x_{ijk} regardless of factor A (and vice versa).

H_o is that all treatment groups have the same mean, whereas H_a indicates that at least one treatment group mean has a different mean. The two assumptions made using this model (Equation 4-50) is that (1) the effects are additive, and (2) the residuals are independent, random variables normally distributed with a zero mean and constant variance across all treatment groups (Snedecor and Cochran, 1980).

Helsel and Hirsch (1995) caution the practitioner that when evaluating data with unequal numbers of observations some smaller statistical packages incorrectly apply the balanced equations (equal number of observations) presented here to unbalanced data sets (unequal number of observations) without notice. Packages such as SAS and Minitab provide options for analyzing unbalanced data sets. Two-way ANOVA can be performed for two cases, one in which there is no interaction between the two variables and one in which there is an interaction between the two variables. The sum of squares for factor A (SSA), factor B (SSB), and the interaction between A and B (SSI) for a balanced data set including interaction can be computed using Equations 51 through 55 (Helsel and Hirsch, 1995).

Table 4-24 is an ANOVA table that incorporates the above equations into the second column, presents the degrees of freedom in column three, and provides the equations for the mean squared error terms and F statistics in the fourth and fifth columns.

$$SSA = \sum_{i=1}^{a} \frac{\left(\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{bn} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \tag{4-51}$$

$$SSB = \sum_{j=1}^{b} \frac{\left(\sum_{i=1}^{a}\sum_{k=1}^{n} x_{ijk}\right)^2}{an} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \tag{4-52}$$

$$SSI = Total\ SS - SSA - SSB - SSE \tag{4-53}$$

where

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (x_{ijk})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(\sum_{k=1}^{n} x_{ijk}\right)^2}{n} \tag{4-54}$$

and

$$Total\ SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (x_{ijk})^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}\right)^2}{abn} \tag{4-55}$$

Table 4-24. Common two-way ANOVA output format.

Source of Variation	SS	df	MS	F	p-value	F criteria
Factor A	SSA	a-1	MSA = SSA/ (a-1)	MSA/MSE	p	F value for selected α
Factor B	SSB	b-1	MSB = SSB/ (b-1)	MSB/MSE	p	F value for selected α
Interaction (Factor AxFactor B)	SSI	(a-1)x (b-1)	MSI = SSI/ [(a-1)x(b-1)]	MSI/MSE	p	F value for selected α
Error	SSE	ab(n-1)	MSE = SSE/ [ab(n-1)]			
Total	Total SS	abn-1				

To demonstrate this procedure, the two-way ANOVA procedure is applied to the data in Table 4-25. This data set includes the trout population data from three streams and three regions (i.e., $a = b = 3$). This test could reflect, for example, the hunch that regional effects on trout population differ across streams (e.g., perhaps the streams are impacted differently by point and nonpoint sources). In this experimental design, factor *A* is the region and factor *B* is the stream. Using standard statistical software, Table 4-26 presents the results of the two-way ANOVA calculations. The *p* values for the region, stream, and region x stream factors are 1.0×10^{-10}, 0.001, and 0.1458, respectively. Using $\alpha = 0.05$, H_o is rejected; there is a significant difference between treatment group means due to region and stream. The interaction of region and stream is not significant at the 95 percent confidence level. Based on this analysis, it is acceptable to perform a multiple comparisons analysis for regions and streams.

This ANOVA discussion is simple in many respects. For example, a balanced data set and a fixed effects model were analyzed. In situations

Table 4-25. Stream trout population.

		Site				
		1	2	3	4	5
Stream	**Region**	**Trout Population (Pounds/Acre/Year - Year Class 2)**				
Black Creek	Mountain	75	70	65	72	68
	Piedmont	68	72	70	70	67
	Coastal Plain	60	65	64	63	58
Blue Creek	Mountain	70	76	69	67	74
	Piedmont	64	66	60	69	62
	Coastal Plain	49	60	54	58	57
Red Creek	Mountain	68	70	63	65	70
	Piedmont	62	66	58	69	67
	Coastal Plain	50	56	51	60	52

Table 4-26. Two-way ANOVA of trout population data using an interaction term.

Source of Variation	SS	df	MS	F	p-value	F crit (α=0.05)
Region	1213.73	2	606.87	46.60	1.0E-10	3.26
Stream	219.73	2	109.87	8.44	0.0010	3.26
Region x Stream	94.93	4	23.73	1.82	0.1458	2.63
Error	468.80	36	13.02			
Total	1997.20	44				

where multiple variables are examined, a balanced data set is not likely to be feasible or economical. A key limitation of the fixed effects model is that inferences cannot be made beyond the groups being tested. In the trout population example, only statements about the three streams and three regions analyzed can be made. Nothing about a fourth stream or region can be inferred. If the three streams had been randomly selected from across the state with the intent of determining whether there was a spatial difference in trout population, the stream factor would have been a random factor rather than a fixed factor, and the calculation of the *F* statistics would be different. If both factors were random, the *F* statistics would use the mean squares for interaction (*MSI*) rather than the *MSE* as the denominator. If there were a mixture of fixed and random factors, the *F* statistic for the fixed factor would be computed with the *MSI* and the random factor would be computed with the *MSE* in the denominator (Helsel and Hirsch, 1995).

Ranked transformed ANOVA

To perform the ANOVA described in Section 4.6.2, the data in each treatment group must be normally distributed with a constant variance. If the data do not meet this requirement, it is possible to use transformations of the data such as logarithms to convert the data to a normal distribution with constant variance. The use of logarithms implies that the influences of each factor are multiplicative in the original units (Helsel and Hirsch, 1995; Snedecor and Cochran, 1980). Alternatively, the data can be rank-transformed (i.e., a rank from 1 to *N* can be assigned to the data) and a two-way ANOVA can be performed on the ranks. Rejection of H_o using an ANOVA on the rank-transformed data indicates that the medians differ between treatment groups. Helsel and Hirsch (1995) state that "rank transformation results in tests which are more robust to non-normality, and resistant to outliers and non-constant variance, than is ANOVA without transformations."

4.6.3 Matched Data

Collecting paired data to mask or block out unwanted noise due to meteorological or geographical differences is a common practice when comparing "before" and "after" data. Comparing just two groups was described in Section 4.5.1. Comparing matched data with more than two groups is described here. In this case, the objective is to compare one factor (referred to as the treatment) while blocking out the other factor (referred to as the block).

The linear model for this analysis is (Helsel and Hirsch, 1995)

$$x_{ij} = \mu + \alpha_j + \beta_i + \varepsilon_{ij} \qquad (4\text{-}56)$$

where $j = 1, ..., k$ and $i = 1, ..., n$. In this case, each observed value is the sum of an overall mean (μ), the influence of the jth group effect (α_j), the influence of the ith block effect (β_i), and a residual error (ε_{ij}). In addition to the two-way ANOVA without replication and the Friedman test described here, Helsel and Hirsch (1995) also describe the median polish and the median aligned-ranks ANOVA.

Two-way ANOVA without replication

In the ANOVA model, ε_{ij} is assumed to be normally distributed. The sums of squares for the two-way ANOVA without replication are computed using Equations 57 through 60 (Helsel and Hirsch, 1995). Table 4-27 presents a common format for a two-way ANOVA without replication. Removing the block effect from the calculation of the *SSE* results in a higher *F* statistic, thus improving the detection of significant differences

$$SST = \frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{n} x_{ij}\right)^2}{n} - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \tag{4-57}$$

$$SSB = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} x_{ij}\right)^2}{k} - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \tag{4-58}$$

$$SSE = Total\ SS - SST - SSB \tag{4-59}$$

where

$$Total\ SS = \sum_{j=1}^{k}\sum_{i=1}^{n} (x_{ij})^2 - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn} \tag{4-60}$$

Table 4-27. Common two-way ANOVA without replication output format.

Source of Variation	SS	df	MS	F	p-value	F Criteria
Treatment	SST	k-1	MST = SST/ (k-1)	MST/MSE	p	F value for selected α
Block	SSB	n-1	MSB = SSB/ (n-1)			
Error	SSE	(k-1) x (n-1)	MSE = SSE/ [(k-1)x(n-1)]			
Total	Total SS	kn-1				

between groups. H_o is rejected if the computed F is greater than the critical F value from Table D6 with (k-1) and (k-1)(n-1) degrees of freedom.

Friedman test

The Friedman test is the most common nonparametric test for randomized complete block designs. It is an extension of the sign test (Helsel and Hirsch, 1995). H_o is that the median value of the k groups are identical, whereas H_a states that at least one median is different. To compute the test statistic, the following steps are used:

- Rank the data in each block from 1 to k.
- Compute the average rank for each group ($\overline{R}_j$).
- Compute χf using the following formula, which accounts for ties:

$$\chi f = \frac{12n}{k(k+1) - \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} (t_{ij}(j^3 - j))} \sum_{j=1}^{k} \left[\overline{R}_j - \frac{k+1}{2} \right]^2 \qquad (4\text{-}61)$$

where t_{ij} equals the number of ties of the extent j in block i. For $k+n \leq 9$, exact tables should be used (see Helsel and Hirsch, 1995). Otherwise, H_o is rejected if χf greater than or equal to the critical F value from Table D6 with (k-1) and (n-1)(k-1) degrees of freedom and p = 1-α.

4.6.4 Multiple Comparisons

All of the hypothesis tests featured to this point allow the analyst to determine whether at least one treatment results in a mean or median that is significantly different from that which results from the other treatments. It does not indicate which treatment is different or whether there are multiple differences. Multiple comparisons should be done only if the analyses performed under Section 4.6.1, 4.6.2, or 4.6.3 indicate a significant difference.

Two key features distinguish multiple comparisons: (1) whether α is based on a pairwise or overall comparison and (2) whether the test is a multiple-stage test (MST) or a simultaneous inference method (SIM). An important distinction should be made about whether a pairwise or overall α is used. The α level indicates the probability of making an incorrect comparison. Helsel and Hirsch (1995) cite an example of a one-factor analysis with six groups (in which there are 15 pairwise comparisons). If α = 0.05, the potential for making at least one error is equal to 1-(1-.05)^{15} or 0.54, a 54 percent chance of making one error. MSTs are valid for groups with constant sample size, whereas SIMs are valid for equal and unequal sample sizes.

For these reasons, Helsel and Hirsch (1995) recommend using Tukey's method, which uses an overall α and is a SIM. Other tests include the Bonferroni t tests, Duncan's multiple range test, Gabriel's multiple-comparison procedure, the Ryan-Einot-Gabriel-Welsch (REGW) multiple F test, the REGW multiple range test, Scheffe's multiple-comparison procedure, and the Waller-Duncan k-ratio test. The reader should consult statistics texts (e.g., Snedecor and Cochran, 1980) to learn more about these procedures, with preference given to Tukey's method for equal or unequal sample sizes and the REGW tests when the sample sizes are equal. If a nonparametric analysis was performed, the most appropriate approach is to rank-transform the data and apply a test based on the above discussion.

Tukey's method indicates that the mean between two groups can be considered different if (Helsel and Hirsch, 1995)

$$\left| \overline{x}_i - \overline{x}_j \right| > q_{(1-\alpha),k,N-k} \sqrt{MSE \, \frac{n_i + n_j}{2 n_i n_j}} \qquad (4\text{-}62)$$

where

q	=	studentized range statistic from Neter, Wasserman, and Kutner (1985);
α	=	overall significance level;
k	=	number of treatment group means compared;
N-k	=	MSE's degrees of freedom; and
n_i, n_j	=	sample size of group i and j respectively.

Typically, the results of a multiple comparison are commonly displayed using letters to distinguish groups. For example,

$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
A	AB	B	C

indicates that $\bar{x}_1 \approx \bar{x}_2$; $\bar{x}_1 < \bar{x}_3$; $\bar{x}_2 \approx \bar{x}_3$; and $\bar{x}_4$ is greater than $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$. The letter groupings could also be placed in boxplots.

4.7 REGRESSION TECHNIQUES

4.7.1 Overview

Regression can be used to model or predict the behavior of one or more variables. The general regression model, where ε is an error term, is given as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon \qquad (4\text{-}63)$$

In this equation, the behavior of a single dependent variable (y) is modeled with one or more independent variables (x_1, ..., x_n). The x's may be linear or nonlinear (e.g., x_i can represent x^2, x^3, x^{-1}, etc.). β_0, ..., β_n are numerical constants that are computed using equations described later. Nonlinear models are commonly applied to physical systems, but they are somewhat more difficult to analyze because iterative techniques are involved when the model cannot be transformed to a linear model. The use of two or more independent variables (x) in a linear function to describe the behavior of y is referred to as multiple linear regression. In either case, regression techniques attempt to explain as much of the variation in the dependent variable as possible.

In nonpoint source analyses, linear regression is often used to determine the extent to which the value of a water quality variable (y) is influenced by land use or hydrologic factors (x) such as crop type, soil type, percentage of land treatment, rainfall, or stream flow, or by another water quality variable. Practical applications of these regression results include the ability to predict the water quality impacts due to changes in the independent variables.

In developing a regression model, the analyst will want to select from a set of variables, normally selecting those independent variables that are most strongly correlated with the dependent variable. To begin, therefore, the analyst might want to compute correlation coefficients between numerous monitored variables at the exploratory phase of the analysis (sometimes referred to as correlation matrices). In fact, determining which variables are most strongly correlated might be the entire goal of the analysis. The correlation matrix can then be used to guide the analyst, to some extent, in selecting appropriate independent variables. In adding additional variables to a model, the analyst must be aware of correlations among different independent variables (multicollinearity) that can mask the relationship of one x to the y variable due to the correlation of this independent variable with another in the model.

In contrast to the univariate models discussed above, where only one dependent response variable (y) is involved, multivariate models can have several dependent variables. Multivariate analyses (which include MANOVA and principal component analysis, among others) are designed to take into account the correlation structure of the x's and y's to reduce the overall variance. For

example, a nonpoint source application might be to examine the effect of different BMP implementation programs on several water quality parameters.

Analysts are encouraged to read the detailed discussion of regression in statistics texts such as Snedecor and Cochran (1980), Cochran (1977), and Srivastava and Khatri (1979) for a more complete discussion of this important statistical procedure.

4.7.2 Simple Linear Regression

The simplest form of regression is to consider just one dependent variable and one independent variable using

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4-64}$$

where y is the dependent variable, x is the independent variable, and β_0 and β_1 are numerical constants representing the y-intercept and slope, respectively. Helsel and Hirsch (1995) summarize the key assumptions regarding application of linear regression (Table 4-28). The uses of a regression analysis should not be extended beyond those supported by the assumptions that are met. Note that the normality assumption (assumption 5) can be relaxed when testing hypotheses and estimating confidence intervals if the sample size is relatively large.

The first step in applying linear regression is to examine the data to see if linear regression makes sense—that is, to use a bivariate scatter plot to see if the points approximate a straight line. If they fall in a straight line, linear regression makes sense; if they do not, data transformation might be needed, or perhaps a nonlinear relationship should be used.

To illustrate the use of linear regression, the data in Table 4-29, which are a subset of calibration data for a plot-size runoff sampler (Dressing et al., 1987), can be used. In this data set the sampling percentage (split) was measured for a range of flow rates. The scatter plot in Figure 4-18 shows that linear regression can be applied to the data.

Presuming that the data are representative (assumption 2 in Table 4-28), the next step is to develop the regression line using the method of least squares, which minimizes the sum of the squares of the vertical deviations from the points to the line (Freund, 1973). To determine the values of β_0 and β_1 in Equation 4-64, the following equations can be used (Helsel and Hirsch, 1995):

$$\beta_1 = \frac{S_{xy}}{SS_x} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \tag{4-65}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{4-66}$$

For the data in Table 4-29, the above equations were used to compute a slope of -0.0119 and an intercept of 3.1317. Thus, the linear model for predicting split versus flow rate is

Split = 3.1317 - 0.0119 • Flow rate

Assumption evaluation

The top section of Table 4-30 provides the same information along with additional characteristics about the β_0 and β_1 that were computed using standard spreadsheet software. Before looking at these additional characteristics, the analyst must make sure that β_0 and β_1 make sense. In this case, perhaps the best approach is to plot the regression line with the raw data as shown in Figure 4-18. The bottom portion of Table 4-30 contains the predicted split (data for the regression line in

Table 4-28. Assumptions necessary for the purposes of linear regression.

Assumption	Purpose			
	Predict y given x	**Predict y and a variance for the prediction**	**Obtain best linear unbiased estimator of y**	**Test hypotheses, estimate confidence or prediction intervals**
(1) Model form is correct: y is linearly related to x	✓	✓	✓	✓
(2) Data used to fit the model are representative of data of interest	✓	✓	✓	✓
(3) Variance of the residuals is constant and does not depend on x or anything else		✓	✓	✓
(4) The residuals are independent			✓	✓
(5) The residuals are normally distributed				✓
✓ Indicates that assumption is required.				

Source: Helsel and Hirsch, 1995.

Figure 4-18) for each flow rate as well as the residual, e_i, defined as $y_i - \hat{y}_i$.

Residuals plotted as a function of predicted values of *y* and time, and normal probability plots of residuals, are the most effective approaches to evaluate the last three assumptions listed in Table 4-28, respectively. As shown in Case A of Figure 4-19, the plot of residuals versus predicted values of *y* or time should appear to be a uniform band of points around 0 (Ponce, 1980a). The analyst should look for two types of patterns when evaluating assumption 3 from Table 4-28 (e.g., constant variance). The first is a pattern of increasing or decreasing variance with predicted values of *y*, as depicted in Case B of Figure 4-19. The second is a pattern (e.g., a trend, a curved line) of the residual with predicted values of *y*. Both characteristics are usually assessed based on a review of the residual plots and professional judgment alone. The analyst may also need to examine variables other than predicted values of *y* to fully evaluate assumption 3.

Independence of residuals (assumption 4 from Table 4-28) can be evaluated by examining residuals plotted as a function of time. The analyst should look for the same patterns as before. As an alternative for evaluating independence, the analyst can also plot the *i*th residual, e_i, as a function of the (*i*-1)th residual, e_{i-1}. One word of caution is in order when reviewing any residual plot:

Table 4-29. Runoff sampler calibration data.

X Flow Rate (gpm)	Y Split (%)	X Flow Rate (gpm)	Y Split (%)
52.1	2.65	17.6	2.84
19.2	3.12	37.6	2.60
4.8	3.05	41.4	2.54
4.9	2.86	40.1	2.58
35.2	2.72	47.4	2.49
44.4	2.70	35.7	2.60
13.2	3.04	13.9	3.19
25.8	2.83		

n = 15	Σx	=	433.30	Σy	= 41.81
	$\bar{x}$	=	28.89	$\bar{y}$	= . 2.79
	Σx^2	=	15,940.33	Σy^2	= 117.25
	Σxy	=	1,166.93		

S_{xy}	=	- 40.817533
SS_x	=	3423.73733
SS_y	=	0.70929333

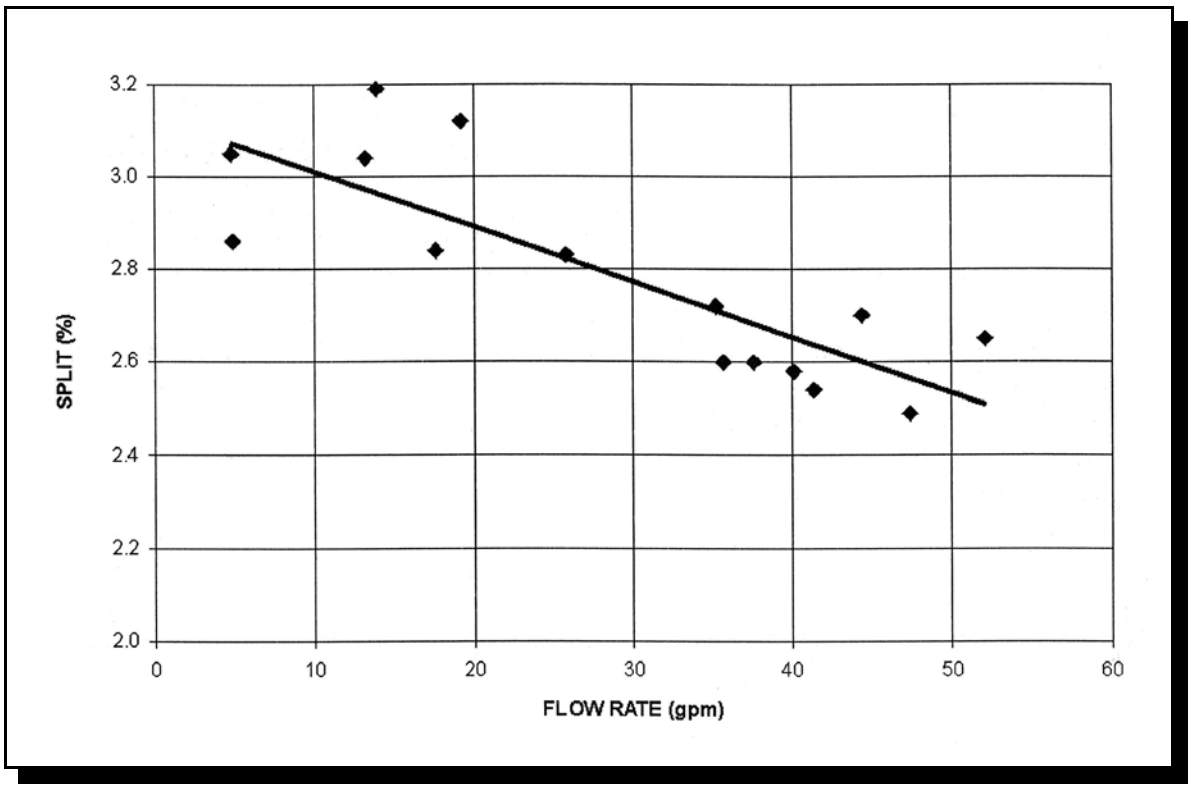


Figure 4-18. Split versus flow rate.

Table 4-30. Regression analysis of runoff sampler calibration data.

	Coefficients	Standard Error	*t* Statistic	*p* Value	Lower 95%	Upper 95%
Intercept (β_0)	3.1317	0.072914	42.950756	2.14E-15	2.97420	3.28924
Flow Rate (β_1)	-0.0119	0.002237	-5.330126	0.00014	-0.01675	-0.00709

	df	SS	MS	F	Significance F
Regression	1	0.486623	0.486623	28.410248	0.0001366
Residual	13	0.222670	0.017128		
Total	14	0.709293			

Flow Rate (gpm)	Split (%)	Predicted Split	Residual $e_i=y_i-\hat{y}_i$
52.10	2.65	2.5106	0.1394
19.20	3.12	2.9028	0.2172
4.80	3.05	3.0745	-0.0245
4.90	2.86	3.0733	-0.2133
35.20	2.72	2.7121	0.0079
44.40	2.70	2.6024	0.0976
13.20	3.04	2.9743	0.0657
25.80	2.83	2.8241	0.0059
17.60	2.84	2.9219	-0.0819
37.60	2.60	2.6835	-0.0835
41.40	2.54	2.6382	-0.0982
40.10	2.58	2.6536	-0.0736
47.40	2.49	2.5666	-0.0766
35.70	2.60	2.7061	-0.1061
13.90	3.19	2.9660	0.2240

If there are more points in a certain section of the residual plot, the residuals might not appear to be a uniform band of points around 0 (as suggested in Case A of Figure 4-19); instead, that section might have a somewhat wider band (Helsel and Hirsch, 1995). This is an expected result.

The normality of residuals can be assessed by examining a probability plot. Two problems with non-normal residuals are the loss of power in subsequent hypothesis tests and increased prediction intervals together with the impression of symmetry (Helsel and Hirsch, 1995).

Figure 4-20 displays all three of these plots for the split data analyzed from Table 4-29. From Figure 4-20, A and B, the split residuals appear to be independent of predicted values of *y* and time as well as having a constant variance. The regression meets assumptions 3 and 4 listed in Table 4-28. In this analysis, testing for residual independence is important since the testing apparatus was calibrated initially. The pumps or other equipment could have differed in performance over time, which in turn would affect the results. Figure 4-20C, the probability plot, suggests that the data might not rigorously follow the normality assumption, although by inspection any normality

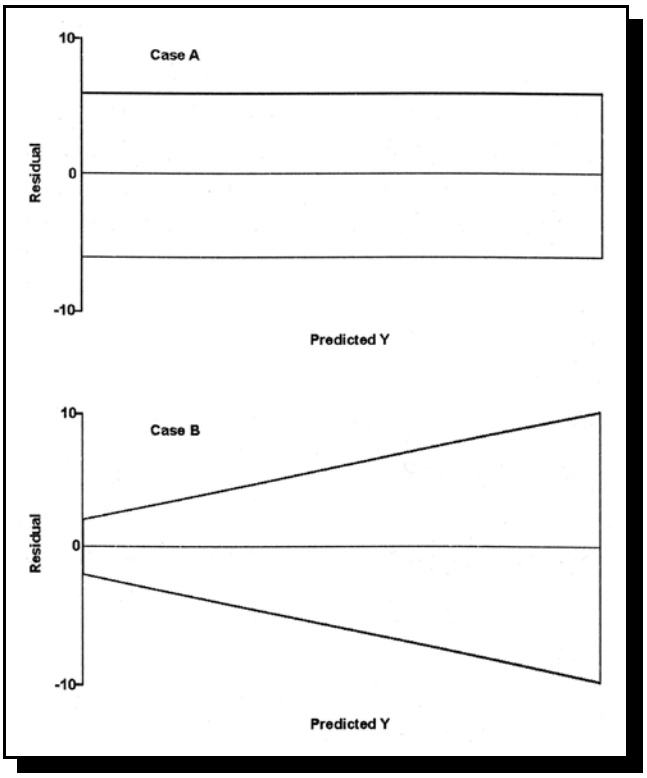


Figure 4-19. Plot of residuals versus predicted values. (Source: Ponce, 1980a)

violation is believed to be relatively minor. To check, the Shapiro-Wilk *W* statistic (see Section 4.4.1) is computed as 0.935. Comparing 0.935 to the test statistic (with p=0.95, n=15) from Table D5, 0.98, the split residuals can be accepted as being normally distributed. (Note that accepting H_o in this case might be due to small sample size and resulting lack of power.) Had this analysis violated any of these assumptions, using a different regression technique, transforming the data, or adding additional variables to the regression would have to be considered. Alternatively, the use of the regression results could be limited to those identified in Table 4-28 as restricted by the assumptions met.

Model evaluation

To determine how well the regression line fits the data, several things can be evaluated:

- Evaluate the proportion of variation in y explained by the model.
- Test whether β_0 is zero.
- Test whether $\beta 1$ is zero.
- Compute the confidence interval for β_0.
- Compute the confidence interval for β_1.

The coefficient of determination, R^2, can be used to evaluate what proportion of the variation can be explained by the model (Gaugush, 1986). R^2 can be computed as (Helsel and Hirsch, 1995)

$$R^2 = \frac{[SS_y - s^2(n-2)]}{SS_y} = 1 - \frac{SSE}{SS_y} = \frac{S_{xy}^2}{SS_x S_{xy}} \tag{4-67}$$

where

$$SS_y = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2 \tag{4-68}$$

and

$$SSE = \sum_{i=1}^{n} e_i^2 \tag{4-69}$$

The residual, e_i, is defined as y_i - $\hat{y}_i$. S_{xy} and SS_x can be computed from Equation 4-65. Values for R^2 range between 0 and 1, with 1 representing the case where all observed y values are on the regression line. The correlation coefficient, r, measures the strength of linear relationships (Freund, 1973) and is computed as the square root of R^2. The sign of r should be the same as the sign of the slope. It ranges from -1 to 1, with the extreme values representing the strongest association and 0 representing no correlation.

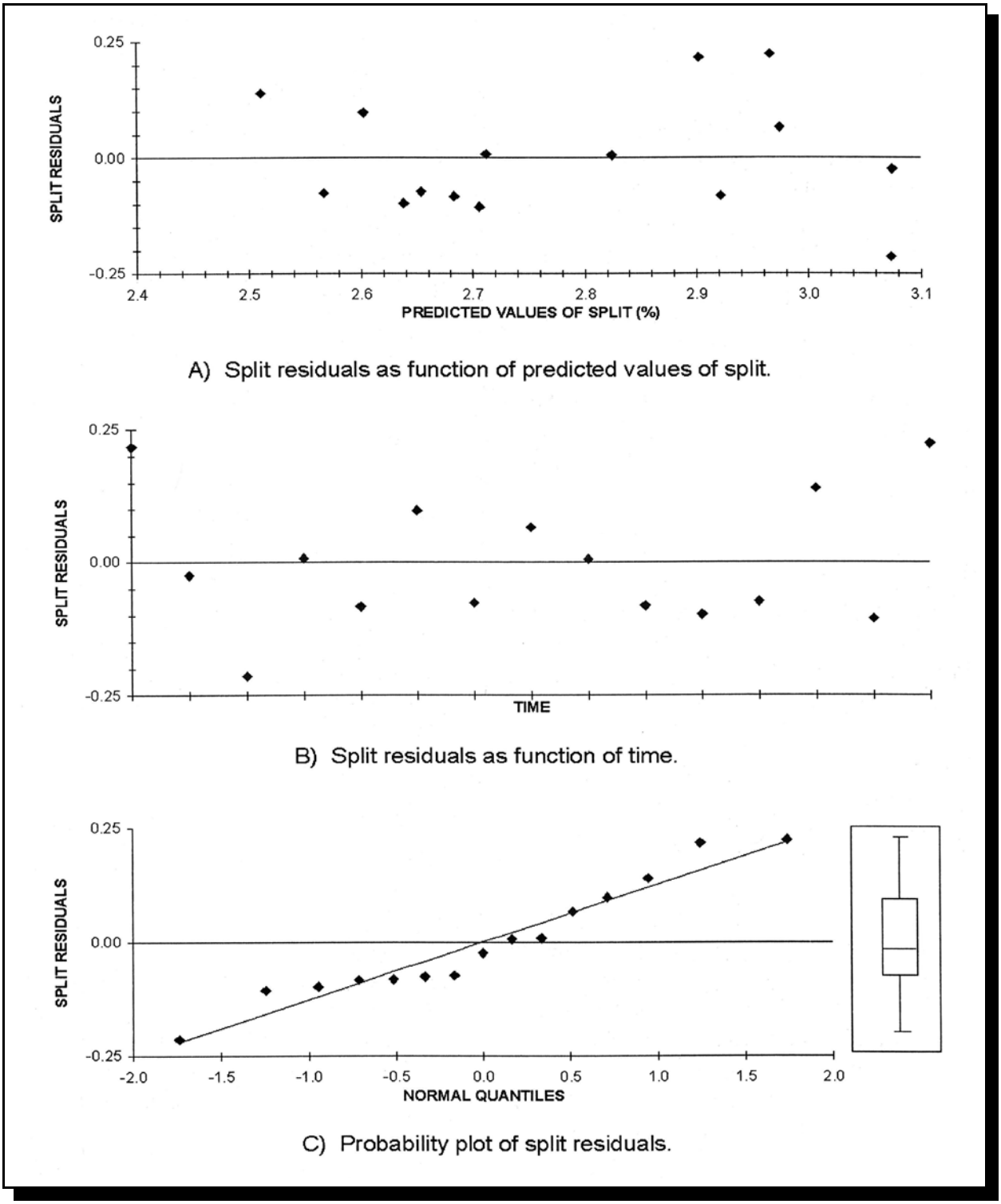


Figure 4-20. Plot of split residuals.

Using the split data from above, the sum of residuals-squared (*SSE*) is equal to 0.2227; thus, R^2 is equal to 1 - (0.2227/0.7093) = 0.686, or 68.6 percent of the variance is explained by the model. (The 0.7093 is from Table 4-29.) The overall model can also be evaluated with the *F* statistic (28.41), which is computed in Table 4-30. The *F* statistic is a measure of the variability in the data set that is explained by the regression equation in comparison to the variability that is not explained by the regression equation. Since the *p* value of 0.0001366 is less than 0.05, the overall model is significant at the 95 percent confidence level.

Are β_0 and β_1 significantly different from zero? The standard error for β_0 and β_1 in the top portion of Table 4-30 can be calculated as (Helsel and Hirsch, 1995)

$$SE(\beta_0) = s\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{SS_x}} \tag{4-70}$$

$$SE(\beta_1) = \frac{s}{\sqrt{SS_x}} \tag{4-71}$$

where

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2} \tag{4-72}$$

The value *s* is equal to the standard error of the regression (which is the same as the standard deviation of the residuals). The corresponding *t* statistics (with *n* - 2 degrees of freedom) for β_0 and β_1 are then equal to β_0 and β_1 divided by their respective standard error. The *t* statistic can then be compared to values from the *t* distribution to determine whether β_0 or β_1 are significantly different from zero. In this case β_0 and β_1 are both significantly different from zero based on inspection of their associated *p* values in Table 4-30. The overall model can also be evaluated with the *F* statistic computed in the middle portion of Table 4-30. This portion of Table 4-30 has the same format as the ANOVA tables described in the previous section. The values in this table are computed using the equations summarized in Table 4-31. Verification of the results in Table 4-30 is left to the reader.

The confidence intervals for β_0 and β_1 can be computed using the following formulas (Helsel and Hirsch, 1995):

$$\beta_0 \pm t_{\alpha/2,n-2}\, SE(\beta_0) \tag{4-73}$$

$$\beta_1 \pm t_{\alpha/2,n-2}\, SE(\beta_1) \tag{4-74}$$

where $t_{\alpha/2,n-2}$ is from Table D2. The lower and upper 95 percent confidence limits for β_0 and β_1 are provided in the top portion of Table 4-30, from which $t_{\alpha/2,n-2}$ was obtained as 2.1604.

The correlation coefficient, *r*, calculated from sample data, is an estimate of the corresponding population parameter, ρ, referred to as the population correlation coefficient. Establishing a confidence interval for ρ requires that *x* also be a normally distributed random variable (Freund, 1973). The Shapiro-Wilk *W* statistic for *x* (the flow rate data in Table 4-29) is 0.931. Comparing 0.931 to the test statistic of 0.98, obtained earlier, the data can be accepted as normally distributed. Using Table D9 (Remington and Schork, 1970) the 95 percent and 99 percent confidence limits for ρ can be obtained knowing *n* and *r*. For the data in Table 4-29, *n* is 15 and *r* is -0.828. So, the 99 percent confidence limits from Table D9 are approximately -0.95 to -0.50.

A *t* test can also be used to test H_o that ρ is zero. The *t* statistic (with *n* - 2 degrees of freedom) for this test is (Freund, 1973)

Table 4-31. Common ANOVA output format for linear regression.

Source of Variation	SS	df	MS	F	Significance F
Regression	SSR = $(S_{xy})^2/SS_x$	1	MSR = SSR/1	MSR/MSE	p
Residual	SSE	n-2	MSE = SSE/(n-2)		
Total	SSR + SSE	n-1			

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{4-75}$$

For the above data *t* would be -5.33. From Table D2 the two-sided *t* value for 95 percent significance (df = 13) is -2.1604. Therefore, H_o ($\rho = 0$) is rejected and H_a that ρ is not zero is accepted.

The Fisher *Z* transformation can be used to test H_o of ρ equal to values other than zero (Freund, 1973). For this test, *r* is changed into a *Z* value using (Freund, 1973)

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{4-76}$$

Freund (1973) provides a table of *Z* values to simplify this procedure. The test statistic is (Freund, 1973)

$$z = (Z - \mu_z)\sqrt{n-3} \tag{4-77}$$

where μ_z corresponds to the *Z* value for the nonzero value of ρ being tested for. For illustration, H_o that ρ is equal to -0.8 for the regression performed can be tested using the data from Table 4-29. Equation 4-76 yields -1.1827 for $r = -0.828$ and -1.0986 for $r = -0.8$. Substituting these values in Equation 4-77 yields

$$z = (-1.1827 + 1.0986)\sqrt{15-3} = -0.2913$$

The two-sided *z* statistic at 95 percent significance (Table D1) is -1.96, H_o is accepted, $\rho = -0.8$.

A confidence interval for ρ can be determined by calculating an interval for μ_z, and then retransforming the confidence interval from *Z* values to ρ (Freund, 1973). The formula for the confidence interval for μ_z is (Freund, 1973)

$$Z - \frac{z_{\alpha/2}}{\sqrt{n-3}} \le \mu_z \le Z + \frac{z_{\alpha/2}}{\sqrt{n-3}} \tag{4-78}$$

Again using the sample data, the 95 percent confidence interval for μ_z becomes

$$-1.1827 - \frac{1.96}{\sqrt{15-3}} \le \mu_z \le -1.1827 + \frac{1.96}{\sqrt{15-3}}$$

$$-1.7485 \le \mu_z \le -0.6169$$

Solving for ρ,

$$\begin{matrix}\textit{Lower limit}\\ \textit{of } \rho\end{matrix} = -1.7485 = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$\begin{matrix}\textit{Upper limit}\\ \textit{of } \rho\end{matrix} = -0.6169 = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$-0.94 \le \rho \le -0.55$$

Using the regression line

The most obvious use of the regression line is to predict *y* values for selected values of *x*. For example, using the regression equation

Split = 3.1317 - 0.0119 x Flow Rate,

the split for any flow rate can be estimated. (It is not good practice, however, to predict values beyond the range of test conditions.) For a flow rate of 10 gpm, the predicted split is 3.01 percent; for a flow rate of 50 gpm, the predicted split is 2.53 percent.

Since in most cases the regression line will not fit the data perfectly, the uncertainty associated with the predicted values should be quantified. The regression line can be used either to establish the confidence interval for the population mean of *y* or to determine the prediction interval for a single value of *y*. The limits for the single value of *y* are wider than the corresponding limits on the mean of *y* (Remington and Schork, 1970) because single observations vary more than means.

The equation for the confidence interval for the population mean *y* at $x = x_0$ is (Helsel and Hirsch, 1995)

$$\hat{y} \pm t_{\alpha/2,n-2}\, s\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_x}} \qquad (4\text{-}79)$$

This interval is most narrow at $\bar{x}$ and widens as x_0 moves farther from $\bar{x}$. By calculating the interval at each point along the regression line, a curve such as the dashed line in Figure 4-21 for the example data can be plotted. The equation for the prediction interval for individual values of *y* at $x = x_0$ is (Helsel and Hirsch, 1995)

$$\hat{y} \pm t_{\alpha/2,n-2}\, s\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_x}} \qquad (4\text{-}80)$$

Figure 4-21 also shows this interval for the example data.

One of the simplest (in theory) nonpoint source control applications of linear regression is the regression of a water quality indicator against an implementation indicator. For example, flow-adjusted total suspended solids (TSS) concentration could be regressed against a sediment control variable such as the total combined erosion rate of all cropland for which delivery to the stream is likely to be 50 percent or greater. A significant negative slope would suggest (but not prove) that water quality has improved because of implementation of sediment control practices.

Another possible use of simple linear regression is to model a water quality parameter versus time. In this application a significant slope would indicate change over time. The sign of the slope would indicate either improvement or degradation depending on the parameter used. For nonpoint source studies, a simple regression versus time will most likely be confounded by the variability in precipitation and flows. Thus, considerable data manipulation (transformations, stratification, etc.) might be required before regression analysis can be successfully applied. In these cases, it might be more appropriate to apply one of the alternatives to regression described by Helsel and Hirsch (1995).

In many cases water quality parameters are regressed against flow. This is particularly relevant in nonpoint source studies. In analysis of covariance, regressions against flow are often performed prior to an ANOVA (Spooner et al., 1985). One of the implicit goals of nonpoint source control is to change the relationship

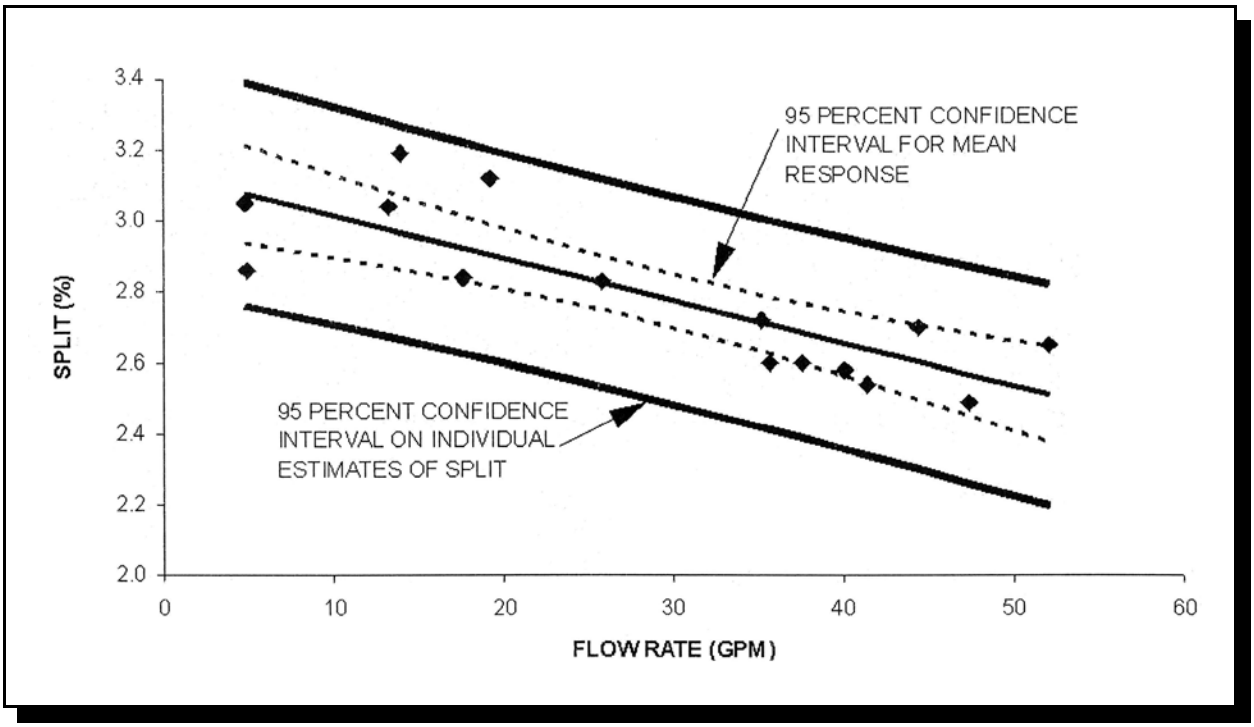


Figure 4-21. Plot of split versus flow rate with confidence limits for mean response and individual estimates.

between flow and pollutant concentration. This will be discussed in greater detail under analysis of covariance.

In paired watershed studies, measured parameters from paired samples are often regressed against each other to compare the watersheds. These regression lines can be compared over time to test for the impact of nonpoint source control efforts (Spooner et al., 1985). This will be discussed in greater detail under analysis of covariance.

4.7.3 Nonlinear Regression and Transformations

The discussion of nonlinear or curvilinear regression is limited to cases where the nonlinear relationship can be transformed into a linear relationship for which simple linear regression can be performed. Data inspection should indicate to the analyst the nature of the relationship between the dependent and independent variables. Possible curvilinear relationships include exponential curves (semi-log), power functions (log-log), and parabolas, among others (Freund, 1973).

Nonlinear regression (as discussed here) involves transformation to linear equations, followed by simple linear regression. Helsel and Hirsch (1995) provide a detailed discussion on transformations using the "bulging rule" described by Mosteller and Tukey (1977), which can be used to select appropriate transformations. Crawford et al. (1983) list the numerous regression models most often applied by the U.S. Geological Survey for flow-adjusting concentrations. The selection of which transformation to use is ultimately based on an inspection of the residuals and whether the assumptions described earlier are met. Typical transformations include x^2, x^3, ln x, $1/x$, $x^{0.5}$, etc.

When the residuals do not exhibit constant variance (heteroscedasticity), one of several common transformations should be used. Logarithmic transformations are used when the standard deviation in the original scale is proportional to the mean of *y*. Square root transformations are used when the variance is proportional to the mean of *y*. In many instances, the right transformation will "fix" the nonlinear and heteroscedastic problem. With data that are percentages or proportions (between the values of 0 and 1), the variances at 0 and 1 are small. The arcsin of the square root of the individual values is a common transformation that helps spread out the values near 0 and 1 to increase their variance (Snedecor and Cochran, 1980).

There are several disadvantages when applying transformations to regression applications. The most important issue is that the regression line and confidence intervals are symmetric in the transformed form of the variables. When these lines are transformed back to their normal units, the lines will no longer be symmetrical. The most notable time in hydrology when this creates a problem is when estimating mass loading. To estimate the mass, the means for short time periods are regressed and summed to estimate the total mass over a longer period. This approach is acceptable if no transformations are used—the analyst is summing the means. However, if a log transformation is used, summing the mass over the back-transformed values results in summing the median, which will result in an estimate that is biased low for the total mass (Helsel and Hirsch, 1995).

As an example of nonlinear regression, consider a common relationship that is used to describe load (L) as a function of discharge (Q):

$$L = aQ^b \tag{4-81}$$

Taking the logarithms of both sides yields

$$\ln(L) = \ln(a) + b\ln(Q) \tag{4-82}$$

which has the same form as Equation 4-64, introduced at the beginning of this section, where $\ln(L)$ corresponds to y, $\ln(a)$ corresponds to β_0, b corresponds to β_1, and $\ln(Q)$ corresponds to x. By taking the logarithms of both sides, the nonlinear problem has been reduced to a simple linear model. The only additional step that the analyst must perform is to convert L and Q to $\ln(L)$ and $\ln(Q)$ before using standard software. The analyst should be aware that all of the confidence limits are in transformed units; when they are plotted in normal units, the confidence intervals will not be symmetric.

Figure 4-22 demonstrates how transforming the data may improve the regression analysis. In Figure 4-22A, sulfate concentrations (in milligrams per liter) are plotted as a function of stream flow (in cubic feet per second). The apparent downward trend is typical of a stream dilution effect; however, the trend is clearly nonlinear. The trend line plotted in this figure, as well as the residuals plotted in Figure 4-22C, demonstrate that a linear model would tend to over- and underestimate sulfate concentrations depending on the flow. Figure 4-22B displays the same data after computing the logarithms (base 10) of the sulfate and flow data. A trend line fitted to these data and the residual plot (Figure 4-22D) clearly demonstrate that applying linear regression after log transformation would be appropriate for these data.

4.7.4 Multiple Regression

Multiple regression is applied to quantify a relationship between a dependent variable and more than one independent variable (Gaugush, 1986). The assumptions made for simple linear regression also apply to multiple regression (Ponce, 1980a). The method of least squares is also used to determine the best multiple

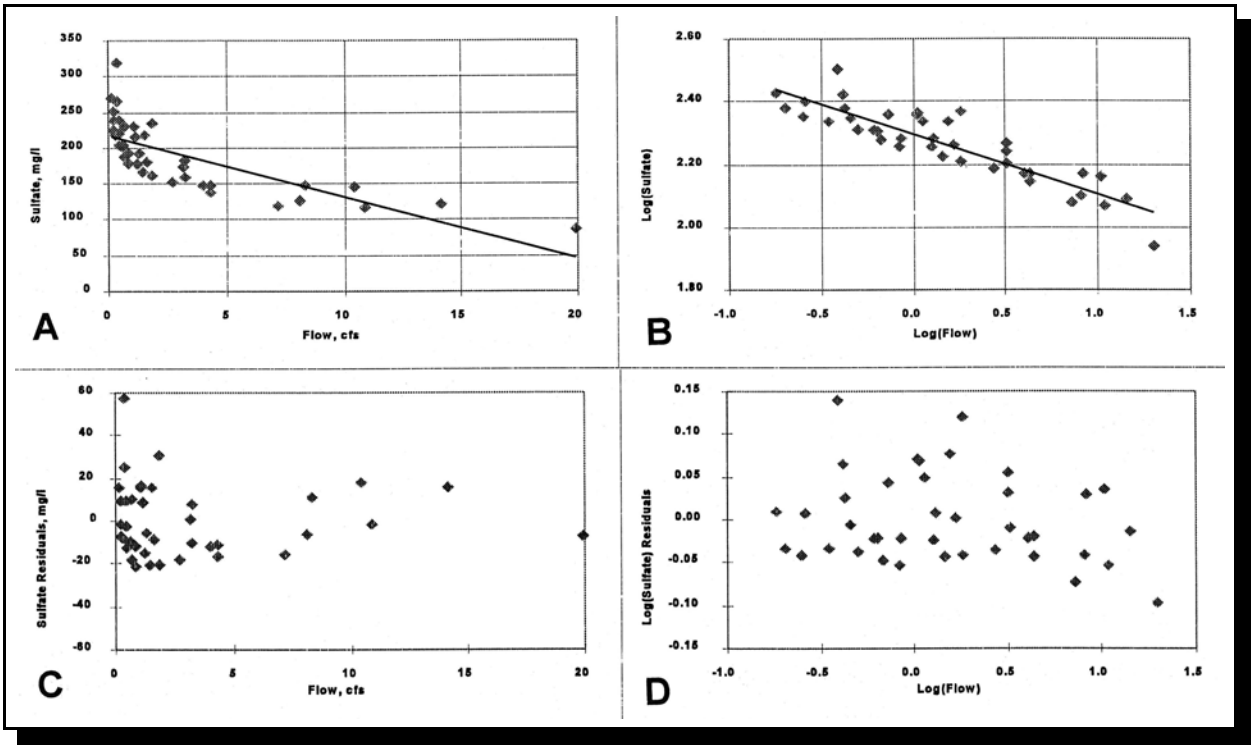


Figure 4-22. Comparison of regression analyses using raw and log-transformed data.

regression line. The general linear model to consider is (Ponce, 1980a)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \tag{4-83}$$

The corresponding normal equations are presented below (Ponce, 1980a).

After solving for the $\beta_1, \ldots, \beta_n$, β_0 can be calculated from (Ponce, 1980a):

$$\left(\sum x_1^2\right)\beta_1 + \left(\sum x_1 x_2\right)\beta_2 + \left(\sum x_1 x_3\right)\beta_3 + \ldots + \left(\sum x_1 x_n\right)\beta_n = \sum x_1 y \tag{4-84}$$

$$\left(\sum x_1 x_2\right)\beta_1 + \left(\sum x_2^2\right)\beta_2 + \left(\sum x_2 x_3\right)\beta_3 + \ldots + \left(\sum x_2 x_n\right)\beta_n = \sum x_2 y \tag{4-85}$$

$$\left(\sum x_1 x_3\right)\beta_1 + \left(\sum x_2 x_3\right)\beta_2 + \left(\sum x_3^2\right)\beta_3 + \ldots + \left(\sum x_3 x_n\right)\beta_n = \sum x_3 y \tag{4-86}$$

$$\ldots$$

$$\left(\sum x_1 x_n\right)\beta_1 + \left(\sum x_2 x_n\right)\beta_2 + \left(\sum x_3 x_n\right)\beta_3 + \ldots + \left(\sum x_n^2\right)\beta_n = \sum x_n y \tag{4-87}$$

$$\beta_0 = \bar{y} - \beta_1\overline{x_1} - \beta_2\overline{x_2} - \ldots - \beta_n\overline{x_n} \qquad (4\text{-}88)$$

Ponce (1980a) presents a hand-computed example of multiple regression using three independent variables. The reader is encouraged to follow through that example to develop an understanding of multiple regression before using computerized procedures. Gaugush (1986) states that multiple regression with two independent variables can be performed using textbook formulas, but that matrix algebra is required for broader appli-cations. Winer (1971) provides a matrix algebra approach to multiple regression, but the discus-sion is complicated and probably not critical to appropriate use of multiple regression techniques (especially when the analyst consults a statis-tician).

Gaugush (1986) also provides an example of multiple regression in which the SAS procedure GLM (SAS Institute, Inc., 1985b) is used. This example relates pollutant level to three indepen-dent variables—distance from source, tempera-ture, and discharge. An interpretation of the SAS output is also provided.

Key points made in the examples above include:

- An *F* test indicates the significance of the regression.
- The coefficient of multiple determination (R^2), which is calculated as in simple linear regres-sion, shows the proportion of variation in *y* explained by the model.
- Computerized output such as that from SAS can be used to refine the model for subsequent runs.

As a further note regarding use of SAS, the RSQUARE procedure (SAS Institute, Inc., 1985b) can be used in an exploratory fashion to perform all possible multiple regressions for subsets of independent variables, listing the models in decreasing order of R^2 magnitude. Thus, the model with the largest R^2 value will be listed first. The STEPWISE procedure allows five approaches to stepwise regression for users who wish to determine which variables should be included in a regression model (SAS Institute, Inc., 1985b). However, this procedure is not guaranteed to identify the model with the largest R^2. Other computer software packages, such as SPSS (Statistical Package for the Social Sciences), can also be used for multiple regression (Ingwersen, 1980).

The following discussion of R^2, taken largely from a technical nonpoint source newsletter (Spooner, 1984), emphasizes proper interpretation of R^2 values.

The purpose of regressing a response variable (*y*) on one or more independent variables (*x*) is to "explain" some of the variation observed in the measured values in *y*. The *F* tests for each individual *x* variable can be used to determine whether they are individually important to the regression on *y*. R^2 is a measure of the fraction of variation in *y* explained by the linear regression on $x_1, x_2, \ldots, x_n$ variables in the model. Specifically, R^2 is the fraction of the sum of squares (SS) of the deviations of *y* from its mean that is attributed to the regression. R^2 values range from 0 (model useless) to 1 (model perfect)(Equation 4-87).

H_o that

$R^2 = 0$ (i.e., $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$)

can be tested using the *F* statistic to determine whether the regression model explains any of the variation in *Y*. The *F* statistic is $(n\text{-}k\text{-}1)\, R^2/(k\text{-}1)(1\text{-}R^2)$ with $(k\text{-}1)$ and $(n\text{-}k\text{-}1)$ degrees of freedom. It should be noted that $(k\text{-}1)$ is the degrees of freedom for the regression

model SS and (n-k-1) is the degrees of freedom for the error SS.

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - F)^2}{\sum_{i=1}^{n} (Y_i - F)^2} = \frac{SS\ Regression}{SS\ Total}$$

$$= 1 - \frac{SS\ Error}{SS\ Total} \qquad (4\text{-}89)$$

A small R^2 might be significantly different from zero if n is large. Conversely, a large R^2 might be insignificant if n is small compared to the number of x's in the model.

If R^2 is small, most of the variation in Y is unexplained by the linear regression model. This remaining "noise" might be random variation, or it might be due to other independent variables not considered in the regression. If these other variables are added to the regression, the relationships among the x's already included might change.

When new variables are added to the model, R^2 always increases although the adjusted R^2 might not increase. This explains why a large R^2 might not be meaningful when the sample size is small. Also, it is not legitimate to compare two models with different numbers of x's solely by their R^2 values. However, R^2, adjusted for the degrees of freedom, may be used to compare models, where adjusted R^2 is

$$R_a^2 = (1 - R^2)\ (n - 1)\ /\ (n-k-1) \qquad (4\text{-}90)$$

How does one test whether a new variable added to a model adds significant information to explain further the variation in y (i.e., is the increase in R^2 significant)? In SAS, for example, the "type III SS or IV SS" (also known as the partial sum of squares) and their associated F tests can be used. These statistics measure the amount of variation in y explained by the addition of an individual x after all other x's are in the model. An equivalent method is to compare the SSE (sum of squares due to error) from "full" and "reduced" models (i.e., SSE from models with and without, respectively, the extra term in question). If the SSE is reduced significantly by the addition of a new variable to the model, the variable is important. The F statistic is

$$F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \qquad (4\text{-}91)$$

where df_R and df_F are the degrees of freedom for the reduced model SS and full model SS, respectively.

4.7.5 Multivariate Regression

Multivariate regression can be a very useful technique in nonpoint source monitoring and evaluation efforts. It involves the development of a linear model to relate two or more dependent variables to two or more independent variables. A detailed discussion of the theory behind multivariate regression is beyond the scope of this document. Readers are referred to statistics texts (e.g., Srivastava and Khatri, 1979) for more on multivariate regression. Multivariate regressions are designed to take into account the correlation structure of the x's and y's to reduce the overall variance.

Users of SAS (SAS Institute, Inc., 1985b) can use the REG procedure for multivariate regression. An example of the MODEL statement used in this procedure is the following (SAS Institute, Inc., 1985b):

MODEL Y1 Y2 = X1 X2 X3

where

Y1 and Y2 are the dependent variables and X1, X2, and X3 are the independent variables.

Within this procedure the MTEST statement can be used to test hypotheses regarding the multivariate regression model. *F* values are calculated for the following procedures (SAS Institute, Inc., 1985b):

- Wilks' lambda
- Pillai's trace
- Hotelling-Lawley trace
- Roy's maximum root

4.8 ANALYSIS OF COVARIANCE

Suppose an analyst is interested in evaluating BMPs by comparing data collected from a paired watershed design. Data are collected from two watersheds during two periods—calibration and treatment. During calibration, neither watershed has a BMP in place, while during the later period, one of the two watersheds has a BMP installed. A natural extension of the regression techniques described in Section 4.7 is to compare regression equations between the treatment watershed and the control watershed, with one regression equation developed during the calibration phase and the second regression equation developed during the treatment phase. The analysis of covariance (ANCOVA), a procedure that combines features of ANOVA and regression, can be used to evaluate this situation. ANCOVA can also be used to test for differences in the average value for a dependent variable (e.g., sediment concentration) between the levels of a group variable (e.g., seasons or years) after adjusting for an independent variable (e.g., flow or upstream concentration).

A typical ANCOVA model in which the slopes and intercepts for the two groups are suspected to be different can be represented as (Helsel and Hirsch, 1995)

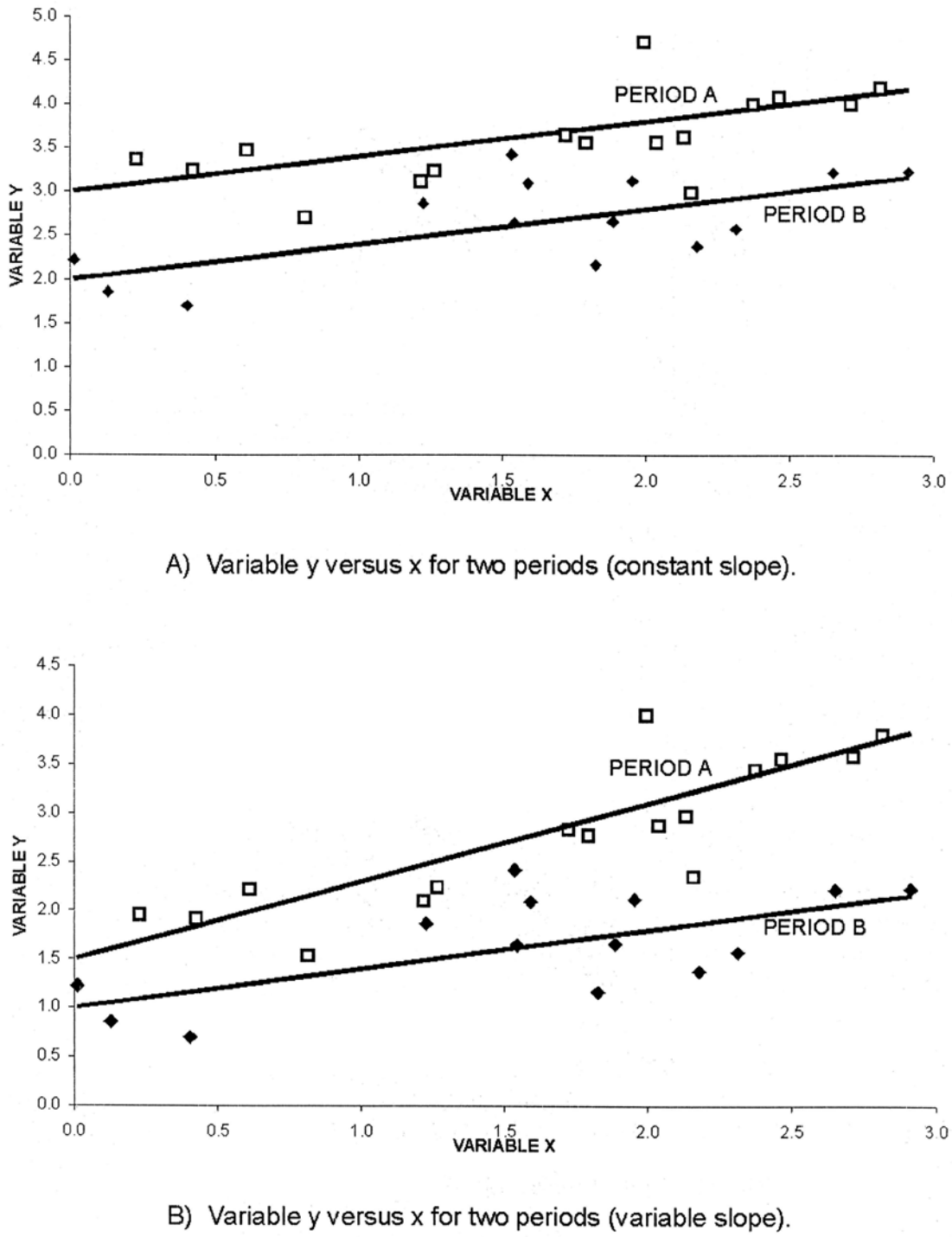
$$y = (\beta_0 + \beta_2 Z) + (\beta_1 + \beta_3 Z)x + \varepsilon \qquad (4\text{-}92)$$

where *Z* is a binary variable that is equal to 0 or 1 depending on which group *x* and *y* are from. For example, *Z* could be 0 during calibration and 1 during treatment of a paired watershed analysis. In this case, β_0 and $\beta_0+\beta_2$ are the intercepts during the calibration and treatment periods, respectively. β_1 and $\beta_1+\beta_3$ are the slopes during the calibration and treatment periods, respectively. If β_2 is nonzero and β_3 is zero, the regression produced by Equation 4-92 would be a pair of parallel lines (Figure 4-23A). If β_2 and β_3 are nonzero, the regression produced by Equation 4-92 would be a pair of lines like those presented in Figure 4-23B.

The remainder of this discussion follows an analysis performed for field runoff (cm) during the conversion from conventional to conservation tillage in Vermont (USEPA, 1993c). Two watersheds were monitored during a calibration period during which 49 (n_1) paired observations of runoff were made. Figure 4-24A is a bivariate log-log plot of storm runoff for the treatment watershed as a function of storm runoff for the control watershed. Based on an inspection of this plot, it seems reasonable to perform the analyses using log-transformed (base 10) data.

A regression analysis was performed on these data to determine whether there was a significant relationship between the watersheds, whether enough data had been collected during calibration, and whether the residual errors were smaller than the expected BMP effect. A summary of the regression ANOVA is provided in Table 4-32 (with n_1 = 49, SS_y = 148.441, SS_x = 70.933, and S_{xy} = 78.463). (Equations 4-65 through 4-67 and Table 4-31 can be used to hand-check the table entries.) The *p* value associated with the resulting *F* statistic indicates that the model explains a significant proportion of the variation.

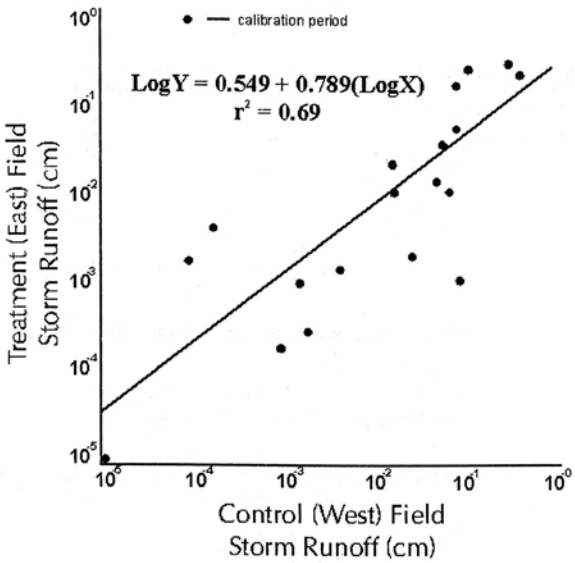
To determine whether enough calibration data have been collected, the ratio of the MSE to the



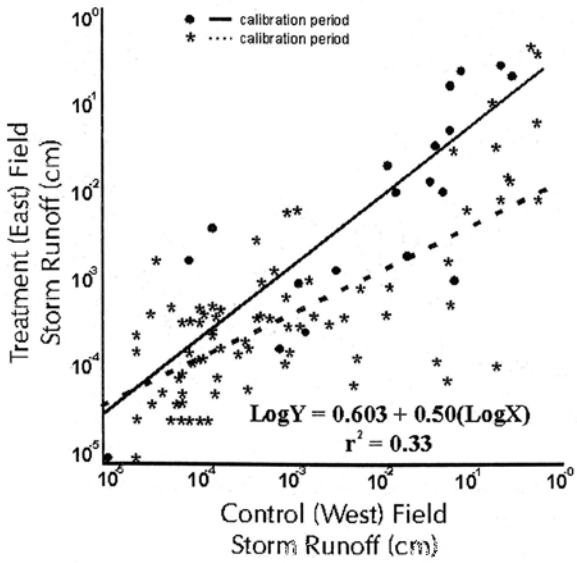
A) Variable y versus x for two periods (constant slope).

B) Variable y versus x for two periods (variable slope).

Figure 4-23. Comparison of regression equations for data from two periods.



Calibration Period Regression



Treatment and Calibration Period Regression

Figure 4-24. Storm runoff from calibration and treatment periods in Vermont. (Source: EPA, 1993c).

Table 4-32. ANOVA for regression of treatment watershed runoff on control watershed runoff during calibration.

Source of Variation	SS	df	MS	F	Significance F
Regression	86.792	1	86.79	66.17	0.0001
Residual	61.649	47	1.31		
Total	148.441	48			

smallest worthwhile difference (d) can be compared using the following formula (EPA, 1993c):

$$\frac{MSE}{d^2} = \frac{n_1 n_2}{n_1 + n_2}\left(\frac{1}{F\left(1 + \frac{F}{n_1 + n_2 - 2}\right)}\right) \quad (4\text{-}93)$$

where n_1 and n_2 are the number of observations in the calibration and treatment periods, respec-tively, and F is from Table D6 with 1 and $n_1 + n_2 - 3$ degrees of freedom. If the treatment period has not been initiated, assume that $n_1 = n_2$. Using the example data where $\bar{x}$ of the log-transformed data is -2.518, the number of observations necessary to detect a 20 percent change can be estimated. The left side of the above equation would be equal to $1.31/(0.2 \times -2.518)^2$ or 5.2. With $n_1 = n_2 = 49$ and $F = 3.94$ ($p = 0.95$, 1 and 95 df), the right side of the above equation can be evaluated as 6.0. Since the left side of the equation is less than the right side, there would be enough samples to detect a 20 percent change in discharge. Equation 4-79 can be used to determine the confidence bands for the regression equation, which allow determining the level of change needed to have a significant treatment effect.

Once the treatment period data have been collected, the same type of regression analysis is performed. Following this step, the significance of an overall regression (which combines calibration and treatment data) can be evaluated and the difference between the individual slopes and intercepts can be evaluated. Continuing with the example, a summary of the regression ANOVA for the treatment period is provided in Table 4-33 (with $n_2 = 114$, $SS_y = 135.0$, $SS_x = 227.43$, and $S_{xy} = 101.32$). The p value associated with the resulting F statistic indicates that the model explains a significant proportion of the variation.

The ANCOVA can be performed by combining the results from Tables 4-32 and 4-33. Table 4-34 demonstrates the general format for performing ANCOVA hand calculations. Note that Σ indicates summation of terms. This approach is applied to the example data with the results presented in Table 4-35. Table 4-36 presents the same calculations performed with SAS. (An appropriate SAS program is provided at right.) The ANCOVA indicates that the overall treatment and calibration regressions were significantly different and that the slopes and intercepts of the equations were also different. The difference in slopes is evident from Figure 4-24B. The small differences between the calculations in Tables 4-35 and 4-36 are due to rounding errors. If

Table 4-33. ANOVA for regression of treatment watershed runoff on control watershed runoff during treatment.

Source of Variation	SS	df	MS	F	Significance F
Regression	45.13	1	45.13	56.25	0.0001
Residual	89.87	112	0.80		
Total	135.00	113			

Table 4-34. ANCOVA for comparing regression lines.

Source	df	Ss_x	S_{xy}	Ss_y	β_1	df	SS (res.)	MS	F
Within									
Calibration	n_1-1	Eq. 4-65	Eq. 4-65	Eq. 4-68	S_{xy}/SS_x	n_1-2	$SS_y-(S_{xy})^2/SS_x$	SS/df	--
Treatment	n_2-1	Eq. 4-65	Eq. 4-65	Eq. 4-68	S_{xy}/SS_x	n_2-2	$SS_y-(S_{xy})^2/SS_x$	SS/df	--
					Error:	Σ	Σ	SS/df	
Slopes	n_1+n_2-2	Σ	Σ	Σ	S_{xy}/SS_x	n_1+n_2-3	$SS_y-(S_{xy})^2/SS_x$	SS/df	
				Slope difference:		1	Slope SS-Error SS	SS/df	MS/Error MS
						1	Comb. SS- Slope SS	SS/df	MS/Slope MS
Intercepts	n_1+n_2-1	combined data				n_1+n_2-2	$SS_y-(S_{xy})^2/SS_x$		

Table 4-35. ANCOVA for comparing regression lines from calibration and treatment (hand calculations).

Source	df	Ss_x	S_{xy}	Ss_y	β_1	df	SS (res.)	MS	F
Within									
Calibration	48	70.933	78.463	148.441	1.106	47	61.650	1.3117	--
Treatment	113	227.430	101.315	135.00	0.445	112	89.866	0.8024	--
					Error:	159	151.516	0.9529	
Slopes	161	298.363	179.778	283.441	0.603	160	175.116	1.0945	
				Slope difference:		1	23.600	23.600	24.77*
						1	5.8453	5.8453	5.34†
Intercepts	162	311.671	178.762	283.492	--	161	180.961		

* Significant at p = 0.001
† Significant at p = 0.05

Table 4-36. ANCOVA for comparing regression lines from calibration and treatment (computerized software).

Source of Variation	df	MS	F	Significance F
Model	3	43.99	46.17	0.001
Error	159	0.95		
Overall	1	103.09	108.18	0.0001
Intercept	1	5.47	5.74	0.0178
Slope	1	23.42	24.58	0.0001

SAS Program to Compare Regression Lines

```
PROC GLM;
  CLASS PERIOD;
  MODEL LOGFLOW2  = LOGFLOW1  PERIOD
    LOGFLOW1*PERIOD;
  RUN;

/* LOGFLOW1 = log-transformed data from control
watershed */
/* LOGFLOW2 = log-transformed data from treatment
watershed */
/* PERIOD = indicator for whether paired data were
from the calibration or treatment period */
```

there are not significant differences between the slopes, all of the periods can be represented by a common slope and the relationship between *y* and *x* is constant over the tested period.

4.9 Evaluation of Time Series

In nonpoint source data analysis, we often want to know whether there is a tendency for a pollutant concentration to increase or decrease over time. If such a tendency exists, we say there is a trend. Trend analysis is often used to determine whether the implementation of a BMP actually reduces the pollutants in a stream, or whether the development of an urban area is causing the deterioration of water quality downstream, as well as maintaining a status of ambient water quality conditions. A trend can be visually examined by plotting the observed data versus time. A statistical test is required to analyze the trend. This section describes statistical procedures for detecting and evaluating monotonic (continuously nonincreasing or nondecreasing) trends in a single time series (e.g., 10 years of monthly TSS at a single station) and presents several methods for evaluating temporal correlation.

The first issue to consider is when a monotonic trend test should be used. The most important factor before beginning the analysis is to assess whether any interventions or activities led to the hypothesis that a shift in water quality might have occurred. For example, suppose a BMP to reduce sediment loadings was installed during the course of the monitoring program. A shift in TSS concentration (hopefully downward) after BMP installation would be expected. In this case, it is more appropriate to divide the data into “before” and “after” groups and analyze the data using the two independent random sample procedures described in Section 4.5. On the other hand, if a series of BMPs are being implemented across a watershed over several years and monitoring is being performed in a downstream estuary, the changes would be expected to be gradual. In this case a monotonic trend test might be more appropriate. If there is no hypothesis to naturally divide the data, it is also best to use a monotonic trend test. Concentration data should not be used to determine data groupings for the purposes of developing hypotheses or selecting between a two-sample or monotonic trend test.

The second issue to consider is the case where sampling was interrupted for several years in the middle of a 10-year monitoring effort. It is suggested that if the data gap is greater than one-third of the total data record, it is better to use a two-sample test (Helsel and Hirsch, 1995). A similar issue to consider is the case where several data records will be examined, but they have different starting and stopping points. Helsel and Hirsch (1995) suggest that the analyst divide the data record into three periods of equal length; if any third of the record has more than 20 percent missing values, that record should not be used.

The final issue to consider is whether to account for exogenous variables (e.g., flow, temperature, rainfall) before testing for trends. A common example is the approach used by the USGS to account for flow variability in its National Stream Quality Accounting Network (NASQAN) stations. In USGS analyses, water quality variable concentrations are adjusted to account for flow.

The flow-adjusted concentrations are then evaluated for trends. These adjustments can be made using simple linear regression analyses, as discussed in Section 4.7 or the nonparametric procedures (e.g., locally weighted scatter plot smoothing) discussed by Helsel and Hirsch (1995). The purpose of adjusting the data for an exogenous variable is to reduce the background noise so that the detection of time trends is more powerful.

There are several methods to detect monotonic trends in time series. Regression analyses have already been discussed in Section 4.7. To apply linear regression where time is the independent variable, all of the assumptions listed in Table 4-28 are necessary. If these assumptions are met, linear regression is an acceptable approach. Significant trends are declared when the slope term, β_1, is significantly different from zero. Multivariate regression procedures that model the water quality variable as a function of an exogenous variable (e.g., flow) and time simultaneously can also be used to detect trends if the regression assumptions are met. When evaluating several data sets for a single report, these assumptions are rarely met for all of the data sets. In these cases, nonparametric procedures are recommended. This is not to say that data transformations for nonparametric tests are not desirable, as will be discussed later. Since simple linear and multivariate regression have been discussed, this section is limited to discussing the Mann-Kendall τ and the Seasonal Kendall tests. Both are nonparametric.

Following the monotonic trend discussion, procedures for computing the autocorrelation coefficient and Spearman's rho are provided. These procedures are useful for evaluating whether the data are truly independent, one of the fundamental assumptions in the procedures described next. If the data are serially correlated, it is possible to systematically sample from the data set, to group the data into time periods and use a summary statistic (e.g., time- or volume-weighted mean or median), or to use more advanced time series analysis procedures (Helsel and Hirsch, 1995) to analyze these data.

4.9.1 Monotonic Trends

Regression

Refer to Section 4.7 for a discussion on simple linear and multivariate regression.

Mann-Kendall τ Test

The Mann-Kendall τ test analyzes the sign of the difference between later-measured data and the earlier-measured data. Each later-measured datum is compared to all data measured earlier. This approach results in a total of n(n-1)/2 possible pairs of data, where n is the total number of observations in the time series. The Mann-Kendall τ test assumptions include the typical requirements that the data be independent and that one value can be declared larger than, smaller than, or equal to another value. The third assumption is similar to the regression requirements that the residuals must have a constant variance, but no distribution requirements are necessary.

The usual hypotheses for a Mann-Kendall τ test is whether y tends to increase or decrease with time (Helsel and Hirsch, 1995):

Mann-Kendall τ Test Assumptions

- The random variables $y_1, y_2, ..., y_i, ..., y_j, ..., y_n$ are mutually independent.
- The measurement scale of the data is at least ordinal (i.e., y_i can be declared as <, >, or = y_j).
- The data are identically distributed with only a shift in the central location if there is a trend.

Two-sided test
H_o: Prob $[y_j > y_i] = 0.5$ where $t_j > t_i$
H_a: Prob $[y_j > y_i] \neq 0.5$ (two-sided test)

One-sided test
H_o: Prob $[y_j > y_i] = 0.5$ where $t_j > t_i$
H_a: Prob $[y_j > y_i] > 0.5$ (one-sided test, increasing trend)

The next step is to compute the difference between the later-measured value and all earlier-measured values, $(y_j - y_i)$, where $j > i$ and assign the integer value of 1, 0, -1 to positive differences, no differences, and negative differences, respectively. The test statistic, S, is then computed as the sum of the integers:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(y_j - y_i) \qquad (4\text{-}94)$$

where sign(•) is equal to 1, 0, or -1 as indicated above. This task is most easily accomplished assuming the data are ordered in increasing time order. When S is a large positive number, later-measured values tend to be larger than earlier-measured values and there might be an upward trend. When S is a large negative number, later-measured values tend to be smaller than earlier-measured values and there might be a downward trend. When the absolute value of S is small, there might be no trend. The test statistic, τ, can be computed as

$$\tau = \frac{S}{n(n-1)/2} \qquad (4\text{-}95)$$

which has a range of -1 to 1 and is analogous to the correlation coefficient in regression analyses. Computing S or τ becomes tedious when n is large. Gilbert (1987) provides a FORTRAN program to alleviate the computation effort. S and τ are invariant to transformations such as logs (i.e., S and τ will be the same value whether the raw or log-transformed data are used).

For sample sizes greater than 10, the large sample approximation can be used to compute a test statistic that can be compared to a normal distribution using the following equation:

$$Z_S = \begin{cases} \dfrac{S-1}{\sigma_S} & if\ S>0 \\ \dfrac{S+1}{\sigma_S} & if\ S<0 \end{cases} \qquad (4\text{-}96)$$

where

$$\sigma_S = \sqrt{\frac{n}{18}(n-1)(2n+5)} \qquad (4\text{-}97)$$

for when there are no ties or

$$\sigma_S = \sqrt{\frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i(i)(i-1)(2i+5)}{18}} \qquad (4\text{-}98)$$

for when there are ties, where Z_S is zero if S is zero and t_i is equal to the number of ties of extent i. Z_S is compared to the critical z value from Table D1. For a two-sided 95 percent confidence level, the critical z value would be ±1.96. If Z_S is not contained within this range, reject H_o. See Helsel and Hirsch (1995) for sample sizes of 10 or less. To determine t_i, consider the following 20 observations that are in ascending order:

<1, <1, <1, 4, 4, 6, 6, 8, 8, 10, 11, 11, 11, 11, 16, 19, 20, 22, 32, 45

In this example there are seven ties of extent 1 (i.e., no ties), three ties of extent 2 (4, 4, 6, 6, 8, 8), one tie of extent 3 (<1, <1, <1), one tie of extent 4 (11, 11, 11, 11), and zero ties of extent 5 and greater. Thus, the summation term that includes t_i from above can be evaluated as

7x1x0x7 + 3x2x1x9 + 1x3x2x11 + 1x4x3x13
= 276

Table 4-37 presents a list of annual rainfall for 21 years. Table 4-38 presents the intermediate calculations for computing S. The top portion of Table 4-38 is a table of the differences $y_j - y_i$, for example $y_2 - y_1 = 13.2$. Observations y_8 through y_{16} were omitted from Table 4-38 for presen-tation purposes. The bottom portion of Table 4-38 presents the intermediate calculations for sign($y_j - y_i$). Summing these values (including those not presented in this table) yields a value of 12. Since there were no ties, $Z_S = (12-1)/(1096.7)^{0.5}$ or 0.33, H_o is accepted— there is no trend in the rainfall data.

Table 4-37. Annual total rainfall for 21 years.

Year	Rainfall (in.)	Year	Rainfall (in.)
1	40.2	12	51.2
2	53.4	13	54.3
3	43.5	14	41.5
4	37.7	15	44.8
5	50.2	16	46.7
6	38.7	17	51.8
7	47.8	18	49.5
8	39.5	19	34.1
9	44.9	20	33.2
10	41.7	21	53.7
11	36.4		

Had there been a significant trend in the data, the Sen slope estimator could be estimated as (Helsel and Hirsch, 1995)

$$\hat{\beta}_1 = median\left(\frac{y_j - y_i}{x_j - x_i}\right) \qquad (4\text{-}99)$$

for all $i < j$ and $i = 1, 2, ..., n$-1 and $j = 2, 3, ..., n$; in other words, computing the slope for all pairs of data that were used to compute S. The median of these slopes is the Sen slope estimator. Using the rainfall data as an example, the slope between y_4 and y_2 is equal to (37.7-53.4)/(4-2) or -7.9. Had there been a significant trend, this process would have been carried out for the remaining pairs of observations and the median slope selected as the Sen slope estimator.

As might be expected, any linear slope estimator is a poor choice when the apparent slope is exponential. In Section 4.7.3, transformations to reduce the analysis to a linear problem were discussed. These same approaches are also appropriate here. So while it does not matter for computing S or τ that the trend be linear, transforming the data prior to computing the slope estimator might be useful. For example, if the data were transformed using natural logs, the percentage change from year to year in the above example would be estimated as $(e^{\beta_1} - 1) \times 100$ (Helsel and Hirsch, 1995).

Seasonal Kendall test

In the nonpoint source area, many data follow seasonal patterns. The decision to use a seasonal Kendall test (Hirsch et al., 1982) can usually be made by examining boxplots by season. The test statistic is computed by performing a Mann-Kendall calculation for each season and then combining the results for each season. That is, if sampling is monthly, January observations are compared only to other January observations, etc. Thus S_k is computed as the sum of the S from each season (Helsel and Hirsch, 1995):

$$S_k = \sum_{i=1}^{m} S_i \qquad (4\text{-}100)$$

Table 4-38. Analysis of rainfall data using Mann-Kendall τ test.

	y_j	y_1	y_2	y_3	y_4	y_5	y_6	y_7	...	y_{17}	y_{18}	y_{19}	y_{20}	y_{21}
	y_i	40.2	53.4	43.5	37.7	50.2	38.7	47.8	...	51.8	49.5	34.1	33.2	53.7
y_1	40.2		13.2	3.3	-2.5	10.0	-1.5	7.6	...	11.6	9.3	-6.1	-7.0	13.5
y_2	53.4			-9.9	-15.7	-3.2	-14.7	-5.6	...	-1.6	-3.9	-19.3	-20.2	0.3
y_3	43.5				-5.8	6.7	-4.8	4.3	...	8.3	6.0	-9.4	-10.3	10.2
y_4	37.7					12.5	1.0	10.1	...	14.1	11.8	-3.6	-4.5	16.0
y_5	50.2						-11.5	-2.4	...	1.6	-0.7	-16.1	-17.0	3.5
y_6	38.7							9.1	...	13.1	10.8	-4.6	-5.5	15.0
y_7	47.8									4.0	1.7	-13.7	-14.6	5.9
...	...										...	...	...	...
y_{17}	51.8										-2.3	-17.7	-18.6	1.9
y_{18}	49.5											-15.4	-16.3	4.2
y_{19}	34.1												-0.9	19.6
y_{20}	33.2													20.5
y_{21}	53.7													
		40.2	53.4	43.5	37.7	50.2	38.7	47.8	...	51.8	49.5	34.1	33.2	53.7
y_1	40.2		1	1	-1	1	-1	1	...	1	1	-1	-1	1
y_2	53.4			-1	-1	-1	-1	-1	...	-1	-1	-1	-1	1
y_3	43.5				-1	1	-1	1	...	1	1	-1	-1	1
y_4	37.7					1	1	1	...	1	1	-1	-1	1
y_5	50.2						-1	-1	...	1	-1	-1	-1	1
y_6	38.7							1	...	1	1	-1	-1	1
y_7	47.8									1	1	-1	-1	1
...	...										...	...	...	...
y_{17}	51.8										-1	-1	-1	1
y_{18}	49.5											-1	-1	1
y_{19}	34.1												-1	1
y_{20}	33.2													1
y_{21}	53.7													

where S_i is S from the ith season and m is the number of seasons. Z_{Sk} is estimated as

$$Z_{Sk} = \begin{cases} \dfrac{S_k - 1}{\sigma_{Sk}} & if\ S_k > 0 \\ \dfrac{S_k + 1}{\sigma_{Sk}} & if\ S_k < 0 \end{cases} \qquad (4\text{-}101)$$

or Z_{Sk} is zero if S_k is zero and

$$\sigma_{Sk} = \sqrt{\sum_{i=1}^{m} \frac{n_i}{18}(n_i - 1)(2n_i + 5)} \qquad (4\text{-}102)$$

where n_i is the number of observations in the ith season.

4.9.2 Correlation Coefficients

Spearman's rho

Spearman's rho test is used to detect whether there is a correlation between paired data. Spearman's rho is computed as (Conover, 1980)

$$\rho = \frac{\sum_{i=1}^{n}\left[R(x_i) - \frac{n+1}{2}\right]\left[R(y_i) - \frac{n+1}{2}\right]}{n(n^2-1)/12} \qquad (4\text{-}103)$$

where R(•) represents the rank of the observation and n is the number of observations. If there are ties, Equation 4-104 may be used.

The resulting value of ρ is then compared to critical values in Table D10. Spearman's rho can be used in the same manner as the τ statistic computed in Section 4.9.1. Spearman's rho can also be used to evaluate serial correlation by setting $y_i = x_{i+k}$ to determine the lag-k autocorrelation. For $k = 1$, the first observation is compared to the second observation, the second observation to the third observation, and so on.

Using the rainfall data, Table 4-39 presents the intermediate calculations for Spearman's rho for $k = 1$. Notice that $y_i = x_{i+1}$ and that there are only 20 observations in this analysis. The third and fourth represent the ranks of x_i and y_i, respectively. The remaining three columns are intermediate calculations for the numerator of the above equation. Finally, ρ is equal to -126/ [(20(20^2-1)/12] or -0.19. Assuming a two-sided hypothesis, the critical value from Table D10 (with $n = 20$ and $\alpha = 0.05$) is ±0.4451; the rainfall data are not correlated at lag-1. This result cannot be compared with the previous example. In the previous example the correlation between annual rainfall and time was evaluated. In this example, "this year's annual rainfall" is compared to "next year's annual rainfall."

Autocorrelation coefficient

The analyst may also use the correlation coefficient, r. Salas et al. (1980) provided the formula for the lag-k autocorrelation coefficient as:

$$r = \frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{n-k}(x_t-\bar{x})^2} \qquad (4\text{-}105)$$

Anderson (1941) gave the limit

$$r_k = \frac{-1 \pm 1.96(n-k-1)^{0.5}}{n-k} \qquad (4\text{-}106)$$

for the 95 percent probability levels for the lag-k autocorrelation coefficient where n is the sample size.

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}\left(\sum_{i=1}^{n} R(y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{0.5}} \qquad (4\text{-}104)$$

Table 4-39. Analysis of rainfall data using Spearman's rho.

x_i	y_i	$R(x_i)$	$R(y_i)$	$R(x_i)$-(n+1)/2	$R(x_i)$-(n+1)/2	Numer.
40.2	53.4	7	18	-3.5	7.5	-26.25
53.4	43.5	19	9	8.5	-1.5	-12.75
43.5	37.7	10	4	-0.5	-6.5	3.25
37.7	50.2	4	15	-6.5	4.5	-29.25
50.2	38.7	16	5	5.5	-5.5	-30.25
38.7	47.8	5	13	-5.5	2.5	-13.75
47.8	39.5	14	6	3.5	-4.5	-15.75
39.5	44.9	6	11	-4.5	0.5	-2.25
44.9	41.7	12	8	1.5	-2.5	-3.75
41.7	36.4	9	3	-1.5	-7.5	11.25
36.4	51.2	3	16	-7.5	5.5	-41.25
51.2	54.3	17	20	6.5	9.5	61.75
54.3	41.5	20	7	9.5	-3.5	-33.25
41.5	44.8	8	10	-2.5	-0.5	1.25
44.8	46.7	11	12	0.5	1.5	0.75
46.7	51.8	13	17	2.5	6.5	16.25
51.8	49.5	18	14	7.5	3.5	26.25
49.5	34.1	15	2	4.5	-8.5	-38.25
34.1	33.2	2	1	-8.5	-9.5	80.75
33.2	53.7	1	19	-9.5	8.5	-80.75
33.2	-	-	-	-	-	-
					Sum	-126.00

4.10 MULTIVARIATE ANALYSES

There are several multivariate procedures in addition to the multivariate regression discussed in 4.7.4. Mathematical descriptions of these procedures are beyond the scope of this guidance, but researchers should consult a statistician to assess the opportunities for using these procedures. In general, the multivariate procedures described in this section have not found wide usage in day-to-day applications.

With the current availability of computerized statistical procedures (e.g., SAS, SPSS), it is possible to perform multivariate analyses with ease, requiring of the researcher only that he or she understands and meets the assumptions of the particular test and knows how to interpret correctly the results of the test. It is extremely important that a qualified statistician be consulted regarding the assumptions involved and the appropriate interpretation of test results. Without such precautions, our current computer technology will only facilitate the proliferation of misguided analyses and misinterpreted results.

The multivariate analyses described briefly in this guidance include canonical correlation, cluster analysis, principal components and factor analysis, and discriminant analysis. These procedures were selected for discussion based on the work of Gaugush (1986), which should be reviewed in addition to the detailed discussions provided in statistics texts for a better understanding of these multivariate analyses.

4.10.1 Canonical Correlation

Canonical correlation is a technique for analyzing the relationship between two sets of variables, with each set able to contain several variables (SAS Institute, Inc., 1985b). It follows that simple and multiple correlation are special cases of canonical correlation in which one or both sets of variables contain only one variable (SAS Institute, Inc., 1985b).

Gaugush (1986) states that "[c]anonical correlation is used to identify and estimate a linear function (called a canonical variate) of one set of variables that is maximally correlated with a linear function of a second set of variables." The SAS CANCORR procedure (SAS Institute, Inc., 1985b) finds as many canonical variates as there are variables in the smaller set of variables. The first and subsequent canonical variates are uncorrelated, with the first having the highest correlation coefficient, followed by the second-highest correlation coefficient for the second canonical variate, etc. It should be noted that "the first canonical correlation is at least as large as the multiple correlation between any variable and the opposite set of variables" (SAS Institute, Inc., 1985b).

Gaugush (1986) notes that the information resulting from canonical correlation is largely descriptive and therefore the procedure has not been used as much as other multivariate procedures that support hypothesis testing and/or prediction.

Gaugush (1986) promotes the use of canonical correlation to, for example, "describe the strength of a relationship between a linear combination of nutrient variables and a linear combination of biomass-related variables." The strength of such a relationship is estimated by the canonical correlation coefficient.

Another use of canonical correlation is in determining how many "common elements" are contained within two sets of variables (Gaugush, 1986). The percent overlapping variance (i.e., the squared canonical correlation coefficient) can be used to indicate the relative importance of each canonical variate (Gaugush, 1986).

To use canonical correlation in hypothesis testing, it is important that the assumption of multivariate normality is satisfied (Gaugush, 1986). Snedecor and Cochran (1980) discuss the multivariate normal distribution briefly and state its property that "any variable has a linear regression on the other variables (or on any subset of the other variables), with deviations that are normally distributed." Gaugush (1986) notes that the assumption of multivariate normality is often satisfied by "creating data distributions that are approximately normal."

To satisfy the assumptions of canonical correlation, Gaugush (1986) recommends:

- Use transformations if needed to create roughly symmetric univariate data distributions.
- Carefully examine the validity of outliers and run analyses with and without outliers to document their impact on the correlations.
- Transform data if necessary to create linear relationships among the variables in each set of variables.

Finally, Gaugush (1986) gives an example application of canonical correlation using the SAS CANCORR procedure described above.

4.10.2 Cluster Analysis

Cluster analysis is a classification method for placing "objects into groups or clusters suggested by the data, not defined a priori, such that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar" (SAS Institute, Inc., 1985b). SAS offers several clustering options under the CLUSTER procedure (SAS Institute, Inc., 1985b). It is important to recognize that numerous methods come under the heading of cluster analysis and these methods will give different results. The types of cluster analysis include the following (SAS Institute, Inc., 1985b):

- Disjoint clusters, which place each object in one and only one cluster.
- Hierarchical clusters, in which one cluster may be contained entirely within another cluster, but for which no other kind of overlap is allowed.
- Overlapping clusters with or without constraints placed on the number of objects that belong to two clusters.
- Fuzzy clusters, which are defined by a probability of membership of each object in each cluster. (These can be disjoint, hierarchical, or overlapping.)

Example analyses include the following:

- Gaugush (1986) used Ward's method of cluster analysis to group reservoirs based on similarity in log total phosphorus concentration, log total nitrogen concentration, log Secchi disk depth, and log chlorophyll *a* concentration.
- Kimball (1986) used cluster analysis to group wells based on mean nitrate, well depth, maximum nitrate, coefficient of variation of nitrate, and variance of nitrate. Mean nitrate and coefficient of variation of nitrate yielded the most information. A major conclusion made from this investigation of wells in South Dakota was that "classification of ground water sample locations by geologic environment and depth is crucial to understanding the system."

4.10.3 Principal Components and Factor Analysis

Principal component analysis (PCA) is a multivariate procedure for examining relationships among several quantitative variables (SAS Institute, Inc., 1985b). PCA is used with factor analysis to "create a relatively small number of new variables (called 'factors') from a larger number of original variables" (Gaugush, 1986). The primary use of these procedures is exploratory analysis; that is, hypothesis testing is not normally performed (SAS Institute, Inc., 1985b).

Gaugush (1986) notes that PCA is usually performed before factor analysis. Principal components are linear combinations of the original variables. The first principal component explains the most variability associated with the data, while the second principal component explains the second-most variability associated with the data and is not correlated to the first principal component. As an example, Gaugush (1986) describes how PCA can be used to develop a trophic state index from biological, nutrient, and physical data. It is sometimes helpful to prepare a scatter plot of the data using the first two principal components for exploratory analysis.

Factor analysis is then used to enhance the scientific interpretation of the principal components developed. Factor analysis can then be used to redefine the factors (i.e., the linear functions of one or more of the original variables) so that they can be interpreted in more scientific

terms. That is, factor analysis can be used to reshape a principal component such that the factors match more closely a researcher's intuitive (or research-based) model of the relationships among the variables.

Although hypothesis testing is not normally performed on the results of PCA and factor analysis, Gaugush (1986) recommends that data distributions be approximately symmetric with no outliers. As in other cases, data transformations might be needed to meet these recommendations. Because of problems of scale, Gaugush (1986) recommends that PCA and factor analysis be based on the correlation matrix unless the variables are all of approximately the same magnitude. In cases where the variables are of the same magnitude, the covariance matrix can be used.

This discussion of PCA and factor analysis is intended only to familiarize the water quality researcher with the general use of these techniques. Gaugush (1986) goes several steps farther in describing these procedures, including an illustrative example. SAS gives a fairly detailed mathematical description of PCA and factor analysis (SAS Institute, Inc., 1985b) and offers procedures for performing both (PRINCOMP and FACTOR procedures).

4.10.4 Discriminant Analysis

Discriminant analysis resembles regression analysis, but with a major difference in that the dependent variable in discriminant analysis is categorical, whereas the dependent variable in regression analysis is often continuous (Gaugush, 1986). An example application of discriminant analysis might be to predict the presence or absence of brook trout based on pH and aluminum concentration. Researchers are encouraged to follow the descriptions of discriminant analysis offered by SAS (SAS Institute, Inc., 1985b) and Gaugush (1986) before using the procedure. The following are some of the uses for discriminant analysis (SAS Institute, Inc., 1985b):

- To find a mathematical rule (or "discriminant function") for predicting to which class an observation belongs, given data for the independent quantitative variables.
- To find linear combinations of the independent quantitative variables that best reveal the differences between the classes.
- To find a subset of the independent quantitative variables that best shows the differences between the classes.

Discriminant analysis requires prior knowledge of all classes (e.g., a sample), whereas cluster analysis has no such requirement (SAS Institute, Inc., 1985b). In fact, cluster analysis is used to *define* the classes. Gaugush (1986) also cautions that outliers can adversely affect the results of discriminant analysis and that the predictor variables should follow a multivariate normal distribution within each group, with variance-covariance matrices that are constant across groups. There is, however, at least one procedure (NEIGHBOR procedure) that can be used for non-normal data (SAS Institute, Inc., 1985b).

4.11 Extreme Events

One of the key characteristics that separate environmental, and in particular nonpoint source-influenced data, is the presence of extreme events. The majority of nonpoint source pollution entering streams occurs during runoff from precipitation events. This section presents an approach for estimating annual precipitation and storm events, describes the approach used by EPA's DESCON model for estimating design flows, and concludes with statistical methods appropriate for evaluating water quality extreme events. Earlier sections describe methods for summarizing average

conditions and determining changes. This section also describes methods for evaluating extreme conditions in water quality variables. This is important for evaluating standard violations or evaluating peak concentrations to determine if a BMP was effective.

4.11.1 Rainfall Analyses

Annual precipitation

Chow (1951) presents a method for computing annual precipitation for a variety of return periods. This method is outlined below assuming that the annual rainfall is available for *n* years.

- Compute the mean and standard deviation for the *n* years of data. Also compute the coefficient of variation (*CV*).
- Use *CV* to estimate the log-probability frequency factor, *K*, for a given return period (Table 4-40).
- Compute the annual precipitation (X_c) for different return periods using Equation 4-107.

$$X_c = \bar{x}[1 + (CV)(K)] \qquad (4\text{-}107)$$

For the rainfall data presented in Table 4-37, $\bar{x}$ and *CV* are equal to 44.5 inches and 0.15, respectively. From Table 4-40, the value of *K* corresponding to a 2 year return period is -0.09. Substituting this value into the above equation yields X_c equal to 44.5(1+(0.15)(-0.09)) or 43.9 inches. The 100 year annual precipitation would be equal to 44.5(1+(0.15)(2.70)) = 62.5 inches. The adequacy of the record length can be evaluated using (Mockus, 1960):

$$Y = (4.30\,t\log_{10}R)^2 + 6 \qquad (4\text{-}108)$$

where *Y* is the minimum record length in years, *t* is the Student's *t* quantile (Table D2) at the 90% level with *Y*-6 degrees of freedom, and *R* is the ratio of the 100 year event to the 2 year event.

To solve the above equation, an iterative approach is necessary. Using an initial guess of *Y* equal to 15 years, *t* is equal to 1.8331, while *R* is equal to 62.5/43.9 or 1.42. Substituting these values into the above equation yields $Y = [(4.3)(1.8331)(.1534)]^2+6$ or 7.5. Adjusting our guess of *Y* to 9 years, *t* is equal to 2.3534 and $Y = [(4.3)(2.3534)(.1534)]^2+6$ or 8.4 years (which is close enough to our initial guess). Since the actual length of record is 21 years, our 100 year return annual precipitation estimate of 62.5 inches can be expected to be reasonable.

Storm return period

The method developed by Hershfield (1961) is the most usually applied method in the field today and is commonly referred to as "TP40." The method is based on interpolating the design storm from four figures (Figures 4-25 through 4-28) and applying the following equation (Weiss, 1962):

$$I = 0.0256(C-A)x + 0.000256[(D-C)-(B-A)]xy + 0.01(B-A)y + A \qquad (4\text{-}109)$$

where *I* is the rainfall amount (in inches); *A* is the 2-year, 1-hour rainfall (in inches) interpolated from Figure 4-25; *B* is the 2-year, 24-hour rainfall (in inches) interpolated from Figure 4-26; *C* is the 100-year, 1-hour rainfall (in inches) interpolated from Figure 4-27; and *D* is the 100-year, 24-hour rainfall (in inches) interpolated from Figure 4-28. The return period, *x*, and duration, *y*, are taken from Table 4-41 and 4-42, respectively.

Table 4-40. Theoretical log-probability frequency factors.

	Return Period (years)					
	1.01	2	5	20	100	
	Probability (%) equal to or greater than the given variate					
C_s	99	50	20	5	1	C_v
0.0	-2.33	0.0	0.84	1.64	2.33	0.0
0.5	-1.98	-0.09	0.80	1.77	2.70	0.166
1.0	-1.68	-0.15	0.75	1.85	3.03	0.324
1.139	-1.61	-0.16	0.73	1.86	3.11	0.363
1.4	-1.49	-0.19	0.69	1.88	3.26	0.436
1.5	-1.45	-0.20	0.68	1.89	3.31	0.462
2.0	-1.28	-0.24	0.61	1.89	3.52	0.596
3.0	-1.04	-0.28	0.51	1.85	3.78	0.818
4.0	-0.90	-0.29	0.42	1.78	3.91	1.000

Source: Chow, 1951

Table 4-41. Linearized rainfall frequency variate for equation 4-109.

Return Period (in years)	1	2	5	10	25	50	100
Linearized Variate (x)	-6.93	0	9.2	16.1	25.3	32.1	39.1

Source: Weiss, 1962

Table 4-42. Linearized rainfall duration variate for equation 4-109.

Duration (hours)	0.17	.033	0.5	0.67	1
Linearized Variate (y)	-37	-24	-15.6	-9.4	0
Duration (hours)	2	3	6	12	24
Linearized Variate (y)	17.6	28.8	49.9	73.4	100.0

Source: Weiss, 1962

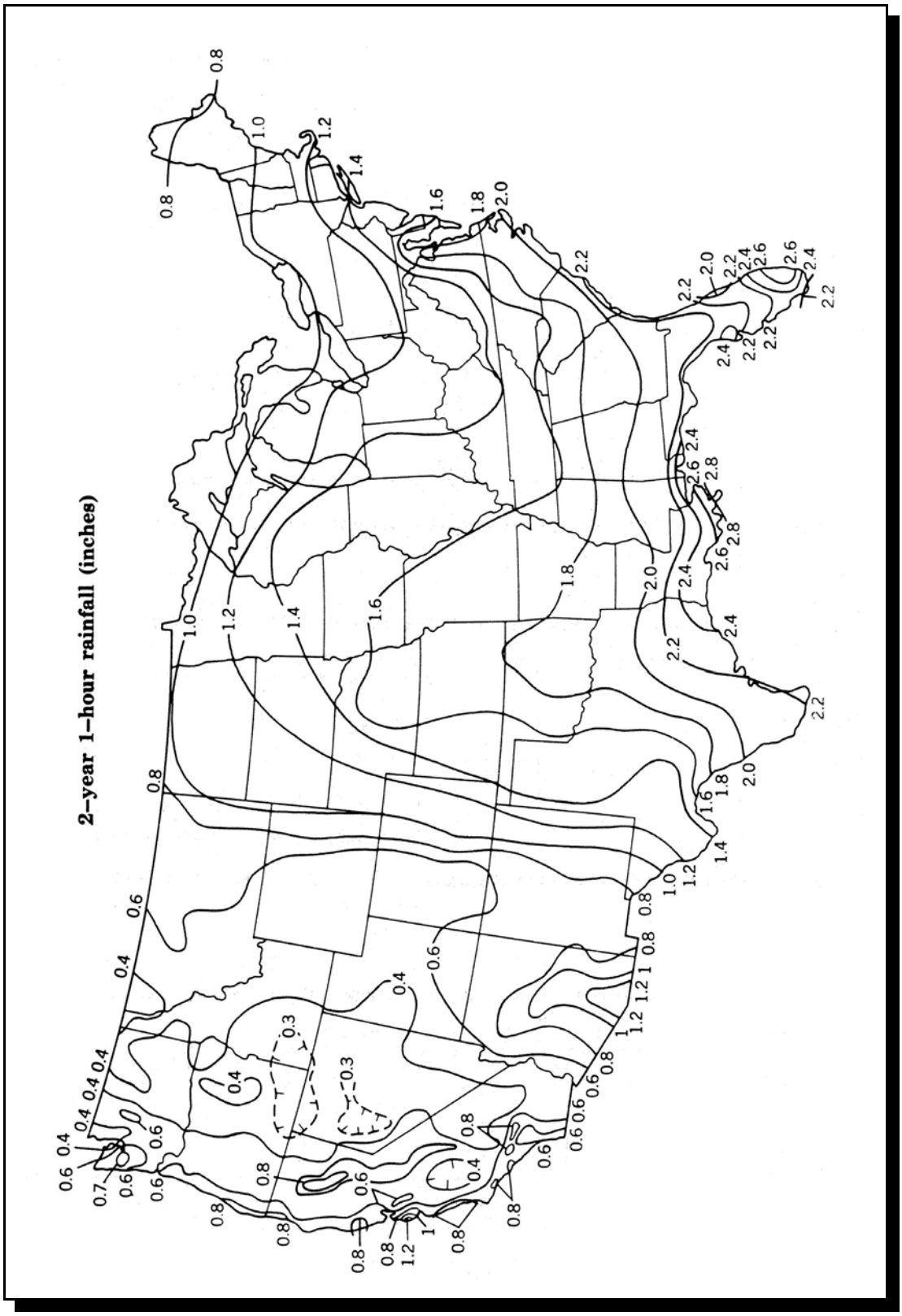


Figure 4-25. One-hour rainfall to be expected at a return period of 2 years (Source: Schwab et al., 1981)

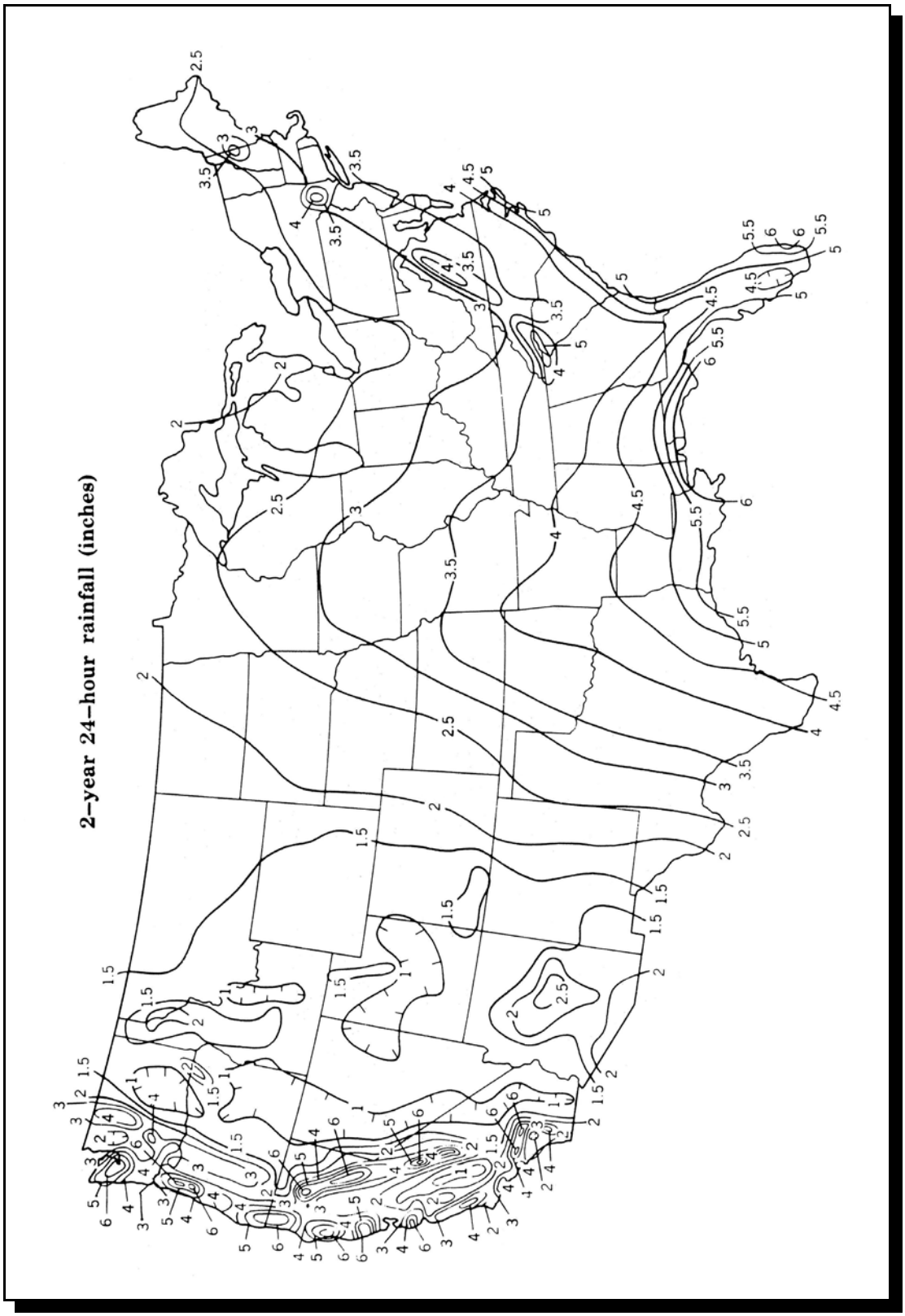


Figure 4-26. 24-hour rainfall to be expected at a return period of 2 years (Source: Schwab et al., 1981)

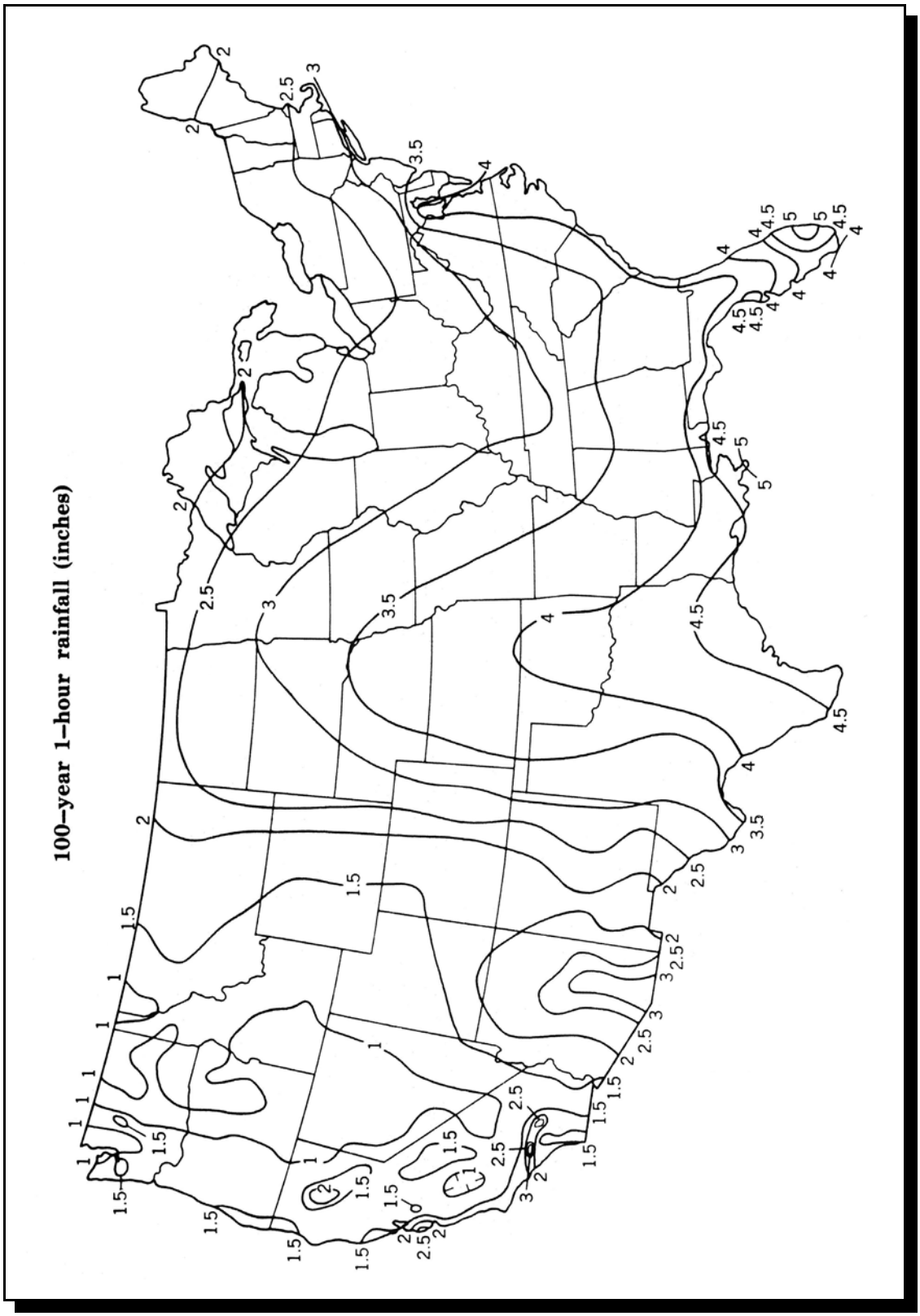


Figure 4-27. One-hour rainfall to be expected at a return period of 100 years (Source: Schwab et al., 1981)

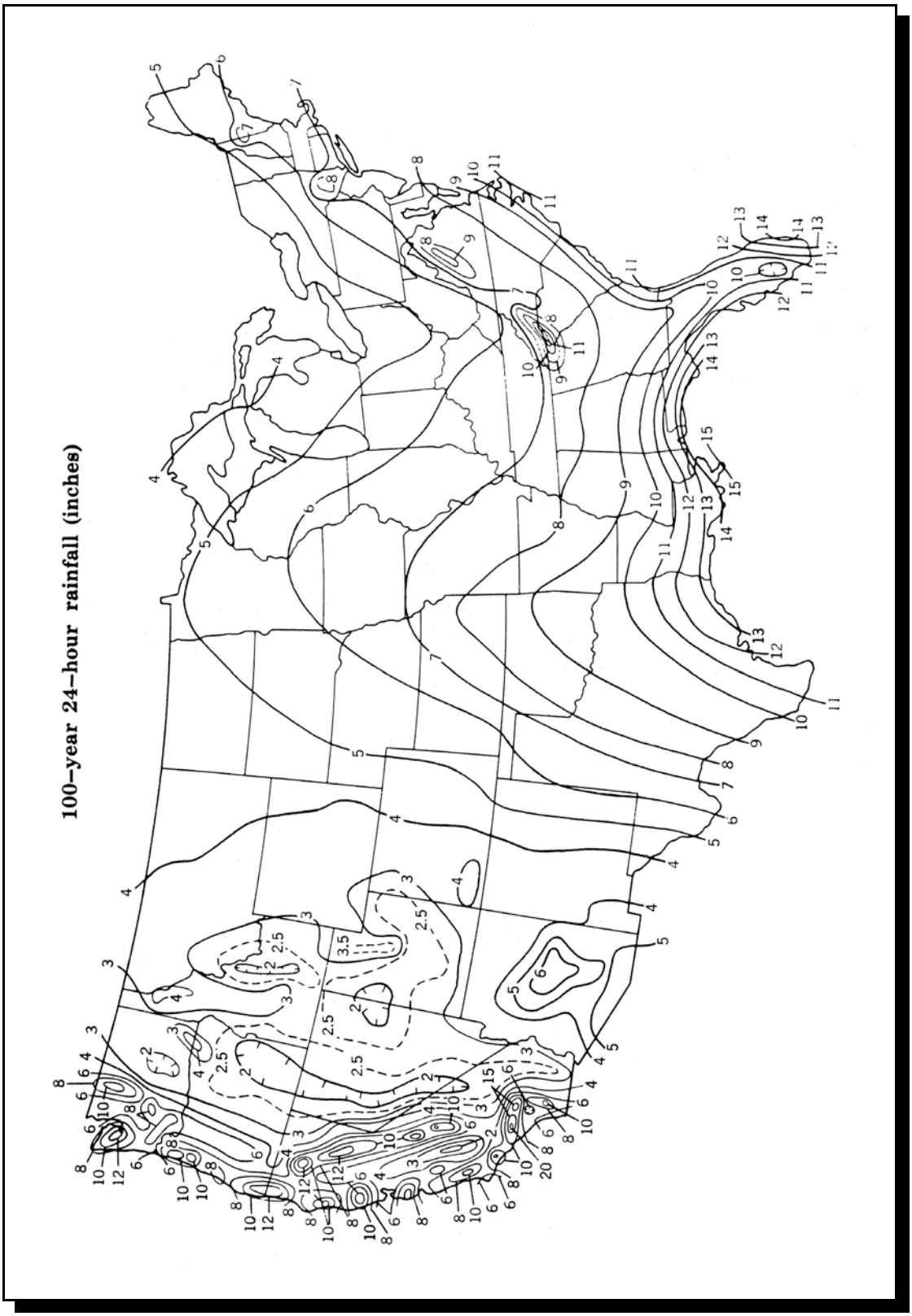


Figure 4-28. 24-hour rainfall to be expected at a return period of 100 years (Source: Schwab et al., 1981)

Suppose the analyst is interested in estimating the 1-year, 2-hour storm in El Paso, Texas. From Figures 4-25 through 4-28, *A*, *B*, *C*, and *D* are estimated as 0.8, 1.5, 2.0, and 4.0, respectively. From Table 4-41, *x* is equal to -6.93 and from Table 4-42, *y* is equal to 17.6. Substituting these values into Equation 4-109 yields a 1-year, 2-hour storm equal to 0.7 inches.

4.11.2 Design Flows

This section describes the computational steps employed by DFLOW and DESCON for each of the three types of design flows considered and has been extracted and adapted from Rossman (1990). It begins with the extreme value design flow, since this type of design flow also serves as a starting point in computing the biologically-based design flow.

Extreme value design flow (low flows)

The extreme value design flow is computed from the sample of lowest *m*-day average flows for each year of record, where "*m*" is the user-supplied flow averaging period. Established practice uses arithmetic averaging to calculate these *m*-day average flows. A log Pearson Type III probability distribution is fitted to the sample of annual minimum *m*-day flows. The design flow is the value from the distribution whose probability of not being exceeded is 1/*R*, where *R* is the user-supplied return period. The procedure is modified slightly to accommodate situations where some annual low flows are zero.

STEP 1. Initialize each element of a vector *X* of daily flow values to UNKNOWN (i.e., a very large number such as 1x 10^{20}).

STEP 2. Read in daily flow values from the retrieved STORET flow file into *X*, where *X*(1) corresponds to the first day of record. (Note: February 29th of leap years is ignored.)

STEP 3. Create *m*-day running arithmetic averages from the daily flows in *X*, and replace the daily flows of *X* with these values. The running average of *X*(*i*), *X*(*i*+1), ..., *X*(*i*+*m*-1) is placed in *X*(*i*).

STEP 4. Find the lowest *m*-day running average value for each water year recorded in *X* (where a water year begins on April 1) and store the resulting values in vector *Y*. Let *NY* denote the number of entries in *Y*.

STEP 5. Let *N* be the number of non-zero entries in *Y*. Assume that these *Y*-values are a sample drawn from a log Pearson Type III probability distribution. The design flow is the value from this distribution whose probability of not being exceeded is 1/*R*, where *R* is the user-supplied return period. Use the following procedure to find the design flow:

STEP 5a. Find the mean (*U*), standard deviation (*S*), and skewness coefficient (*G*) of the natural logarithms of the non-zero entries in *Y*.

STEP 5b. Let *F*0 be the fraction of entries in *Y* that are zero:

$$F0 = (NY - N)/NY \qquad (4\text{-}110)$$

Let *P* be the cumulative probability corresponding to the user-supplied return period of *R* years, adjusted for the presence of zero-flow years:

$$P = (1/R - F0)/(1 - F0) \qquad (4\text{-}111)$$

In other words, if *F*0 is the probability of having a year with zero stream flow, and 1/*R* is the allowed probability of a year with an excursion below the design flow, then *P* is the corresponding excursion probability in years with non-zero flows.

STEP 5c. Let *Z* be the standard normal deviate corresponding to cumulative probability *P*. *Z* can be computed using the following formula (Joiner and Rosenblatt, 1971):

$$Z = 4.91(P^{0.14} - (1 - P)^{0.14}) \quad (4\text{-}112)$$

TEP 5d. Compute the gamma deviate, *K*, corresponding to the standard normal deviate *Z* and skewness *G* using the Wilson-Hilferty transformation (Loucks et al., 1981):

$$K = (2/g)((1 + GZ/6 - G^2/36)^3 - 1) \quad (4\text{-}113)$$

STEP 5e. Compute the design flow as

$$\exp(U + KS) \quad (4\text{-}114)$$

Biologically-based design flow

Biologically-based design flows are computed by starting with a trial design flow, then counting how often this flow is not exceeded by *m*-day average flows in the historical record. (In contrast with the traditional method of computing extreme value design flows, the *m*-day flow averages are harmonic means, not arithmetic ones. This count is compared to the allowed number of such occurrences, and the trial design flow is adjusted accordingly. The specific computational steps involved are as follows:

STEP 1. Initialize each element of a vector *X* of daily flow values to UNKNOWN (i.e., a very large number such as 1x 10^{20}).

STEP 2. Read in daily flow values from the retrieved STORET flow file into *X*, where *X*(1) corresponds to the first day of record. (Note: February 29th of leap years is ignored.)

STEP 3. Create *m*-day running harmonic averages from the daily flows in *X*, and replace the daily flows of *X* with these values. The running average of *X*(*i*), *X*(*i*+1), ..., *X*(*i*+*m*-1) is placed in *X*(*i*) and is computed as follows:

Define *B*(*j*) as 1/*X*(*i*+*j*-1) if *X*(*i*+*j*-1 > 0, and 0 otherwise, for *j* = 1 to *m*. Let DSUM be the sum of *B*(*j*) for *j* = 1 to *m* and *m*0 be the number of *B*(*j*) values that equal 0. Then replace *X*(*i*) with *X*(*i*) = (*m*-*m*0)/DSUM*(*m*-*m*0)/*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 percent of FL, then end with DFLOW = FU. If XL is within 0.5 percent of *A*, then end with DFLOW = FL. If XU is within 0.5 percent of *A*, then end with DFLOW = FU. Otherwise proceed to the next step.

STEP 7c. Interpolate between the bounds to find a new trial design flow, FT:

$$FT = FL + (FU - FL)(A - XL)/(XU - XL) \tag{4-115}$$

and compute the number of excursions (XT) occurring for this flow (see procedure described below).

STEP 7d. Update the bounds based on the value of XT: If XT ≤ *A*, then set FL = FT and XL = XT. Otherwise set FU = FT and XU = XT. Then return to the convergence check of step 7b.

The process used to count the number of flow excursions for a given design flow proceeds in two phases. The first phase identifies all excursion periods in the period of record. An excursion period is a sequence of consecutive days where each day belongs to an *m*-day running average flow that is below the given design flow. Recall that "*m*" is the flow averaging period set by the user. Phase two groups these excursion periods into excursion clusters and counts up the total number of excursions occurring within all clusters. An excursion cluster consists of all excursion periods falling within a prescribed length of time from the start of the first period in the cluster (120 days is the default cluster length). The number of excursions counted per cluster is subject to an upper limit whose default value is 5.

Before describing the detailed procedures for each of these phases a simple numerical example will be used to illustrate the method. Suppose that the design flow under consideration is 100 cfs and that the period of record yields a sequence of 4-day running average flows as detailed in Box 1.

The first flow excursion period for this record consists of the 4-day averages occurring on days 1,2 and 3. Thus the period extends from day 1 to day 6 (days 4, 5 and 6 belong to the averaging period that begins on day 3). There are two other excursion periods consisting of days 13 to 18 and 513 to 548. Under the default clustering parameters, there are 2 excursion clusters; cluster 1 contains periods 1 and 2, and cluster 2 contains period 3. The number of excursions in each cluster is detailed in Box 2.

Note that the number of excursions in each period equals the period length divided by the averaging period. The nominal number of excursions in cluster 2 is 9, and since this exceeds the limit of 5, only 5 are counted. The total number of excursions for the design flow of 100 cfs in this example is 3 + 5 = 8.

The detailed procedure for counting biologically-based flow excursions under a specified design flow is as follows:

PHASE 1

<u>Define</u>:

P1(*i*) = day which begins excursion period *i*,
P2(*i*) = day which ends excursion period *i*,
XP(*i*) = number of excursions in period *i*,
XKLmax = maximum cluster length (e.g., 120 days).
t = current day of record.

STEP 1. Set *i* = 0, P2(0) = 0, and *t* = 1.

STEP 2. If the *m*-day running average beginning on day *t* is greater or equal to the specified design flow then proceed to Step 5.

Day	4-Day Average Flow (cfs)	Day	4-Day Average Flow (cfs)
1	34	513-545	< 100
2	65		
3	25	546-end	> 100
4-12	> 100		
13	57		
14	34		
15	26		
16-512	> 100		

Box 1

Cluster	Period	Start Day	Length (days)	No. of Excursions in Period	No. of Excursions in Cluster
1	1	4	6	6/4 = 1.5	3.0
	2	13	6	6/4 = 1.5	
2	3	513	36	36/4 = 9.0	5.0

Box 2

STEP 3. If the current day t is more than a day beyond the end of the current excursion period (t > P2(i) + 1), or if the length of the current excursion period equals XKLmax then begin a new excursion period by setting:

$i = i + 1$
P1(i) = t
P2(i) = m - 1
XP(i) = 0.

STEP 4. Update the ending day of the current excursion period and the excursion count for this period:

P2(i) = P2(i) + 1
XP(i) = (P2(i) - P1(i)) / m.

STEP 5. Proceed to the next day of record ($t = t$ + 1). If not at the end of the record then return to Step 2. Otherwise proceed to phase 2.

PHASE 2

Define:

i = current excursion period,
k = current excursion cluster,
K1 = day of record which begins cluster k,
XK(k) = number of excursions in cluster k,
Xkmax = maximum number of excursions counted per cluster (e.g., 5),

STEP 1. Set i = 1, k = 0, and K1 = a large negative number.

STEP 2. If the length of the current cluster is greater than the maximum length (i.e., P2(i) - K1 > XKLmax) then begin a new cluster with excursion period i, i.e.,

$k = k + 1$
K1 = P1(k)
XK(k) = 0.

STEP 3. Update the excursion count for the current cluster,

XK(k) = minimum(XK(k) + XP(i), XKmax).

STEP 4. Proceed to the next excursion period ($i = i + 1$) and return to Step 2. If no more excursion periods remain, then total up the number of excursions in each cluster (XK(1) +XK(2) + ... + XK(k)) to determine the total number of excursions.

4.11.3 Frequency of Extreme Events

This section describes methods for evaluating extreme conditions in water quality variables. This is an important consideration for evaluating standard violations or evaluating peak concentrations to determine if a BMP was effective. Gilbert (1987) presents an approach for evaluating proportions. The method is based on computing the number of observations exceeding a threshold value X_c. The proportion of observations, p, exceeding X_c can be computed as

$$p = u/n \tag{4-116}$$

where u is the number of observations exceeding X_c and n is the number of observations. For $n \leq 30$, Table D11 can be used to develop nonparametric 90th or 95th percentile confidence limits. For $n > 30$, Equations 4-117 and 4-118 may be used. The lower limit is equal to 0 if u is 0 and the upper limit is 1 if the u is equal to n.

If np and $n(1-p)$ are greater than 5 (some authors suggest a value of 10), then Gilbert (1987) suggests that the normal approximation can be used to compute the upper and lower limits with the following equation:

$$p \pm Z_{1-\alpha/2}\left[\frac{p(1-p)}{n}\right]^{1/2} \tag{4-119}$$

The confidence intervals can be used to evaluate one-sample hypotheses such as

H_o: $p = 0.10$
H_a: $p \neq 0.10$

$$\text{Lower limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{(u-0.5) + \frac{Z_{1-\alpha/2}^2}{2} - Z_{1-\alpha/2}\left[(u-0.5) - \frac{(u-0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4}\right]^{1/2}\right\} \tag{4-117}$$

$$\text{Upper limit} = \frac{1}{n+Z_{1-\alpha/2}^2} \times \left\{(u+0.5) + \frac{Z_{1-\alpha/2}^2}{2} + Z_{1-\alpha/2}\left[(u+0.5) - \frac{(u+0.5)^2}{n} + \frac{Z_{1-\alpha/2}^2}{4}\right]^{1/2}\right\} \tag{4-118}$$

If the 95 percent confidence intervals include 0.10, we accept the null hypothesis. Otherwise the null hypothesis is rejected.

An evaluation of proportions can also be used to determine the necessary sample size to ensure that *q* percent of the population is less than the largest randomly sampled observation. This approach provided by Conover (1980) is demonstrated with the next example.

Example:

Determine the number of random samples that would be required to ensure with a 95 percent probability (α=0.05) that 90 percent of the population is less than the largest observation.

Solution:

Enter Table D11 with *q* equal to 0.9 and 1-α equal to 0.95 and directly read a sample size of 29. Therefore, it would require 29 samples to ensure that the largest observation is greater than 90 percent of the population.

Application of this example is similar to quality control processes. In this case, once 29 samples have been collected, the upper bound is set equal to the largest observation. From then on, we would expect that only 10 percent of the future samples would exceed the upper bound with 95 percent confidence. If more than 10 percent of future observations exceeded the upper bound, we would infer that some change has occurred (Ward et al., 1990).

It is also possible to compare the proportions p_1 and p_2 between two samples with sample sizes equal to n_1 and n_2. For example, it may be appropriate to compare the percent of standard violations from before and after. In this case, the null and two-sided alternative hypothesis are

H_o: $p_1 = p_2$
H_a: $p_1 \neq p_2$

Moore and McCabe (1989) provide the test statistics as

$$z = \frac{p_1 - p_2}{s_p} \tag{4-120}$$

where s_p and p are given by

$$s_p = \sqrt{p\ (1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \tag{4-121}$$

$$p = \frac{u_1 + u_2}{n_1 + n_2} \tag{4-122}$$

Moore and McCabe (1989) suggest that n_1p, $n_1(1 - p)$, n_2p, and $n_2(1 - p)$ all be greater than or equal to 5 for application. If the absolute value of *z* is greater than the associated normal deviate (e.g., 1.96 for a two-sided test with α equal to 0.05), then H_o is rejected.

5. QUALITY ASSURANCE AND QUALITY CONTROL

5.1 INTRODUCTION

Quality assurance (QA) and quality control (QC) are commonly thought of as procedures used in the laboratory to ensure that all analytical measurements made are accurate. Yet QA and QC extend beyond the laboratory and are essential components of all phases and all activities within each phase of a nonpoint source (NPS) monitoring project. This section defines QA and QC, discusses their value in NPS monitoring programs, and explains EPA's policy on these topics. The following sections provide detailed information and recent references for planning and ensuring quality data and deliverables that can be used to support specific decisions involving nonpoint source pollution.

5.1.1 Definitions of Quality Assurance and Quality Control

Quality assurance:

An integrated system of management procedures and activities used to verify that the quality control system is operating within acceptable limits and to evaluate the quality of data (Taylor, 1993; USEPA, 1994c).

Quality control:

A system of technical procedures and activities developed and implemented to produce measurements of requisite quality (Taylor, 1993; USEPA, 1994c).

QC procedures include the collection and analysis of blank, duplicate, and spiked samples and standard reference materials to ensure the integrity of analyses and regular inspection of equipment to ensure it is operating properly. QA activities are more managerial in nature and include assignment of roles and responsibilities to project staff, staff training, development of data quality objectives, data validation, and laboratory audits. Table 5-1 lists some common activities that fall under the headings of QA and QC. Such procedures and activities are planned and executed by diverse organizations through carefully designed quality management programs that reflect the importance of the work and the degree of confidence needed in the quality of the results.

5.1.2 Importance of QA/QC Programs

Although the value of a QA/QC program might seem questionable while a project is under way, its value should be quite clear after a project is completed. If the objectives of the project were used to design an appropriate data collection and analysis plan, all QA/QC procedures were followed for all project activities, and accurate and complete records were kept throughout the project, the data and information collected from the project will be adequate to support a choice from among alternative courses of action. In addition, the course of action chosen will be defensible based on the data and information collected. Development and implementation of a QA/QC program can require up to 10 to 20 percent of project resources (Cross-Smiecinski and Stetzenback, 1994), but this cost can be recaptured in lower overall costs due to the project's being well planned and executed. Likely problems are anticipated and accounted for before they arise, eliminating the need to spend countless hours and dollars resampling, reanalyzing data, or mentally reconstructing portions of the project to determine where an error was introduced. QA/QC procedures and activities are cost-effective measures used to determine how to allocate project energies and resources toward improving the quality of research and the usefulness of project results (Erickson et al., 1991).

Table 5-1. Common QA and QC activities.

QA Activities

- Organization of project into component parts
- Assignment of roles and responsibilities to project staff
- Use of statistics to determine the number of samples and sampling sites needed to obtain data of a required confidence level
- Tracking of sample custody from field collection through final analysis
- Development and use of data quality objectives to guide data collection efforts
- Audits of field and laboratory operations
- Maintenance of accurate and complete records of all project activities
- Personnel training to ensure consistency of sample collection techniques and equipment use

QC Activities

- Collection of duplicate samples for analysis
- Analysis of blank and spike samples
- Replicate sample analysis
- Regular inspection and calibration of analytical equipment
- Regular inspection of reagents and water for contamination
- Regular inspection of refrigerators, ovens, etc. for proper operation

Adapted from Drouse et al., 1986, and Erickson et al., 1991.

This chapter discusses many elements and aspects of QA/QC programs that do not differ significantly from one type of program to another—for instance, from a point source permit compliance sampling program to an NPS best management practice effectiveness monitoring program. Therefore, much of the following discussion is not specific to NPS projects. This does not, however, mean that a well-designed and well-implemented QA/QC program is not necessary for an NPS project. It is hoped that the following discussion will convey to the reader the importance of QA and QC to the success of every project involving the collection and analysis of environmental data.

5.1.3 EPA Quality Policy

EPA has established a QA/QC program to ensure that data used in research and monitoring projects are of known and documented quality to satisfy project objectives. The use of different methodologies, lack of data comparability, unknown data quality, and poor coordination of sampling and analysis efforts can delay the progress of a project or render the data and information collected from it insufficient for decision making. QA/QC practices should be used as an integral part of the development, design, and implementation of an NPS monitoring project to minimize or eliminate these problems (Erikson et al., 1991; Pritt and Raese, 1992; USEPA, 1994d).

EPA's mandatory agency-wide Quality System policy requires each office or laboratory generating data to implement minimum procedures to ensure that precision, accuracy, completeness, comparability, and representativeness of data are known and documented (Erickson et al., 1991; USEPA, 1984c). This policy is now based on the quality system standard developed by the American Society of Quality Control (ASQC, 1994). Each office or laboratory is required to specify the quality levels that data must meet to be acceptable and satisfy project objectives. This requirement applies to all environmental monitoring and measurement efforts mandated or supported by EPA through regulations, grants, contracts, or other formal agreements. To ensure that this responsibility is met uniformly across EPA, each organization performing work for EPA must document in a Quality Management Plan (QMP) that is approved by its senior management how it will plan, implement, and assess the effectiveness of QA and QC operations applied to environmental programs (USEPA, 1994d). In addition, each non-EPA organization must have a well-documented Quality Assurance Project Plan (QAPP) that covers each monitoring or measurement activity associated with a project (Erickson et al., 1991; USEPA, 1983c, 1994).

The purpose of writing a QAPP prior to undertaking an NPS monitoring project is to establish clear objectives for the program, including the types of data needed and the quality of the data generated (accuracy, precision, completeness, representativeness, and comparability). This information is used to design the program to meet these objectives. Developing a QAPP prior to undertaking the NPS monitoring project also establishes the boundaries of the project, in terms of the time allotted to it and the decisions that can realistically be made from the data and information that will be collected.

The QAPP should specify the policies, organization, objectives, functional activities, QA procedures, and QC activities designed to achieve the data quality goals of the project. It should be distributed to all project personnel, and they should be familiar with the policies and objectives outlined in the QAPP to ensure proper interaction of the sampling and laboratory operations and data management. All persons involved in an NPS monitoring project who either perform or supervise the work done under the project are responsible for ensuring that the QA/QC procedures and activities established in the QAPP are adhered to.

The QMP and each QAPP must be submitted for review to the EPA organization responsible for the work to be performed, and they must be approved by EPA or its designee (e.g., federal or state agency) as part of the contracting or assistance agreement process before the work can begin. In addition, it is important to note that the QMP and QAPP are "live" documents and programs in the sense that once they have been developed they cannot be placed on a shelf for the remainder of the project. All QA/QC procedures should be evaluated and plans updated as often as necessary during the course of a project to ensure that they are in accordance with the present project direction and efforts (Knapton and Nimick, 1991; USEPA, 1994c).

5.2 Data Quality Objectives (DQOs)

Before collecting environmental data in support of an NPS project, it is important to determine the type, quantity, and quality of data needed to meet the project objectives and support a specific decision based on the results of the project. Not doing so creates the risk of expending too much effort on data collection (i.e., more data are collected than necessary), not expending enough effort on data collection (i.e., more data are necessary than were collected), or expending the wrong effort (i.e., the wrong data were collected).

Proper planning and execution of a data collection effort can prevent these problems. EPA has developed the Data Quality Objectives Process as a flexible planning tool that should be used to prepare for a data collection activity. The information compiled in this effort is then used to develop the QAPP (USEPA, 1994e).

5.2.1 The Data Quality Objectives Process

The Data Quality Objectives (DQO) process takes into consideration the factors that will depend on the data (most importantly, the decision(s) to be made) or that will influence the type and amount of data to be collected (e.g., the problem being addressed, existing information, information needed before a decision can be made, and available resources). From these factors the qualitative and quantitative data needs are determined. The purpose of the DQO process is to improve the effectiveness, efficiency, and defensibility of decisions made based on the data collected, and to do so in a resource-effective manner (USEPA, 1994e).

DQOs are qualitative and quantitative statements that clarify the study objective, define the most appropriate type of data to collect, and determine the most appropriate conditions under which to collect them. DQOs also specify the minimum quantity and quality of data needed by a decision maker to make any decisions that will be based on the results of the project. By using the DQO process, investigators can ensure that the type, quantity, and quality of data collected and used in decision making will be appropriate for the intended use. Similarly, efforts will not be expended to collect information that does not support defensible decisions. The products of the DQO process are criteria for data quality and a data collection design that ensures that data will meet the criteria.

The DQO process consists of seven steps, described below. The process is iterative. As one step of the process is completed, its outputs might lead to reconsideration of previous steps. The previous steps should then be repeated. Optimization of the design (the last step) should begin only when all previous steps have been completed. When the optimization step is reached, as at any time during the DQO process, it might be necessary to reconsider earlier steps (i.e., to reiterate part or all of the process) to determine the optimum design.

A brief description of each step of the DQO process and a list of activities that are part of each step follow. For a detailed discussion of the DQO development process, refer to EPA's *Guidance for the Data Quality Objectives Process* (USEPA, 1994e), from which the following information was taken. This reference contains a case study example of the DQO process. A computer program, *Data Quality Objectives Decision Error Feasibility Trials* (EPA QA/G-4D), is also available to help the planning process by generating cost information about several simple sampling designs based on the DQO constraints before the sampling and analysis design team begins developing a final sampling design in the last step of the DQO process. (Contact EPA's Quality Assurance Management Staff, 202 260-9464).

(1) State the problem

In this first step the problem to be studied is described concisely. A review of prior studies and existing information is important during this step to gain a sufficient understanding of the problem in order to define it. The specific activities to be completed during this step (outputs) are:

- Identify members of the planning team.

- Identify the primary decision maker of the planning team and define each member's role and responsibilities during the DQO process.
- Develop a concise description of the problem.
- Specify the available resources and relevant deadlines for the study.

(2) Identify the decision

Identify what questions the study will attempt to resolve and what actions might be taken based on the study. This information is used to prepare a "decision statement" that will link the principal study question to one or more possible actions that should solve the problem. Possible options include take no action, take action, or modify an action. A decision statement might be phrased as follows: *Determine whether [or which] NPS impacts require taking [one of the alternative actions]*. For example, if the question to be addressed is "Are nutrients from agricultural runoff contributing to the growth of algal mats in the river?" and the alternative actions are "require vegetation buffers along streams" or "take no action," the decision statement is "Determine whether nutrients from agricultural runoff are contributing to algal growth and require regulation." The specific activities to be completed during this step are:

- Identify the principal study question.
- Define the alternative actions that could result from resolution of the principal study question.
- Combine the principal study question and the alternative actions into a decision statement.
- If applicable, organize multiple decisions to be made by priority.

(3) Identify the inputs to the decision

Identify the information that needs to be obtained and the measurements that need to be taken to resolve the decision statement. The specific activities to be completed during this step are:

- Identify the information that will be required to resolve the decision statement.
- Determine the sources for each item of information identified above.
- Identify the information that is needed to establish the threshold value that will be the basis of choosing among alternative actions.
- Confirm that appropriate measurement methods exist to provide the necessary data.

(4) Define the study boundaries

Specify the time periods and spatial area to which decisions will apply and determine when and where data should be collected. This information is used to define the population(s) of interest. The term *population* refers to the total collection or universe of objects from which samples will be drawn. The population could be the concentration of a pollutant in sediment, a water quality parameter, algae in the river, or bass in the lake. It is important to define the study boundaries to ensure that data collected are representative of the population being studied (since every member of a population cannot be sampled) and will be collected during the time period and from the place that will be targeted in the decision to be made. The specific activities to be completed during this step are:

- Specify the characteristics that define the population of interest.
- Identify the geographic area to which the decision statement applies (such as a county)

and any strata within that area that have homogeneous characteristics (e.g., recreational waters, dairy farms).

- Define the time frame to which the decision applies.
- Determine when to collect data.
- Define the scale of decision making, or the actual areas that will be affected by the decision (e.g., first-order streams, dairy farms with streams running through them, a county).
- Identify any practical constraints on data collection.

(5) Develop a decision rule

Define the statistical parameter of interest, specify the threshold at which action will be taken, and integrate the previous DQO outputs into a single statement that describes the logical basis for choosing among alternative actions. This statement is known as a *decision rule*. It is often phrased as an "If...then..." statement. For example, "If the mean concentration of contaminant X in the water downstream from farm Y exceeds 0.5 μg/L, then vegetation will be planted; otherwise, no action will be taken." The specific activities to be completed during this step are:

- Specify the statistical parameter that characterizes the population (the parameter of interest), such as the mean, median, or percentile.
- Specify the numerical value of the parameter of interest that would cause a decision maker to take action, i.e., the threshold value.
- Develop a decision rule in the form of an "if...then..." statement that incorporates the parameter of interest, the scale of decision making, the threshold level, and the actions that would be taken.

(6) Specify limits on decision errors

Define the decision maker's tolerable limits of making an incorrect decision (or decision error) due to incorrect information (i.e., measurement and sampling error) introduced during the study. These limits are used to establish performance goals for the data collection design. Base the limits on a consideration of the consequences of making an incorrect decision. The decision maker cannot know the true value of a population parameter because the population of interest almost always varies over time and space and it is usually impractical or impossible to measure every point (sampling design error). In addition, analytical methods and instruments are never absolutely perfect (measurement error). Thus, although it is impossible to eliminate these two errors, the combined total study error can be controlled to reduce the probability of making a decision error. The specific activities to be completed during this step are:

- Determine the possible range (likely upper and lower bounds) of the parameter of interest.
- Identify the decision errors and choose the null hypothesis. Decision errors for NPS pollution problems might take the general form of deciding there is no impact when there is [a false positive, or type I error], or deciding there is an impact when there is none [a false negative, or type II error].
- Specify the likely consequences of each decision error. Evaluate their potential severity in terms of ecological effects, human health, economic and social costs, political and legal ramifications, and other factors.

- Specify a range of possible parameter values where the consequences of decision errors are relatively minor (gray region). The boundaries of the gray region are the threshold level and the value of the parameter of interest where the consequences of making a false negative decision begin to be significant.

- Assign probability limits to point above and below the gray region that reflect the tolerable probability for the occurrence of decision errors.

(7) Optimize the design

Evaluate information from the previous steps and generate alternative data collection designs. The designs should specify in detail the monitoring that is required to meet the DQOs, including the types and quantity of samples to be collected; where, when, and under what conditions they should be collected; what variables will be measured; and the QA/QC procedures that will ensure that the DQOs are met. The QA/QC procedures are fully developed when the QAPP is written (see below). Choose the most resource-effective design that meets all of the DQOs. The specific activities to be completed during this step are:

- Review the DQO outputs and existing environmental data.

- Develop general data collection design alternatives.

- Formulate the mathematical expressions needed to solve the design problem for each data collection design alternative. This involves selecting a statistical test method (e.g., Student's *t* test), developing a statistical model that relates the measured value to the "true" value, and developing a cost function that relates the number of samples to the total cost of sampling and analysis.

- Select the optimal sample size that satisfies the DQOs for each data collection design alternative.

- Select the most resource-effective data collection design that satisfies all of the DQOs.

- Document the selected design's key features and the statistical assumptions of the selected design. It is particularly important that the statistical assumptions be documented to ensure that, if any changes in analytical methods or sampling procedures are introduced during the project, these assumptions are not violated.

The DQO process should be used during the planning stage of any study that requires data collection, and before the data are collected. EPA's policy is to use the DQO process to plan all data collection efforts that will require or result in a substantial commitment of resources. The DQO process is applicable to all studies, regardless of size; however, the depth and detail of the DQO development effort depends on the complexity of the study. In general, more complex studies benefit more from more detailed DQO development.

5.2.2 Data Quality Objectives and the QA/QC Program

The DQOs and the quality objectives for measurement data that will be specified in the QAPP are interdependent. The DQOs identify project objectives; evaluate the underlying hypotheses, experiments, and tests to be performed; and then establish guidelines for the data collection effort needed to obtain data of the quality necessary to achieve these objectives (Erickson et al., 1991; USEPA, 1994e). The QAPP presents the policies, organization, and objectives of the data collection effort and explains how particular QA and QC activities will be

implemented to achieve the DQOs of the project, as well as to determine what future research directions might be taken (Erickson et al., 1991; USEPA, 1994e). At the completion of data collection and analysis, the data are validated according to the provisions of the QAPP and a Data Quality Assessment (DQA), using statistical tools, is conducted to determine:

- Whether the data meet the assumptions under which the DQOs and the data collection design were developed.
- Whether the total error in the data is small enough to allow the decision maker to use the data to support the decision within the tolerable decision error rates expressed by the decision maker (USEPA, 1994e).

Thus, the entire process is designed to assist the decision maker by planning and obtaining environmental data of sufficient quantity and quality to satisfy the project objectives and allow decisions to be made (USEPA, 1994c, 1994e). The DQO process is the part of the quality system that provides the basis for linking the intended use of the data to the QA/QC requirements for data collection and analysis (USEPA, 1994e).

5.3 Elements of a Quality Assurance Project Plan

QAPPs must be prepared according to guidance provided in *EPA Requirements for Quality Assurance Project Plans for Environmental Data Objectives* (USEPA, 1994c). EPA requires that four types of elements be discussed in a Quality Assurance Project Plan (QAPP). These elements are listed in Table 5-2 and discussed briefly below. (For complete descriptions and requirements, be sure to see USEPA (1994c)). Additional information on the contents of a QAPP is contained in Drouse et al. (1986), Erickson et al. (1991), and Cross-Smiecinski and Stetzenback (1994). Drouse et al. (1986) and Erickson et al. (1991) are examples of EPA QAPPs prepared under previous guidance.

The elements described below should always be addressed in the QAPP, unless otherwise directed by the overseeing or sponsoring EPA organization(s). The types, quantity, and quality of environmental data collected for each project could be quite different. As noted in USEPA (1994c), "the content and level of detail in each QAPP will vary according to the nature of the work being performed and the intended use of the data." If an element is not applicable or required, then this should be stated in the QAPP. For some complex projects, it might be necessary to add special requirements to the QAPP. Again, the QAPP must be approved by the sponsoring EPA organization before work can begin, and it should be reviewed annually (for multiyear projects) and updated and reapproved as often as necessary during the project.

5.3.1 Group A: Project Management

These elements cover basic project management, including project history and objectives, roles and responsibilities of participants, and other factors to ensure that the project has a defined goal understood by all the participants and that all planning activities have been documented.

A1 Title and Approval Sheet

Provide the title of the plan; name of organization(s) implementing the project; and names, titles, and signatures of the appropriate approving officials and their approval dates.

A2 Table of Contents

List sections, figures, tables, references, and appendices. If document control format is required, see Cross-Smiecinski and Stetzenback (1994) and USEPA (1994e).

Table 5-2. Elements required in an EPA Quality Assurance Project Plan. (USEPA, 1994b)

	QAPP Element
A1	Title and Approval Sheet
A2	Table of Contents
A3	Distribution List
A4	Project/Task Organization
A5	Problem Definition/Background
A6	Project/Task Description
A7	Quality Objectives and Criteria for Measurement Data
A8	Project Narrative (ORD only)
A9	Special Training Requirements/Certification
B1	Sampling Process Design
B2	Sampling Methods Requirements
B3	Sampling Handling and Custody Requirements
B4	Analytical Methods Requirements
B5	Quality Control Requirements
B6	Instrument/Equipment Testing, Inspection, Maintenance Requirements
B7	Instrument Calibration and Frequency
B8	Inspection/Acceptance Requirements for Supplies and Consumables
B9	Data Acquisition Requirements (Nondirect Measurements)
B10	Data Management
C1	Assessments and Response Action
C2	Reports to Management
D1	Data Review, Validation, and Verification Requirements
D2	Validation and Verification Methods
D3	Reconciliation and User Requirements

A3 Distribution List

List all individuals and organizations who will receive copies of the approved QAPP and subsequent revisions.

A4 Project/Task Organization

Discuss the specific roles and responsibilities of all individuals or organizations participating in the project.

A flow chart or box diagram is useful for depicting project organization and responsibilities (Figure 5-1). Using the diagram, explain the rationale for the organization (e.g., to maximize 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

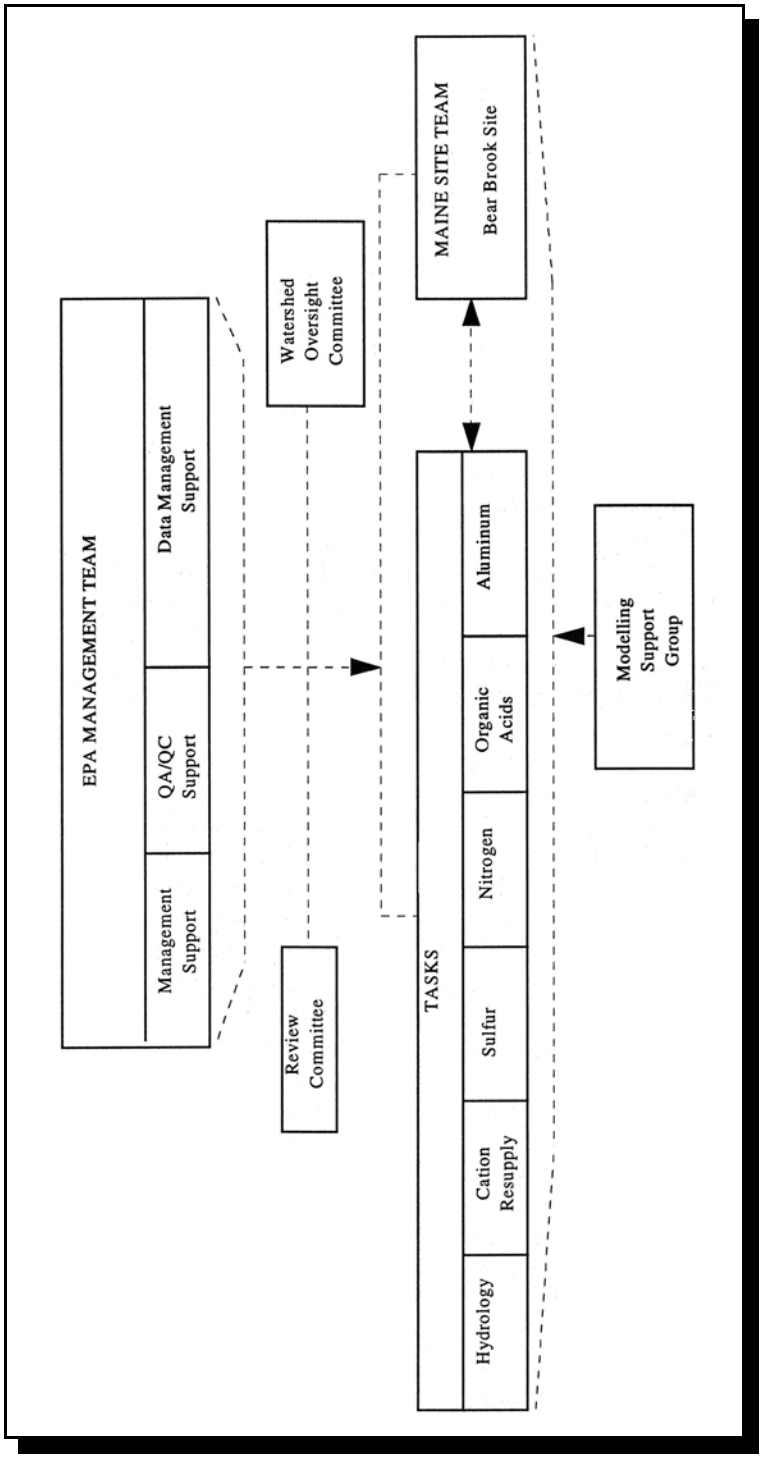


Figure 5-1. Sample organization chart for a quality assurance project plan. (Erickson et al., 1991)

affiliations with participating organizations. The program manager, managers or coordinators of any specific tasks, directors of technical tasks to be conducted, and any organizations or agencies that will be involved in the project should be identified. Also identify the specific roles and responsibilities (such as field sampling, laboratory analyses, and report preparation) that will be conducted by each person and organization involved in the project.

A5 Problem Definition/Background

State the problem to be solved or the decision to be made and describe its history for this particular project.

A6 Project/Task Description

Describe the work to be performed (measurements to be made, applicable quality standards, any special personnel or equipment requirements, assessment tools needed, records and reports needed) and the schedule for its implementation.

A7 Quality Objectives and Criteria for Measurement Data

The DQO process will provide this information, or state the project quality objectives and measurement performance criteria that are necessary to support the management decision(s) to be made based on the result(s) of the project. State quality objectives in terms of project requirements, preferably in quantitative terms, rather than in terms of analytical or sampling method capabilities. Then, with the quality objectives stated, select the appropriate methods to achieve the requirements (Cross-Smiecinski and Stetzenback, 1994). The quality of data should be expressed in terms of precision, accuracy, comparability, representativeness, and completeness (defined below). A table of quality objectives, like that in Figure 5-2, is helpful.

Definitions of data quality terms

Precision (reproducibility)

(a) Precision is a measure of mutual agreement among individual measurements of the same property. The coefficient of variation (CV), also known as the percent relative standard deviation (RSD), is used to express precision (Erickson et al., 1991).

$$CV = \left(\frac{s}{\bar{x}}\right) 100$$

where

s = sample standard deviation and
x = arithmetic mean.

(b) Precision is an expression of mutual agreement of multiple measurement values of the same property conducted under prescribed similar conditions. It is evaluated by recording and comparing multiple measurements of the same parameter on the same exact sample under the same conditions. Relative percent difference (RPD) is a measure of precision and is calculated with the following formula (Cross-Smiecinski and Stetzenback, 1994):

$$RPD = \frac{2\,(x_1 - x_2)}{x_1 + x_2}\,(100)$$

where

x_1 = analyte concentration of first duplicate and
x_2 = analyte concentration of second duplicate.

Parameter	Units	Expected Range	Accuracy	Precision	Completeness
Particulate NO_3/SO_4					
SSI[a]	$\mu g/m^3$	10-1000	10%	20%	70%
47 mm TF/PC[b]	$\mu g/m^3$	10-1000	10%	20%	90%
NO_3					
47 mm TF/PC	$\mu g/m^3$	1 to 25	20%	20%	90%
SO_2					
47 mm TF/PC	$\mu g/m^3$	1 to 75	20%	20%	90%
Meteorological					
Wind speed	m/s	0 to 75	2%	2%	90%
Wind direction	deg	0 to 360	2%	2%	90%
Dew point	0°C	-30 to 70	2%	5%	90%
Solar radiation	watts/m^2	-	-	-	-
Ambient temperature	°C	-20 to 50	1°C	2°C	90%

[a] PM_{10} Size Selective Inlet High Volume Sampler
[b] Teflon/Polycarbonate Filter

Figure 5-2. Sample quality assurance objectives. (Erickson et al., 1991).

Accuracy (bias)

(a) Accuracy is the degree of agreement of a measurement (or an average of measurements), *X*, with an accepted reference or true value, *T*. Accuracy is expressed as the percent difference from the true value {100 [(X-T)/T]} unless spiking materials are used and percent recovery is calculated (Erickson et al., 1991).

(b) Accuracy is the correctness of the value obtained from analysis of a sample. It is determined by analyzing a sample and its corresponding matrix spike. Accuracy can be expressed as percent recovery and calculated using the following formula (Air National Guard, 1993):

$$\%R = \frac{A - B}{C}(100)$$

where

A = spiked sample result;
B = sample result; and
C = spike added.

Comparability

(a) Comparability is defined as the confidence with which one data set can be compared to another (Erickson et al., 1991).

(b) Comparability is the quality that makes data obtained from one study comparable to data from other studies. Consistent sampling methodology, handling, and analyses are necessary to ensure comparability. Also, assurance that equipment has been calibrated properly and analytical solutions prepared identically is necessary to attain data comparability (Air National Guard, 1993).

Representativeness

(a) Representativeness can be defined both qualitatively and quantitatively; it depends on the experimental design and choice of sampling methods. The desired degree of representativeness is important in planning for the collection of samples and the subsequent uses of the data. A relevant sampling design issue, for example, is to determine how a sample will be collected to ensure it is representative of the desired characteristic (Erickson et al., 1991).

(b) Representativeness is a measure of how representative the data obtained for each parameter is compared with the value the same parameter has within the population being measured. Since the total population cannot be measured, sampling must be designed to ensure that the samples are representative of the population being sampled (Air National Guard, 1993).

Completeness

(a) Completeness is defined as the amount of valid data obtained from a measurement system compared to the amount that was expected to be obtained under anticipated sampling/analytical conditions (Erickson et al., 1991).

(b) Completeness is the amount of valid data obtained from the measurement system (field and laboratory) versus the amount of data expected from the system. An assessment of the completeness of data is performed at the end of each sampling event, and if any omissions are apparent an attempt is made to resample the parameter in question, if feasible. Data completeness should also be assessed prior to the preparation of data reports that check the correctness of all data. An example of a formula used for this purpose is

$$\% C = 100\left[\frac{V}{n}\right]$$

where

%C = percent complete;
V = number of measurements judged valid; and
n = total number of measurements necessary to achieve a specified level of confidence in decision making (Cross-Smiecinski and Stetzenback, 1994).

A8 Project Narrative

This is a narrative description of work to be performed that will demonstrate to technical or QA reviewers that the project or task will achieve its quality objectives. See USEPA (1994e) for complete details of what should be included in a project narrative.

A9 Special Training Requirements/Certification

If personnel will require any specialized training or certification to successfully complete the project, discuss how this training will be obtained and documented.

A10 Documentation and Records

Itemize all of the information and records (e.g., raw data, field logs, instrument printouts, results of calibration and QC checks, analytical laboratory case narratives) that must be included in a data report package, and describe the desired report format and final disposition of records and documents.

5.3.2 Group B: Measurements and Acquisition

The Project/Task Description element (A6) contains a summary of this information, which should be provided in detail in this section. Methods that have been well documented and are available to all participants can merely be cited; for those not well documented, detailed copies of the methods and/or Standard Operating Procedures (SOPs) must be provided in the QAPP.

B1 Sampling Process Design (Experimental Design)

Explain the experimental design or data collection design, including types and numbers of samples required, sampling locations and frequencies, sampling screening criteria (if applicable), sample matrices, measurement parameters of interest, and the rationale for the design. As with all information contained in a QAPP, recording information such as the reasoning behind decisions will make the data more defensible in the future. Statistics can play an important part in determining the sampling strategy. Therefore, record all statistical procedures that will be used to determine the sampling strategy. Two basic sampling decisions that must be made are the types and numbers of quality control samples to be collected (Keith, 1988). See USEPA (1994e) for additional details on what to include in this element of the QAPP.

B2 Sampling Methods Requirements

Identify and describe all procedures for collecting samples for each sampling method, as well as what should be done when a sampling or measuring failure occurs and who is responsible for taking corrective action. Other aspects pertinent to sampling, such as record keeping, sample storage, and transport to laboratories, should also be described in this section (Cross-Smiecinski and Stetzenback, 1994).

B3 Sample Handling and Custody Requirements

Describe all aspects of sample handling and custody. Sample custody is a documentation of where and with whom samples are at all times from the moment they are collected in the field to when they are analyzed in the laboratory. A sample is considered to be under custody if (1) it is in your actual possession; (2) it is in your view, after being in your physical possession; (3) it was in your physical possession and then you locked it up to prevent tampering; or (4) it is in a designated and identified secure area (Air National Guard, 1993). Special tracking procedures called “chain-of-custody” procedures are used whenever samples are collected for use in an enforcement action or when demonstrating compliance with a regulatory requirement (e.g., NPDES). Chain-of-custody forms should be printed on multipart carbonless paper for tracking custody and should have, at a minimum, space for recording date, time, name of person accepting samples, sample numbers, and remarks (Figure 5-3). Copies of the form must be completed in the field, and signed by the fieldteam when they transfer custody of the samples to the shipper. Upon receipt in the laboratory, the laboratory signs the remaining copies, indicating they have accepted custody of the samples. Each time the form is signed, the person signing the form retains the bottom copy and passes the remaining copies along with the samples. The

CHAIN-OF-CUSTODY RECORD

PROJ NO		PROJECT NAME	No. of Containers	TAG NUMBERS	PARAMETERS				REMARKS
SAMPLERS (*Signature*)									
DATE	TIME	DESCRIPTION							

Relinquished by: (*Signature*)	Date	Time	Received by: (*Sign.*)	Relinquished by (*Sign.*)	Date	Time	Received by (*Sign.*)
Print Name:			Print Name	Print Name			Print Name

Relinquished by: (*Signature*)	Date	Time	Received by: (*Sign.*)	Date	Time	Remarks
Print Name:			Print Name			

Figure 5-3. Sample custody chart. (After Cross-Smiecinski and Stetzenback, 1994)

laboratory should return at least one copy of the completed chain-of-custody record to the client, or proceed as directed in the QAPP.

B4 Analytical Methods Requirements

Describe the analytical methods and equipment required for both field and laboratory activities, waste disposal requirements, and specific performance requirements, as well as what should be done when a failure in the analytical system occurs and who is responsible for taking the corrective action. Also include information on any supporting methods or documents used to collect field or laboratory data. For instance, if identifications of benthic invertebrates are made, include information on the source(s) used to verify identifications; if the amount of riparian vegetation cover is estimated, describe the method used to arrive at the estimate.

Analytical methods: Describe the methods that will be used for the project. If the methods to be used are published (e.g., by the U.S. Geological Survey, EPA, or ASTM) it is sufficient to indicate what methods will be used and where descriptions of them can be found. If the best methods to be used cannot be completely ascertained until some samples have been analyzed, indicate the order of preference for use of the methods. Any modifications to published or standard methods or variations of them must be documented, and the variations must be verified as providing data of acceptable quality.

Method validation: Method validation accounts for and documents, at a minimum, the following characteristics: known and possible interferences; method precision; method accuracy, bias, and recovery; method detection level, and method comparability to superseded methods, if any (Pritt and Raese, 1992). All methods chosen for use in the project must be validated.

Generally, laboratories with their own QA/QC procedures will be used for sample analyses. The methods to be used in the laboratory must be acceptable to project managers. All potential laboratory facilities to be used in the project should be extensively evaluated before their selection and throughout their participation in the project.

B5 Quality Control Requirements

Identify the QC procedures (types, frequency, and control limits of QC checks) needed for each sampling, analysis, or measurement technique. (They might have to be modified to suit each project.) Also state what corrective action is required when control limits are exceeded. Data collected as part of field sampling and laboratory measurements must be verified as accurate. Thus, some samples are taken or measurements made to check for accuracy rather than to collect additional data. Specify what means will be used to check the accuracy of samples and measurements. Field blanks, duplicate samples, replicate samples, spiked samples, and spiked blanks are commonly used methods. Describe precisely how these control samples will be prepared for analysis.

Standard reference materials (SRMs) should be used periodically in any measurement system to monitor for changes to the system that might go unnoticed. SRMs should be used when a measurement change is noted to verify that the change is not due to a change in the measurement system. The optimum frequency of use of SRMs and also of replicates of actual test samples depends on the integrity of the measurement system and the magnitude of the errors involved when the system ceases to give predictable results. All measurements from last-known-in-control sample to first-known-out-of-control sample are suspect, so the length of the period between these two samples must be calibrated to be appropriate to the measurements being made (Taylor, 1993).

B6 Instrument/Equipment Testing, Inspection, and Maintenance Requirements

This section should include descriptions of the types of preventive maintenance for equipment that will be used to ensure that research schedules are adhered to and project objectives are completed on schedule. The section should include the following: a schedule of preventive maintenance, an inventory of critical spare parts and supplies, maintenance contract information, location of important manuals and instructions, record keeping requirements, and training of instrument and equipment operators (Cross-Smiecinski and Stetzenback, 1994). Some aspects of training can be considered a part of preventive maintenance. Describe in this section general safety precautions that will be part of project operations. Examples include materials handling, transportation of chemicals, hazardous waste disposal procedures, emergency procedures, standard safety operations,

chemical hygiene, hazard communication, hazardous waste management, waste disposal, location of safety equipment, tour of facilities, and annual classes in cardiopulmonary resuscitation and standard first aid (Pritt and Raese, 1992).

B7 Instrument Calibration and Frequency

Describe the procedures used for equipment calibration, the frequency of calibration of each piece of equipment, and the results of calibration procedures. Record any problems encountered and corrective actions taken. This section should identify each tool, gauge, instrument, or other sampling, measuring, and test equipment used for data collection activities for which quality must be controlled and which must be calibrated to maintain performance within specified limits.

B8 Inspection/Acceptance Requirements for Supplies and Consumables

Supplies and consumables to be used in the project must be inspected and accepted, according to specified criteria, for use in the project. Identify who will perform the inspections and how they will be conducted.

B9 Data Acquisition Requirements (Nondirect Measurement)

Data obtained from noninstrument sources such as computer databases, spreadsheets, and programs and literature files need to be identified and acceptance criteria established for the use of the data. Also discuss any limitations resulting from uncertainty in the quality of the data and the impact of adding more error to the results.

B10 Data Management

This section should describe all aspects of data management, from their generation in the field or laboratory to final use or storage. Discuss the control mechanisms (and provide examples of forms or checklists) for detecting and correcting errors and for preventing loss of data during data reduction. This discussion should also include all data handling equipment and procedures that will be used to process, compile, and analyze the data (hardware and software).

5.3.3 Group C: Assessment/Oversight

The purpose of these elements is to ensure that the QAPP will be implemented as prescribed; they describe the activities for assessing the effectiveness of the implementation of the QAPP and its associated QA/QC program.

C1 Assessments and Response Actions

Assessments can include a variety of activities, such as surveillence, peer review, management systems review, technical systems audit, or performance evaluation. Audits are assessments of the extent to which QA procedures and QC activities are being adhered to. They may be performed by an internal (i.e., within the project structure) but independent audit team or by an external audit team. Audits may be performed before, during, and/or after the project is performed. Audit frequency, intensity, and type should be determined, and the audit(s) should be scheduled as part of the overall program QA effort. This section of the QAPP should describe the audits to be performed and the process and procedures for responding to problems raised during audits (Cross-Smiecinski and Stetzenback, 1994).

This section should also describe actions to be taken if and when unexpected problems arise during the course of the study. Problems that can be foreseen, such as running low on commonly used laboratory supplies, should be addressed as SOPs. Many problems, however, are encountered so infrequently or are unpredictable enough that

SOPs will not be prepared for them. Special or emergency procedures address these types of problems. It is difficult to address unanticipated problems before they arise, but the QAPP should specify who is responsible for handling problems that arise from different aspects of the project (e.g., field sampling, laboratory analysis, audits). It is helpful to categorize problems based on their impact on the project (e.g., critical, important, noncritical, unimportant) and to specify the type of corrective action necessary based on the problem's category. A critical problem, for instance, would be one that would affect obtaining data of the necessary quality or quantity. If a critical problem arises, a critical-problem response by project staff would be required. This should be specified in the QAPP.

C2 Reports to Management

This section specifies the type and frequency of reports to be prepared and submitted to project management, as well as the chain of responsibility for ensuring that reports are prepared and submitted. The preparer of the reports and recipients of each report should be identified. Any required report contents and format should also be specified.

5.3.4 Group D: Data Validation and Usability

After the data collection has been completed, the data must be examined to determine whether they conform to the specified criteria and will satisfy project objectives.

D1 Data Review, Validation, and Verification Requirements

The requirements used to review and accept, reject, or qualify data should be identified, including any project-specific calculations or algorithms.

D2 Validation and Verification Methods

This section should describe each of the elements defined below in enough detail to support use of the data for their intended purpose and for comparability to past, present, and future studies (Cross-Smiecinski and Stetzenback, 1994). If computer software is used in data manipulations, record which software is to be used. Software that performs complex manipulations might have to be verified before its use to ensure that it functions properly (Cross-Smiecinski and Stetzenback, 1994).

Data reduction: The transformation of raw data into a more useful form, calculations.

Data verification: A routine activity conducted by technical, laboratory, and clerical personnel on small sets of the data to determine whether data have been accurately quantified, recorded, and transcribed; whether data have been collected and analyzed in accordance with prescribed, approved procedures; whether the data appear suitably complete; and whether the data appear to be reasonable and consistent, based on prior knowledge of the research.

For example, it is a good practice to enter data into the database twice and scan them for outlying values. This helps to detect and eliminate transcription errors. Range checks, internal consistency checks, and quality assurance evaluations should also be included for data certification (Drouse et al., 1986).

Data validation: The process by which a sample, measurement method, or datum is deemed useful for a specified purpose; an independent, timely review of a body of verified data against a predetermined set of qualitative and quantitative criteria to evaluate their adequacy for their intended use.

Data reporting: Specify any special forms or formats (e.g., tables and figures) that are to be used, as well as who is responsible for data reporting, due dates, etc.

D3 Reconciliation with User Requirements

The precision, accuracy, completeness, representativenesss, and comparability of data must be assessed using appropriate techniques. This section should give details of the formulas, statistical techniques, and procedures that will be used to assess the data. The methods used to assess the data must be in agreement with the DQOs. The terms *precision, accuracy, completeness, representativeness*, and *comparability* are defined on page 5-11, and some sample data assessment formulas are given.

The following sections provide more specific information for preparing QAPPs with respect to field and laboratory operations, and data and reporting requirements.

5.4 Field Operations

Field operations are an important activity in an NPS monitoring program. Field operations involve the organization and design of the field operation, selection of sampling sites, selection of sampling equipment, sample collection, sample handling and transport, and safety and training issues. For the purposes of QA and QC, the process of conducting field operations should be broken down into as many separate steps as are necessary to ensure complete consideration of all of the elements and processes that are a part of field activities. Field operations described in this section have been broken down into the phases mentioned above, but individual monitoring programs might require the use of more or fewer phases. For example, if the sample collection phase is very complex or if it is anticipated that sample collection will often be done under inclement weather conditions when field personnel might experience discomfort and feel rushed, it is advisable to break sample collection into separate preparation, sampling, and termination phases and discuss QA and QC for each of the phases separately. This will ensure that no details are omitted. Table 5-3 summarizes many important items that should be considered in the field operations portion of a QA/QC program.

5.4.1 Field Design

Adherence to the procedures specified in the QAPP for field operations and documentation of their use for all aspects of field operations are extremely important if the data obtained from the project are to be useful for decision making, supportable if questioned, and comparable for use by future researchers (Knapton and Nimick, 1991). Data sheets prepared beforehand, with quality reminders included where appropriate, will help ensure that all data are collected and QA/QC procedures are followed during all field activities.

General information that should be included in the documentation of the design for field operations includes the scale of the operations (laboratory, plot, hillslope, watershed); size of plots/data collection sites; designation of control sites; basin characteristics; soil and vegetation types; maps with the location of plots/data collection sites within the basin/catchment; weather conditions under which sampling is conducted; equipment and methods used; problems that might be encountered during sampling; dates of commencement and suspension of data collection; temporal gaps in data collection; frequency of data collection; intensity of data collection; and sources of any outside information (e.g., soil types, vegetation identifications) (Erickson et al., 1991). Some of these aspects are discussed in greater detail in the following sections.

Table 5-3. Checklist of items that should be considered in the field operations section of a QA/QC program.

Field Operations			
Element	**Specifics**	**Check-off**	**Responsibility**
Organization	Field organization chart created . . . . . . Staff duties and responsibilities defined . . . . . . Communication lines within and with other units established . . . . . . Project documents made available to all staff . . . . . . Staff qualifications established	_____ _____ _____ _____ _____	
Field Logistics	Sampling sites investigated and selected . . . . . . Means of access to sampling sites determined . . . . . . Sample transport and shipping procedures specified . . . . Field sample handling areas selected . . . . . . Chain-of-custody for samples established . . . . . . Field equipment selected and supplied . . . . . . Procedures for decontamination of sampling equipment established	_____ _____ _____ _____ _____ _____ _____	
Monitoring Equipment	Equipment installation procedures specified . . . . . . Equipment maintenance and control schedules established . . . . . . Equipment maintenance manual updated and distributed Trouble shooting and corrective action manual updated and distributed	_____ _____ _____	
Quality Control Samples	Type(s) of control samples (blanks, duplicates, spikes, analytical standards, reference materials) to be used have been determined . . . . . . Frequency of control sample use has been determined	_____ _____	
Field Audits	QA field auditor designated . . . . . . Aspects of field operations that will undergo quality assessments as part of field audits have been determined Acceptance criteria for compliance with SOPs and the QAP set for field events and activities . . . . . . Field audit forms, with investigations to be conducted and data to be collected, prepared . . . . . . Person(s) to review field audit records designated	_____ _____ _____ _____ _____	

Table 5-3. (continued)

Health and Safety	Field personnel properly trained	_____	
	Proper field gear and clothing issued to field personnel	_____	
SOPs	Sample management	_____	
	Sample collection procedures	_____	
	Reagent preparation	_____	
	Decontamination	_____	
	Equipment calibration and maintenance	_____	
	Corrective action	_____	
	Waste disposal	_____	
	Health and safety	_____	
	Field measurements	_____	
	Reagent/standard preparation	_____	
	Equipment calibration and maintenance	_____	
	Data reduction and validation	_____	
	Reporting	_____	
	Corrective action	_____	
	Waste disposal	_____	
	Health and safety	_____	
	Records management	_____	
	Project-specific records	_____	
	Field operations records	_____	

Source: USEPA, 1990b.

5.4.2 Sampling Site Selection

The selection of sampling sites is important to the validity of the results. Sites must be selected to provide data to meet the goals/objectives of the project. The QAPP should provide detailed information on sampling site locations (e.g., latitude and longitude); characteristics that might be important to data interpretation (e.g., percent riparian cover, stream order); and the rationale for selecting the sites used (Knapton and Nimick, 1991). Sites from other studies can be convenient to use due to their familiarity and the availability of historical data, but such sites should be scrutinized.

carefully to be certain that data obtained from them will serve the objectives of the project. If during the course of the project it is found that one or more sampling sites are not providing quality data, alternative sites might be selected and the project schedule adjusted accordingly. The adequacy of the sampling locations and the sampling program should be reviewed periodically by project managers, as determined by data needs (Knapton and Nimick, 1991).

Sampling sites should be visited before sampling begins. It is important to verify that the sites are accessible and are suitable for collection of the data needed. Consideration should be given to accessibility in wet or inclement weather if samples will be taken during such conditions. The sites should be visited, if possible, in the type(s) of weather during which sampling will occur. Plastic- laminated pictures of each sampling site

with an arrow pointing to each monitoring location can assist field personnel in finding the sites during inclement weather when the sites might appear different.

If permission to access a site is needed (for instance, if one or more sites are on or require passage through private property), such permission must be obtained before sampling begins. The person(s) granting the permission should be fully informed about the number of persons who will be visiting during each sampling event, frequency of sampling, equipment that will have to be transported to the sampling site(s), any hazardous or dangerous materials that will be used during sampling, and any other details that might affect the decision of the person(s) to grant access permission. A lack of full disclosure of information to gain access permission creates a risk of the permission's being revoked at some point during the project.

5.4.3 Sampling Equipment

Equipment for field operations includes field-resident equipment such as automatic samplers and stage-level recorders and nonresident sampling equipment such as flow, pH, and conductivity meters; equipment needed to gain access to sampling sites such as boats; and equipment for field personnel health and safety, such as waders, gloves, and life vests. The condition and manner of use of the field equipment determines the reliability of the collected data and the success of each sampling event. Therefore, operation and maintenance of the equipment are important elements of field QA and QC. All measurement equipment must be routinely checked and calibrated to verify that it is operating properly and generating reliable results (Spooner, 1994), and all access and health and safety equipment should be routinely checked to be certain that it will function properly under all expected field conditions.

A manual with complete descriptions of all field equipment to be used should be available to all field personnel. The manual should include such information as model numbers for all measurement equipment, operating instructions, routine repair and adjustment instructions, decontamination techniques, sampling preparation instructions (e.g., washing with deionized water), and use limitations (e.g., operating temperature range). If any samples are to be analyzed in the field, the techniques to be used should be thoroughly described in the manual.

5.4.4 Sample Collection

The process of sample collection should be described with the same amount of detail as the equipment descriptions. A thorough description of the sample collection process includes when the sampling is to be done (e.g., time of day, month, or year; before and/or after storms); the frequency with which each type of sample will be collected; the location at which samples are to be taken (i.e., depth, distance from shore, etc.); the time between samples (if sampling is done repetitively during a single sampling site visit); and how samples are to be labeled. Each field person must be thoroughly familiar with the sampling techniques (and equipment) prior to the first sampling event. Holding practice sampling events prior to the commencement of actual sampling is an excellent way to prepare all field personnel and will help to identify potential problems with the sampling sites (access, difficulty under different weather conditions), sampling equipment, and sampling techniques.

Quality control activities for field operations must ensure that all field operations are conducted so that sampling is done in a consistent manner and that all generated information is traceable and of known and comparable quality. Each field activity should be standardized. Standard operating procedures (SOPs) for field sampling have been

developed and might be required depending on the agency for which the sampling is being conducted. Elements of the field operations section of a QAPP should include clear statements of the regulatory requirements applicable to the project (Spooner, 1994). Any SOPs that are part of regulatory requirements should be followed precisely. The pictures taken of each sampling site to aid in locating the sampling sites also help ensure consistency of field monitoring across time and personnel by ensuring that the same spot is used at each sampling event (Spooner, 1994).

Depending on the DQOs and data requirements of the program (type of data and frequency of collection), additional quality control samples might be needed to monitor the performance of various field (as well as laboratory) operations including sampling, sample handling, transportation, and storage.

As the samples are collected, they must be labeled and packaged for transport to a laboratory for analysis (or other facility for nonchemical analyses). Computer-generated sample bottle labels prepared before the sampling event and securely attached to each bottle help minimize mistakes. Sampling location and preservation, filtration, and laboratory procedures to be used for each sample should be recorded on each label (Spooner, 1994). Be sure these labels are printed with waterproof ink on waterproof paper, and use a No. 2 pencil or waterproof/solvent-resistant marker to record information.

5.4.5 Sample Handling and Transport

Once samples have been collected, they must be analyzed, usually in a laboratory. Handling and transport of sampling containers and custody of sample suites is also a part of field operations. Sample transport, handling, and preservation must be performed according to well-defined procedures. The various persons involved in sample handling and transport should follow SOPs for this phase of the project. This will help ensure that samples are handled properly, comply with holding time and preservation requirements, and are not subject to potential spoilage, cross-contamination, or misidentification.

The chain of custody and communication between the field operations and other units such as the analytical laboratory also need to be established so that the status of the samples is always known and can be checked by project personnel at any time. The chain of custody states who the person(s) responsible for the samples are at all times. It is important that chain of custody be established and adhered to so that if any problem with the samples occurs, such as loss, the occurrence can be traced and possibly rectified, or it can be determined how serious the problem is and what corrective action needs to be taken. Field data custody sheets are essential for this effort (Cross-Smiecinski and Stetzenback, 1994; Spooner, 1994). Chain-of-custody seals must be applied to sample containers and shipping containers.

5.4.6 Safety and Training

When dealing with NPS monitoring, sampling activities often occur during difficult weather and field conditions. It is necessary to assess these difficulties and establish a program to ensure the safety of the sampling personnel. The following types of safety issues, at a minimum, should be considered and included in training and preparation activities for sampling: exposure, flood waters, debris in rivers and streams, nighttime collecting, criminal activity, and first aid for minor injuries. The trade-off between the need for data quality and the safety of personnel is a factor that project staff should consider collectively.

Finally, the QAPP for the field operations should include provisions for dealing with any foreseeable problems such as droughts, floods, frozen water, missing samples, replacement personnel during sickness or vacation, lost samples, broken sample containers, need for equipment spare parts, and other concerns (Spooner, 1994).

5.5 Laboratory Operations

Laboratory operations should be conducted with the same attention to detail as field operations. Often, an independent laboratory conducts sample analyses, so QA and QC for the laboratory are not under the direct control of project personnel. However, it is important that project personnel are certain that the laboratory chosen to do analyses follows acceptable QA/QC procedures so that the data produced meet the DQOs established for the project. Laboratories should be selected based on quality assurance criteria established early in the project. The Quality Assurance Officer for the project should be certain that these criteria are used for selecting a laboratory to perform any necessary analyses for the project and that any laboratories selected meet all criteria. Laboratories can be evaluated through the following measures (Air National Guard, 1993):

- Performing proficiency testing through analysis of samples similar to those which will be collected during the project.
- Performing inspections and audits.
- Reviewing laboratory QA/QC plans.
- One or more of these measures should be used by the project manager, and the laboratories should be visited before entering into a contract for sample analyses.

5.5.1 General Laboratory QA and QC

Numerous references are available on laboratory QA/QC procedures, and one or more should be consulted to gain an understanding of laboratory QA and QC requirements if project personnel are not familiar with them already. The details of a laboratory's QA/QC procedures must be included in the QAPP for the NPS monitoring project. Some elements to look for in a laboratory QA/QC plan include (Cross-Smiecinski and Stetzenback, 1994):

- How samples are received
- Proper documentation of their receipt
- Sample handling
- Sample analysis
- QC requirements (procedures and frequencies of QC checks, criteria for reference materials, types of QC samples analyzed and frequencies)
- Waste disposal
- Cleanliness and contamination
- Staff training and safety
- Data entry and reporting
- Confidentiality

This section provides some information on laboratory QA/QC procedures to which managers of monitoring programs should pay particular attention when deciding to use a particular laboratory for sample analysis (Table 5-4). More detailed references on laboratory QA and QC should be consulted for further information.

Table 5-4. Checklist of items that should be considered in the laboratory operations section of a QA/QC program.

Laboratory Operations			
Element	**Specifics**	**Check-off**	**Responsibility**
Sample Management	Sample receipt . . . . . . Sample storage . . . . . . Sample handling . . . . . . Sample scheduling	_____ _____ _____ _____	
Equipment	Equipment calibration and maintenance	_____	
SOPs	Sample management . . . . . . Analytical methods . . . . . . Sample preparation and analysis procedures . . . . Reagent/standard preparation . . . . . . Raw data requirements . . . . . . Data reduction and validation . . . . . . Precision, accuracy, and method detection/reporting limits . . . . . . Reporting . . . . . . Corrective actions	_____ _____ _____ _____ _____ _____ _____ _____ _____	
Records Management	Project-specific records . . . . . . Laboratory operations records	_____ _____	
QC Procedures	Control samples . . . . . . Method blanks . . . . . . Matrix spikes . . . . . . Matrix duplication/matrix spike duplicates	_____ _____ _____ _____	
Audits	Laboratory audits schedule	_____	
Health and Safety	Fire and emergency equipment . . . . . . Fire and emergency equipment inspection . . . . . . Health equipment (masks, gloves, ...) . . . . . . Waste disposal	_____ _____ _____ _____	

Source: USEPA, 1990b.

5.5.2 Instrumentation and Materials for Laboratory Operations

The laboratory chosen to do chemical analyses should have all equipment necessary to perform the analyses required, including organic analysis, inorganic analysis, and assessments of precision and accuracy. If any specialized analyses are required (e.g., microbiology, histopathology, toxicology), be certain that the laboratory has the appropriate equipment and that laboratory staff are adequately trained to perform the desired analyses. As noted in the elements of the QAPP, periodic calibration checks that are conducted to ensure that measurement systems (instruments, devices, techniques) are operating properly should be described in the QAPP, including procedures and frequency (Cross-Smiecinski and Stetzenback, 1994).

5.5.3 Analytical Methods

The laboratory chosen for sample analysis should use analytical methods approved by the agency for which the sampling is being conducted or by project personnel, as appropriate. Standard methods include those published by the U.S. Geological Survey, the U.S. Environmental Protection Agency, and the American Society for Testing and Materials, or those published in *Standard Methods for the Analysis of Water and Wastes* (Clesceri et al., 1989). If any methods to be used are not published, they should first be validated and verified as acceptable for the project. Each approved and published method should be accompanied by an SOP that is followed rigorously by the laboratory (Pritt and Raese, 1992).

5.5.4 Method Validation

The laboratory chosen for sample analysis should have well-developed procedures for method validation. Method validation should account for and document the following (at a minimum): Known and possible interferences; method precision; method accuracy, bias, and recovery; method detection level; and method comparability to superseded methods, if applicable (Pritt and Raese, 1992).

5.5.5 Training and Safety

An analytical laboratory should be able to ensure its customers that its personnel are adequately trained to perform the necessary analyses. Individual laboratory staff should be independently certified for each of the analyses they will be allowed to perform in the laboratory. Selection of a laboratory for sample analysis should be based on queries about how often training is conducted, whether employees are limited to using equipment for which they have been adequately trained, whether the training program is independently certified, who conducts the training, how the staff's competence with individual instruments is measured, and other factors (Pritt and Raese, 1992).

Safety for staff is an important consideration when choosing a laboratory because, aside from the paramount concern for human well-being, accidents can seriously delay sample analyses or create a need for resampling. Prospective laboratories should be inspected for their attention to safety procedures, including the availability of safety equipment such as fire extinguishers, safety showers and eyewashes, fume hoods, and ventilation systems; use and disposal practices for hazardous materials; and compliance with environmental regulations. Safety equipment should be tested on a regular basis (Pritt and Raese, 1992).

Additionally, laboratory safety includes procedures for ensuring that the laboratory is accessible only to authorized personnel to ensure confidentiality of the data. The laboratory should have a system for

accounting for and limiting (or denying) laboratory access to all visitors, including persons affiliated with projects for which the laboratory is analyzing samples (Pritt and Raese, 1992).

5.5.6 Procedural Checks and Audits

A laboratory should have established procedures (SOPs) for conducting internal checks on its analyses and taking corrective action when necessary. If more than one laboratory is used for sample analyses, it will be important to know that the data obtained from the two are of the same quality and consistency. A protocol for conducting interlaboratory comparisons should also be an element of a laboratory's QA/QC plan. For many projects occasional samples are analyzed by a second laboratory to determine whether there is any bias in the data associated with the primary laboratory's analyses.

Laboratory audits by independent auditors are normally conducted on a prescribed basis to ensure that laboratory operations are conducted according to accepted and acceptable procedures (Cross-Smiecinski and Stetzenback, 1994). Determination that a laboratory undergoes such audits and reviews audit results might be sufficient to determine that a laboratory will be adequate for conducting analyses of samples generated by the NPS monitoring project.

5.6 Data and Reports

It is essential during the conduct of an NPS monitoring project to document all data collected and used, to document all methods and procedures followed, and to produce clear, concise, and readable reports that will provide decision makers with the information they need to choose among alternative actions, as described in the DQOs.

5.6.1 Generation of New Data

All data generated during the project, whether in the field, laboratory, or some other facility, should be recorded. Include with the data any reference materials or citations to materials used for data analyses. These include computer programs, and all computer programs used for data reduction should be validated prior to use and verified on a regular basis. Calculations should be detailed enough to allow for their reconstruction at a later date if they need to be verified (Cross-Smiecinski and Stetzenback, 1994). Data generated by a laboratory should be accompanied by pertinent information about the laboratory, such as its name, address, and phone number, and names of the staff who worked directly with the project samples.

5.6.2 Use of Historical Data

Historical data are data collected for previous projects that concerned the same resource in the same area as the project to be implemented. Historical data sometimes contain valuable information, and their use can save time and effort in the implementation and/or data analysis phases of a new project. Before new data are collected, all historical data available should be obtained and their validity and usability should be assessed. *Data validity* implies that individual data points are considered accurate and precise because the field and laboratory methods used to generate the data points are known. *Data usability* implies that a database demonstrates an overall temporal or spatial pattern, though no judgment of the accuracy or precision of any individual data point is made (Spreizer et al., 1992). The validity of historical data can be difficult to ascertain, but data usability can be assessed through a combination of graphical and statistical techniques (Spreizer et al., 1992).

Specifically, historical data that can be shown to be either valid or usable can be applied to a new project in the following ways (Coffey, 1993; Spreizer et al., 1992; USEPA, 1994c):

- If the quality (i.e., accuracy and precision) of historical data is sufficiently documented, the data can be used alone or in combination with new data. The quality of historical data must be determined absolutely, generally with the help of a statistician.
- Characteristics derived from the historical data, such as the variability or mean of data, can be used in the development or selection of a data collection design. Knowledge of expected variability assists in determining the number of samples needed to attain a desired confidence level, the length of monitoring program necessary to obtain the necessary data, and the required sampling frequency.
- Spatial analysis of historical data can indicate which sampling locations are most likely to provide the desired data.
- Historical data can provide insights about past impacts and water quality that can be useful in defining an NPS pollution problem.
- Past trends can be ascertained, and the present tendency of water quality characteristics (degrading, stable, or improving) can be established for trend analysis.

5.6.3 Documentation and Record Keeping

All information and records related to the NPS monitoring project should be kept on file and kept current. This documentation should include:

- A record of decisions made regarding the monitoring project design
- Records of all personnel, with their qualifications, who participated in the project
- Intended and actual implementation schedules, and explanations for any differences
- A description of all sampling sites
- Field records of all sampling events, including any sampling problems and corrective actions taken
- Copies of all field and laboratory SOPs
- Equipment manuals and maintenance schedules (intended and actual, with explanations for any discrepancies)
- Printouts from any equipment
- Sample management and custody records
- Laboratory procedures
- Copy of the laboratory QA/QC plan
- Personnel training sessions and procedures, including any training manuals or other materials
- All data generated during the project in hard copy and electronic forms
- All correspondence related to the project
- Project interim and final reports

5.6.4 Report Preparation

The original project description should include a schedule and required format for required reports, including the final report. Adherence to this schedule is important to provide information and

documentation of project progress, problems encountered, and corrective actions taken. Reports are also valuable for supporting continuation of a project if at any point during the project its continuation is scrutinized or if additional funding must be secured to ensure its completion. Reports can also become the primary sources of historical information on projects if there are changes in project personnel during the project. Project managers should decide on the necessary content and format of all reports prior to commencement of the project, and these will differ depending on funding and intended audience.

Standard Operating Procedures

SAMPLING

Collecting Benthic Macroinvertebrate Samples and Associated Physical and Chemical Data for Ambient Bioassessments in California (updated 02/01/07)
A PDF document that includes two procedures (i.e., targeted-riffle composite, reachwide benthos) for sampling benthic macroinvertebrate (BMI) assemblages for ambient bioassessments. This document also contains procedures for measuring instream and riparian habitats and ambient water chemistry associated with BMI samples. Bioassessment Field Data Sheets with Algae for the SWAMP v2.5 database are also available.

Note: SWAMP issued two policy memos describing additional details for SWAMP bioassessment methods.

Memo 1: SWAMP Standard Operating Procedure (SOP) and Interim Guidance on Quality Assurance for SWAMP Bioassessments (updated 05/21/07)

Memo 2: Amendment to SWAMP Interim Guidance on Quality Assurance for SWAMP Bioassessments (updated 09/17/08)

Collecting Stream Algae Samples and Associated Physical Habitat and Chemical Data for Ambient Bioassessments in California (updated 06/29/10)
A PDF document that includes the procedure for sampling stream algae assemblages for ambient bioassessment. This document also contains procedures for measuring instream and riparian habitats and ambient water chemistry associated with algae samples. Bioassessment Field Data Sheets with Algae for the SWAMP v2.5 database are also available.

Note: In May 2010, SWAMP updated the document about the procedure for sampling stream algae assemblages for ambient bioassessment. There are three main updates to this document: (1) The labels for biomass and taxonomic identification have been updated to include additional information (page 20, figure 6); (2) If presence/absence of a microalgal layer cannot be determined with certainty, the scoring of microalgal thickness is now "UD" (page 39) instead of "Z"; and (3) the bioassessment field data sheets with algae are no longer included in the document. Please download the most recent version of the datasheets.

Note: The SWAMP Conversion and Data Reporting for Benthic Chlorophyll a, Pheophytin and AFDM_Algae (updated 01/26/12) document is

now available along with the corresponding webinar training.

MPSL-DFG SOP for Conducting Field Measurements and Field Collections of Water and Bed Sediment Samples in the Surface Water Ambient Monitoring Program (updated 10/15/07)
A PDF document that describes field measurement and collection procedures used by the Marine Pollution Studies Lab – Department of Fish and Game (MPSL-DFG) for SWAMP water and bed sediment samples. This SOP is for reference and information purposes only and is not required by SWAMP. Water Quality Field Data Sheets for the SWAMP v2.5 database are also available.

TAXONOMY new!

Standard Operating Procedures for Laboratory Processing and Identification of Benthic Macroinvertebrates in California (updated 08/17/12) new!
A PDF document that describes the requirements and recommendations for all laboratories performing SWAMP benthic macroinvertebrate (BMI) taxonomic identifications, and those wishing to be SWAMP-comparable. This SOP also documents the full procedures of the SWAMP Quality Assurance (QA) BMI referee laboratory, the Department of Fish and Game Aquatic Bioassessment Laboratory (DFG-ABL). A presentation on the SWAMP SOP for Benthic Macroinvertebrate Laboratories, presented by Melinda Woodard of the SWAMP QA Team at the California Aquatic Bioassessment Workgroup (CABW) 19th Annual Meeting (November 8, 2012; 9:00 am), is available for viewing: PDF Presentation and Video Presentation (available soon).

Data Entry and Submittal
SWAMP maintains various tools for entering and submitting taxonomy data to SWAMP and CEDEN (California Environmental Data Exchange Network). Access to the forms, additional documentation, and training webinars can be found on the Tools page. BMI samples directly funded by SWAMP should report their data to SWAMP in coordination with the project manager and Data Management Team (DMT) liaison. BMI laboratories processing samples not funded by SWAMP can use SWAMP tools for entering data but data submission should be to CEDEN. If one uses the SWAMP MS Access Taxa Entry database form for entering taxonomy data, Appendix B details the process followed by DFG-ABL with additional documentation provided through the Tools link.

Supporting DFG-ABL Worksheets and Datasheets
The following documents are used by DFG-ABL in their BMI processing and identification procedures, and for their specific data entry requirements. Those documents labeled "ABL" are not for use by non-ABL laboratories. Other documents may serve as examples for non-ABL laboratories and may be used or modified by other laboratories as is relevant. None of these documents are required for non-ABL laboratories performing SWAMP BMI work except for the Corrective Action Process Document.

Forms for Chain of Custody and Sample Log-in Procedures

DFG Pollution Enforcement BMI sample COC Form (PDF) (updated 10/31/11)

ABL Sample Log-in Form – SWAMP Database (PDF)

Forms for the Subsampling and Sorting Procedures

ABL Subsampling worksheet (PDF)

ABL Sorting worksheet (PDF)

Forms for the Taxonomic Identification and Internal QC Procedures

ABL Taxa Entry Form – SWAMP Database (PDF)

Internal QC Sample selection Form (PDF)

Internal QC worksheet (PDF)

Corrective Actions

SWAMP Corrective Action Process Document

DATABASE

Field Data Verification of the Surface Water Ambient Monitoring Program Database (updated 12/17/04)
A PDF document that describes the process used by both the agency responsible for the capture and entry of field data and the SWAMP Data Management Team (DMT) to verify field data in the SWAMP database. This SOP may be used as a guide for other related programs.

Verification of the Surface Water Ambient Monitoring Program Database (updated 03/23/11)
A PDF document that describes the process used by the SWAMP DMT to verify analytical chemistry and bacteria data in the SWAMP database. This SOP may be used as a guide for other related programs.

Toxicity Data Verification of the Surface Water Ambient Monitoring Program Database (updated 02/24/05)
A PDF document that describes the process used by the SWAMP DMT to verify toxicity data in the SWAMP database. This SOP may be used as a guide for other related programs.

Search for:

Search

SWAMP website

For comprehensive SWAMP content, visit the SWRCB's official website

CEDEN

California Environmental Data Exchange Network CEDEN Link

FTP

Allows file sharing when conducting State and Regional Board SWAMP business on the FTP site

---

Contact Webmaster