Monitoring and Evaluating Nonpoint Source Watershed Projects

May 2016

Developed under Contract to U.S. Environmental Protection Agency by Tetra Tech, Inc.
GS Contract #GS-10F-0268K
Order # EP-G135-00168

Authors: Dressing, S.A., D.W. Meals, J.B. Harcum, J. Spooner[1], J.B. Stribling, R.P. Richards[2], C.J. Millard, S.A. Lanberg, and J.G. O'Donnell
[1]North Carolina State University, Raleigh, NC
[2]Heidelberg University, Tiffin, OH.

United States Environmental Protection Agency
Office of Water
Nonpoint Source Control Branch
Washington, DC 20460
EPA 841-R-16-010
May 2016

This document is available at: https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed

This page intentionally left blank.

Foreword

The diffuse nature of nonpoint sources and the variety of pollutants generated by them create a challenge for their effective control requiring a systematic approach based on assessment, planning, implementation, and evaluation. Monitoring is an important component in all four of these activities. While substantial progress has been made since 1972 in the protection and enhancement of water quality, much work is still needed to identify nonpoint source management strategies that are both effective and economically achievable under a wide range of conditions. Lack of adequate information on best management approaches is the major obstacle in developing effective watershed management strategies. We are relearning previous lessons because we have failed to institutionalize previous lessons learned from intensive monitoring efforts from 1970 to the present. This version of the nonpoint source monitoring guide (guide) incorporates the monitoring lessons learned from the Rural Clean Water Program (RCWP), the Clean Water Act Section 319 National Nonpoint Source Monitoring Program (NNPSMP), and other efforts to provide a state-of-the-reference for monitoring nonpoint source projects. Monitoring plays an important role in addressing the need to evaluate our watershed management efforts and document the lessons learned so we can use them as a foundation for future management efforts.

This guide is written primarily for those who develop and implement monitoring plans for watershed management projects, but it can also be used by those who wish to evaluate the technical merits of monitoring proposals they might sponsor. It is an update to the 1997 *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* (EPA 841-B-96-004) and includes many references to that document.

The style and technical level of this guidance are intended to make it accessible to both beginners and experts alike. Numerous real-world examples from RCWP and NNPSMP projects are provided to give the reader a true sense of the challenges faced by those who have monitored waters impacted by nonpoint sources. Included in the guidance document are many references to other related resource materials for those seeking additional or more detailed information.

This guidance begins with an overview of the extent and types of nonpoint source problems reported by the States and Tribes. The overview is intended to provide perspective and set the stage for the chapters that follow. Subsequent chapters describe the basic steps involved in designing a nonpoint source monitoring plan, including sections and chapters devoted to biological, photopoint, and land use monitoring. A chapter that focuses on ways to address the many unique challenges associated with nonpoint source monitoring is also included. The chapter on data analysis describes and illustrates techniques ranging from exploratory data analysis to advanced statistical approaches for assessing the effectiveness of both individual best management practices and watershed projects. Pollutant load estimation methods are also described in detail. A chapter on quality assurance and quality control is then followed by a chapter addressing monitoring costs.

Good monitoring design begins with a clear monitoring objective and an understanding of the water quality problem or concern addressed. Because problems and objectives vary, there is no single approach that can be applied to nonpoint source monitoring efforts. It is hoped this guidance provides a foundation that allows practitioners to design monitoring programs that meet their unique needs.

Readers are encouraged to consult the many resources listed in this document. In addition to these resources, readers are urged to contact monitoring and quality assurance experts in academia and at the local, State, Tribal, and federal levels for assistance in developing monitoring plans and analyzing the collected data.

Acknowledgments

This document has been reviewed by EPA and approved for publication. It was developed by Tetra Tech Inc. under the direction of Mr. Thomas Davenport and Mr. Paul Thomas of EPA Region 5.

The authors gratefully acknowledge the helpful technical reviews provided by Dr. Brian Fontenot of EPA Region 6, Dr. Marty Kelly of Atkins North America, and Mr. John McCoy of the Columbia Association in Maryland. In addition, the authors thank the many individuals who have contributed to the knowledge base on nonpoint source monitoring and data analysis over the past quarter century or more. The references contained in this document only begin to recognize the contributions of others.

Inspiration for this document was provided long before the 1997 version for which this serves as an upgrade. Mr. James W. Meek, former Chief of the Nonpoint Source Control Branch at EPA Headquarters, was particularly inspirational in his support for developing and documenting improved methods to demonstrate the effectiveness of nonpoint source control measures and programs. The late Dr. Frank J. Humenik, Professor in the Department of Biological and Agricultural Engineering at North Carolina State University, was instrumental in the promotion of long-term monitoring projects to evaluate the effectiveness of approaches to solve water quality problems at the watershed level. Finally, Mr. Thomas Davenport of EPA Region 5 has been the driving force behind EPA's continued involvement in nonpoint source watershed projects that began in earnest with the Model Implementation Program, Nationwide Urban Runoff Program, and Rural Clean Water Program. Mr. Davenport has led the effort to document the effectiveness of nonpoint source pollution control efforts through sound scientific approaches, and he has been the major proponent of developing this upgraded nonpoint source monitoring guidance.

Contents

Figures

Tables

This page intentionally left blank.

Acronym List

AA	atomic absorption
ac	acre
ac/ft	acre-foot
ACF	autocorrelation function
ADCP	acoustic Doppler current profiler
AFDM	ash-free dry mass
Ag	silver
Al	aluminum
ANCOVA	analysis of covariance
ANOVA	analysis of variance
APA	acid/alkaline phosphatase activity
ARIMA	autoregressive integrated moving average
As	arsenic
ATTAINS	assessment TMDL tracking & implementation system
Au	gold
BCG	biological condition gradient
BEACH	beaches environmental assessment, closure and health
BioK	biological/habitat with kick net
BMP	best management practice
BOD	biochemical oxygen demand
BOD_5	5-day biochemical oxygen demand
CADDIS	causal analysis/diagnosis decision information system
CAFO	concentrated animal feeding operation
CCA	canonical correlation analysis *and* canonical correspondence analysis
C-CAP	coastal change analysis program
Cd	cadmium
CEAP	conservation effects assessment project

cfs	cubic feet per second
CI	confidence interval
Cl^{-1}	chloride
cm	centimeter
cms	cubic meters per second
Co	cobalt
COD	chemical oxygen demand
Cu	copper
CV	coefficient of variation
CWA	Clean Water Act
DBI	diatom bioassessment index
DD	detectable difference
DEM	digital elevation model
d.f.	degree of freedom
DIA	digital image analysis
DL	detection limit
DO	dissolved oxygen
DQO	data quality objective
EDA	exploratory data analysis
EDI	equal discharge interval
EMC	event mean concentration
EMMA	environmental monitoring and measurement advisor
EPA	U.S. Environmental Protection Agency
EPT	Ephemeroptera-Plecoptera-Trichoptera
EWI	equal width interval
Fe	iron
FSA	Farm Service Agency
ft	feet
ft^3/s	cubic feet per second
GIS	geographic information system

GPS	global positioning system
H_2SO_4	sulfuric acid
ha	hectare
HBI	Hilsenhoff Biotic Index
Hg	mercury
HNO_3	nitric acid
IBI	Index of Biological Integrity
ICP	inductively coupled plasma
in	inch
IQR	interquartile range
IR	integrated reporting
IWL	Izaak Walton League
kg	kilogram
KS	Kolmogorov-Smirnov
L	liter
Li	lithium
LA	load allocation
LIA	line-intersect analysis
LID	low impact development
LiDAR	light detection and ranging
LOWESS	locally weighted scatterplot smoothing
LS-means	least square means
LSD	least significant difference
LULC	land use/land cover
m	meter
m^3/s	cubic meters per second
MA	moving average
MAI	macroinvertebrate aggregated index
MBI	macroinvertebrate biotic index
MDC	minimum detectable change

MDNR	Maryland Department of Natural Resources
mg	milligram
mi	mile
ml	milliliter
MLE	maximum likelihood estimation
mm	millimeter
MMI	multimetric index
Mn	manganese
MOS	margin of safety
MQO	measurement quality objective
mRPD	median relative percent difference
N	nitrogen
NAWQA	national water-quality assessment program
NELAC	national environmental laboratory accreditation conference
NEMI	national environmental methods index
NGO	non-governmental organization
NH_3-N	ammoniacal nitrogen
Ni	nickel
NLCD	national land cover dataset
NNPSMP	national nonpoint source monitoring program
NO_3	nitrate nitrogen
NPDES	national pollution discharge elimination system
NPS	nonpoint source
NRCS	Natural Resources Conservation Service
NRI	national resources inventory
NRSA	national rivers and streams assessment
NSC	nutrient and sediment grab samples
NSL	nutrient and sediment loads
NWQI	national water quality initiative
O/E	observed/expected

P	phosphorus
PACF	partial autocorrelation function
Pb	lead
PCA	principal component analysis
PDTG	percent dominant taxa (generic level)
PGDER	Prince George's County Department of Environmental Resources
PIBI	potential index of biological integrity
POCIS	polar organic chemical integrative samplers
PPCC	probability plot correlation coefficient
PROC AUTOREG	SAS procedure to estimate and forecast linear regression models for time series data
QAP	quality-assurance plan (USGS)
QAPP	quality assurance project plan
QHEI	qualitative habitat evaluation index
QL	quantitation limit
QMP	quality management plan
RCB	randomized complete block
RCWP	rural clean water program
ROS	regression on order statistics
RPD	relative percent difference
RUSLE	revised universal soil loss equation
SA	subjective analysis
SAP	sampling and analysis plan
SAS	SAS Institute, Inc.
Sb	antimony
SCC	suspended sediment concentration
SIMPLE	spatially integrated models for phosphorus loading and erosion
SNT	sondes for nutrients and turbidity
SO_4^{-2}	sulfate
SOP	standard operating procedure

SPARROW	spatially referenced regressions on watershed attributes watershed modeling technique
SRP	soluble reactive phosphorus
SSC	suspended sediment concentration
STEPL	spreadsheet tool for estimating pollutant load
STORET	EPA's storage and retrieval database for water quality, biological, and physical data
SWAT	soil and water assessment tool
SWM	statewide monitoring network
SWP	stormwater detention/retention pond
TDS	total dissolved solids
TIGER	topologically integrated geographic encoding and referencing
TIR	thermal infrared
TKN	total Kjeldahl nitrogen
TMDL	total maximum daily load
TNTC	too numerous to count
TP	total phosphorus
TSS	total suspended solids
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
VIF	variance inflation factor
W/D	width-to-depth ratio
WLA	waste load allocation
WQS	water quality standards
WRTDS	weighted regressions on time, discharge, and season
WWTP	wastewater treatment plant
Zn	zinc

1 Overview of the Nonpoint Source Problem

By S.A. Dressing, D.W. Meals, J.B. Harcum, and J. Spooner

1.1 Definition of a Nonpoint Source

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

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

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

Stormwater can be classified as a point or nonpoint source of pollution. Stormwater is classified as a point source when it is regulated through the National Pollution Discharge Elimination System (NPDES) Stormwater Program. An NPDES stormwater permit is required for medium and large municipal separate storm sewer systems (MS4s) of incorporated areas and counties with populations of more than 100,000, certain industrial activities, and construction activities disturbing five ac or more. An NPDES permit is also required for small MS4s in "urbanized areas" and small construction activities disturbing between one and five acres (ac) of land. The NPDES permitting authority may also require operators of small MS4s not in urbanized areas and small construction activities disturbing less than one ac to obtain an NPDES permit based on the potential for contribution to a violation of a water quality standard. Detailed information on the NPDES Storm Water Program is available at http://www.epa.gov/npdes/npdes-stormwater-program. If stormwater originates from a location that does not fall within the NPDES permit requirements, it is considered to be nonpoint source pollution (USEPA 2005). Concentrated animal feeding operations (CAFOs) are also classified as point sources and regulated under the NPDES program (USEPA 2012b). Despite differing regulatory requirements, monitoring issues and concepts encountered for permitted stormwater and CAFOs are similar to those of nonpoint sources.

1.2 Extent of Nonpoint Source Problems in the United States

During the last three decades, significant achievements have been made nationally in the protection and enhancement of water quality. Much of this progress, however, has resulted from controlling point sources of pollution. Pollutant loads from nonpoint sources continue to present problems for achieving water quality goals and maintaining designated uses in many parts of the United States. Nonpoint sources are generally considered the number one cause of water quality problems reported by states, tribes, and territories.

Categories of nonpoint source pollution affecting waterbodies include agriculture, atmospheric deposition, channelization, construction, contaminated sediment, contaminated ground water, flow regulation, forest harvesting (silviculture), ground water loading, highway maintenance/runoff, hydrologic and habitat modification, in-place contamination, land development, land disposal, marinas, onsite disposal systems, recreational activities, removal of riparian vegetation, resource extraction, shoreline modification, streambank destabilization, and unspecified or other nonpoint source pollution.

Nonpoint sources can generate both conventional pollutants (e.g., nutrients, sediment) and toxic pollutants (e.g., pesticides, petroleum products). Even though nonpoint sources can contribute many of the same kinds of pollutants as point sources, these pollutants are usually generated in different timeframes, volumes, combinations, and concentrations.

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

Nutrient pollution (i.e., nitrogen [N] and phosphorus [P]) is often associated with NPS and has received increasing attention as algal blooms and resulting hypoxic or “dead” zones caused by the decay of algae have negatively affected waterbodies around the country (NOAA 2012). Various other pollutants contributed by NPS include sediment, pathogens, salts, toxic substances, petroleum products, and pesticides. Each of these pollutants, as well as habitat alteration and hydrologic modification, can have adverse effects on aquatic systems and, in some cases, on human health.

- Waste from livestock, wildlife, and pets contain bacteria that contaminate swimming, drinking, and shellfishing waters, as well as oxygen-demanding substances that deplete dissolved oxygen (DO) levels in aquatic systems. Suspended sediment generated by construction, overgrazing, logging, and other activities in riparian areas, along with particles carried in runoff from cropland, highways, and bridges, reduces sunlight to aquatic plants, smothers fish spawning areas, and clogs filter feeders and fish gills.
- Salts from irrigation water become concentrated at the soil surface through evapotranspiration and are carried off in return flow from surface irrigation. Road salts from deicing accumulate along the edges of roads and are often carried via storm sewer systems to surface waters. Salts cause the soil structure to break down, decrease water infiltration, and decrease the productivity of cropland. Salts can also be toxic to plants at high concentrations.

- Some pesticides are persistent in aquatic systems and biomagnify in animal tissue (primarily fish tissue) as they are passed up through the food chain. Biomagnification has detrimental physiological effects in animals and negative human health impacts. Herbicides can be toxic to aquatic plants and therefore remove a food source for many aquatic animals. Herbicides can also kill off the protective cover that aquatic vegetation offers to many organisms.
- Finally, the trampling of stream bottoms by livestock and equipment; stream bank erosion caused by grazing, logging, and construction; conversion of natural habitats to agricultural, urban, and other land uses; flow regulation; and activities in riparian areas (e.g., tree removal, buffer removal) can reduce the available habitat for aquatic species, increase erosion, increase water temperature via reduced shading, and create flow regimes that are detrimental to aquatic life.

Every two years, states and territories are required to submit a 305(b) report that describes the status of all assessed waters and a 303(d) report that lists the impaired waters, the causes of impairment and the status of their restoration. In 2001, the U.S. Environmental Protection Agency (EPA) issued guidance to the states encouraging submitting one electronic, integrated water monitoring and assessment report. This report is currently expected to include the 305(b), 303(d), and 314 (Clean Lakes Program) assessments (Keehner 2011). Currently there are no plans to release a new National Water Quality Inventory Report to Congress. The last Report to Congress was released in 2009 and provided a synopsis of 2004 data. Information on Integrated Reporting, including the guidance issued by EPA, is available at www.epa.gov/tmdl/integrated-reporting-guidance. The Assessment TMDL Tracking & Implementation System (ATTAINS) provides the most current 305(b) and 303(d) information available for all 50 states and territories. ATTAINS summarizes state-reported data for the nation, individual states, individual waters and the 10 EPA regions.

A national summary of assessment data submitted by the states from 2004 through 2014 (with over 80 percent for the period 2010–2014) documents the extent of the nonpoint source problem (USEPA 2016). The share of waters assessed by the states in these reports was 32 percent of river miles (mi); 45 percent of lake, reservoir, and pond acreage; 40 percent of bay and estuary square mileage; 14 percent of coastal shoreline mi; 3 percent of ocean and near coastal water square mileage; 1 percent of wetlands acreage; 85 percent of Great Lakes shoreline; and 88 percent of Great Lakes open water square mileage. For these assessed waters, Table 1-1 shows national totals for causes of impairments or threats to impairment that are often associated with nonpoint sources. A wide range of causes frequently associated with nonpoint sources are at the top of the list for rivers and streams, including pathogens, sediment, nutrients, organic enrichment/oxygen depletion, temperature, metals, habitat and flow alterations, and turbidity. Nutrients, organic enrichment/oxygen depletion, turbidity, metals, and sediment are also leading causes of impairments and threats to lakes, while pathogens and organic enrichment/ depleted oxygen are among the top causes of problems identified in bays and estuaries and coastal shoreline. Organic enrichment/ depleted oxygen is the largest cause of impairment to wetlands, with metals, pathogens, and nutrients also among the leading causes of impairment. Pesticides were found to be a significant cause of problems in Great Lakes open waters and along the Great Lakes shoreline, while organic enrichment/ depleted oxygen was the largest cause of impairment to ocean and near coastal waters.

Table 1-1. National causes of impairment (excerpted from USEPA 2016)

Cause of Impairment Group	Size of Assessed Waters with Listed Causes of Impairment							
	Rivers and Streams (Miles)	Lakes, Reservoirs, and Ponds (Acres)	Bays and Estuaries (Square Miles)	Coastal Shoreline (Miles)	Ocean and Near Coastal (Square Miles)	Wetlands (Acres)	Great Lakes Shoreline (Miles)	Great Lakes Open Water (Square Miles)
Algal Growth	6,013	908,513	1,474	93		4,631	191	
Ammonia	11,673	214,501	41	22	1	31		
Flow Alteration(s)	42,694	190,228	3			4,387	202	
Habitat Alterations	67,242	319,965	2			2,104	170	
Metals (other than Mercury)	89,069	1,304,587	1,878	60	15	94,630		
Nutrients	117,412	3,586,616	3,605	131	7	67,955	380	
Oil and Grease	3,014	44,285	101	95				
Organic Enrichment/Oxygen Depletion	99,578	1,697,788	5,421	437	579	462,402	138	13,867
Pathogens	178,219	549,515	7,034	1,056	80	72,385	621	
Pesticides	19,565	494,613	1,847	36	52	169	2,483	29,661
Sediment	145,289	788,465	224	5		10,786	319	
Temperature	93,513	240,684	145	96	1	14,900		
Turbidity	47,854	1,341,862	899	331	24	3,915		

Table 1-2 summarizes the extent to which sources often associated with NPS are responsible for documented impairments and threats to impairment for different waterbody types. Agriculture is the top source reported for river and stream problems, with a range of other sources associated with nonpoint sources also contributing significantly, including hydromodification, habitat alteration, urban-related runoff/stormwater, unspecified NPS, forestry, mining, and construction. The states reported that agriculture is the third leading source causing problems in lakes behind atmospheric deposition and unknown sources, with unspecified NPS, hydromodification, and urban-related runoff/stormwater also major sources. Problems in the Nation's bays and estuaries are more commonly associated with unknown sources and atmospheric deposition, but unspecified NPS, urban-related runoff/stormwater, agriculture, habitat alteration, and hydromodification are also significant contributors to these problems according to the states. Urban sources play a substantial role in the problems reported for coastal shoreline, ocean and near coastal waters, and open Great Lakes waters, whereas agriculture is also an important source for impairments and threats to wetlands and Great Lakes shoreline and open waters.

Finally, Table 1-3 shows the number of TMDLs (total maximum daily loads) written since October 1, 1995, for various pollutants. Pathogens, which come from both point and nonpoint sources, are second to mercury at the top of the list, and metals, nutrients, sediment, and other pollutants often associated with NPS have also been the focus of many TMDLs. The figures in Table 1-3 are based on a total of 69,173 TMDLs written to address 72,618 causes of impairment.

Table 1-2. National probable sources contributing to impairments (excerpted from USEPA 2016)

Probable Source Group	Size of Assessed Waters with Probable Sources of Impairments							
	Rivers and Streams (Miles)	Lakes, Reservoirs, and Ponds (Acres)	Bays and Estuaries (Square Miles)	Coastal Shoreline (Miles)	Ocean and Near Coastal (Square Miles)	Wetlands (Acres)	Great Lakes Shoreline (Miles)	Great Lakes Open Water (Square Miles)
Agriculture	148,728	1,241,455	3,056	113		201,786	620	4,373
Construction	21,527	336,942	1	4	4	1,000	18	
Habitat Alterations (Not Directly Related to Hydromodification)	66,932	273,438	2,231			33	90	
Hydromodification	92,067	762,274	1,717	140	7	6,762	231	
Recreational Boating And Marinas	138	38,743	789	106	8	72,320		
Resource Extraction	33,873	524,820	320			32,112		
Silviculture (Forestry)	40,637	162,244	0					
Unspecified NPS	54,142	847,767	3,363	103	4	1,324	6	
Urban-Related Runoff/Stormwater	61,984	744,646	3,086	268	379	54	99	13,867

Table 1-3. National cumulative TMDLs by pollutant (excerpted from USEPA 2016)

Pollutant Group	Number of TMDLs	Number of Causes of Impairment Addressed
Pathogens	13,263	13,572
Metals (other than Mercury)	9,955	10,153
Nutrients	6,154	7,520
Sediment	3,941	4,591
Temperature	2,305	2,315
Organic Enrichment/Oxygen Depletion	2,191	2,315
Turbidity	1,603	1,829
Pesticides	1,351	1,514
Ammonia	1,131	1,230
Algal Growth	95	103
Habitat Alterations	83	84
Oil and Grease	14	14

Many other measures and indicators of the extent of the NPS problem are also available, including the National Rivers and Streams Assessment (NRSA), under which 1,924 river and stream sites were sampled during the summers of 2008 and 2009 (USEPA 2013). This study was based on a robust, commonly used index that combines different measures of the condition of aquatic benthic macroinvertebrates. The draft report indicates that 21 percent of the nation's river and stream length is in good biological condition, 23 percent is in fair condition, and 55 percent is in poor condition (no data for 1 percent). Of the four chemical stressors assessed in this study (total P [TP], total N [TN], salinity, and acidification), it was concluded that P and N are by far the most widespread. It was found that 40 percent of the nation's river and stream length has high[1] levels of P and 28 percent has high levels of N. Poor biological condition (for macroinvertebrates) was found to be 50 percent more likely in rivers and streams with high levels of P and 40 percent more likely in rivers and streams with high levels of N. Four indicators of physical habitat condition (excess streambed sediments, riparian vegetative cover, riparian disturbance, and in-stream fish habitat) were also assessed for the study. Results indicated that poor riparian vegetative cover and high levels of riparian disturbance are the most widespread physical stressors, reported in 24 percent and 20 percent of the nation's river and stream length, respectively. Excess levels of streambed sediments, however, were reported in 15 percent of river and stream length and were found to have a greater impact on biological condition. The study concluded that poor biological condition is 60 percent more likely in rivers and streams with excessive levels of streambed sediments. While this study was not designed to identify the sources of stressors, other research has shown that nonpoint sources are often contributors to both the chemical and physical stressors described here. The draft report was released for comment on March 25, 2013, and is currently undergoing final revision.

EPA also performed a National Wetland Condition Assessment (NWCA) to determine the ecological integrity of wetlands at regional and national scales through a statistical survey approach. Field data were collected in 2011 and a draft report was released for public comment through January 6, 2016 (USEPA 2015c). Draft findings indicate that nationally, 48% of the wetland area is in good condition, 20% is in fair condition and the remaining 32% of the area is in poor condition. The study also assessed a number of physical, chemical, and biological indicators of stress that reflect potential negative impact to wetland condition. These indicators were assigned to "low," "moderate," or "high" stressor levels depending on criteria established for each indicator. Of the six physical indicators, vegetation removal and hardening (e.g., pavement, soil compaction) stressors were assessed as high for 27% of wetland area nationally, while the ditching stressor was high for 23% of wetland area. Both of the chemical indicators (a heavy metal index and soil P concentration) were low for the majority of wetland area nationally, but at variable levels across the four aggregated ecoregions created for the study. A Nonnative Plant Stressor Indicator developed for NWCA was used to assess the level of biological stress in wetlands. Nationally, 61% of wetland area had low stressor levels for nonnative plants, but results varied across aggregated ecoregions.

Still, other reports indicate the pervasive nature of NPS pollution and the need to document and solve the many problems it causes. For example:

- Based on the sampling of over 1,000 lakes across the country in 2007, it was determined that poor lake physical habitat is the biggest problem affecting biological condition, followed by high nutrient levels (USEPA 2009). This statistical survey found that lakes with excess nutrients (i.e., a "poor" stressor condition) are two-and-a-half times more likely to have poor biological health[2].
- EPA's 2012 National Coastal Condition Report noted that U.S. coastal areas are facing significant population pressures and associated higher volumes of urban nonpoint source runoff with 53 percent of the U.S. population living in coastal areas that comprise only 17 percent of the total conterminous U.S. land area (USEPA 2012a). This report rated the U.S. coasts as "fair" on a scale

[1] Thresholds for high, medium and low values were set on a regional basis relative to the least-disturbed reference sites for each of the nine NRSA ecoregions.)
[2] This likelihood is expressed relative to the likelihood of Poor response condition in lakes that have Not-Poor stressor condition (USEPA 2010).

of good, fair, or poor. Dissolved inorganic P levels, one of the five components of the water quality index, was also rated "fair."

- Nonpoint sources, particularly from the agricultural areas north of the confluence of the Ohio and Mississippi Rivers, contribute most of the N and P loads to the Gulf of Mexico (Goolsby et al. 1999). The nitrate load to the Gulf approximately tripled from 1970 to 2000, with the greatest sources believed to be basins in southern Minnesota, Iowa, Illinois, Indiana, and Ohio that drain agricultural land (Goolsby et al. 2001).
- In 2015, the Gulf of Mexico hypoxic zone measured 6,474 square miles (4.14 million ac), larger than the state of Hawaii (USEPA 2015f). The greatest source of pollution causing the hypoxic zone in the Gulf of Mexico is nonpoint source runoff from agriculture. It has been estimated that corn and soybean cultivation contributes 52 percent of the N delivered to the Gulf from the Mississippi River Basin, with other cropland, manure on pasture and rangeland, and forest contributing 14, 5, and 4 percent, respectively (Alexander et al. 2008). It was also estimated that animal manure on pasture and rangeland, corn and soybeans, other cropland, and forest contribute 37, 25, 18, and 8 percent of the P, respectively.

1.3 Major Categories of Nonpoint Source Pollution

1.3.1 Agriculture

The 2012 Census of Agriculture reported that there are 2,109,303 farms covering 914,527,657 acres (ac) in the U.S. (USDA-NASS 2014). Approximately 1.5 million farms grew crops on 390 million ac, and there were about 415 million ac of permanent pasture and range on nearly 1.2 million farms. Woodland covered 77 million acres, while other agricultural features (e.g., farmsteads, buildings, livestock facilities, ponds, and roads) accounted for 32 million ac of farmland. Animal agriculture included nearly 90 million cattle and calves on approximately 900 thousand farms, 66 million hogs and pigs on 63 thousand farms, and 1.5 billion broilers on 42 thousand farms.

The primary agricultural nonpoint source pollutants are inorganic and organic nutrients (N and P), sediment, organic matter and pathogens from animal waste, salts, and agricultural chemicals. Agriculture and agricultural activities can also have direct impacts on aquatic habitat. N and P are applied to agricultural land in several different forms and come from various sources, including commercial fertilizer, manure from animal production facilities, municipal and industrial treatment plant sludge and/or effluent applied to agricultural lands, legumes and crop residues, irrigation water, and atmospheric deposition.

Land disturbance and clearing for agricultural operations can increase sediment loadings in runoff and surface waters. In addition, increased instream flows resulting from this land clearing can also contribute to accelerated stream bank erosion. Sediment loss and runoff are especially high if it rains or if high winds occur while the soil is being disturbed or soon afterward.

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

Large amounts of salt can be added to agricultural soils by irrigation water that has a natural base load of dissolved mineral salts, regardless of whether the water is supplied by ground water or surface water sources. Irrigation water is consumed by plants and lost to the atmosphere by evaporation, and the salts in the water remain on and become concentrated in the soil. Salt accumulation leads to soil dispersion, soil compaction, and possible toxicity to plants and soil fauna. Salt can also be carried from fields in irrigation return flows.

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

Riparian vegetation and its pollutant buffering capacity are lost when crops are planted too close to surface waters. Livestock grazing can cause loss of cover vegetation on pasturelands, resulting in erosion, loss of plant diversity on pasturelands, and adverse impacts on stream courses and surface waters. Cattle with access to streams can directly deliver fecal contamination to waterbodies, trample riparian vegetation and disturb stream bank soils, leading to bank erosion. In addition, grazing can alter riparian vegetation species composition.

1.3.2 Urban Sources

The most common pollutants coming from stormwater sources include sediment, pathogens, nutrients, and metals (USEPA 2015b). Other pollutants in runoff from urban areas include oil, grease and toxic chemicals from motor vehicles; pesticides and nutrients from lawns and gardens; viruses, bacteria and nutrients from pet waste and failing septic systems; road salts; heavy metals from roof shingles, motor vehicles and other sources; and thermal pollution from impervious surfaces such as streets and rooftops (USEPA 2015e). Research has indicated that the unit area contribution of pesticides to watersheds by urbanized areas (e.g. golf courses and home lawn care) may be greater than that from agriculture (Steele et al. 2010).

Urbanization converts large portions of vegetated land to unvegetated, impervious land, thus changing the extent to which the land can absorb and filter rainfall and runoff before it enters waterbodies. The amount of impervious surface in urban areas—such as rooftops, roads, parking lots, and sidewalks—can range from 35 percent or lower in lightly urbanized areas to nearly 100 percent in heavily urbanized areas. These changes to the landscape increase pollutant loadings, stormwater runoff volumes, and peak flow rates in urban streams. Pollutants carried in urban runoff often reach surface waters without treatment.

The impacts of urbanization on local hydrology can be particularly acute. Urban streams are frequently flashy, meaning that discharge rates increase rapidly in response to storms, followed by a quick return to normal after the storm passes. A study in the Piedmont of western Georgia, for example, showed that high flow pulses and elevated peak discharges were more frequent in urban watersheds than any other land cover, and baseflow inputs in urban streams were lower than other watersheds (Schoonover et al. 2006). Streams in urbanized areas are also often characterized by accelerated bank erosion, channel widening, and sedimentation (Roy et al. 2010), with much of this due to the destructive energy of large volumes of rapidly moving stormwater runoff. The frequency of flooding is also increased in many cases, particularly during spring snowmelt and rain-on-snow events (Buttle and Xu 1988, Pitt and McLean 1992). The combination of pollutants and hydrologic impacts in urban settings tends to produce biotic assemblages of low diversity dominated by tolerant and nonnative species (Roy et al. 2010). Wide-ranging research relating impervious cover to stream quality has been incorporated within the Impervious Cover Model

(ICM), a watershed planning model that predicts that most stream quality indicators decline when watershed impervious cover exceeds 10 percent, with severe degradation expected beyond 25 percent impervious cover (CWP 2003). Urbanization can change in-stream processing of nutrients and other elements through the combined impacts of changes to stream hydrology, sediment texture, organic matter levels, and stream flora and fauna (Steele et al. 2010).

1.3.3 Removal of Streamside Vegetation

Riparian zones are transitional areas, containing elements of both aquatic and terrestrial habitats (Knutson and Naef 1997). Riparian habitat performs many functions, including (Knutson and Naef 1997, USDOI 1991):

- providing shade to cool stream waters;
- stabilizing stream banks and controlling erosion and sedimentation;
- rebuilding floodplains; and
- contributing leaves, twigs, and insects to streams, thereby providing basic food and nutrients that support fish and aquatic wildlife.

Fish also benefit from large trees that fall into streams creating pools, riffles, backwater, small dams, and off-channel habitat. In addition, riparian areas filter sediments and pollutants from runoff and moderate stream volumes by reducing peak flows and slowly releasing water to maintain base flows.

Losses of riparian or streamside vegetation are attributed to conversion to farmland, drainage for agriculture, forest harvesting, channelization, damming, creating impoundments, irrigation diversions, ground water pumping, and overgrazing (Brinson et al. 1981). Riparian vegetation is also lost due to urbanization (MSD 2012, Ozawa and Yeakley 2007).

Removal of riparian vegetation cuts off the natural supply of nutrients and energy to biological communities in low-order streams (USEPA 1991). Terrestrial and aquatic habitat available for shelter, forage, and reproduction is destroyed, and canopy removal results in increased stream temperatures and greater temperature fluctuations. Streambank stability is reduced and erosion and sedimentation are increased when the rooting systems of riparian vegetation are destroyed or removed (Brinson et al. 1981). In addition, stream flow buffering is reduced, flooding may increase, and in-stream sedimentation and pollutant loads may increase, all of which can cause severe stress to aquatic plant and animal communities.

1.3.4 Hydromodification

Hydromodification is the alteration of the hydrologic characteristics of coastal and non-coastal waters, which in turn could cause degradation of water resources (USEPA 2007). It includes channelization or channel modification and flow alteration. Channel modification is river and stream channel engineering undertaken for the purpose of flood control, navigation, drainage improvement, or reduction of channel migration potential (Brookes 1990). Examples of channel modification include straightening, widening, deepening, or relocating existing stream channels; excavation of borrow pits, canals, underwater mining, and other practices that change the depth, width, or location of waterways or embayments in coastal areas; and clearing or snagging operations. Channel modification typically results in more uniform channel cross sections, steeper stream gradients, and reduced average pool depths.

Flow alteration describes a category of hydromodification activities that results in either an increase or a decrease in the usual supply of fresh water to a stream, river, wetland, lake, or estuary. Flow alterations include diversions, withdrawals, and impoundments. In rivers and streams, flow alteration can also result from transportation embankments, tide gates, sluice gates, weirs, and the installation of undersized culverts. Levees and dikes are also flow alteration structures.

Channel modification can deprive wetlands and estuarine shorelines of enriching sediment; change the ability of natural systems to absorb hydraulic energy and filter pollutants from surface waters; increase transport of suspended sediment to coastal and near-coastal waters during high-flow events; increase instream water temperature; and accelerate the discharge of pollutants (Sherwood et al. 1990). Channelization can also increase the risk of flooding by causing higher flows during storm events (USEPA 2007). Hydromodification often diminishes the suitability of instream and riparian habitat for fish and wildlife through reduced flushing, lowered DO levels, saltwater intrusion, interruption of the life cycles of aquatic organisms, and loss of streamside vegetation. Dams, for example, can change water temperatures and impact fish spawning (USEPA 2007).

1.3.5 Mining

Much of the environmental damage caused by mining occurred prior to passage of the Surface Mining Control and Reclamation Act (SMCRA) of 1977, when standards for environmental protection during mining operations and the means for reclaiming abandoned mines were generally lacking (Demchak et al. 2004). For example, past practices used to mine silver (Ag) and gold (Au) from low-grade ore generated large volumes of waste material (spoil) that were dumped at the heads of drainages, potentially serving as sources of sediment to streams as they weathered over time (Sidle and Sharma 1996). Mercury (Hg) was used to separate Au and Ag from ore in the past and is contained in waste piles from the amalgamation process (Oak Ridge National Laboratory 1993). Numerous pollutants are released from coal and ore mining. Acid drainage from coal mining contains sulfates, acidity, heavy metals, ferric hydroxide, and silt (USEPA/USDOI 1995, Stewart and Skousen 2003). The heavy metals released from mining activities include Ag, arsenic (As), copper (Cu), cadmium (Cd), Hg, lead (Pb), antimony (Sb), and zinc (Zn) (Horowitz et al. 1993).

While modern-day mining practices are much improved, there remains a need to address the environmental impacts of past mining practices in many locations. For example, two Section 319 National Nonpoint Source Monitoring Program (NNPSMP) projects were designed to monitor the effects of restoration activities on water quality in areas impacted by past mining activities. In Pennsylvania, monitoring was carried out to determine the effectiveness of remediation efforts designed to counter the impact of abandoned anthracite mines on the aquatic ecosystem and designated beneficial uses of Swatara Creek (Cravotta et al. 2010). Impairments were caused both by acid mine drainage and losses of surface water to the abandoned underground mines. In Michigan's Keweenaw Peninsula efforts are underway to address problems caused by fine-grained stamp sands from historic copper mining operations (Rathbun 2007). These sands erode into streams and wetlands and degrade fish and macroinvertebrate communities by smothering aquatic habitat features and leaching copper into the water column.

While remediation efforts often result in water quality improvements, solutions are sometimes more complicated than initially envisioned. For example, acid mine drainage resulting from Cu mining in the Ducktown Mining District of Tennessee introduced significant amounts of toxic trace metals into tributaries of the Ocoee River (Lee et al. 2008). Downstream neutralization of acidic water resulted in the precipitation of iron hydroxides and the sorption of trace metals to the suspended particulates which were then transported downstream to a lake where they settled on the lake bottom. This sediment layer contains elevated levels of Fe, Al, Mn, and trace metals such as Cu, Zn, Pb, Ni, and Co. Study results have shown that even a modest decrease in pH of the sediment pore water from 6.4 to 5.9 caused significant release of trace metals to the environment, creating a risk of ingestion by bottom-dwelling aquatic species.

1.3.6 Forestry

Forestry operations can degrade water quality in several ways, with sediment, organic debris, nutrients, and silvicultural chemicals the major pollutants of concern (Binkley et al. 1999, Michael 2003, Ryan and Grant 1991). Construction of forest roads and yarding areas, as well as log dragging during harvesting, can accelerate erosion and sediment deposition in streams, thus harming instream habitats (Ryan and Grant 1991, USEPA 2015a). Road construction and road use are the primary sources of NPS pollution on forested lands, contributing up to 90 percent of the total sediment from forestry operations (USEPA 2015a). Removal of overstory riparian shade can increase stream water temperatures (USEPA 2015a). Harvesting operations can leave slash and other organic debris to accumulate in waterbodies, resulting in depleted dissolved oxygen (DO) and altered instream habitats. Fertilizer applications can increase nutrient levels and accelerate eutrophication, whereas pesticide applications can lead to adverse wildlife and habitat impacts (Brown 1985). Herbicides can be applied with reduced or shorter-term environmental impact, however, in situations where macroinvertebrate recolonization is rapid and herbicide concentrations are low and short-lived because of acidic soil and water conditions (Michael 2003).

A review of forest fertilization studies around the world concluded that, in general, peak stream concentrations of nitrate-N increase after forest fertilization, with a few studies reporting concentrations as high as 10-25 milligrams (mg) nitrate (NO_3)-N/L (lithium) (Binkley et al. 1999). In addition, the highest reported annual average NO_3-N concentration found was 4 mg N/L. The higher nitrate concentrations were related to repeated fertilization, use of ammonium nitrate instead of urea, and fertilization of N-saturated hardwood forests. It was found that phosphate fertilization could create peak concentrations exceeding 1 mg P/L, but annual averages remain below 0.25 mg P/L. A study of the effects of fertilizer addition to an artificially drained North Carolina pine plantation resulted in the flushing out of all excess nutrients by three major rain events within 47 days of application (Beltran et al. 2010). Researchers considered this to be a worst-case scenario, however, noting that N concentrations did not exceed EPA’s drinking water standard of 10 mg N/L and loading rates returned to pretreatment or lower levels as soon as 90 days after fertilization. Still, the results point out the importance of timing of fertilizer applications to reduce potential losses.

The use of forest lands for application of biosolids and animal wastes has received increased attention in the literature, reflecting concerns that such applications could increase nutrient loadings from these lands. For example, a study designed to evaluate the potential for using loblolly pine stands for poultry litter application in the South indicated that moderate application rates (~20 kilograms [kg] N/ hectare (ha), ~92 kg P/ha) can increase tree growth with minimal impacts to water quality (Friend et al. 2006). Higher application rates (800 kg N/ha, 370 kg P/ha), however, resulted in soil water nitrate levels exceeding 10 mg N/L and P buildup in soils. A study examining surface runoff of N and P in a small, forested watershed in Washington yielded no evidence of direct runoff of N or P from biosolids into surface waters (Grey and Henry 2002). This study illustrated the importance of best management practices (BMPs) as N-based application rates were used and a 20-meter (m) buffer was established around the creek and all ephemeral drainages. Only 40 percent of the watershed received nutrient applications (700 kg N/ha, 500 kg P/ha) and the acidic soils were expected to reduce P mobility. Before biosolids application, however, there was no relationship between discharge and nitrate-N concentration, but within nine months of application discharge and nitrate-N concentrations were positively correlated, indicating the potential for impacts to water quality with continued biosolids applications.

1.3.7 Construction

Stormwater runoff from construction activities can have a significant impact on water quality (USEPA 2015g). As stormwater flows over a construction site, it can pick up pollutants like sediment, debris, and chemicals and transport these to a nearby storm sewer system or directly to a river, lake, or coastal water. Although construction activities are generally temporary at any given location, polluted runoff from construction sites can harm or kill fish and other wildlife. Sedimentation can destroy aquatic habitat, high volumes of runoff can cause stream bank erosion, and debris can clog waterways.

Potential pollutants associated with construction activities include sediment, suspended solids, nutrients, chemicals, petroleum products, fuel, fertilizers, pesticides, and pH modifying contaminants (e.g., bulk cement) (WA DOE 2014). The variety of pollutants present and the severity of their effects depend on the nature of the construction activity, the physical characteristics of the construction site, and the proximity of surface waters to the construction area.

Soil loss rates from construction sites can be 1,000 times the average of natural soil erosion rates and 20 times that from agricultural lands (Keener et al. 2007). Even with control measures, waters discharged from disturbed lands often contain higher than desired concentrations of suspended solids, particularly the finer particles (Przepiora, et al. 1998). Ehrhart et al. (2002) investigated the effects of sedimentation basin discharges on receiving streams at three construction sites, reporting that stream sediment concentrations increased significantly with high levels persisting for at least 100 m below the basin discharge. A two-year study of runoff from three residential construction sites in Wisconsin showed that pollutant loads (suspended solids and nutrients) from these sites are variable and site dependent (Daniel et al. 1979). Compared to an adjacent watershed in dairy agriculture, however, the annual yield of suspended solids from the construction sites was considerably higher (19.2 vs. < 1 metric ton/ha). Similar differences in total nutrient yields were also observed between the construction and agricultural sites.

The 10-year Jordan Creek (CT) NNPSMP project compared stormwater runoff from three urban watersheds using a paired-watershed design (Clausen 2007). The watersheds were: a developed watershed serving as the control, a watershed being developed using traditional practices and subdivision requirements, and a watershed developed using a BMP approach (e.g., alternative driveway pavement treatments). The volume of stormwater runoff from the BMP watershed decreased (-97%) during the construction period compared to the control watershed while stormwater runoff from the traditional watershed increased compared to the control watershed. The concentrations of total suspended solids (TSS), NO_3-N, NH_3-N, total Kjeldahl nitrogen (TKN), and TP increased during construction in the BMP watershed, with peaks associated with turfgrass development. Because of the decreased stormwater runoff volume, however, exports from the BMP watershed generally did not change during the construction period, except for TSS and TP which increased and Zn which decreased. In the traditional watershed, concentrations either did not change or, for TKN and TP, declined during construction. Because of the increased stormwater runoff volume, however, exports from the traditional site increased for all variables during construction despite the observation that the erosion and sediment controls used during construction appeared to work.

Chemical pollutants, such as paints, acids for cleaning masonry surfaces, cleaning solvents, asphalt products, soil additives used for stabilization, pollutants in wash water from concrete mixers, and concrete-curing compounds, can also be carried in runoff from construction sites. When eroded sediment is transported to nearby surface waters, it can carry with it fertilizers, pesticides, fuels, and other contaminants and substances that readily attach to soil particles (Keener et al. 2007). Pollutants attached to sediment from construction sites can become desorbed quickly and transported in their soluble form which is often more reactive and bioavailable to organisms (Faucette et al. 2009).

Petroleum products used during construction include fuels and lubricants for vehicles, power tools, and general equipment maintenance. Asphalt paving also can be harmful because it releases various oils for a considerable time period after application. Solid waste on construction sites includes trees and shrubs removed during land clearing and structure installation, wood and paper from packaging and building materials, scrap metal, sanitary wastes, rubber, plastic, glass, and masonry and asphalt products.

1.3.8 Marinas

Because marinas are located right at the water's edge, there is a high potential for marina waters to become contaminated with pollutants generated from the various activities that occur there, such as boat cleaning, fueling operations, and marine head discharge, or from the entry of stormwater runoff from parking lots and hull maintenance and repair areas into marina basins (USEPA 2015d). Chemicals used to maintain and repair boats, such as solvents, oils, paints, and cleansers, may spill into the water, or make their way into waterbodies via runoff (NOAA 2013). Spilling fuel (gasoline or oil) at marinas or discharging uncombusted fuels from engines also contribute to NPS pollution (McCoy and Johnson 2010). In addition, poorly maintained sanitary waste systems aboard boats or poorly maintained pump-out stations at marinas can significantly increase bacteria and nutrient levels in the water.

Studies have shown that boats can be a source of fecal coliform bacteria in estuaries with high boat densities and poor flushing (Fisher et al. 1987, Gaines and Solow 1990, Milliken and Lee 1990, NCDEM 1990, Sawyer and Golding 1990, Seabloom et al. 1989). Fecal coliform levels in marinas and mooring fields become most elevated during periods of high boat occupancy and usage, such as holiday weekends. In addition, DO levels in marina basins can be lowered by inadequate water circulation and the decomposition of organic materials from sources such as sewage and fish waste.

Both the construction and design of marina or port construction can negatively affect the ecology of an area; effects include loss of habitat and alterations to local hydrodynamics. Protective measures like bulkheads, breakwaters and jetties are built near marinas to prevent damage to boats and shoreline structures, but these structures can have unintended water quality impacts. Both the attenuation of waves by in-water structures and the creation of waves by the increased boat traffic in marinas and ports affect shoreline processes, often result in increased turbidity, resuspension of sediment-bound pollutants, and increased shoreline erosion (USFWS 1982).

Metals and metal-containing compounds are contained in fuel additives, antifouling paints, ballast, and other marina structures. Arsenic is used in paint pigments, pesticides, and wood preservatives. Zn anodes are used to deter corrosion of metal hulls and engine parts (McCoy and Johnson 2010). Cu and tin (Sn) are used as biocides in antifoulant paints (McCoy and Johnson 2010). Other metals (Fe, chrome, etc.) are used in the construction of marinas and boats. These metals are released to marina waters through spillage, incomplete fuel combustion, wear on boat hulls and marina structures, and boat bilge discharges (McCoy and Johnson 2010, NCDEM 1990). Elevated levels of Cu, Zn, Cd, Cr, Pb, Sn, and PCBs have been found in oysters, other bivalves, and algae in some marinas (CARWQCB 1989, Marcus and Stokes 1985, McMahon 1989, NCDEM 1990, Nixon et al. 1973, SCDHEC 1987, Wendt et al. 1990, Young et al. 1979).

1.4 Solving the Problem

A wide range of federal, state, and local efforts with varying objectives, methods, and resources have been employed over the past few decades to address NPS problems at the local to national levels. A program central to many of these efforts is EPA's NPS program authorized under Section 319 of the CWA. Under this program, states, territories and tribes receive grant money that supports a wide variety of activities including technical assistance, financial assistance, education, training, technology transfer, demonstration projects and monitoring to assess the success of specific NPS implementation projects. Federal funds allocated under Section 319(h) of the CWA are distributed based on a state-by-state allocation formula to implement approved nonpoint source management programs. Section 319 funding grew from its initial funding level of $37 million in FY1990 to $238.5 million in FY2003, dropping back to $164.5 million in FY2012. Additional information on Section 319, including success stories, is available at EPA's website.

While the Section 319 program is a very important part of efforts to solve the NPS problem, there are numerous other programs and activities that are carried out in conjunction with or separate from Section 319 to address various aspects of the problem. Information about state programs can be found at EPA's NPS program website. Other examples include:

- The new Urban Waters Federal Partnership was designed to reconnect urban communities with their waterways by improving coordination among federal agencies and collaborating with community-led revitalization efforts to improve our nation's water systems and promote their economic, environmental and social benefits (USEPA 2015h). Stormwater runoff is one of several sources of pollution in urban settings creating public and environmental health hazards such as lowered drinking water quality and water bodies that are unsafe for swimming.
- The U.S. Department of Agriculture (USDA) Natural Resources Conservation Service's (NRCS) Environmental Quality Incentives Program (EQIP) is a voluntary program that provides financial and technical assistance to agricultural producers through contracts up to a maximum term of 10 years in length. These contracts provide financial assistance to help plan and implement conservation practices that address natural resource concerns and for opportunities to improve soil, water, plant, animal, air and related resources on agricultural land and non-industrial private forestland. In addition, a purpose of EQIP is to help producers meet federal, state, tribal and local environmental regulations.
- Under the USDA's National Water Quality Initiative, the NRCS works with farmers and ranchers in small watersheds throughout the nation to improve water quality where this is a critical concern. In 2013, NRCS will provide nearly $35 million in financial assistance to help farmers and ranchers implement conservation systems to reduce N, P, sediment and pathogen contributions from agricultural land. This is the second year of the initiative; NRCS provided $34 million in 2012.
- Efforts that help define the problem also support NPS programs. For example, in 2011, numeric nutrient water quality standards were established for lakes and flowing waters in Florida to address harmful algal blooms caused by excess nutrients from fertilizer, stormwater, and wastewater runoff (FLDEP 2015).
- Hundreds of local projects across the nation are addressing various NPS problems. For example, alum (aluminum sulfate) treatments and upland nutrient management practices have been employed in the Grand Lake St. Marys watershed in western Ohio to address hypereutrophic conditions caused by high inflows of P (Tetra Tech 2013).

1.5 References

Alexander, R.B., R.A. Smith, G.E. Schwarz, E.W. Boyer, J.V. Nolan, and J.W. Brakebill. 2008. Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin. *Environmental Science and Technology* 42: 822-830.

Beltran, B.J., D.M. Amatya, M. Youssef, M. Jones, T.J. Callahan, R.W. Skaggs, and J.E. Nettles. 2010. Impacts of fertilization on water quality of a drained pine plantation: a worst case scenario. *Journal of Environmental Quality* 39:293–303

Binkley, D., H. Burnhamb, and H.L. Allen. 1999. Water quality impacts of forest fertilization with nitrogen and phosphorus. *Forest Ecology and Management* 121:191-213.

Brinson, M.M., B.L. Swift, R.C. Plantico, and J.S. Barclay. 1981. *Riparian Ecosystems: Their Ecology and Status*. FWS/OBS-81/17. U.S. Fish and Wildlife Service, National Water Resources Analysis Group, Eastern Energy and Land Use Team, Kearneysville, WV.

Brookes, A. 1990. Restoration and enhancement of engineered river channels: some European experiences. *Regulated Rivers: Research and Management* 5:45-56.

Brown, R.G. 1985. Effects of an urban wetland on sediment and nutrient loads in runoff. *Wetlands* 4:147-158.

Buttle, J.M., and F. Xu. 1988. Snowmelt runoff in suburban environments. *Nordic Hydrology* 19:19-40.

CARWQCB (California Regional Water Quality Control Board). 1989. *Nonpoint Source Experience in the Santa Clara Valley*. California Regional Water Quality Control Board, Santa Clara Valley Water District, Santa Clara, CA.

Clausen, J.C. 2007. *Jordan Cove Watershed Project Final Report*. University of Connecticut, College of Agriculture and Natural Resources, Department of Natural Resources Management and Engineering. Accessed January 8, 2016. http://jordancove.uconn.edu/jordan_cove/publications/final_report.pdf.

Cravotta III, C.A., R.A. Brightbill, and M.J. Langland. 2010. Abandoned mine drainage in the Swatara Creek Basin, Southern Anthracite coalfield, Pennsylvania, USA: 1. stream water quality trends coinciding with the return of fish. *USGS Staff -- Published Research*. Accessed January 8, 2016.

CWP (Center for Watershed Protection). 2003. *Impacts ofImpervious Cover on Aquatic Systems*. Center for Watershed Protection, Ellicott City, MD.

Daniel, T.C., P.E. McGuire, D. Stoffel, and B. Miller. 1979. Sediment and nutrient yield from residential construction sites. *Journal of Environmental Quality* 8(3):304-308.

Demchak, J., J. Skousen, and L.M. McDonald. 2004. Longevity of acid discharges from underground mines located above the regional water table. *Journal of Environmental Quality* 33:656–668.

Ehrhart, B.J., R.D. Shannon, and A.R. Jarrett . 2002. Effects of construction site sedimentation basins on receiving stream ecosystems. *Transactions of the American Society of Agriculture and Biological Engineers* 45(3):675–680.

Faucette, L.B., F.A. Cardoso-Gendreau, E. Codling, A.M. Sadeghi, Y.A. Pachepsky, and D.R. Shelton. 2009. Storm water pollutant removal performance of compost filter socks. *Journal of Environmental Quality* 38:1233–1239.

Fisher, J.S., R.R. Perdue, M.F. Overton, M.D. Sobsey, and B.L. Sill. 1987. *A Comparison of Water Quality at Two Recreational Marinas During a Peak-Use Period*. North Carolina State University, Raleigh, NC. Prepared for the National Oceanic and Atmospheric Administration.

FLDEP (Florida Department of Environmental Protection). 2015. *Numeric Nutrient Standards for Florida Waters*. Florida Department of Environmental Protection, Tallahassee, FL. Accessed January 11, 2016. http://www.dep.state.fl.us/water/wqssp/nutrients/.

Friend, A.L., S.D. Roberts, S.H. Schoenholtz, J.A. Mobley, and P.D. Gerard. 2006. Poultry litter application to loblolly pine forests: growth and nutrient containment. *Journal of Environmental Quality* 35:837–848.

Gaines, A.G., Jr., and A.R. Solow. 1990. *The Distribution of Fecal Coliform Bacteria in Surface Waters of the Edgartown Harbor Coastal Complex and Management Implications*. Edgartown Harbor Project, Interim Report No. 5. Edgartown Harbor Association, Inc.

Goolsby, D.A., W.A. Battaglin, G.B. Lawrence, R.S. Artz, B.T. Aulenbach, R.P. Hooper, D.R. Keeney, and G.J. Stensland. 1999. *Flux and Sources of Nutrients in the Mississippi-Atchafalaya River Basin: Topic 3 Report for the Integrated Assessment on Hypoxia in the Gulf of Mexico*. Decision Analysis Series No. 17. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Coastal Ocean Program, Silver Spring, MD.

Goolsby, D.A., W.A. Battaglin, B.T. Aulenbach, and R.P. Hooper. 2001. Nitrogen input to the Gulf of Mexico. *Journal of Environmental Quality* 30:329-336.

Grey, M. and C. Henry. 2002. Phosphorus and nitrogen runoff from a forested watershed fertilized with biosolids. *Journal of Environmental Quality* 31:926–936.

Horowitz, A.J., K.A. Elrick, J.A. Robbins, and R.B. Cook. 1993. *The Effect of Mining and Related Activities on the Sediment Trace Element Geochemistry of Lake Coeur D'Alene, Idaho, USA Part II: Subsurface Sediments*. Open File Report 93 656. U.S. Geological Survey, Denver, CO.

Keehner, D. 2011. Memorandum: information concerning 2012 Clean Water Act Sections 303(d), 305(b), and 314 integrated reporting and listing decisions. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 8, 2016. http://www.epa.gov/sites/production/files/2015-10/documents/final_2012_memo_document.pdf.

Keener, H.M., B. Faucette, and M.H. Klingman. 2007. Flow-through rates and evaluation of solids separation of compost filter socks versus silt fence in sediment control applications. *Journal of Environmental Quality* 36:742–752.

Knutson, K.L. and V.L. Naef. 1997. *Management Recommendations for Washington's Priority Habitats: Riparian*. Washington Department of Fish and Wildlife, Olympia, WA. Accessed January 8, 2016. http://wdfw.wa.gov/publications/00029/wdfw00029.pdf.

Lee, G., J.M. Bigham, and D.J. Williams. 2008. Metal release from bottom sediments of Ocoee Lake No. 3, a primary catchment area for the Ducktown Mining District. *Journal of Environmental Quality* 37:344–352.

Marcus, J.M., and T.P. Stokes. 1985. Polynuclear aromatic hydrocarbons in oyster tissue and around three coastal marinas. *Bulletin of Environmental Contamination and Toxicology* 35:833-844.

McCoy, J.A. and L.T. Johnson. 2010. *Boating Pollution Economics & Impacts*. UCCE-SD Fact Sheet 2010-2. University of California, Cooperative Extension, County of San Diego. Accessed January 8, 2016. http://ucanr.org/sites/coast/files/59476.pdf.

McMahon, P.J.T. 1989. The impact of marinas on water quality. *Water Science and Technology* 21(2):39-43.

Michael, J.L. 2003. Environmental fate and impacts of sulfometuron on watersheds in the southern United States. *Journal of Environmental Quality* 32:456–465

Milliken, A., and V. Lee. 1990. *Pollution Impacts from Recreational Boating: a Bibliography and Summary Review*. Rhode Island Sea Grant, Narragansett, RI.

MSD (Metropolitan Sewer District). 2012. *Problems Facing Urban Streams*. Metropolitan Sewer District, Louisville, KY. Accessed January 8, 2016. http://www.msdlouky.org/insidemsd/wqurban.htm.

NCDEM (North Carolina Division of Environmental Management). 1990. *North Carolina Coastal Marinas: Water Quality Assessment.* North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC.

Nixon, S.W., C.A. Oviatt, and S.L. Northby. 1973. *Ecology of Small Boat Marinas*. Rhode Island Sea Grant, Narragansett, RI.

NOAA (National Oceanic and Atmospheric Administration). 2012. *Pollutants from Nonpoint Sources: Nutrients*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, NOAA Ocean Service Education. Accessed January 11, 2016. http://oceanservice.noaa.gov/education/kits/pollution/010nutrients.html.

NOAA (National Oceanic and Atmospheric Administration). 2013. *Nonpoint Source Pollution: Marinas and Boating Activities*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Ocean Service Education. Accessed January 8, 2016. http://oceanservice.noaa.gov/education/kits/pollution/09activities.html.

Oak Ridge National Laboratory. 1993. *Public Health Assessment for Carson River Mercury Site, Moundhouse, Lyon County, Nevada. Preliminary Report*. Oak Ridge National Laboratory, Oak Ridge, TN.

Ozawa, C.P. and J.A. Yeakley. 2007. Performance of management strategies in the protection of riparian vegetation in three Oregon cities. *Journal of Environmental Planning and Management* 50(6): 803-822.

Pitt, R., and J. McLean. 1992. Stormwater, Baseflow, and Snowmelt Pollutant Contributions from an Industrial Area. In *Water Environment Federation 65th Annual Conference & Exposition, Surface Water Quality & Ecology Symposia*, New Orleans, Louisiana, September 20-24, Volume VII.

Przepiora, A., D. Hesterberg, J.E. Parsons, J.W. Gilliam, D.K. Cassel, and W. Faircloth. 1998. Field evaluation of calcium sulfate as a chemical flocculant for sedimentation basins. *Journal of Environmental Quality* 27:669-678.

Rathbun, J. 2007. *Quality Assurance Project Plan for the Central Mine Site Stamp Sand Remediation Project – Version 1; August 8, 2007*. Michigan Department of Environmental Quality, Water Bureau – Nonpoint Source Unit, Lansing, MI.

Roy, A.H., M.J. Paul, and S.J. Wenger. 2010. Urban stream ecology. In *Urban Ecosystem Ecology*, Agronomy Monographs 55, ed. J. Aitkenhead-Peterson and A. Volder. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Ryan, S.E. and G.E. Grant. 1991. Downstream effects of timber harvesting on channel morphology in Elk River Basin, Oregon. *Journal of Environmental Quality* 20:60-72.

Sawyer, C.M., and A.F. Golding. 1990. Marina pollution abatement. In *1990 Environmental Management for Marinas Conference*. International Marina Institute, Wickford, RI.

SCDHEC (South Carolina Department of Health and Environmental Control). 1987. *Heavy Metals and Extractable Organic Chemicals from the Coastal Toxics Monitoring Network 1984-1986*. Technical Report No. 007-87. South Carolina Department of Health and Environmental Control, Columbia, SC.

Schoonover, J.E., B. Graeme Lockaby, and B.S. Helms. 2006. Impacts of land cover on stream hydrology in the West Georgia Piedmont, USA. *Journal of Environmental Quality* 35:2123–2131.

Seabloom, R.A., G. Plews, F. Cox, and F. Kramer. 1989. *The Effect of Sewage Discharges from Pleasure Craft on Puget Sound Waters and Shellfish Quality*. Washington State Department of Health, Shellfish Section, Olympia, WA.

Sherwood, C.R., D.A. Jay, R.B. Harvey, P. Hamilton, and C.A. Simenstad. 1990. Historical changes in the Columbia River Estuary. *Progress in Oceanography* 25:299-352.

Sidle, R.C. and A.Sharma 1996. Stream channel changes associated with mining and grazing in the Great Basin. *Journal of Environmental Quality* 25:1111-1121.

Steele, M.K., W.H. McDowell, and J.A. Aitkenhead-Peterson. 2010. Chemistry of urban, suburban, and rural surface waters. In *Urban Ecosystem Ecology*, Agronomy Monographs 55, ed. J. Aitkenhead-Peterson and A. Volder. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Stewart, J. and J. Skousen 2003. Water quality changes in a polluted stream over a twenty-five-year period. *Journal of Environmental Quality* 32:654–661.

Tetra Tech. 2013. *Preliminary Assessment of Effectiveness of the 2012 Alum Application—Grand Lake St. Marys*. Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. http://www.lakeimprovement.com/sites/default/files/GLSM%20Alum%20Report%2002202013.pdf.

USDA-NASS (U.S. Department of Agriculture-National Agricultural Statistics Service). 2014. *2012 Census Of Agriculture - United States Summary and State Data.* Geographic Area Series, Volume 1, Part 51, AC-12-A-51. U.S. Department of Agriculture, National Agricultural Statistics Service, Washington, DC. Accessed January 8, 2016. http://www.agcensus.usda.gov/Publications/2012/.

USDOI (U.S. Department of the Interior). 1991. *Riparian-Wetland Initiative for the 1990's*. BLM/WO/GI-91/001+4340. U.S. Department of the Interior, Bureau of Land Management, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1991. *Characteristics of Successful Riparian Restoration Projects in the Pacific Northwest*. EPA 910/9-91-033. U.S. Environmental Protection Agency, Region 10, Water Division, Nonpoint Sources Section, Seattle, WA.

USEPA (U.S. Environmental Protection Agency). 2005. *National Management Measures to Control Nonpoint Source Pollution from Urban Areas*. EPA-841-B-05-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 8, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/urban-runoff-national-management-measures.

USEPA (U.S. Environmental Protection Agency). 2007. *National Management Measures to Control Nonpoint Source Pollution from Hydromodification*. EPA 841-B-07-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 8, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/hydromodification-and-habitat-alteration-national.

USEPA (U.S. Environmental Protection Agency). 2009. *National Lakes Assessment: a Collaborative Survey of the Nation's Lakes*. EPA-841-R-09-001. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/national-aquatic-resource-surveys/national-lakes-assessment-2007-results.

USEPA (U.S. Environmental Protection Agency). 2010. *National Lakes Assessment Technical Appendix: Data Analysis Approach*. EPA-841-R-09-001a. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/national-aquatic-resource-surveys/national-lakes-assessment-2007-results.

USEPA (U.S. Environmental Protection Agency). 2012a. *National Coastal Condition Report IV*. EPA-842-R-10-003. U.S. Environmental Protection Agency, Office of Research and Development and Office of Water, Washington, DC. Accessed January 11, 2016. http://water.epa.gov/type/oceb/assessmonitor/nccr/upload/0_NCCR_4_Report_508_bookmarks.pdf.

USEPA (U.S. Environmental Protection Agency). 2012b. *NPDES Permit Writers' Manual for Concentrated Animal Feeding Operations*. EPA 833-F-12-001. U.S. Environmental Protection Agency, Office of Water, Office of Wastewater Management, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/npdes/npdes-permit-writers-manual-concentrated-animal-feeding-operations.

USEPA (U.S. Environmental Protection Agency). 2013. *National Rivers and Streams Assessment 2008–2009 a Collaborative Survey: Draft February 28, 2013*. EPA/841/D-13/001. U.S. Environmental Protection Agency, Office of Wetlands, Oceans and Watersheds and Office of Research and Development, Washington, DC. Accessed January 8, 2016. http://water.epa.gov/type/rsl/monitoring/riverssurvey/upload/NRSA0809_Report_Final_508Compliant_130228.pdf.

USEPA (U.S. Environmental Protection Agency). 2015a. *Forestry*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-forestry.

USEPA (U.S. Environmental Protection Agency). 2015b. *Impaired Waters and Stormwater.* U.S. Environmental Protection Agency. Washington, DC. Accessed January 11, 2016. http://www.epa.gov/tmdl/impaired-waters-and-stormwater.

USEPA (U.S. Environmental Protection Agency). 2015c. *National Wetland Condition Assessment 2011: a Collaborative Survey of the Nation's Wetlands-Draft for Public Review and Comment*. EPA-843-R-15-005. U.S. Environmental Protection Agency, Office of Water, Office of Research and Deveopment, Washington, DC. Accessed January 8, 2016. http://www.epa.gov/national-aquatic-resource-surveys/national-wetland-condition-assessment-2011-results.

USEPA (U.S. Environmental Protection Agency). 2015d. *Nonpoint Source: Marinas and Boating Overview*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-marinas-and-boating.

USEPA (U.S. Environmental Protection Agency). 2015e. *Nonpoint Source: Urban Areas*. U.S. Environmental Protection Agency, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-urban-areas.

USEPA (U.S. Environmental Protection Agency). 2015f. *Northern Gulf of Mexico Hypoxic Zone. Mississippi River/Gulf of Mexico Hypoxia Task Force.* U.S. Environmental Protection Agency, Washington, DC. Accessed January 11, 2016. http://water.epa.gov/type/watersheds/named/msbasin/zone.cfm.

USEPA (U.S. Environmental Protection Agency). 2015g. *Stormwater Discharges from Construction Activities*. U.S. Environmental Protection Agency, Office of Wastewater Management, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/npdes/stormwater-discharges-construction-activities#overview.

USEPA (U.S. Environmental Protection Agency). 2015h. *Urban Waters Federal Partnership*. U.S. Environmental Protection Agency, Washington, DC. Accessed January 11, 2016. http://www.epa.gov/urbanwaterspartners.

USEPA (U.S. Environmental Protection Agency). 2016. *National Summary of State Information*. U.S. Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans and Watersheds, Washington, DC. Accessed January 8, 2016. http://ofmpub.epa.gov/waters10/attains_nation_cy.control.

USEPA/USDOI (U.S. Environmental Protection Agency/U.S. Department of the Interior). 1995. *1995 Progress Report. Statement of Mutual Intent Strategic Plan for the Restoration and Protection of Streams and Watersheds Polluted by Acid Mine Drainage from Abandoned Coal Mines*. U.S. Environmental Protection Agency, Region 3 and U.S. Department of the Interior, Office of Surface Mining, Appalachian Regional Coordinating Center, Pittsburgh, PA.

USFWS (U.S. Fish and Wildlife Service). 1982. *Mitigation and Enhancement Techniques for the Upper Mississippi River System and Other Large River Systems*. Resource Publication 149. U.S. Department of the Interior, U.S. Fish and Wildlife Service, Washington, DC.

WA DOE (Washington State Department of Ecology). 2014. *2012 Stormwater Management Manual for Western Washington as Amended in December 2014*. Publication No. 14-10-055. Washington State Department of Ecology, Water Quality Program. Accessed April 21, 2016. https://fortress.wa.gov/ecy/publications/SummaryPages/1410055.html.

Wendt, P.H., R.F. van Dolah, M.Y. Bobo, and J.J. Manzi. 1990. *Effects of Marina Proximity on Certain Aspects of the Biology of Oysters and Other Benthic Macrofauna in a South Carolina Estuary*. Technical Report No. 74. South Carolina Marine Resources Center, Charleston, SC.

Young, D.R., G.V. Alexander, and D. McDermott-Ehrlich. 1979. Vessel-related contamination of southern California harbors by copper and other metals. *Marine Pollution Bulletin* 10:50-56.

This page intentionally left blank.

2 Nonpoint Source Monitoring Objectives and Basic Designs

By S.A. Dressing, D.W. Meals, J.B. Harcum, and J. Spooner

Water quality monitoring is performed to support a wide range of programs. National-level monitoring with continuous water-quality monitors is performed by the USGS, for example, to assess the quality of the Nation's surface waters (Wagner et al. 2006). Studies of large basins such as the Mississippi River Basin and Great Lakes Basin are designed to assess the general condition of waterbodies, track the health of fisheries, identify the causes and sources of designated beneficial use support impairments, and aid in the design of programs and projects to solve identified problems. EPA, states, and tribes conduct a series of surveys of the nation's aquatic resources that can also be used to track changes in condition over time. Each year the survey focuses on a different aquatic resource (i.e., rivers and streams, lakes, wetlands, or coastal waters) to yield unbiased estimates of the condition of the whole water resource being studied.[1] Monitoring of smaller watersheds is done for a number of purposes, including assessing local water quality problems, developing watershed plans to address current and prevent future problems, and educating the public about the water environment. Monitoring of individual practices or BMPs is typically carried out to determine the effectiveness of the particular practices, provide data for the development or validation of watershed modeling tools, document efforts to address watershed-scale problems, and inform stakeholders.

2.1 Monitoring Objectives

Monitoring is an information gathering exercise that is intended to generate data that serve management decision-making needs (USEPA 2003a). The formulation of clear monitoring objectives is an essential first step in developing an efficient and effective monitoring plan. Monitoring supports a range of water quality management objectives including establishing water quality standards, determining water quality status and trends, identifying impaired waters, identifying causes and sources of water quality problems, and evaluating program effectiveness. Specific objectives appropriate for NPS monitoring plans include:

- Identify water quality problems, use impairments, causes, and pollutant sources.
- Develop TMDLS and load or wasteload allocations.
- Analyze trends.
- Assess the effectiveness of BMPs or watershed projects.
- Assess permit compliance.
- Validate or calibrate models.
- Conduct research.

[1] http://water.epa.gov/type/watersheds/monitoring/aquaticsurvey_index.cfm

All monitoring programs should be designed to answer questions. The design process begins with identifying the problem and setting objectives that pose the questions. Then an experimental design appropriate to those objectives is selected and decisions related to sample type, sampling locations, which variables to monitor, and how to collect and analyze the samples that need to be made. Because the purpose of NPS monitoring is often to evaluate practice effectiveness and programs taking place on the land, land use and land management monitoring is an integral part of the overall effort. Finally, management and analysis of data gathered through the monitoring program must also be incorporated.

The specific steps essential in designing a monitoring program to meet objectives are discussed in detail in this chapter and chapter 3.

Overview of Steps in Monitoring Design

1. Identify problem(s)
2. Form objectives
3. Design experiment
4. Select scale
5. Select variables
6. Choose sample type
7. Locate stations
8. Determine sampling frequency
9. Design stations
10. Define collection/analysis methods
11. Define land use monitoring
12. Design data management

(USDA-NRCS 2003)

EPA and states need comprehensive water quality monitoring and assessment information to help set levels of protection in water quality standards. This information will also help to identify emerging problem areas or areas that need additional regulatory and non-regulatory actions to support water quality management decisions such as TMDLs, NPDES permit limits, enforcement, and NPS management (USEPA 2003a). Statewide monitoring to assess the degree to which designated beneficial uses (e.g., drinking, swimming, aquatic life) are supported is an essential component of efforts to achieve CWA goals. For example, CWA §106(e)(1) and 40 CFR Part 35.168(a) provide that EPA award Section 106 funds to a state only if the state has provided for or is carrying out the establishment and operation of appropriate devices, methods, systems and procedures necessary to monitor and to compile and analyze data on the quality of its navigable waters. States must also update the data in the Section 305(b) report.

In accordance with EPA water quality standards regulations, states designate uses for each waterbody or waterbody segment and establish criteria (e.g., dissolved oxygen [DO] levels, temperature, metals concentrations) that must be met through their water quality standards programs. Monitoring of the criteria parameters is then performed to assess whether the criteria are being met (USEPA 2003b). Biological monitoring of aquatic systems has been increasingly used to assess aquatic life use support.

Because of the magnitude of the task, states generally monitor portions of the state on a rotating basis (USEPA 2011). Ohio, for example, has a 15-year plan for monitoring all 8-digit HUCs in the state (OEPA 2013). Each year Ohio EPA collects data from streams and rivers in five to seven different areas of the state. A total of 400 to 450 sampling sites are examined, and each site is visited more than once per year. During these studies, Ohio EPA technicians collect chemical samples, examine and count fish and aquatic insects, and take measurements of the stream. There are three major objectives for the studies:

- Determine how the stream is doing compared to goals assigned in the Ohio Water Quality Standards;
- Determine if the goals assigned to the river or stream are appropriate and attainable; and
- Determine if the stream's condition has changed since the last time the stream was monitored.

The findings of each biological and water quality study can be used in Ohio EPA regulatory actions. The results are incorporated into Water Quality Permit Support Documents, State Water Quality Management Plans, and the Ohio Nonpoint Source Assessment. This information also provides the basis for the Integrated Water Quality Monitoring and Assessment Report – a biennial statewide report on the condition of Ohio's waters and the list of impaired and threatened waters required by sections 303(d) and 305(b) of the Clean Water Act.

The growing linkage between the national NPS and TMDL programs has resulted in a greater emphasis on estimating pollutant loads (USEPA 2003b). Specific considerations and recommendations for monitoring to estimate pollutant loads are presented in section 3.8. Development of load-duration curves has become an important step in developing TMDLs for many watersheds, and this topic is addressed in detail in section 8.9.4.

Well-formulated monitoring objectives drive the rest of the monitoring study design and are critical to a successful monitoring project (USDA-NRCS 2003). It is also important that monitoring objectives are directly linked to overall program or project objectives that depend on the monitoring data. Table 2-1 illustrates this important linkage between program and monitoring objectives.

Table 2-1. Complementary program and monitoring objectives

Program Objective	Complementary Monitoring Objective
Reduce annual P loading to lake by at least 15 percent in 5 years with nutrient management	Measure changes in annual P loading to lake and link to management actions
Reduce *E. coli* load to stream to meet water quality standards in 3 years	Measure changes in compliance with water quality standard for *E. coli*

A good example to illustrate the development of program goals, supporting monitoring goals, and specific monitoring designs is the USGS National Water-Quality Assessment Program (NAWQA). NAWQA was designed to assess the status and trends in the nation's ground- and surface-water resources and to link the status and trends with an understanding of the natural and human factors that affect water quality (Gilliom et al., 1995). The study design balanced the unique assessment requirements of individual hydrologic systems with a nationally consistent design structure that incorporated a multi-scale, interdisciplinary approach. The Occurrence and Distribution Assessment, the most important component of the first intensive study phase in each of the NAWQA study units, was intended to characterize, in a nationally consistent manner, the broad-scale geographic and seasonal distributions of water-quality conditions in relation to major contaminant sources and background conditions. General goals for study-units were:

- Identify the occurrence of water quality conditions that are significant issues.
- Characterize the broad-scale geographic and seasonal distribution of a wide range of water quality conditions in relation to natural factors and human activities.
- Evaluate study-unit priorities and required study approaches for effectively assessing long-term trends and changes.
- Evaluate geographical and seasonal distribution in greater detail and in relation to the sources, transport, fate and effects of contaminants for water quality conditions of greatest importance. Determine the priorities and design for follow-up case studies.

USGS developed a study design to address each of the above goals for stream monitoring, employing three interrelated components: water-column, bed sediment and tissue, and ecological studies. Ecological studies incorporate three strategies:

- Fixed-Site Reach Assessments provide nationally consistent ecological information at water column sites as part of an integrated physical, chemical, and biological assessment of water quality.
- Intensive Ecological Assessments assess spatial and temporal variability associated with biological communities and habitat characteristics.
- Ecological Synoptic Studies provide improved spatial resolution of selected ecological characteristics in relation to land uses, contaminant sources, and habitat conditions.

While the NAWQA program is an exceptionally large monitoring program in both scope and scale, the approach of linking monitoring to program goals and developing clear and specific monitoring goals and objectives to drive monitoring program design is applicable to most if not all monitoring efforts.

2.2 Fundamentals of Good Monitoring

Water quality monitoring is a complex and demanding enterprise. Conducted well, monitoring can provide fundamental information about the water resource and its impairments. Monitoring data can allow managers to document changes through time, show response to NPS pollution reduction practices and programs, and confirm achievement of management objectives. Conducted poorly, monitoring can fail to meet objectives, create confusion, leave critical questions unanswered and waste time and money. It is essential that monitoring be designed to meet project and program objectives efficiently. The purpose of this section is to present key elements of good monitoring design and execution.

2.2.1 Understand the System

When little is known about a watershed, monitoring may be used to assess the problem. For this purpose, monitoring requires a fairly general approach, e.g., reconnaissance or synoptic sampling (see section 2.4.2.1). When designing a monitoring program to assess a response to nonpoint source control programs, a thorough understanding of the system (e.g., a farm, an urban catchment, or a rural watershed) is required. An early step in watershed planning and management is characterization of the watershed, which includes topography, geology, climate, soils, hydrology, biota, land use, infrastructure, and cultural resources (USEPA 2008b). This is essential information about the system and will assist in the design of the monitoring system. The first of EPA's Nine Key Elements of watershed restoration plans requires identification of the causes of impairment and the sources of pollutants that will need to be controlled to achieve the goals of the watershed plan (USEPA 2008b). Basic questions that should be addressed when characterizing a watershed include:

- What are the critical water quality impairments or threats?
- What are the key pollutants involved?
- What are the sources of these pollutants?
- How are pollutants transported through the watershed?
- What are the most important drivers of pollutant generation and delivery?

2.2.1.1 Causes and Sources

Decisions on where, when, and how often to sample, what to measure, and other elements of monitoring design depend on knowledge of the watershed being monitored. The source of pollutants is an obvious issue. For example, suspended sediment exported from a watershed may arise from upland erosion from agricultural fields, streambank and streambed erosion, or a combination of both. Knowledge of the source of suspended sediment measured at the watershed outlet is essential to designing a system to monitor watershed response to implementation of erosion control measures (as well as in selection of the appropriate control measures). In Minnesota, for example, Sekely et al. (2002) estimated that streambank sources accounted for 31 percent to 44 percent of the suspended sediment load at the mouth of the Blue Earth River. Allmendinger et al. (2007) reported that upland sediment production and sediment from enlargement of stream channels were approximately equal sources of sediment yield in an urbanizing watershed in Maryland. Unusually high mean NO_3-N concentrations (about 5 mg/l in the major tributaries) in watersheds that drain into the western and central basins of Lake Erie are thought to reflect the extensive use of tile drainage systems in the region (Baker 1988).

Indicator bacteria like fecal coliform or *E. coli* commonly cause impairment of recreation and shellfish harvesting in U.S. waters. These organisms can arise from numerous sources in a watershed, including wildlife, livestock, pets, and human wastes. It is essential to know the source(s) of these organisms both to apply the right control measures and to monitor response to control programs. For example, in the Oak Creek Canyon (AZ) NNPSMP project seasonal increases in fecal indicator bacteria were initially believed to originate primarily from poor sanitation practices by recreational users and a lack of adequate restroom facilities (Donald et al. 1998). Bacteriological water quality failed to improve after sanitation BMPs were installed, however, and analysis of fecal coliform levels in sediments versus those in the water column led to the conclusion that resuspension of contaminated sediments by recreational users and major storm events was the major cause of water quality violations. While sediment was identified as the major reservoir of bacteria, this analysis fell short of identifying the primary source(s) of the bacteria found in the sediment. DNA fingerprinting was used to determine the relative contributions of human, livestock, and wildlife sources, resulting in the finding that wildlife contributed a greater share of fecal pollution than humans. Still, at the conclusion of the NNPSMP project questions remained regarding the major sources of fecal contamination in the watershed (Donald et al. 1998, Spooner et al. 2011). A subsequent study by Poff and Tecle (2002) suggested that domesticated and wild animals, residential homes (septic systems), and business establishments were probably greater sources of E. coli than recreational visitors. Monitoring and BMP implementation have continued in the Oak Creek watershed, with establishment of a TMDL for *E. coli* in 2010 and completion of a watershed implementation plan in 2012 (OCWC 2012). Implementation efforts are now focused on education and outreach as the top priority, followed by septic systems, stormwater, recreation, and agriculture (OCWC 2012). Continued uncertainty regarding sources of fecal pollution is reflected by plan recommendations for additional monitoring of Oak Creek sediment to identify *E. coli* sediment reservoir hot spots and locate up-gradient sources of *E. coli*.

2.2.1.2 Pollutant Transport

It is absolutely critical to understand the mode of pollutant transport through the watershed from source to receiving water before setting up a monitoring system. Particulate pollutants, such as sediment and attached substances, generally move in surface waters. Monitoring for sediment or particulate phosphorus is generally best focused on surface runoff and streamflow. Dissolved pollutants, like nitrate-nitrogen (NO_3-N), are transported in both surface and ground waters. The relative importance of these distinct pathways needs to be understood to decide where and when to sample. For example in the Chesapeake Bay watershed, it is estimated that 40 percent of the annual N load to the Bay is delivered by groundwater

(STAC 2005). As much as 80 percent of annual export of NO_3-N, sulfate (SO_4^{-2}), and chloride (Cl^{-1}) from small Iowa watersheds occurred in baseflow (Schilling 2002).

In many cases, additional details regarding pollutant pathways must be understood to fine tune monitoring plans. For example, decisions on whether to focus on high-flow events (e.g., for particulate pollutants delivered episodically in surface runoff or storm flows) or base flows (e.g., for dissolved pollutants that tend to be delivered continuously via ground water) require an understanding of how pollutants move through the system. The Warner Creek (MD) NNPSMP discovered that base flow contributed 76 percent of total streamflow and that subsurface flow, including a substantial portion from outside the surface watershed boundaries, was the major pathway for transport of NO_3-N to the stream (Shirmohammadi et al. 1997, Shirmohammadi and Montas 2004). This complex hydrology contributed to the failure of an above/below monitoring design for this project.

The timing of sampling during storm events can also be informed by knowledge of pollutant pathways. For example, analysis of long-term data collected in the Lake Erie basin showed that peak concentrations of particulate and soluble pollutants occurred during different parts of the storm hydrograph (Baker et al. 1985, Baker 1988). TP and sediment concentrations reached their peak early in the runoff event before peak discharge, and decreased faster than the discharge decreased. The TP concentration did not decline as rapidly as the sediment concentration due to the presence of soluble P and an increasing ratio of particulate P to sediment as the sediment concentration decreased. The atrazine concentration pattern followed the hydrograph very closely, indicating movement off the fields with surface runoff water. Nitrate (NO_3) increased during the falling limb of the hydrograph due to tile drainage and interflow.

2.2.1.3 Seasonality

Seasonal patterns are often critical factors in monitoring design because NPS pollution is highly weather-driven. In northern regions, snowmelt and spring rains are often the dominant hydrologic feature of the annual cycle and a majority of the annual pollutant load may be delivered in a few weeks. A seven-year study on corn-cropped watersheds in southwestern Iowa, for example, showed that most of the average annual total N and P losses occurred during the fertilizer application, seedbed preparation, and crop establishment period from April through June (Alberts et al. 1978). February accounted for 23 percent of the total P load in a two-year study in the Clear Lake watershed in Iowa, indicating that the snowmelt period is a time of significant P loss from fields (Klatt et al. 2003). In Wisconsin, Stuntebeck et al. (2008) stated that it was critical that agricultural NPS monitoring take place year-round to fully characterize sediment and nutrient losses throughout the year rather than just during the growing season.

For herbicides such as atrazine, losses from agricultural fields in humid areas are highly episodic, with the majority of annual losses occurring in transient storm events soon after herbicide application. In a comparative study of agricultural watersheds in different climatic regions, Domagalski et al. (2007) found that stormwater runoff after application was the primary determinant of pesticide loads in humid environments. A significant portion of the load of some pesticide degradation products, however, can be transported under base-flow conditions in humid environments. In such cases, a monitoring effort would need to reliably monitor short, intense and unpredictable events during specific seasons, depending on both seasonal and agronomic factors. Sampling of base flow would be needed to track degradation products. The same comparative study, however, found that irrigation practices and timing of chemical use greatly affected pesticide transport in the semiarid basins, suggesting a monitoring effort that would need to be focused on irrigation events in these regions.

The importance of characterizing seasonality depends on the specific program and monitoring objectives. In cases where available water quality data are not sufficient to assess seasonality in a specific watershed, it may be necessary to perform seasonal synoptic surveys (see section 2.4.2.1), collect year-round samples initially, or rely on watershed modeling to better define seasonality and facilitate fine-tuning of the monitoring design.

2.2.1.4 Water Resource Considerations

Each type of water resource—rivers and streams; lakes, reservoirs, and ponds; wetlands; estuaries; coastal shoreline waters; and ground water—possesses unique hydrologic and biological features that must be considered in the design of a monitoring program. All water resource types exhibit both temporal (long- and short-term) and spatial (small- and large-scale) variability. For example, suspended sediment concentrations vary with depth and location in reservoirs; salinity concentrations in estuaries vary vertically and horizontally, as well as temporally as they are affected by relatively light fresh water flowing over heavier salt water; and ground water quality varies with soil and aquifer type and geozone. Placement of monitoring stations and the timing and duration of sampling are affected by consideration of these and other sources of variability.

2.2.1.4.1 Rivers and Streams

Streams can be classified at various levels of detail using a number of criteria (e.g., Montgomery and Buffington 1997, Rosgen and Silvey 1996), but streams can also be lumped for monitoring considerations into two major groups – intermittent and perennial – based simply on general flow characteristics. Clearly, water quality sampling cannot be conducted in intermittent streams when they do not have flow; however, the ability to measure and sample intermittent flows when they do occur is often critical and usually challenging. Year-to-year variations in precipitation can have major impacts on flow duration and frequency, pollutant loads, and water quality in intermittent streams. Flow in perennial streams and rivers is also affected by seasonal rainfall and snowfall patterns, reservoir discharge management, and irrigation management. Pollutant loads and concentrations, in turn, are affected by these patterns as the highest concentrations of suspended sediment and nutrients often occur during spring runoff, winter thaws, or intense rainstorms.

Because good flow measurement is essential to estimating pollutant loads, it is also important to understand spatial flow patterns in the monitored stream or river. Water velocities vary both horizontally (e.g., outside vs. inside meander bends) and vertically with depth (USDA-NRCS 2003, Meals and Dressing 2008). In addition, the complexity of currents at obstructions and points of constriction (e.g., bridges) makes them poor monitoring sites (Meals and Dressing 2008). Rudimentary stream classification can be very helpful in predicting a river's behavior from its appearance, which, in turn, can be useful in identifying locations for fixed sampling stations. Flow measurement is discussed in greater detail in section 3.1.3.1.

Flow patterns often play a significant role in determining the variability of water quality, both within a stream cross-section and throughout a stream reach. Figure 2-1 illustrates the relationship between pollutant concentrations and the vertical variability of stream velocity (Brakensiek et al. 1979). The effects of tributary flows must be considered when designing a stream or river monitoring program. Such flows can add pollutant loads, dilute pollutant loads, and create horizontal gradients. In some cases mixing below tributary junctions might be incomplete, with tributary flow primarily following one bank or forming spatially and temporally persistent plumes or bands (Sommer et al. 2008). If a representative sample of a river is required, it is important to select a sampling point where the flow is uniform and well-mixed, without sharp flow variations or distinct tributary inflow plumes. If more detail is required, segmentation of a stream into fairly homogeneous segments prior to monitoring might be necessary, with one to several monitoring stations located in each segment (Coffey et al. 1993). When dividing a stream into homogeneous segments, both land use and drainage area should be considered because both affect the quantity and quality of flows.

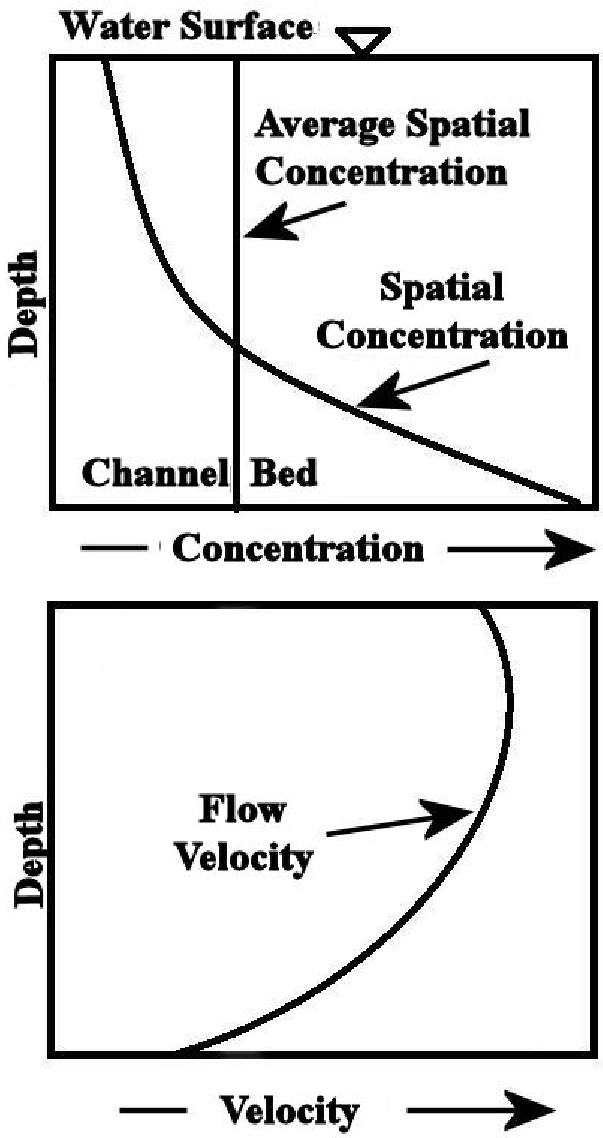


Figure 2-1. Vertical sediment concentration and flow velocity distribution in a typical stream cross section (after Brakensiek et al. 1979)

Vertical variability is particularly important during runoff events and in slow-moving streams because suspended solids, dissolved oxygen (DO), and algal productivity can vary substantially with depth (Figure 2-2) (Brakensiek et al. 1979). Levels of contaminants in bed sediment also vary horizontally and vertically, as deposition and scouring are strongly influenced by water velocity.

Biological communities in stream systems vary with a number of factors including landscape position, type of substrate, light, water temperature, current velocity, and amount and type of aquatic and riparian vegetation. Monitoring of aquatic communities is discussed in chapter 4.

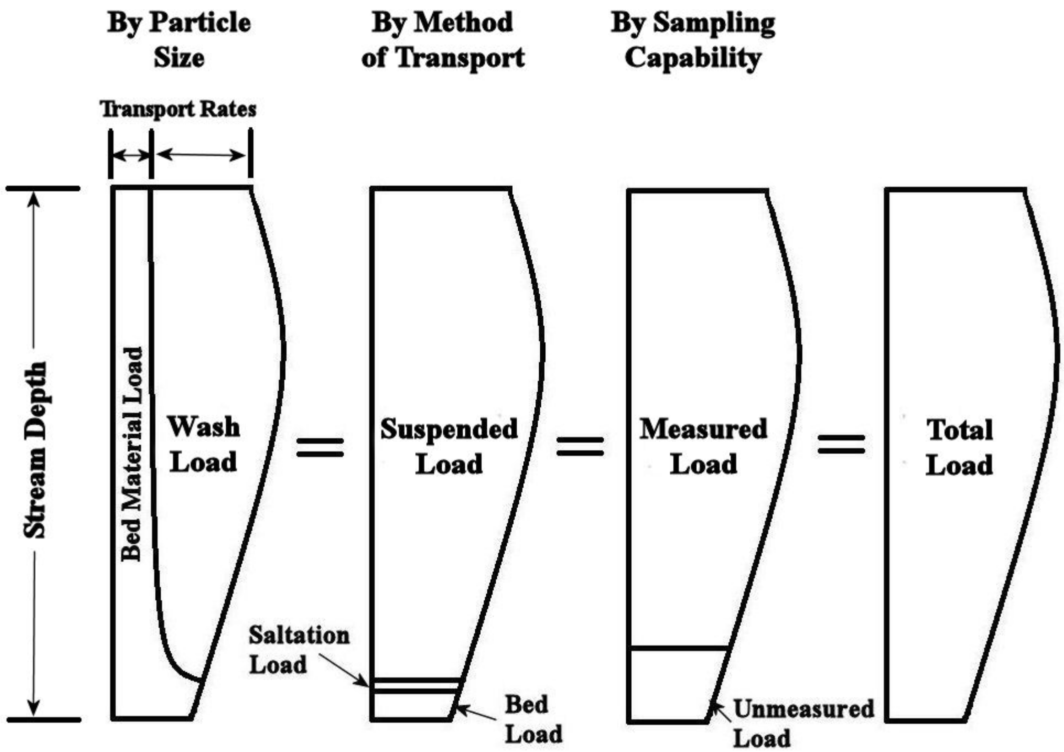


Figure 2-2. Schematic diagram of stream vertical showing position of sediment load terms (after Brakensiek et al. 1979)

2.2.1.4.2 Lakes, Reservoirs, and Ponds

Lakes are defined here as natural standing or slow-moving bodies of water. Reservoirs are considered to be human-made lakes typically created by impounding a river or stream. Ponds can be either natural or human-made, and are generally much smaller and shallower than lakes. The following discussion focuses primarily on lakes and reservoirs with lake referring to both types of water bodies.

Lakes are more than simple bowls of water. The physical, chemical, and biological characteristics of lakes vary horizontally, vertically, seasonally, and throughout the day. In addition, reservoirs can exhibit characteristics of both rivers and lakes, with the upstream section more river-like and downstream areas near the dam more lake-like. The balance between river and lake characteristics can vary widely among reservoirs with some more river-lake throughout. This variability must be understood and considered when designing a lake monitoring program.

Hydrology and geomorphology are strong determinants of the physical, chemical, and biological characteristics of lakes (Wetzel 1975). Lakes can be classified based on how water enters and exits the lake: seepage lakes, spring lakes, groundwater drained lakes, drainage lakes, and impoundments (WAL 2009). Knowledge of the primary sources of water and the presence or absence of inlets and outlets is essential to determining options for an effective monitoring plan.

Lake shape has major implications for monitoring design. Lakes and ponds with simple, rounded shapes may tend to be well-mixed at most times and might require only a single sampling station to provide an accurate representation of water quality. Lakes with complex interconnected basins or with dendritic shapes like reservoirs tend to exhibit significant spatial variability as mixing is inhibited; such lakes may require numerous sampling stations to represent more uneven water quality characteristics (USEPA

1990). In Lake Champlain (VT-NY-Quebec) for example, the lake's complex geometry of bays, islands, and bathymetry generally divide the lake into five distinct regions for monitoring (Figure 2-3).

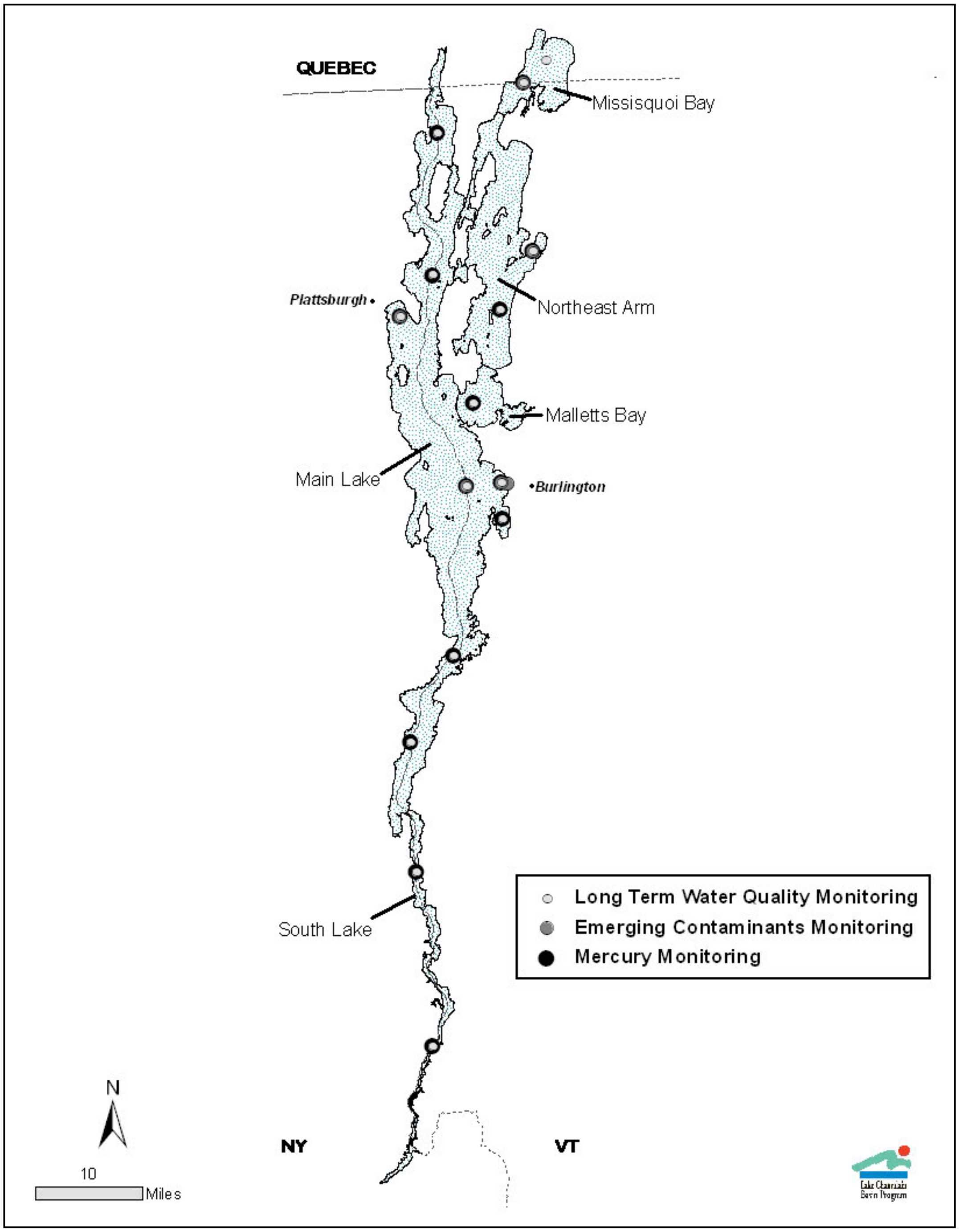


Figure 2-3. Map of water quality monitoring stations in Lake Champlain lake regions (Lake Champlain Basin Program)

Tributary inflows and effluent discharge points also contribute to horizontal variations in lake water quality. Localized inputs of large water and/or pollutant loads – e.g., suspended sediment from a large tributary river basin draining agricultural land or a nutrient load from a WWTP – can influence localized water quality, especially in a confined bay. Locations of such discharges are key factors in placing monitoring stations – either to deliberately sample them to represent important localized impairments or distinct components of total lake inputs, or to deliberately avoid them as unrepresentative of the broad lake, depending on program objectives.

Vertical variability in lakes can affect water quality and consequently monitoring design choices. Uniformly shallow lakes such as Grand Lake St. Marys in Ohio (GLWWA 2009) tend to be well-mixed vertically and have extensive photic zones, yielding a fairly homogeneous water column that can be effectively sampled at a single depth. Deeper lakes tend to stratify seasonally because of the temperature-density properties of water (Figure 2-4). Vertical stratification in lakes and reservoirs depends largely on depth, temperature, and seasonality, all of which should be included as covariates when monitoring lakes. When stratification is strong, the upper waters (epilimnion) may exhibit water quality characteristics (e.g., warm temperatures, high DO, low dissolved P) very different from those of the lower waters (hypolimnion) (e.g., cold temperatures, low DO, high dissolved P) because the two layers do not mix readily for long periods of time. This stratification breaks down in many lakes during fall and spring, when the water column mixes due to wind (turnover) and water quality is more uniform vertically. Depending on study objectives (e.g., monitoring algae populations in the epilimnion or measuring oxygen depletion in the hypolimnion), monitoring at different points with depth during periods of peak stratification may be appropriate. Alternatively, sampling during the periods when the water column is completely mixed (e.g., at spring or fall turnover) may yield information on the general character of the lake for that year. Some mass-balance lake P models, for example, use P concentration at spring turnover to represent the overall nutrient status of the lake.

Vertical variability is also important in lake biological monitoring. Chlorophyll levels and phytoplankton populations are naturally concentrated in the upper waters where sunlight can penetrate (the photic zone). However, zooplankton are mobile and show diurnal vertical migrations, moving up in the water column at night to feed and down during the day to avoid predators (Lampert 1989, Stich and Lampert 1981, Zaret and Suffern 1976).

Lake currents (primarily wind and inflow-driven) influence the dispersal of pollutants in a lake. In a reservoir, pollutant concentrations may exhibit a longitudinal gradient as circulation is dominated by inflow from the main tributary and outflow at the dam. Conditions in small embayments can be very different from conditions in open water. These conditions are due to circulation patterns caused by prevailing winds if currents tend to retain pollutants in the bay and inhibit mixing with the main lake waters.

Finally, sediment/water interactions exert strong controls on some pollutant dynamics in lakes. Concentrations of pollutants like P or toxic compounds that are strongly adsorbed to sediment particles can vary across the lakebed as sediments delivered from large tributary river basins settle around tributary mouths or are moved by currents into deeper lake regions. These dynamics may lead to hot-spots of high sediment pollutant levels that could be important for biological monitoring. In some cases, bottom sediments store pollutants like P for long periods as particles settle out over time or even sorb P from the water. In other cases, (especially where bottom waters are low in oxygen), P and other pollutants can be released from lake sediments to add to the lake pollutant load. Consequently, sediment remediation is sometimes part of efforts to reduce in-lake P concentrations, e.g., dredging (GLWWA 2009) or alum treatment (Welch and Cook 1999).

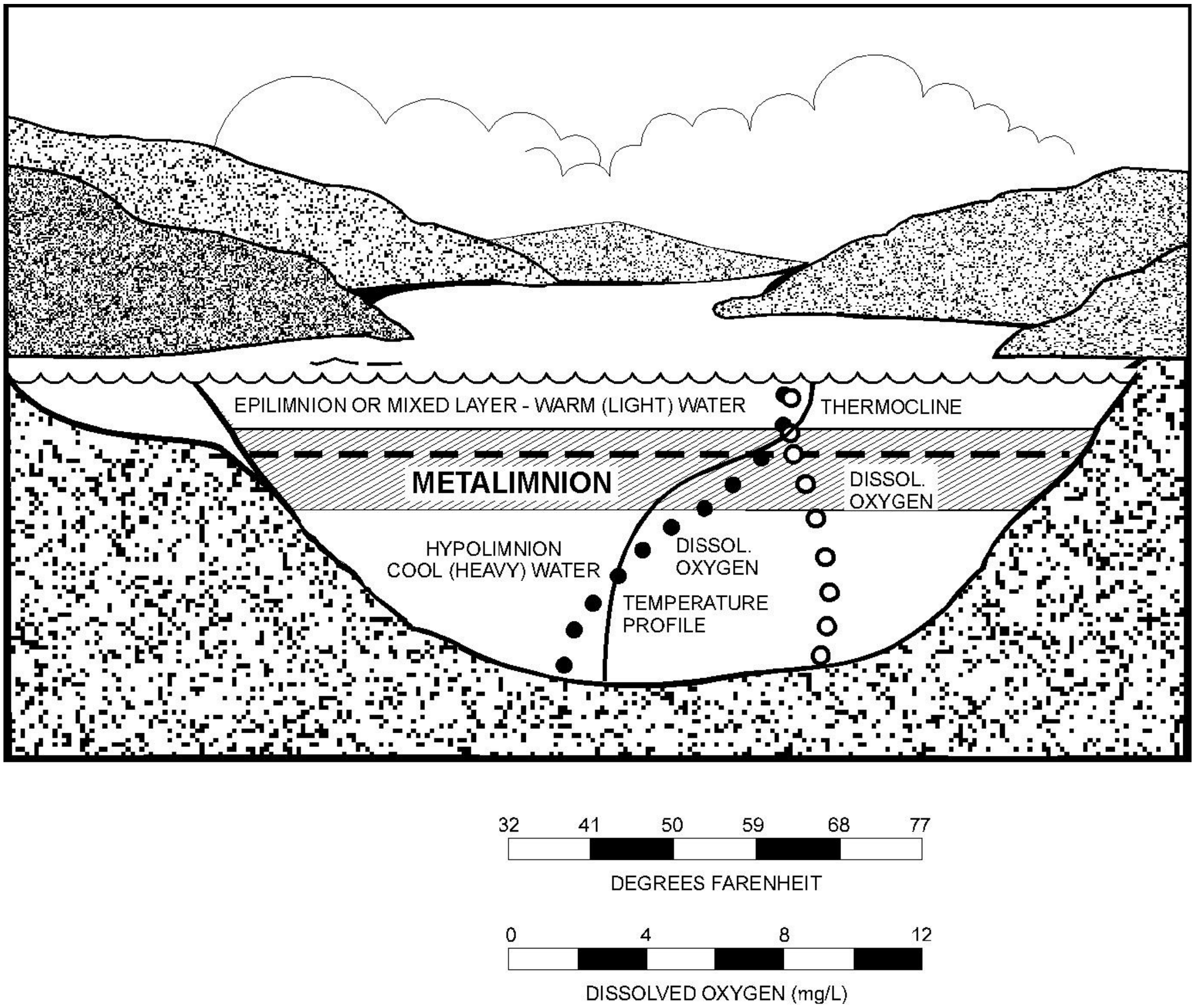


Figure 2-4. Thermally stratified lake in mid-summer (USEPA 1990). Curved solid line is water temperature. Open circles are DO in an unproductive (oligotrophic) lake and solid circles are DO in a productive (eutrophic) lake.

2.2.1.4.3 Wetlands

Since 1979, the Fish and Wildlife Service's definition of a "wetland" has been accepted as a standard for purposes of collecting information on the location, characteristics, extent, and condition of wetlands (Tiner 2002):

> *Wetlands are lands transitional between terrestrial and aquatic systems where the water table is usually at or near the surface or the land is covered by shallow water. For purposes of this classification wetlands must have one or more of the following three attributes: 1) at least periodically, the land supports predominantly hydrophytes (plants adapted to grow in water or hydric soils); 2) the substrate is predominantly undrained hydric soil (waterlogged or flooded soils); and 3) the substrate is nonsoil and is saturated with water or covered by shallow water at some time during the growing season of each year."*

Three factors (hydrology, the presence of hydric soils, and the presence of hydrophytic vegetation) largely determine the characteristics of wetlands, but hydrology is considered the master variable of wetland ecosystems, driving the development of wetland soils and leading to the development of the biotic communities (USEPA 2004). All three factors, however, serve as the foundation of any wetland condition assessment method.

Wetlands can occur at numerous landscape positions (Figure 2-5) and are often classified according to differences in hydrologic conditions (source of water, hydroperiod, hydrodynamics), vegetation (emergent, shrub-scrub), topography (depressional, riverine), and to a lesser degree, soils (muck, peat, unconsolidated) (USEPA 2004). Within the context of assessing wetland condition, classification is intended to reduce variability within a class and enable more sensitivity in detecting differences among impacted and impaired wetlands within the same classification. Different classes of wetlands may be subject to different stressors and may vary in their relative susceptibility to particular stressors.

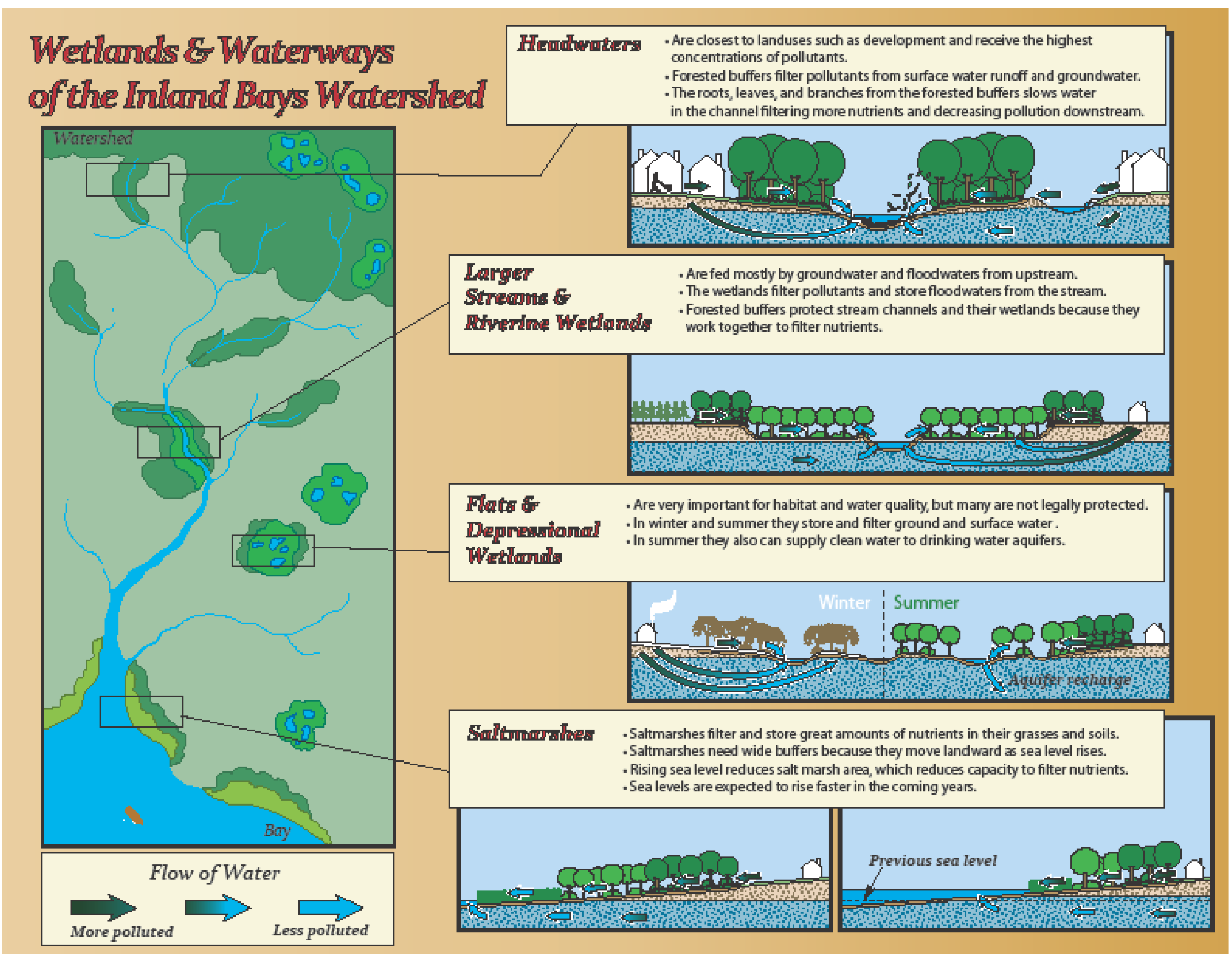


Figure 2-5. Wetlands and waterways of the Inland Bays watershed (DE CIB n.d.)

Due to the tremendous diversity among natural wetlands, a wetland monitoring program needs to be based on a specific wetland's attributes. Strategies for designing an effective monitoring program build from a hierarchy of three levels that vary in intensity and scale, ranging from broad, landscape-scale assessments (Level 1), to rapid field methods (Level 2), to intensive biological and physico-chemical measures at Level 3 (USEPA 2004). One of the key considerations for wetlands monitoring is definition of the assessment area, whether it is the entire wetland or a portion of the wetland. Rapid assessment procedures have been shown to be sensitive tools to assess anthropogenic impacts to wetland ecosystems, and can therefore be used to evaluate best management practices, to assess restoration and mitigation projects, to prioritize wetland related resource management decisions, and to establish aquatic life use standards for wetlands.

USEPA's 2006 *Application of Elements of a State Water Monitoring and Assessment Program for Wetlands* provides states with information to plan a wetland monitoring program and includes a discussion of the selection of indicators and metrics that reflect the unique characteristics of wetlands and their response to human-induced disturbance (USEPA 2006a). Several "modules" have been developed by EPA to support development of biological assessment methods to evaluate the overall condition and nutrient enrichment of wetlands (USEPA 2002b). These modules can be found at http://water.epa.gov/scitech/swguidance/standards/criteria/nutrients/wetlands/index.cfm.

Finally, because they are so biologically productive, wetlands tend to cycle sediment, nutrients, and other pollutants very actively among physical (e.g., sediment), chemical (e.g., water column), and biological (e.g., vegetation) compartments. Therefore, in a wetland monitoring program it may be important to look at each of these compartments, not treat the wetland as a simple input-output box. Moreover, because vegetation is a key element of wetland systems, seasonality of vegetation growth and senescence may be an important driver for nutrient cycling and therefore for monitoring design (USEPA 2002a).

2.2.1.4.4 Estuaries

Estuaries differ from freshwater bodies largely due to the mixing of fresh water with salt water and the influence of tides on the spatial and temporal variability of chemical, physical, and biological characteristics. Incoming tides affect estuaries by pushing salt water shoreward while fresh water is entering from freshwater systems (Figure 2-6). Fresh water is lighter, so it flows over the top of salt water, while the tide forces the salt water shoreward and under the inflowing fresh water. Outgoing tides pull the entire water mass toward the ocean, and the freshwater input fills the gap left by the receding submerged salt water. These processes affect daily and seasonal salinity distributions.

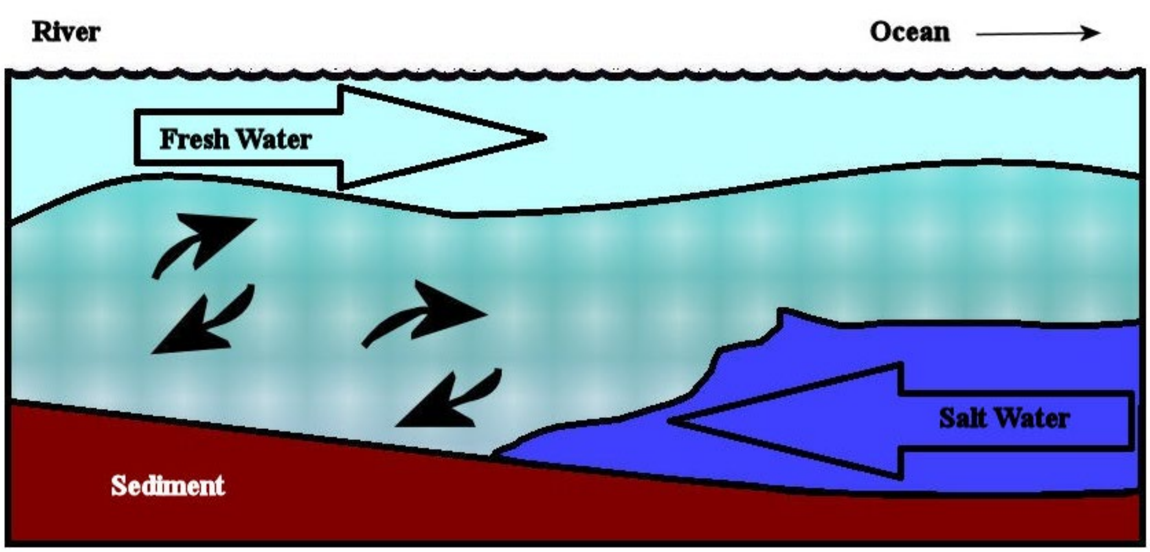


Figure 2-6. Mixing of salt water and fresh water in an estuary (after CBP 1995)

Because of the dynamic interaction of fresh water and salt water, pollutants are not flushed out from estuaries in the same manner as they are in most stream systems. Instead, an estuary often has a lengthy retention period (Ohrel and Register 2006). Consequently, waterborne pollutants, along with contaminated sediment, may remain in the estuary for a long time, magnifying their potential to adversely affect the estuary's plants and animals. This retention period also introduces a lag time that must be factored into monitoring plans intended to measure improvements resulting from restoration or improved land management.

The unique characteristics of each estuary must be recognized and understood when developing a monitoring plan because of their impact on estuarine hydrology, chemistry, and biology. Basin shape, mouth width, depth, area, tidal range, surrounding topography, and regional climate all play important roles in determining the nature of an estuary (Ohler and Register 2006). The earth's rotation (Coriolis effect), barometric pressure, and bathymetry (submerged sills and banks, islands) affect circulation and spatial variability in estuaries. For example, Puget Sound's complex circulation pattern is driven by tidal currents, the surface outflow of freshwater from Puget Sound rivers, the deep inflow of saltwater from the ocean, wind strength and direction, and underwater sills (Gaydos 2009).

Freshwater inflow is a major determinant of the physical, chemical, and biological characteristics of most estuaries. It affects the concentration and retention of pollutants, the distribution of salinity, and the stratification of fresh water and salt water (NOAA 1990). These freshwater inputs typically vary seasonally. For example, Figure 2-7 shows how salinity in the Chesapeake Bay is generally higher in fall and lower in spring due to spring runoff (CBP 1995). Salinity and other characteristics of estuaries may also vary spatially due to the location of freshwater inflows. The temporal variability of estuary condition is also influenced by factors other than freshwater inputs. For example, temperature profiles vary seasonally, and tidal cycles can affect the mixing of fresh and salt waters and the position of the fresh water-salt water interface.

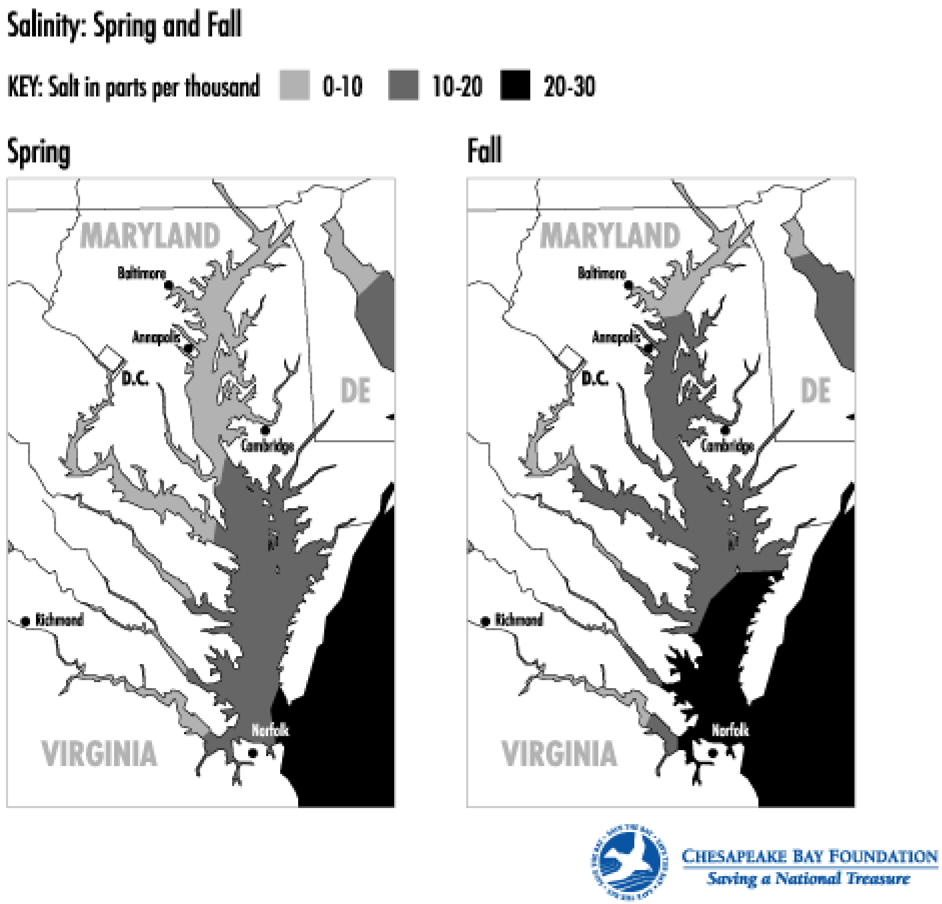


Figure 2-7. Salinity in the fall and spring in the Chesapeake Bay (CBF n.d.)

2.2.1.4.5 Nearshore Waters

The interplay of wind, waves, currents, tides, upwelling, tributaries, and human influences control water quality – and monitoring requirements – in nearshore waters. For the purposes of this guidance, nearshore waters include an indefinite zone extending away from shore, beyond the breaker zone (USEPA 1998); the term applies to both coastal waters and large freshwater bodies such as the Great Lakes. Wind and tides are the primary sources of energy in the coastal nearshore, and waves generated by the wind are largely responsible for currents (SIO 2003). These waves also have a central role in the transport and deposition of coastal sediments as well as the dispersion of pollutants and nutrients.

Upwelling brings cold, nutrient-rich waters to the surface, encouraging biological growth (Gaines and Airame 2010). Upwelling is extremely variable in space and time, depending on factors such as the strength and direction of the winds and the topography of the coastline (Gaines and Airame 2010). The spread of upwelled water down the coast of southern California can vary from a relatively narrow band near the coastline to enormous filaments extending hundreds of miles from shore. Upwelling on the east coast of Florida has been shown to be so dependent on the prevailing winds that it ceases as soon as the driving force is terminated (Taylor and Stewart 1959). Upwelling also occurs in the Great Lakes (Blanton 1975, Plattner et al. 2006). For the period 1992-2000, the magnitude of upwelling events observed in the southern basin of Lake Michigan tended to be greater than in the northern basin because the southern lake surface is typically warmer than in the north, while the temperature of the hypolimnion is more balanced over the extent of the lake (Plattner et al. 2006). In Lake Ontario, upwellings caused rapid shifts in the nearshore species composition of zooplankton and may be a mechanism for transport of certain diatom species to the epilimnion from hypolimnetic waters (Haffner et al. 1984).

Nearshore currents and pollutant transport are also affected by tributaries and human-made structures. Tributaries introduce fresh water to coastal waters and have varying potential to alter nearshore currents depending on factors such as tide stage, wind conditions, and tributary flow rate. Headlands (narrow strips of land that extend seaward), breakwaters (barriers built into the water to break the force of waves), and piers can influence the circulation pattern and alter the direction of nearshore currents (SIO 2003). For example, an obstruction on the down-current side of a linear beach will cause a pronounced rip current to extend seaward.

Current patterns must be sufficiently understood to determine the best locations for monitoring and to establish pollutant pathways and the likely relationships between land-based activities and nearshore water quality. Because circulation and pollutant transport is so variable in nearshore areas, designing monitoring plans based on assumptions about current patterns is not recommended. For example, a study of nearshore coastal circulation at the mouth of the Kennebunk River in Maine showed that currents did not carry river water directly to a local beach as expected (Slovinsky 2008); instead, river outflow extended much farther offshore from the beach. Because the current system of nearshore waters drives the relationship between land-based pollutant sources and receiving water quality, monitoring should include provisions to track variables needed to characterize the current sufficiently to aid interpretation of other chemical, biological, and physical data that are generated. Basic data on salinity, water temperature, and depth are often essential to identify the source of the sampled water and characterizing current patterns. The NOAA (U.S. Department of Commerce, National Oceanic and Atmospheric Administration) Great Lakes Coastal Forecasting System forecasts surface currents, winds, water temperature, and water level displacement, information that could be useful for sampling on any given date (GLERL 2011).

EPA, through its new Beaches Environmental Assessment, Closure and Health (BEACH) Program, is working with state, tribal, and local governmental partners to make nearshore water quality information available to the public. The BEACH Program provides a framework for local governments to develop equally protective and consistent programs across the country for monitoring the nearshore water quality along beaches and posting warnings or closing beaches when pollutant levels are too high. More information on this program can be found at http://water.epa.gov/type/oceb/beaches/beaches_index.cfm.

Note that because nearshore areas tend to be subject to heavy human use (e.g., swimming, boating, shellfish production), special water quality criteria and standards may apply. Fecal bacterial criteria for shellfish production, for example, tend to be far more restrictive than criteria for contact recreation. Such criteria may require special monitoring programs.

2.2.1.4.6 Ground Water

Ground water is the source of much of the Nation's streamflow, and ground water discharges often sustain water levels in lakes and wetlands, particularly during dry periods (Taylor and Alley 2001). The fact that the presence, quantity, and movement of ground water are not readily observable presents special challenges for monitoring design.

Ground water occurs in two general types of aquifers – confined and unconfined. Unconfined (water table) aquifers are in direct contact with the atmosphere through the soil and the elevation of the water table surface (i.e., depth to ground water). They fluctuate freely in response to changes in recharge and discharge (Figure 2-8). Confined (artesian) aquifers are separated from the atmosphere by an impermeable layer (USDA-NRCS 2003) and may be under pressure that results in a flowing (artesian) well if drilled into the aquifer. Perched water is held above the water table by an impermeable or slowly permeable layer below and is often the source of springs.

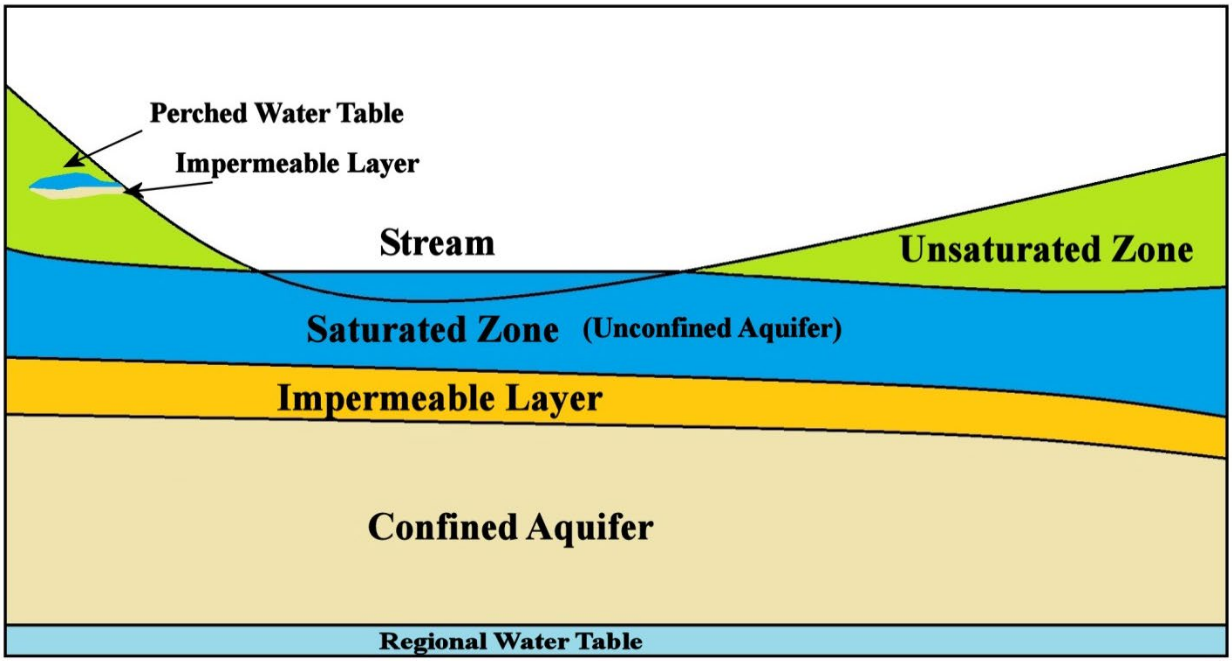


Figure 2-8. Basic aquifer types

Ground water *levels* are controlled by the balance among aquifer recharge, storage, and discharge (Taylor and Alley 2001). This balance is affected by characteristics (e.g., porosity, permeability, and thickness) of the rocks or sediments that compose the aquifer, as well as climatic and hydrologic factors (e.g., the timing and amount of precipitation, discharge to surface-water bodies, and evapotranspiration). Ground water moves along a hydraulic gradient from locations of higher hydraulic head to locations of lower hydraulic head. The rate of ground water movement depends not only on hydraulic head but also on the hydraulic conductivity (permeability) of the aquifer material; movement may be as rapid as 50 – 1000 m/day in a coarse gravel aquifer or as slow as 0.001 – 0.1 m/day in a silt and clay formation. The direction of ground water movement is not always obvious and not always consistent with the land surface topography. Patterns of ground water movement must be determined in the field (usually by measuring hydrologic head in numerous positions across a wide area) before determining sampling locations.

Ground water *quality* is influenced by a range of factors including aquifer type, native geology, precipitation patterns, flow patterns, land use, pollutant sources, and pollutant characteristics such as density and solubility (Scalf et al. 1981). Naturally, these factors can vary widely, even within a small region. A study of two adjacent Maryland watersheds with similar topography, land use and soils found that N yields differed significantly, largely due to the different characteristics of the aquifer underlying the watersheds (Bachman et al. 2002).

A special case of ground water systems is karst topography. Karst is a geologic condition shaped by the dissolution of channels or layers of soluble bedrock due to the movement of water. Karst regions typically display such surface features as sinkholes and disappearing streams and may be underlain by extensive cave systems. Aquifers in karst terrains are very sensitive to contamination because direct and rapid connections exist between the land surface and ground water, via the dissolution channels. Sinkholes are, for example, potentially direct pathways for sediment and chemicals to enter ground water without filtration through the soil. Karst systems present special challenges for ground water monitoring efforts, as sources of aquifer contamination may be widely dispersed and difficult to map.

The uncertainties regarding ground water flow patterns and the composition of underground materials, coupled with seasonal patterns and the interplay of surface water and ground water, require that basic knowledge of the particular ground water system under study be obtained before a monitoring program is designed or initiated.

Regional or statewide ground water level recording and water quality monitoring networks are common across the nation, especially in regions where ground water is a primary source of drinking and irrigation water (FACWI 2013). These networks often detect contaminants via well monitoring and model contaminant transport based on ground-water level data. Watershed-level monitoring of ground water, however, is still relatively rare despite the frequently important interaction between ground water and surface water. The interaction of surface water and ground water can be considered from the perspective of surface water recharging ground water or ground water discharging to a stream or lake (Goodman et al. 1996). The former is important when determining the impact of surface water on a ground water resource, whereas the latter should be a key element of monitoring when ground water comprises a significant portion of the water or contaminant budget of the surface water body (e.g., Schilling and Wolter 2001, Schilling 2002). When conducted well, ground water monitoring data, coupled with agricultural and land use data, can develop convincing evidence of the response of ground water quality to changes in agricultural management (e.g., Exner et al. 2010).

While the collection and analysis of groundwater data are not addressed in detail here, it is important that the role of subsurface waters be factored into watershed-scale and field-scale monitoring efforts described in this guidance. Several guidance documents are available for those seeking additional details regarding ground water monitoring, including guidance on construction of monitoring wells and sampling procedures (Scalf et al. 1981, Wilde 2006). The USGS has produced a series of groundwater technical procedures documents (GWPDs) that describe measurement and data-handling procedures commonly used by the agency in its groundwater monitoring activities (Cunningham and Schalk 2011). These procedures address groundwater-site establishment, well maintenance, water-level measurements; groundwater-discharge measurements, and single-well aquifer tests. In addition, guidance specific to monitoring ground water in NPS studies was developed based on experiences in the Rural Clean Water Program (Goodman et al. 1996).

Ground water monitoring is performed for a number of purposes, including:

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

Successful ground water monitoring design begins with a good understanding of the ground water system and the establishment of specific monitoring objectives. Ground water monitoring often requires a two-stage approach in which the first stage is a hydrogeologic survey to determine ground water surface elevations and flow rates and directions. First-stage surveys require numerous sampling stations because aquifer water quality can vary considerably with depth and location (Figure 2-9). Ground water level data can be used to determine ground water flow patterns as shown in Figure 2-10 (Winter et al. 1998).

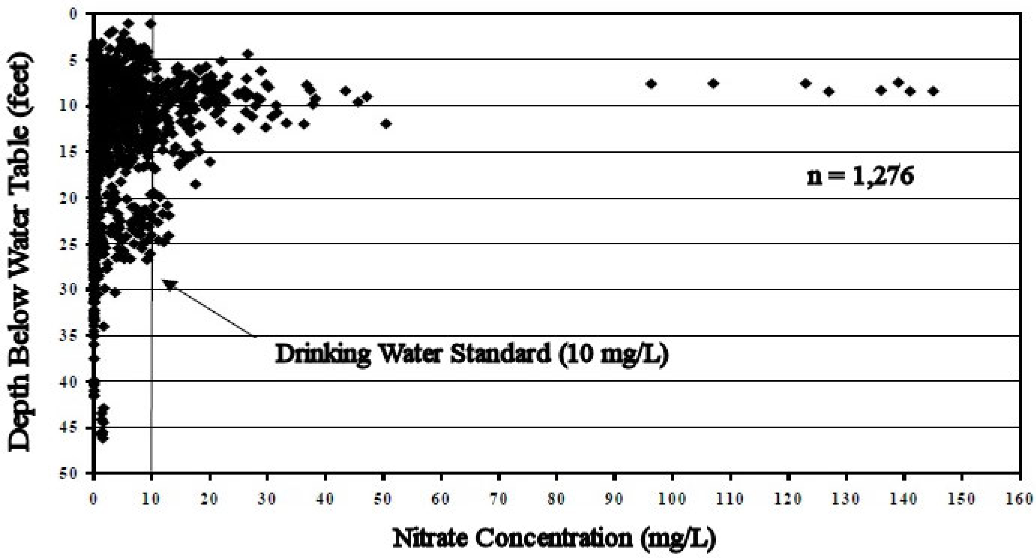


Figure 2-9. NO_3 concentration versus depth to water table (after Rich 2001)

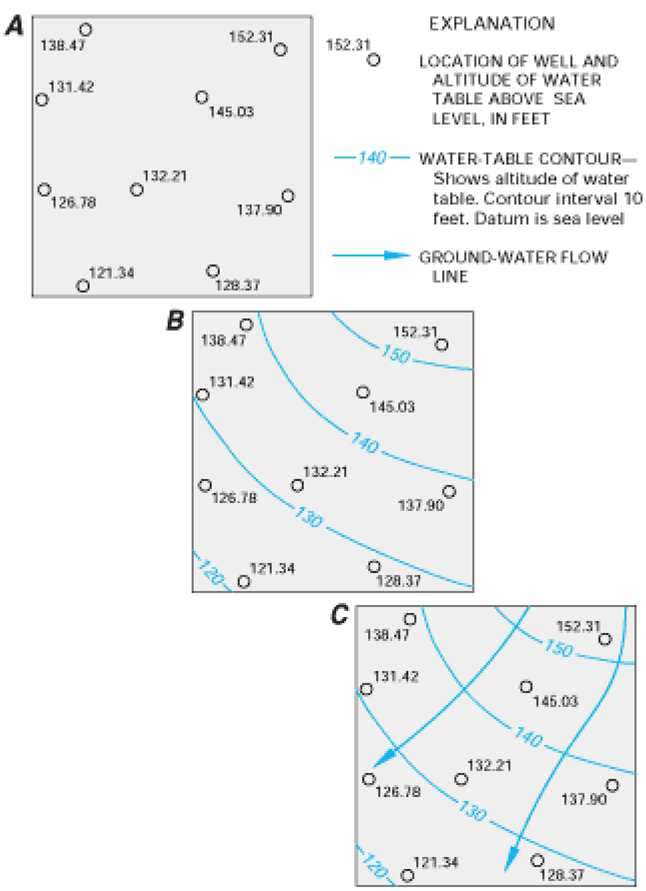


Figure 2-10. Determining ground water flow patterns (Winter et al 1998)

The second stage is an investigation of water quality, with stations selected based on monitoring objectives and the results of the first stage (Goodman et al. 1996, USDA-NRCS 2003). Sampling locations can and should be guided by knowledge of the hydraulic gradient, but the heterogeneous nature of surbsurface environments makes appropriate location of sampling points a complicated and unpredictable task (Scalf et al. 1981). Several sampling locations and sampling from multiple depths may be required to characterize ground water and determine contaminant pathways (Goodman et al. 1996). Ground water investigations in South Dakota, for example, have shown that nested wells may be necessary for adequate examination of shallow aquifer water quality (SDDENR 2001).

In some cases it may be necessary to sample the unsaturated zone to get a true picture of the threat to ground water (Scalf et al. 1981, Goodman et al. 1996). In addition, long-term water level measurements may be needed to show how contaminants are transported from their sources through the groundwater system (Taylor and Alley 2001). It may even be possible to establish relationships between water levels and contaminant concentrations, possibly indicating patterns associated with seasons or rainfall-events.

Because ground water monitoring is both complex and expensive, sophisticated geostatistical techniques (e.g. Chiles and Delfiner 1999, Lee et al. 2005) are increasingly used both to build conceptual hydrogeological models of ground water flow, quality, and contamination and to assess health and environmental risk based on observed sample data (EPA Victoria 2006). Thus, modeling and spatial analysis can be useful in designing ground water monitoring programs and in organizing and interpreting results.

2.2.1.5 Climate

Climate is one of the principal determinants of the basic structure of a monitoring program. The frequency, intensity, and duration of runoff-producing storm events affect sampling frequency and duration, equipment selection, automatic sampler programming, and many other elements of a monitoring program. Freezing conditions can have immense impact on the duration of the sampling season, the design and cost of permanent sampling stations, and the operation and maintenance of sampling equipment. Droughts and floods can be fatal to monitoring programs that have no budget flexibility, and the lag time between BMP implementation and measurable water quality impacts can be changed drastically by persistent changes in weather patterns.

Average precipitation patterns and the resulting average flow conditions are typically used to establish sampling frequencies, the relative emphasis on base-flow and storm-event sampling, the location of biological monitoring sites, and the design and siting of flow gaging stations. Precipitation patterns over any given study period, however, can vary significantly from long-term averages, as evidenced by a seven-year study in Illinois in which annual precipitation was lower than the long-term average in all but one year (Algoazany et al. 2007). An analysis of precipitation in the Minnesota River Basin for the period 1891-2003 showed a slightly increasing trend, with annual totals ranging from well under 400 mm to well above 900 mm (Johnson et al. 2009). In a runoff study on a dairy in the Cannonsville Reservoir watershed in New York, seven of the eight highest event flows occurred in the post-BMP period despite the fact that the pre- and post-BMP study periods exhibited similar scales and frequencies of precipitation and event flow volume (Bishop et al. 2005). Short-term and long-term drought greatly influenced runoff events in an 11-year study in Georgia (Endale et al. 2011). During the 86 months with below-average rainfall there were only 20 runoff events, compared to 54 runoff events during the 46 months with average or greater rainfall.

Monitoring program managers must plan for a wide range of flow conditions, but flow is not the only important consideration when designing a monitoring program. Climatic variability can also influence aquatic organisms and land treatment programs. For example, the growth and development of riparian buffers is dependent on adequate precipitation. No monitoring program can be designed to handle all of the potential impacts of climatic variability, but all monitoring programs should be designed to account for a foreseeable range of conditions. Design concerns can range from determining the size of a flume required to measure edge-of-field runoff to planning budgets and time frames to allow the capture of a sufficient number of high-flow events.

2.2.1.6 Soils, Geology and Topography

Soils, geology, and topography are local or regional features that must be considered in monitoring program design (MacDonald et al. 1991). These characteristics influence the hydrologic budget, potential suspended sediment loading from erosion, background levels of nutrients and dissolved ions in ground and surface waters, and other factors that drive monitoring program design.

The importance of soil groups is illustrated by a Pennsylvania study in which runoff was monitored from two contrasting hillslope soil groups (colluvial and residual) that differed in subsurface morphological characteristics such as the presence of a fragipan, the clay content of argillic horizons, and drainage class (Needelman et al. 2004). Results showed greater runoff from the four colluvial sites for all significant events, and overall runoff yields were also greater from the colluvial sites (average of 2.4 percent) than from the two residual soil sites (average of 0.01 percent).

A study of an agricultural watershed in the coastal plain of Maryland showed the importance of near-stream geomorphology and subsurface geology in determining riparian zone function and delivery of NO_3 to streams (Böhlke et al. 2007). Stream NO_3 levels were higher during high flow conditions when much of the groundwater passed rapidly across the riparian zone in a shallow, oxic aquifer wedge and higher during low flow conditions when stream discharge was dominated by upwelling from the deeper, denitrified parts of the aquifer.

Slope must also be factored into the design of a monitoring program because slope and slope length affect the rate and duration of runoff from a watershed, rate of erosion, depth of soil (steep slopes often have less soil overlying the bedrock), and stream characteristics. Slope also affects the likelihood of landslides and debris flow, erosional processes, and weathering rates.

2.2.2 Monitor Source Activities

NPS pollution is highly variable and is generated by activities on the land that vary in location, intensity, and duration. To make the connection between pollutant sources and water quality observed through monitoring, it is also necessary to monitor the activities on the land that generate NPS pollutants. In the context of linking cause and effect, water quality monitoring data represent the effect, while source activities represent a major component of the cause. Put another way, to fully understand NPS pollution, we must measure both the dependent variables (water quality) and the independent variables (source activities).

In practice, monitoring pollutant source activities usually translates to land use and land management monitoring. This means more than taking a static picture of land use/land cover in a watershed from a satellite image (although that may be very useful for establishing a baseline condition). It means monitoring dynamic pollutant-generating activities in time and space. Examples of NPS pollutant types and common corresponding source activities to be monitored are shown in Table 2-2.

Table 2-2. Selected NPS pollutants and watershed source activities to monitor

NPS Pollutant Type	Potential Variables to Monitor
Suspended sediment (field erosion)	Cropland tillage, planting, harvesting, erosion control BMPs, precipitation.
Suspended sediment (streambank erosion)	Streamflow, stream morphometry, riparian condition, precipitation.
Phosphorus	Manure applications, livestock populations, manure and fertilizer management, soil test P.
Nitrogen	Fertilizer applications, legume cropping, manure and fertilizer management, groundwater movement.
Crop herbicides	Herbicide application rates and timing, precipitation
Pathogens	Livestock populations, grazing practices, riparian condition, pasture fencing, manure land application practices. Pet populations, wildlife/waterfowl activity, septic system maintenance/failure, sewer maintenance, illicit discharge/connections.
Salt	Amount and timing of road salt used for deicing. Road salt contract amounts. Miles and locations of roads salted. Irrigation return flows.
Heavy metals	Vehicle traffic, highway infrastructure, street sweeping, stormwater management structures and activities.
Stormwater flow	Impervious cover, stormwater management facilities, precipitation.

The practice of source activity monitoring is discussed in more detail in section 3.7 of this guidance.

2.2.3 Critical Details

Execution of a monitoring plan requires careful attention to some critical details, as the following discussion reveals.

2.2.3.1 Logistics

Logistics are defined here as matters concerning the management of the flow of materials, information or other resources from the point of origin to the point of use to meet the requirements of an enterprise. In water quality monitoring, logistics refers specifically to supporting the basic functions of data collection.

- Supplying power to field stations.
- Ensuring access to sampling locations for sample collection and field measurements.
- Delivering, maintaining, and retrieving equipment, instruments, and supplies to and from the sampling sites.
- Providing communications and data links between a base and remote sampling stations.
- Having available, well-trained, and on-call field personnel.
- Traveling to and from sampling stations.
- Delivering samples to the laboratory on time and under appropriate chain of custody.

All of these elements must be addressed in the process of developing a monitoring plan. If necessary, how will power (direct AC, solar, or battery) be supplied? Can desired sampling locations be accessed legally and safely under the range of expected conditions (e.g., high flow, inclement weather)? If structures or shelters are necessary, can the property owner's and municipality's permission be obtained? What is the time and cost involved in traveling to and from a network of sampling locations? Is electronic communication between stations and a base (if desired) possible, considering distance and topography? Can samples be delivered to the laboratory within the limits of required holding times?

These and other practical questions of how to carry out the physical tasks of monitoring need to be considered in the planning stage. Some practical guidance for addressing such logistical issues is presented by Harmel et al. (2006).

2.2.3.2 Quality Assurance/Quality Control and the Quality Assurance Project Plan (QAPP)

Data collected by a monitoring program must be of sufficient quality and quantity – with respect to accuracy, precision, and completeness – to meet project objectives. Provisions for ensuring data quality must be made during the monitoring design process, not after a plan is underway. These provisions fall into two main categories. Quality control (QC) refers to a system of technical procedures developed and implemented to produce measurements of requisite quality. QC activities typically include the collection and analysis of blank, duplicate and spiked samples, analysis of standard reference materials, and inspection/calibration/maintenance of instruments and equipment. Quality assurance (QA) is an integrated system of management procedures and activities to verify that the QC system is operating within acceptable limits and to evaluate and verify the quality of data collected. A QA system addresses the roles and responsibilities of monitoring staff, required staff skills and training, tracks sample custody, sets data quality objectives and procedures for data validation, and monitors QC activities, including actions taken to correct problems. In general, each organization that conducts monitoring should ensure that the appropriate QA/QC measures are followed, but may vary among funding organizations.

All organizations conducting environmental programs funded by EPA are required to establish and implement a quality system, a structured system that describes the policies and procedures for ensuring that work processes, products, or services satisfy stated expectations or specifications (USEPA 2001). EPA also requires that all environmental data used in decision making be supported by an approved Quality Assurance Project Plan (QAPP) which documents the planning, implementation, and assessment procedures for a particular project, as well as any specific quality assurance and quality control activities (USEPA 2008b). The purpose of the QAPP is to document planning results for environmental data operations and to provide a project-specific “blueprint” for obtaining the type and quality of environmental data needed for a specific decision or use. In most monitoring programs, an approved QAPP is required before data collection can begin; even in cases where a QAPP is not specifically required, such a document is a valuable resource for documenting consistent monitoring procedures, and therefore useful to prepare even if not required. Quality control, quality assurance and the QAPP process are discussed in detail in chapter 8.

Other agencies including the United States Geological Survey (USGS) have issued guidance and requirements regarding data quality (Wilde 2005). Every USGS study requires a sampling and analysis plan (SAP) and a quality-assurance plan (QAP) that include a description of the objectives, purpose, and scope of the study and its data-quality requirements. In addition, each USGS Water Science Center develops general quality-assurance plans that articulate its policies, responsibilities, and protocols. Specific guidance on obtaining representative samples can be found in USGS’s National Field Manual (Wilde 2006). USGS quality control procedures emphasize generating information on bias and variability because of their importance in proper and scientifically defensible interpretation of collected data.

2.2.3.3 Data Management and Record-keeping

Even short-term monitoring efforts may generate tremendous quantities of data. A system for managing that data stream must be included in an overall monitoring plan. Poorly recorded, misunderstood or even lost data represent an irretrievable loss of information, a waste of resources, and a threat to program objectives. Poor data management can also make the task of data analysis and interpretation more difficult and challenging than necessary.

Good data management begins in the field, where a clear identification system is required to correctly attribute data to their source. Field log sheets or notebooks are valuable tools for initial recording of field data, sample identification and observations that may represent critical knowledge later. A good field log can also serve as a guide or checklist for the field technician.

Chain of custody records are essential where litigation is involved, but also useful for simply tracking delivery of samples to the lab. It is also important to document assignment of lab sample numbers and their correspondence to field identification codes.

The process and schedule of data reporting from the laboratory should be outlined and agreed upon. Timely reporting of data from the lab is essential in providing feedback to the monitoring and land treatment program.

A good data management system should be implemented in a simple, consistent format (e.g., a spreadsheet or a database form) that can accept both manually transcribed data (such as those from field logs or lab data reports) and data already in electronic form (such as downloads from field instruments or data loggers). Electronic data formats should be designed to be consistent with formats used for later analysis (e.g., in a statistics package or uploads to STORET) to avoid the cost and potential errors of transcribing data from one format to another.

Data validation and error checking are essential and should be performed at an early stage. Validation involves checking for correct transcription between data sources and data storage (e.g., between field logs and electronic spreadsheets), checking for typographic errors, looking for extreme or impossible values, and ensuring that all required data have been included. Validation should be performed on 100 percent of the data, not just a spot-check. It is very important that validation be performed early in the process, as it is costly and frustrating to have to repeat data analysis and presentation if errors are discovered late in the process.

Data storage is also an important consideration. Paper records such as field logs or lab data reports should be archived and perhaps scanned for electronic storage. Both original data and data derived from calculations, analysis or other manipulations should be stored. Maintain a metadata file to record important information about the data and the monitoring program, QA/QC results, exceptions or unusual occurrences, and any other important monitoring records. If data are stored in a national repository, such as STORET, download the data and make sure they are identical to the data on your desktop. Electronic data forms should be stored redundantly and protected with frequent backups. For long-term archiving, select the storage medium carefully. Data from a 1985 project stored on 5.25 inch (in) floppy disks may be nearly impossible to access in 2015; data recorded on a CD today may be unreadable in the future.

2.2.3.4 Roles and Responsibilities

Most monitoring programs involve cooperation among several different agencies, offices or individuals. For example, a watershed project might include funding (USEPA), planning and implementation of BMPs (USDA-NRCS, Soil and Water Conservation Districts), flow measurement (USGS), water chemistry sampling and analysis (Health Department), and biomonitoring (state environmental agency). Even within a single activity, such as water chemistry monitoring, different individuals like field technicians, laboratory analysts and graduate students may play different roles. A good monitoring plan needs to specify the roles and responsibility of each participating entity and individual so that all monitoring tasks can be accomplished smoothly. Perhaps even more important, a mechanism for coordinating among the variety of agency and individual roles and responsibilities should be established from the start. Strong leadership from an overall project director/coordinator can facilitate good cooperation among a project team. In addition, frequent contact, progress reports, and regular meetings among all project participants have been shown to be key ingredients for effective coordination.

2.2.3.5 Review of Monitoring Proposals

Monitoring plans may be developed and reviewed under a variety of different templates or formats. Whether for an internal check for completeness or for an external review in an approval process (e.g., for state Section 319 funding), it is often useful to step back from the details and review the contents of a monitoring plan to make sure that all necessary elements have been considered and addressed. Experience of NPS monitoring efforts across the country suggests that confirmation of the following elements is useful in review of monitoring plans:

- Watershed Identification and Characterization
  - Descriptive information on physiographic setting, water resources, land use/management
  - Identification of stakeholders and project participants
- Problem Identification
  - Clear identification of water quality problem(s)
  - Documentation of impairment(s) and supporting data
  - Known or suspected causes and supporting data
  - Known or suspected sources of pollutants and supporting data
- Project Goals and Objectives
  - Quantitative goals for water quality
    - Tied to impairment, restoration of use(s)
    - Including estimated load reductions as appropriate
  - Quantitative goals for land treatment implementation
- Land Treatment(s) to be Implemented
  - Identify critical areas and measures to be implemented
  - Justification for specific practices selected
  - Schedule and interim milestones/indicators of progress
  - Availability of funds, personnel, and other resources
- Monitoring Plan
  - Water quality
    - Design
    - Variables

    - Locations
    - Frequency and duration
    - Sample collection and analysis
  - Land use/land treatment – process and responsibility
  - Availability of funds, personnel, and facilities
- Data Management and Analysis
- Administration/Management/Coordination
- Reporting, Communication, Stakeholder Involvement
- Timetable and List of Deliverables
- Budget

Note that this checklist addresses several elements such as those associated with land treatment that may seem to be outside the immediate realm of water quality monitoring, but these must also be considered and coordinated with other project activities.

2.2.4 Feedback

Although implementation of BMPs on the landscape and monitoring water quality at various locations in the watershed may seem to be separate activities that can proceed independently, successful NPS watershed projects require effective coordination and collaboration among all activities. It is therefore important to facilitate feedback of data and other information among different components of a watershed project. For example, water quality monitoring staff should know where and when BMPs are implemented in the watershed, and land treatment implementation should be guided by water quality data where possible. Even within the monitoring program, it can be critical for biomonitoring staff to know the results of water chemistry monitoring to fully understand what they observe in the biotic community.

Feedback mechanisms should be built into a watershed project from the beginning, not left to chance or put off to the final project report. Frequent examination, presentation and discussion of monitoring data will keep all project participants informed. Regular review of field data and observations can provide evidence of events or conditions in the watershed that reveal small problems before they become large. Similarly, frequent examination of laboratory results can show evidence of analytical or QA/QC problems before they result in major data loss. Feedback between water quality monitoring and land treatment personnel can help fine tune BMP implementation to known water quality problems and can provide land-based data to improve understanding of observed patterns in water quality.

Feedback can be generated by requiring frequent reports (e.g., monthly) and meetings for all project participants. The reports and meetings can be brief and follow a simple formula, but by requiring all sectors to periodically compile, examine and present their data, feedback among all data streams can be guaranteed.

2.2.5 Limitations of Monitoring

In practice, monitoring does not always answer all of the questions or achieve all of the objectives because:

- Available resources may fall short of estimated costs.

- A large number of different situations/scenarios create too many alternatives to evaluate cost-effectively.
- Socio-economic factors may require modification of monitoring or land treatment plans.
- Watershed size, access or other features impose significant logistical limitations.
- Water quality conditions (e.g., range of flows) are too variable to effectively monitor.
- Appropriate data quality cannot be achieved (e.g., volunteer monitoring).
- The actual desired response to treatment cannot be monitored due to flooding or other physical changes in the resource.
- The magnitude of change expected to result from treatment, especially in context of background levels or contributions from other uncontrolled sources, is small.
- Lag time between land treatment and water quality response exceeds the duration of monitoring.
- Random or catastrophic events (e.g., intense storms or chemical spills) overwhelm response to treatment.

In principle, if a project cannot afford monitoring that can be reasonably expected to achieve objectives within the design parameters, it is recommended to forgo the inadequate monitoring that will not serve project needs but will drain budget resources. It may be possible to narrow the monitoring objectives, reduce the required precision, or reduce the scope of the monitoring effort to stay within budget, but such compromises must be made within the context of designing a plan that will meet stated objectives that help the project meet its goals. Modeling may be an effective alternative, especially when numerous alternative scenarios must be considered. However, note that proper model application (including calibration and validation) requires some data and considerable resources. See section 6.3 for ideas on how to integrate monitoring and modeling.

2.3 Monitoring Scale Selection

2.3.1 General Considerations

The scale of a monitoring plan is the size of the area to be monitored, a spatial consideration. Selection of the appropriate scale depends on the study objectives, study duration, type of water resource monitored, the complexity of the project, and available resources (USDA-NRCS 2003). Monitoring scale is generally locked in with the selection of monitoring design.

The choice of scale affects monitoring costs, duration and logistics. The ability to isolate the factors of interest (e.g., BMP effectiveness, transport pathways) generally increases as scale decreases, but the transferability of results generally decreases as scale decreases. Monitoring a set of 1 x 3 m plots, for example, may yield good data on how cover crops reduce soil loss, but such data are very difficult to extrapolate to a watershed-scale because small plots do not always reflect field-scale runoff processes or watershed-scale transport and delivery processes. Analysis of long-term data collected in the Lake Erie basin showed that watershed size had a much greater effect on concentration patterns than on unit area loadings (Baker 1988). The greatest effect was on peak concentrations; as watershed size decreased, peak concentrations of sediments, nutrients, and pesticides increased (Baker et al. 1985, Baker 1988). Pollutant concentrations returned to baseline levels more quickly in smaller watersheds and streams. It was also determined that increasing proportions of the annual load occur in decreasing proportions of time as watersheds become smaller, but that the high rates of export from small watersheds are distributed into larger numbers of individual events compared to larger watersheds. Because of these occurrences, it takes more sampling effort to accurately measure the loads from a smaller watershed, and the likelihood of missing high export rate events is greater (Baker 1988).

2.3.2 Options

Monitoring can be performed at scales ranging from national to single points, but the primary options for the types of NPS monitoring studies addressed in detail by this guidance are the watershed and BMP scales, the latter of which includes field and plot studies described by USDA-NRCS (2003). National and statewide (or regional) monitoring scales are only briefly touched upon here, and point-scale sampling is addressed primarily for explanatory variables. In fact, the data collected from point-scale sampling performed in support of the types of monitoring described in this guidance will generally be extrapolated to larger scales. For example, precipitation data collected at a single point may be applied to a rooftop to estimate flow volume handled by a green roof, or used to represent precipitation in either or both drainage areas in an above/below design. Soil samples collected at single points in a field may be combined for analysis to represent average conditions or analyzed separately with results interpolated to represent varying conditions for plot or field stations. Soils and precipitation data are generally used as explanatory variables in the statistical approaches discussed in chapter 7.

2.3.2.1 Statewide or regional

Statewide monitoring designs generally emphasize larger streams and rivers, public lakes, and the outlets of watersheds. States usually locate some of their monitoring at sites gaged by the USGS to take advantage of the flow data. Some monitoring stations at key locations are equipped with automatic samplers or sondes for continuous monitoring, but cost and logistical constraints limit most monitoring efforts to the collection of grab samples, a few field measurements (e.g., temperature, DO, conductivity), and biological and habitat monitoring. With the exception of the few stations with automatic or continuous sampling, monitoring frequencies are generally low.

Statewide monitoring associated with NPS pollution is generally designed to assess current conditions. It is unlikely that most statewide monitoring efforts can support trend analysis because of the strict requirements (e.g., no breaks in the data, consistent methods over time, collection land use and other covariate data) and the difficulty states have committing to consistent long-term monitoring efforts. Recognizing this and related limitations, Congress in 2005 began appropriating additional funds within Section 106 grants for an initiative to enhance monitoring programs and provide statistically-valid reports on water conditions (USEPA 2006b). In 2008, EPA amended its guidelines for this initiative to provide incentives for states to implement state-wide statistically-valid monitoring surveys (USEPA 2008a).

2.3.2.2 Watershed

Watershed-level monitoring for NPS assessments and the evaluation of project effectiveness has evolved over the years as lessons have been learned from such programs as the Rural Clean Water Program (USEPA 1993a), ACP-Special Water Quality Projects (Davenport 1984), Model Implementation Program (NCSU and Harbridge House 1983), Nationwide Urban Runoff Program (USEPA 1983), and the Section 319 National NPS Monitoring Program (Tetra Tech n.d.). The current emphasis on Total Maximum Daily Loads (TMDL)s has changed watershed-level monitoring even further by placing a greater focus on estimating the pollutant loads from each source category in the watershed, setting numeric targets for pollutant load reductions at a watershed level, and linking NPS and point source control efforts (USEPA 2012a).

Watershed-level monitoring can be triggered by findings from state-level monitoring, but is even more likely to develop from local stakeholder efforts to identify and solve problems that they care about (USEPA 2012b). Infrequent grab samples, basic chemical and physical parameters, and rapid assessments of aquatic biology conditions typically constitute the monitoring performed prior to or in the initial stages of a watershed project (USEPA 2008b). Flow data are generally lacking unless the watershed has a USGS gaging station, and monitoring stations are usually located where convenient for sampling, a situation that can bias results.

As watershed projects evolve, the monitoring approach should change to address project goals (USEPA 2008b). Initial efforts generally focus on refining the problem definition, including characterizing the water quality problem better, determining the major sources and causes of the problem, and providing data to aid in the design of a plan to solve the problems. Monitoring during this phase of a watershed project may include a synoptic survey, tests for toxicity, flow measurements at various points in the watershed to support a load analysis, detailed habitat assessments, and higher level biological assessments. Land use mapping, investigation of permitted discharger reports, and windshield surveys may also take place to better characterize sources. Most of the monitoring during this phase of a watershed project is short-term, with one or two sampling events at most. In rare cases, projects may install "permanent" monitoring stations and factor long-term monitoring considerations into the short-term effort. Such projects should ensure that all design requirements for long-term monitoring can be met, however, before monitoring begins.

Watershed-level monitoring, as described in this guidance, typically begins after a watershed project secures funding to address the identified problems. All too often project implementation begins before or simultaneously with monitoring, complicating efforts to assess the effectiveness of the project. The watershed monitoring plan should be coupled with and complementary to the watershed project or management plan (USDA-NRCS 2003). Depending on the specific objectives, the size and characteristics of the watershed, and the parameters of concern, watershed-level monitoring can take various forms. A key difference between watershed- and state-level monitoring is the narrowing of focus and increased intensity of watershed-level monitoring. Because the questions to be addressed by monitoring are more specific at the watershed level, each watershed monitoring effort is unique.

Watershed-level monitoring to assess project effectiveness generally requires a control condition to serve as a benchmark. This is not unlike the use of reference conditions for biological monitoring, but whereas reference conditions are generally sought in areas with minimal human impact (see chapter 4), the control conditions for watershed-level monitoring are usually found within or very near to the watershed being treated. In some cases, watershed projects will have relatively pristine conditions that serve as the control, but usually the control is an upstream area or a paired watershed within a short drive of the treatment watershed. Data from the control are used to isolate the effects of the BMP implementation in the study or treated area. This type of monitoring is not generally performed in statewide monitoring programs.

Another approach used at the watershed level is long-term trend monitoring at a single station where there is no control condition to serve as a benchmark. As described in section 2.4.2.5 and noted above for statewide monitoring, this approach has a relatively high risk of failure because the requirements of the method are difficult to satisfy over the long term. When performed at the watershed level, rather than the statewide level, however, the risk of failure can be reduced because more frequent sampling is done (e.g., once per two weeks) and the timeframe for demonstrating results is shortened. Still, the risk of failure with this approach remains generally greater than for designs using controls because the timeframe for monitoring at the watershed level can still be on the order of a decade.

One of the biggest challenges at this scale is to determine when a watershed is too large for successful monitoring (USDA-NRCS 2003). The upper limit on watershed size will depend on the monitoring objectives. For trend analysis, there may be no upper size limit if monitoring will continue for decades. If the goal is to attribute trends to changes in land use or land management, the cost of tracking such information at a sufficient level of detail may be prohibitive for extremely large watersheds that cover 100,000 acres or more (~40,500 ha).

Problem assessment can be done at various levels of detail for various purposes. If the assessment is to form the basis for an action plan that will require the obligation of substantial resources, it may be wise to limit the monitored watershed size to 50,000 acres (~20,200 ha) or less (i.e., the HUC-12 level). If, for example, the watershed or basin plan is for an area of 500,000 acres (~202,000 ha), monitoring would be performed at ten subwatersheds of 50,000 acres each. Watersheds larger than 50,000 acres are considered very large and may be inappropriate for assessment monitoring because of their likely heterogeneity in land uses (USDA-NRCS 2003).

Pollutant load measurement can be performed at watersheds of any size, but attribution of those loads to specific sources or source categories becomes more difficult for large watersheds. A limit of 50,000 acres may be appropriate for load estimation within the context of watershed plans or TMDLs where load and wasteload allocations will be made. The exact size limit for any situation will depend on a wide range of factors including whether or not watershed modeling is part of the effort.

The appropriate watershed size for evaluating watershed projects is probably 25,000 acres (~10,000 ha) or less based on experiences in the NNMP and RCWP. The actual size depends on a number of factors including average annual precipitation, the type and degree of use impairment, lag time, study duration, the potential for making improvements with BMPs, the number and location of monitoring stations, and sample type and frequency.

In comparison to statewide monitoring, watershed-level monitoring generally involves a shorter timeframe (3-10 years), fewer stations, more frequent sampling of both storm events and base flow, and a targeted and unique set of monitoring variables. Flow is usually measured in watershed-level monitoring. Due to automation of monitoring stations, primarily to monitor storm events, the annual cost per monitoring station is generally greater than routine grab sampling programs. On the other hand, biological data are typically collected only once or twice per year, and as a result, the timeframe for demonstrating results is quite dependent on the specific problem identified and the treatment plan designed to solve the problem. Where stream and habitat restoration has been done in watershed projects, biological monitoring has often been successful in documenting benefits, but in the various, long-term, watershed-based NPS monitoring programs listed above (e.g., Rural Clean Water Program), there have been very few instances where biological monitoring has demonstrated the effectiveness of watershed projects that did not involve such in-stream work. As a result, this guidance recommends that biological monitoring be coupled with physical/chemical monitoring of the stressors targeted in watershed projects.

2.3.2.3 BMP or practice

Monitoring at the BMP level is generally the most intensive of the levels described here. The scale for BMP monitoring can vary from plot studies to the inflow and outflow of a multi-acre constructed wetland to the influent and effluent of a manufactured stormwater treatment device. Monitoring variables are selected based upon the specific sources treated by the BMP, including such possibilities as metals and organic chemicals in an urban setting; bacteria and BOD near shellfish beds or livestock operations;

nutrients and sediment on cropland; temperature, DO, and biological communities for a stream restoration; and pH in abandoned coal mining areas. Flow needs to be monitored in most BMP studies.

BMP-level monitoring is generally storm-event driven, with little or no base-flow sampling. Exceptions include monitoring of constructed wetlands or wet ponds where it is essential that time-of-travel is assessed to provide for "matching" inflow and outflow samples to estimate pollutant removals. In addition, monitoring of stream restoration often includes both storm events and base flow. Each condition presents a different combination of stressors on the biological community that is typically monitored for such projects.

While sampling is frequent and the schedule hectic, the duration of BMP-level monitoring is usually short, typically ranging from one to two years to no more than five years. Paired studies, random block designs and similar research-type study designs are often used when inflow-outflow monitoring is not possible. The intent of these studies is to eliminate all factors other than the BMP itself, so samples are collected as close as possible to the BMP.

Composite samples are collected in many BMP studies, but because it is often desirable to assess contaminant levels at different stages of the hydrograph, BMP studies may include analysis of several samples per storm event. BMP monitoring is often more expensive per site than either statewide or watershed monitoring.

2.3.2.3.1 Plot

The plot scale is generally used in monitoring designs that feature replication, particularly for research objectives. This scale is not appropriate for problem assessment, pollutant load estimation, or trend analysis but can be used for preliminary assessment of the effectiveness of BMPs.

Monitoring at this scale will focus on storm-event monitoring, generally requiring automatic samplers, continuous flow measurement, and considerable annual expense. Rainfall simulation is often used at this scale to control study conditions. Monitoring duration is generally less than three years (USDA-NRCS 2003).

2.3.2.3.2 Field

Field scale study units are larger than individual plots but can vary considerably in size (USDA-NRCS 2003). Field scale studies in urban settings include parking lots, rooftops and street segments. In agricultural settings, field scale studies include cropland field segments, paddocks and barnyards. A key characteristic of most field studies is that samples are taken from episodic runoff, not from waterbodies.

Monitoring at the field scale is useful for the following objectives:

- Problem assessment, especially source characterization.
- Load allocations.
- BMP effectiveness.

Field scale monitoring is not recommended for trend analysis or determining the effectiveness of watershed projects.

Monitoring at this scale is usually completed in less than five years, and studies of individual BMPs can be completed in three years or less. Sampling is typically focused on storm events, with either discrete or composite sampling depending on the specific needs of the study. Flow measurement is essential to most field scale monitoring plans.

2.3.2.4 Summary

Table 2-3 matches monitoring objectives with the appropriate monitoring scale. The best scales for assessing BMP effectiveness are plot and field, whereas the watershed scale is best for evaluating watershed projects. Both field and watershed scales can be useful for problem assessment and load estimation, whereas the watershed scale is generally best for trend analysis.

Table 2-3. Monitoring scale as a function of objective

	Objective				
Scale	Problem Assessment	TMDL Loads	Trends	BMP Effectiveness	Watershed Project Evaluation
Plot				X	
Field	X	X		X	
Watershed	X	X	X		X

2.4 Monitoring Design Selection

2.4.1 General Considerations

As discussed above, monitoring objectives drive decisions on the details of a monitoring program. There are several experimental designs that can be applied to meet monitoring objectives, and some of the choices are obvious. All else being equal, the monitoring design selected should be the one that best matches available resources and presents the fewest logistical obstacles. Although this may seem obvious, it is important that monitoring design be determined before monitoring begins to ensure that suitable data are collected to meet monitoring objectives. Our discussion addresses monitoring design as a direct function of monitoring objective, modeled after guidance developed for agricultural monitoring projects (USDA-NRCS 2003).

2.4.2 Design Options

The design options discussed in this section are:

- Reconnaissance or synoptic
- Plot
- Paired
- Single watershed before/after
- Single-station long-term trend
- Above/below
- Side-by-side
- Multiple
- Input/output

All of the above designs are targeted designs, which are the most common type of design used to evaluate BMPs or plan and evaluate watershed projects. With the exception of the input/output design, all of the designs are applicable to biological monitoring; however, the discussions that follow focus on the measurement of chemical and physical parameters via in-stream measurements and the collection of water samples. Chapter 4 provides more discussion on biological monitoring and provides some discussion on probabilistic designs, which are particularly useful for providing unbiased assessments of conditions in a waterbody or across a large geographic area. Section 2.5 of the 1997 guidance (USEPA 1997) also provides a detailed discussion of probabilistic monitoring designs. Detailed discussions of statistical tests recommended for each design can be found in chapter 7.

2.4.2.1 Reconnaissance or Synoptic

Reconnaissance or synoptic studies are designed to provide a preliminary, low-cost overview or summary of water quality conditions in the area of interest. Reconnaissance surveys are often used to (USDA-NRCS 2003):

- Determine the magnitude and extent of a problem.
- Obtain preliminary data where none exist.
- Target critical areas.

Data collected from reconnaissance surveys are generally used in the problem assessment and planning phases of watershed projects but can also be used to help design projects to evaluate BMPs. Reconnaissance surveys typically involve a relatively large number of sampling sites distributed across the study area, low sampling frequencies (e.g., one or two samples at each site for high/low flow or seasons), a core set of common monitoring variables with or without additional variables (e.g., pesticide scans) selected based on knowledge of specific problems or sources, and a short study duration (e.g., completed in under 12 months) (Figure 2-11). A common strategy is to sample significant tributaries, longitudinally along primary streams, and at locations indicative of selected land uses (including undisturbed).

Permanent monitoring stations are not installed for reconnaissance surveys and grab sampling is typically used. Integrated grab samples, sediment samples, the use of multi-parameter probes, and instantaneous flow measurements may be part of the sampling scheme depending upon the purpose of the study.

Because sample sizes are generally small, statistical analyses are not usually performed on reconnaissance data. Instead, it is common to gather all potentially useful data on water quality and land use and management, summarize the data in graphic or tabular form, and then interpret the data using best professional judgment. For example, pollutant concentrations can be plotted against water quality criteria to identify potential problem areas. Water chemistry data might be examined to see where pollutant concentrations are highest and lowest and whether any patterns might exist along the course of a stream or in tandem with patterns found in biological monitoring data. By superimposing water quality data summaries on top of a land use map, it may be possible to identify critical areas where BMPs are needed, particularly in cases where information has also been gained from visual surveys.

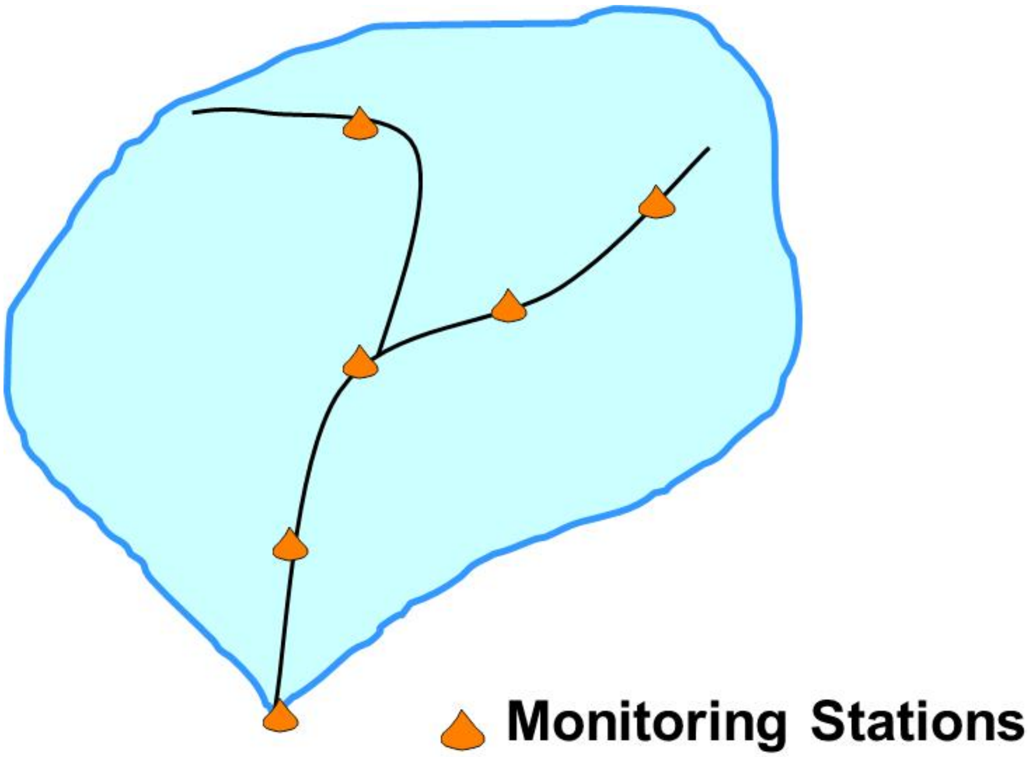


Figure 2-11. Reconnaissance sampling design

2.4.2.2 Plot

Plot studies are often used to:

- Assess soil conditions including nutrient levels.
- Assess pollutant transport pathways.
- Determine the effects of BMPs on pollutant transport.

Plot studies are typically used in research and are sometimes established to provide information that is important to a broader watershed study. For example, the effectiveness of agricultural drainage water management practices, which is important to the broader Mississippi River Basin nutrient management effort (MRGOMWNTF 2008), could be addressed with plot studies. The ability of catch or cover crops to reduce N loading to the Chesapeake Bay is another example of a research question that could be addressed with plot studies. Findings from the plot studies could be incorporated into planning and modeling efforts for the large basins.

Nonpoint source plot studies typically employ a randomized complete block design or other, more complicated statistical design such as the Latin square and split-plot designs, or a factorial arrangement of treatments (USDA-NRCS 2003). Blocking is the arranging of study plots in groups (blocks) that are similar to one another (e.g., slope, soils, pavement type). The blocking factor is generally a source of variability that is not of primary interest to the investigator. The study plots within each block include a control and a treatment that is of primary interest such as different levels or forms of nutrient application, different cover crops, different levels of crop residue, or different street sweeping frequency. Each set of treatments is generally replicated at least three times, and each block is considered a replicate. The number of plots required for the study is:

Plots = t×b

Where t = number of treatments including the control
b = number or blocks

Data from plot studies are typically analyzed using analysis of variance methods which are described both in chapter 7 and in chapter 4 of the 1997 guidance.

Plot studies using randomized complete block designs are common in the literature (Lentz and Lehrsch 2010, Wilson et al., 2010). Runoff plot studies usually require construction of permanent monitoring stations equipped with weather stations, automatic sampling equipment, and continuous flow measurement devices. In some cases remote access is provided. Samplers may be programmed to collect either discrete samples over the course of individual storm events or composite samples. Sampling is intensive and storm-event based, but studies are generally completed in three or fewer years.

Advantages of plot studies include the use of replicates and tight experimental control, but the results of plot studies are not generally widely transferrable (USDA-NRCS 2003).

2.4.2.3 Paired

Paired-watershed design is the most powerful design option that is used with any frequency to evaluate the impacts of BMPs or projects (USDA-NRCS 2003, USEPA 1993b, Hewlett and Pienaar 1973). It has been used with success in a number of Section 319 NNMP projects, including Jordan Cove, CT (Clausen 2007) and Morro Bay, CA (CCRWQCB and CPSU 2003). The basic design requires two watersheds (treatment and control) and two time periods (calibration and treatment) as illustrated in Figure 2-12, but the design can include more than one treatment watershed. The discussion here is limited to a two-watershed design. Paired samples are collected from both watersheds during the calibration period and regression analysis is used to test for a relationship between the paired samples (USEPA 1993b). After the calibration period relationship is established, BMPs are implemented in the treatment watershed, paired samples are collected, and a new relationship is established between paired samples collected during the treatment period. At the end of the treatment period, the significance of the effect of the BMPs is determined using analysis of covariance (see section 7.8.2). A helpful narrated video describing the paired-watershed design can be found at the Jordan Cove, CT project website (Dietz 2006).

Selection of Paired Watersheds

- Similar size and location
- Similar slope, soils, and land cover
- Similar runoff and base flow patterns
- Similar relationships between monitored variables and flow
- Ability to control and document land use and land treatment in both watersheds

A variation on the paired-watershed design is the nested-paired-watershed design in which both monitoring stations are located in the same watershed (Hewlett and Pienaar 1973). In this design, it is preferred that the upper sub-watershed is used as the control, and the lower portion of the watershed is treated. This alignment will reduce the chances that the treatment basin will influence the control as the treatment effect passes through it (Hewlett and Pienaar 1973). The nested-paired watershed design is essentially equivalent to the above/below-before/after design described in section 2.4.2.6.

The first step in establishing a paired study is the selection of two watersheds or study areas that are likely to have qualitatively similar responses to precipitation events. This means that both watersheds will generally have runoff when it rains, base or no flow in the absence of runoff events, and a relationship between the concentrations of measured variables and flow. If these general responses to precipitation patterns differ significantly, good calibration may not be possible. Several attempts at paired studies have failed because the watersheds could not be calibrated. Choosing watersheds that are similar in size, slope, location, soils and land cover will increase the chances of selecting a good pair (USEPA 1993b).

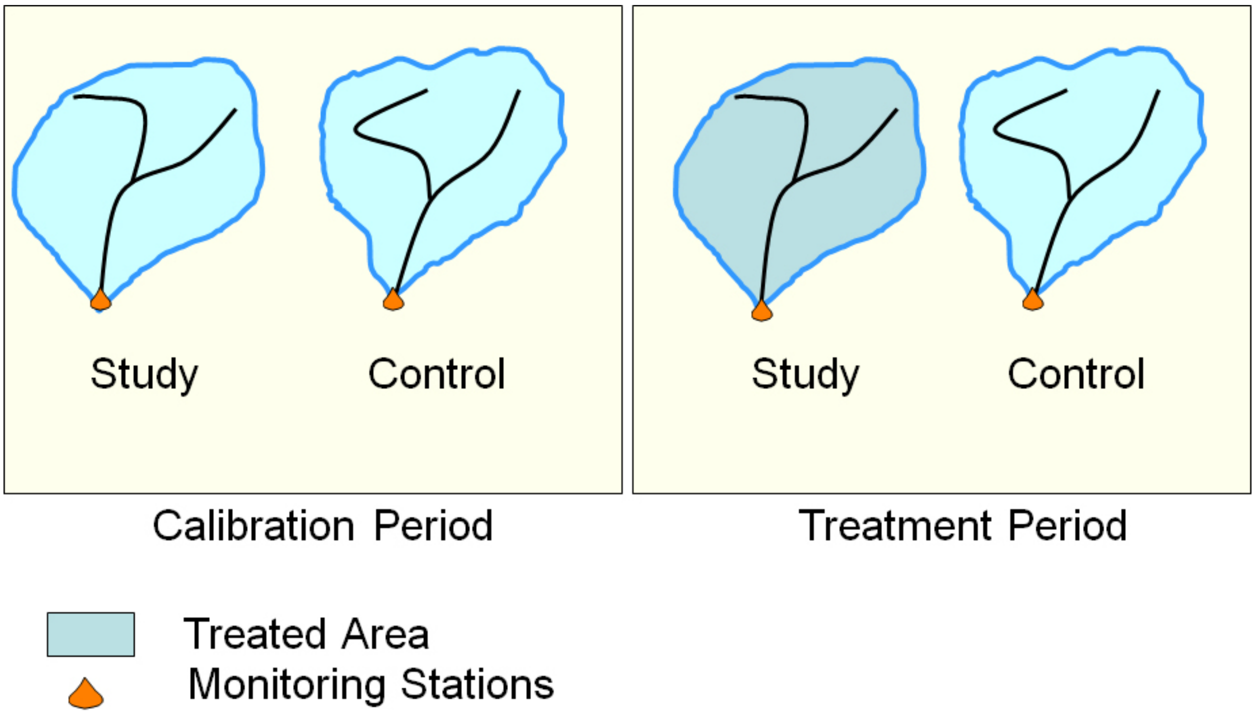


Figure 2-12. Paired sampling design

Note that the control watershed can either be in a similar impaired condition as the treatment watershed or in the improved condition desired for the treatment watershed. In the first case, the expected result is that conditions in the treated watershed get better than in the control watershed, whereas in the second case the expected result is that conditions in the treated watershed become more like those in the control watershed. In VT, for example, Meals (2001) successfully used a highly-impaired agricultural watershed as a control, measuring the beneficial effects of treatment against that watershed. The essential characteristic of a control watershed is that it does not receive BMP treatment during the life of the monitoring project.

After a suitable pair is found, the next important hurdle is to ensure that land use and land treatment activities at both watersheds can be documented and controlled through study's duration. Too often investigators have achieved control over the "treatment" watershed without securing equivalent control over the "control" watershed, only to find that activities in the control watershed (e.g., housing developments springing up on cropland or voluntary adoption of the BMPs under study) have compromised the study.

The New York NNMP project successfully paired a dairy farm watershed with a forested watershed, but concluded that it was important to include matched measurements of flow volumes, and farm watershed measurements of event peak flow and event flow rate to account for inherent differences in watershed characteristics and hydrologic response in the analysis of P load data (Bishop et al. 2005). A forested watershed was selected as the control watershed because no significant changes were expected, as opposed to a farmed watershed where operational practices could be modified or the farm could go out of business during the 8-year study period.

Monitoring for paired studies can range from biological monitoring to grab sampling to automated sampling with permanent monitoring stations. Given the difficulty associated with orchestrating paired studies at the watershed scale, it is recommended that higher-level biological and/or automated sampling be performed to ensure collection of robust datasets. Well-designed paired studies can generally be completed in five to seven years. Projects can be extended or even interrupted in cases where BMP implementation is a lengthy process.

Advantages of this design for evaluating the effectiveness of BMPs and projects include control for hydrologic variation and inherent watershed differences and the potential for clear attribution of water quality response to the BMPs. It is also possible to examine the magnitude of the treatment effect for large versus small events or baseflow conditions, but this study requires adequate sample sizes for each data subset (Lewis 2006). Further, as demonstrated by the New York NNMP project, water quality monitoring can be suspended during BMP implementation without compromising the study design (Bishop et al. 2005). This can help address the problem of diluting the overall effect of treatment by lumping together data collected during the treatment implementation with post-treatment data (Lewis 2006). Land management tracking in both the control and treatment watersheds is recommended during BMP implementation to aid in data interpretation after water quality monitoring is resumed. Depending on the location of monitoring stations, pollutant loads estimated from paired studies could be used to support TMDLs.

Based on the NNMP experience, the greatest practical disadvantage of the paired design is the difficulty finding pairs, particularly at the watershed scale. Even when pairs are found, it is often difficult to control both watersheds, so land use and land management changes occur where and when needed to support the study. This challenge becomes even greater if multiple treatment watersheds are included in the design. A disadvantage of paired studies versus multiple-watershed studies is the conclusions pertain only to the specific watersheds and treatments tested; there is generally no practical ability to predict effectiveness of the treatments or differing levels of treatment in other watersheds from the same population (Lewis 2006). This disadvantage is because descriptors of the treatment (e.g., percent impervious area or percent of cropland in no-till) are generally not included in the statistical model tested.

2.4.2.4 Single Watershed Before/After

In this design, a single monitoring station is located at the outlet of the study area. Sampling is performed before and after the implementation of BMPs. In watersheds subject to TMDLs, this design may be considered to measure pollutant loads before and after implementation of the TMDL to determine if loads have been reduced or TMDL load targets achieved. Typically the investigator is expecting to detect step changes in the target parameter with the monitoring program.

This design is not recommended for BMP effectiveness studies because there are no control stations (as in the paired design described earlier); and BMP effectiveness cannot easily be distinguished from other confounding effects (USDA-NRCS 2003). For example, if the “before” years are relatively dry and the "after" years are relatively wet or vice-versa, the differences in water quality and loads could be due to differences in weather rather than the effects of implemented BMPs. Generally, this design does not collect data that can be used to directly separate the effects of the BMPs from those of climate and is therefore a poor choice for assessing the effectiveness of BMPs or watershed projects. However, in the case of TMDLs, it would remain possible to compare measured loads versus target loads to see if goals have been achieved although attributing success or failure to TMDL implementation would be difficult. A possible exception would be the fortunate situation where water quality improved and pollutant loads were reduced despite increased runoff during the "after" years.

While this design is generally not recommended for BMP effectiveness studies, analysis of covariance (see section 7.8.2) can be used to provide some indication of BMP effects (USDA-NRCS 2003). Under this approach, a water quality variable could be related to a climate variable such as precipitation using a method described by Striffler (1965). A change in this relationship could be attributed to the BMP, but there would be no direct estimate of reduction of the water quality variable of interest. This approach would require a longer calibration period and results are not transferable to other areas (USDA-NRCS 2003).

Depending upon the study purposes, single watershed designs may or may not require construction of permanent monitoring stations, weather stations, automatic sampling equipment or continuous flow measurement devices. For some applications, grab samples will be sufficient, whereas composite sampling with automatic samplers will usually be needed in applications that require pollutant load estimation.

2.4.2.5 Single-Station Long-Term Trend

This design has been a staple of water quality monitoring for decades and can be used to determine changes in water quality or pollutant loads over time. Single-station trend monitoring generally cannot be used to determine if BMPs improve water quality unless a very long water quality and concurrent land use/treatment monitoring program can be sustained.

Trend monitoring is similar to single watershed before/after monitoring except that there is no planned before/after period and the study duration is expected to be much longer (decades versus years). Advantages of this design include the single monitoring station, wide applicability, and the ability to account for lengthy lag times (see section 6.4) and gradual implementation of BMPs (USDA-NRCS 2003). Because of the expected longer study durations, additional considerations may be necessary due to major land use changes and data gaps; long-term commitment of resources; consistency in sampling and analytical methods over time; and tracking land use, land treatment, precipitation, and flow.

Simple statistical analyses can be used to detect trends; however, they do not indicate why the trend exists. Adjusting the trend data set for hydrologic influences can help in that regard. For example, using streamflow as an explanatory variable, it was possible to document a statistically significant reduction in sediment and TP load in Willow Creek, Michigan, storm runoff over the eight years of monitoring (Suppnick 1999). These reductions were then correlated with the percent of land in no-till.

See Meals et al. (2011) and chapter 7 for additional information on trend analysis.

2.4.2.6 Above/Below

Above/below monitoring has been used with success to assess the water quality impact of isolated sources and determine the effectiveness of BMPs at isolated sources.

In this method, design stations are located upstream (or up-gradient) and downstream (or down-gradient) of the area or source that will be treated with BMPs (Figure 2-13). When used in the planning phase of a watershed project, grab samples and instantaneous flow measurements for one or two sampling events may be sufficient, but in most cases permanent monitoring stations are constructed and equipped with automatic sampling equipment, continuous flow measurement devices, and sometimes remote access. Depending on study objectives and flow conditions, samplers could be programmed to collect discrete samples over the course of individual storm events, weekly composite samples, or some other variant. Weather stations are usually included, but one station is often sufficient for both sampling sites. Study duration for evaluating BMP effectiveness is generally three to five years, but potentially longer for larger areas with multiple sources treated with a range of BMPs.

Samples are “paired” in this design, with the intent to sample the same unit of water when it is above and then below the study area. If monitoring is performed both before and after the BMPs are implemented (above/below-before/after design), this design becomes equivalent to a nested-paired-watershed design and can be treated statistically as if it were a paired-watershed design (see 2.4.2.3).

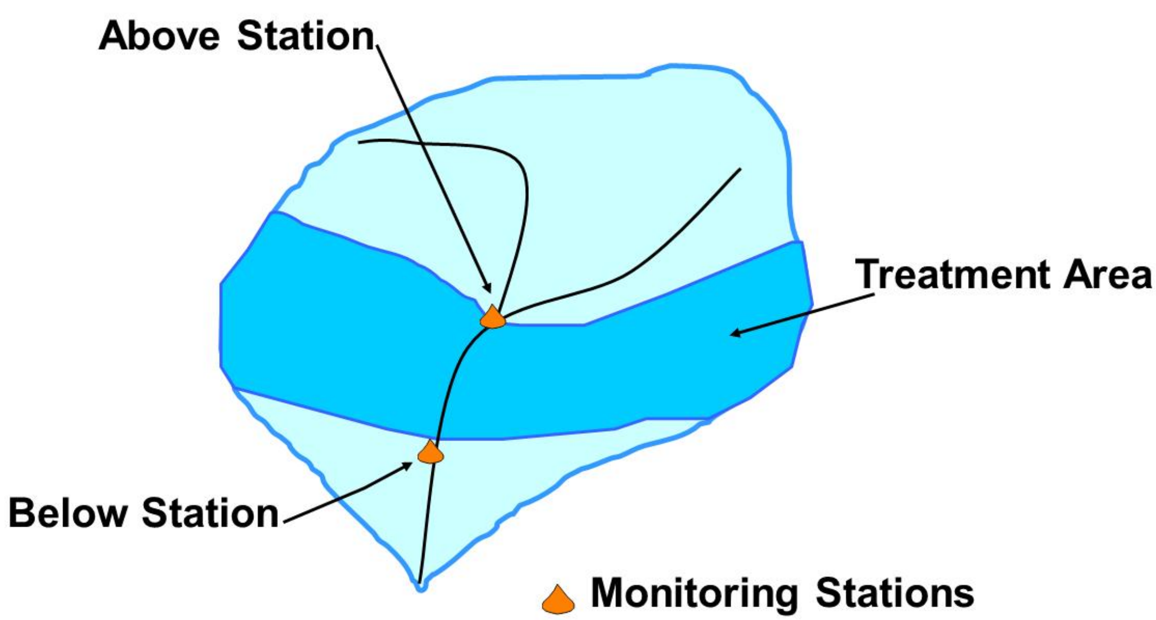


Figure 2-13. Above/below sampling design

Advantages of this design include:

- It is widely applicable.
- It is not as vulnerable to climate variability as the single watershed design.
- It is useful for isolating critical areas in the watershed.
- It can use the same statistical procedures (e.g., analysis of covariance) as a paired-watershed design because monitoring is performed before and after BMP implementation.
- Load measurements can be useful in TMDL watersheds depending on station location.

A significant disadvantage of this design is the potential for upstream conditions (e.g., high stream temperatures or pollutant concentrations) to overwhelm downstream conditions, masking both the inputs of the isolated source or area and the effects of the implemented BMPs. This risk can be addressed by ensuring the isolated area contributes substantially to the downstream flow, but this may not be feasible without sacrificing the extent of the area is isolated. It is recommended that a reconnaissance effort be undertaken with sampling at both low flow and high flow conditions to see if the site is suitable for this design.

Other disadvantages include the possibility that inherent differences between the two stations (e.g., geology) or interactions between the BMPs and the watershed could be causing the measured differences in water quality (USDA-NRCS 2003). Monitoring before and after BMP implementation can help to address these issues. Absent complete knowledge of the interaction between subsurface and stream flows, there is always the risk that unsuspected sources are impacting downstream water quality as was found in the Maryland Section 319 National NPS Monitoring Program (NNMP) project where it was determined that ammonia from the upper portion of the watershed passed through nitrifying zones and was converted to NO_3, thus elevating NO_3 levels at the outlet of the watershed (Shirmohamadi et al. 1997, Shirmohammadi and Montas 2004). A major ramification of this discovery was that nutrient management had to be added to dairy manure management and stream fencing both as a BMP needed to solve the nutrient problems in the watershed and as a factor influencing stream N levels.

2.4.2.7 Side-by-Side Before/After

Monitoring watersheds that are adjacent to each other without calibrating paired samples before treatment is equivalent to having two separate single-watershed studies as described under section 2.4.2.4. This design is not recommended for evaluation of BMPs or watershed projects because it is highly likely that there will be no way to distinguish among causal factors such as BMPs or land treatment, inherent watershed differences, or an interaction between BMPs and watershed differences (USDA-NRCS 2003).

2.4.2.8 Multiple

Where resources allow, monitoring of multiple watersheds (or fields) may be used to demonstrate the effectiveness of BMPs (USDA-NRCS 2003). This design requires that more than two watersheds are selected for monitoring within the geographic area of interest. Two different treatments and perhaps a control are replicated across the monitored watersheds in roughly equal numbers. Treatments may already be in place when the study begins or may be implemented after a pre-treatment monitoring period. An effective design would typically include monitoring over several years.

One advantage of using this approach with a large number of monitored watersheds (e.g., 10 or more) is that variability among watersheds can be estimated. Furthermore, observing water quality changes of similar direction and magnitude occurring with land treatment changes across several watersheds serves to substantiate the evidence for BMP effectiveness and robustness of the BMP over a range of watershed conditions. A major problem with this design is the practical limitations and cost associated with controlling and monitoring many watersheds or fields. It is possible, however, that the increased annual sample size associated with the inclusion of multiple watersheds could reduce the overall timeframe for monitoring.

Lewis (2006) describes a multiple-watershed approach in which 3 of 13 watersheds are used as controls, 5 are fully treated, and 5 are partially treated. He argues that this design has a significant advantage over paired-watershed studies in that it allows for prediction under different conditions or treatment levels, whereas prediction based on paired-watershed study results requires the assumed treatments are identical to the treatments used in the study.

2.4.2.9 Input/Output

Input/output design is used to evaluate the effects of individual BMPs on water quality. It is not generally useful for any of the other objectives addressed in this guidance.

Under this design, paired samples are collected at the inflow and outflow of the BMP as illustrated in Figure 2-14. For some BMPs such as manufactured devices used for urban runoff, the inflow and outflow are clearly defined and collecting paired samples is simple. For practices such as constructed wetlands, however, the inflow and outflow may be clearly defined, but collecting paired samples is challenging because retention time may not be known or may vary between runoff events. Other practices such as rain gardens can be difficult to evaluate because inflow may occur at several points or as sheet flow (Figure 2-15), and outflow may not be directly measurable because underdrains are not used.

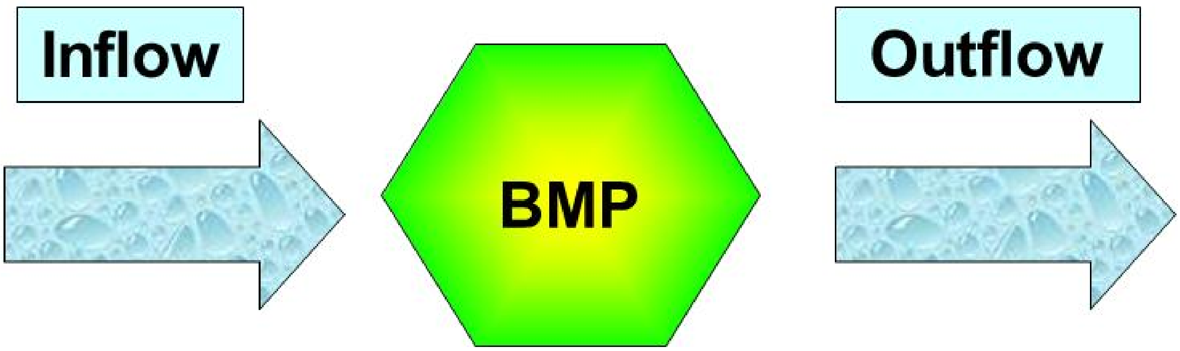


Figure 2-14. Input/output sampling design

Figure 2-15. Multiple input pathways for rain garden

Most sampling under this design will be storm-event based. Grab or automatic samples can be taken depending upon the specific practice evaluated, and either discrete or composite samples could be appropriate based upon specific study objectives. Flow measurements are essential to most BMP evaluations because performance usually varies with both flow rate and influent pollutant concentration.

With the exception of larger practices such as constructed wetlands and animal waste management systems, the possibility of having replicates and controls is an advantage of this design. No calibration period is required for inflow/outflow studies and measured pollutant reductions can be clearly attributed to the practice. A disadvantage of this design is the likelihood that results may not be widely transferable, but the relevance of this disadvantage will vary depending on the practice and study specifics. For example it may be possible to evaluate a manufactured urban stormwater device over a wide range of flow and influent concentrations at a single location, resulting in fairly widespread applicability.

Statistical methods commonly used for input-output studies are described in chapter 7 and include the paired t-test, nonparametric t-tests and the calculation of pollutant removal efficiencies. The effluent probability method is described in section 7.7.2.

2.4.2.10 Summary

Table 2-4 matches monitoring objectives with appropriate monitoring designs. Reconnaissance is best for the assessment phase of a watershed project, but above/below monitoring can also be helpful in providing information about the isolated source or area. The paired, above/below-before/after, plot, and input/output designs are generally the best designs for evaluating the effectiveness of BMPs or watershed projects. All but reconnaissance, plot, and input/output monitoring can provide useful load estimation in support of TMDLs if flow and the relevant variables are monitored. Not surprisingly, the trend design is the best for trend detection.

Some monitoring designs can be used for more than one objective depending upon the location of monitoring stations, the schedule for BMP implementation, and the duration of the monitoring program. Both the single watershed and side-by-side watershed designs could be used for trend detection if monitoring is continued in a consistent manner over a longer than planned timeframe. It may be wise to actually plan for an extension of monitoring duration for trend detection under these two designs if they fail to yield results under a before/after monitoring design. Even above/below and paired designs could be extended for trend detection but the cost associated with continued monitoring over a longer timeframe would be very high; an alternative would be to consider extended monitoring for only the downstream (for above/below) or treatment (for paired) stations with some reduction in the set of monitoring variables or monitoring frequency to reduce costs. Any changes in monitoring frequency, however, would be contingent upon the ability to meet the requirement of consistent methods throughout a trend study.

Table 2-4. Monitoring design as a function of objective

Design	Short Description	Variations	Objective: Problem Assessment	TMDL Loads	Trends	BMP or Project Effectiveness
Reconnaissance	Multiple sites distributed across study area and monitored for a short duration (<12 months)		X			
Plot	Traditional research study design with varying treatments (BMPs) replicated in randomized block design					X
Paired	Treatment and control watersheds monitored during a control and treatment period	Variation: nested-paired watershed—monitoring stations are in the same watershed		X		X
Single watershed before/after	Single station at study area outlet monitored before and after BMP implementation			X	X	
Single-station long-term trend	Single station at study area outlet monitored before and after BMP implementation	Same as single watershed before/after without BMP implementation		X	X	
Above/below-(before/after)	Stations, with paired sampling, located upstream (up-gradient) and downstream (down-gradient) of BMP	Same as nested-paired-watershed design if sampled before and after BMP implementation	X	X		X
Side-by-side before/after	Same as single watershed since there are no calibrating paired samples			X	X	
Multiple	Multiple watersheds monitored in two or more groups: treatment and control					X
Input/output	Stations located at the input and output of an individual BMP					X

2.5 References

ADEQ (Arizona Department of Environmental Quality). 2010. Oak Creek and Spring Creek Verde River watershed total maximum daily loads for *Escherichia coliform*. Open File Report 10-01, Arizona Department of Environmental Quality. Accessed January 26, 2016.

Alberts, E.E., G. E. Schuman and R.E. Burwell. 1978. Seasonal runoff losses of nitrogen and phosphorus from Missouri Valley loess watersheds. *Journal of Environmental Quality* 7(2):203-208.

Algoazany, A.S., P.K. Kalita, G.F. Czapar, and J.K. Mitchell. 2007. Phosphorus transport through subsurface drainage and surface runoff from a flat watershed in east central Illinois, USA. *Journal of Environmental Quality* 36:681–693.

Allmendinger, N.E., J.E. Pizzuto, G.E. Moglen, and M. Lewicki. 2007. A sediment budget for an urbanizing watershed, 1951-1996, Montgomery County, Maryland, U.S.A. *Journal of the American Water Resources Association* 43(6):1483-1498.

Bachman, L.J., D.E. Krantz, and J. Böhlke. 2002. *Hydrogeologic Framework, Ground-Water Geochemistry, and Assessment of Nitrogen Yield from Base Flow in Two Agricultural Watersheds, Kent County, Maryland*. EPA/600/R-02/008. U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, OH. Accessed January 26, 2016. http://water.usgs.gov/nrp/proj.bib/Publications/2002/bachman_krantz_etal_2002.pdf.

Baker, D.B. 1988. *Sediment, Nutrient, and Pesticide Transport in Selected Lower Great Lakes Tributaries*. EPA-905/4-88-001. Great Lakes National Program Office Report No. 1. Prepared for U.S. Environmental Protection Agency, Chicago, Illinois, by Heidelberg College, Water Quality Laboratory, Tiffin, OH.

Baker, D.B., K.A. Krieger, R.P. Richards, J.W. Kramer. 1985. Effects of Intensive Agricultural Land Use on Regional Water Quality in Northwestern Ohio. In Proceedings of Perspectives on Nonpoint Source Pollution, A National Conference, Kansas City, MO, May 19-22, 1985. EPA 440/5-85-001, U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

Bishop, P.L., W.D. Hively, J.R. Stedinger, M.R. Rafferty, J.L. Lojpersberger, and J.A. Bloomfield. 2005. Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. *Journal of Environmental Quality* 34:1087–1101.

Blanton, J.O. 1975. Nearshore lake currents measured during upwelling and downwelling of the thermocline in Lake Ontario. *Journal of Physical Oceanography* 5:111–124.

Böhlke, J.K., M.E. O'Connell, and K.L. Prestegaard. 2007. Ground water stratification and delivery of nitrate to an incised stream under varying flow conditions. *Journal of Environmental Quality* 36:664-680.

Brakensiek, D.L., H.B. Osborn, and W.J. Rawls. 1979. *Field Manual for Research in Agricultural Hydrology*. Agriculture Handbook No. 224. U.S. Department of Agriculture, Science and Education Administration, Washington, DC.

CBF (Chesapeake Bay Foundation). n.d. Spring and Fall Salinity Levels. Chesapeake Bay Foundation. Annapolis, MD. Accessed January 26, 2016.

CBP (Chesapeake Bay Program). 1995. *Chesapeake Bay: Introduction to an Ecosystem*. Chesapeake Bay Program, Annapolis, MD.

CCRWQCB (Central Coast Regional Water Quality Control Board) and CPSU (California Polytechnic State University). 2003. *Morro Bay National Monitoring Program: Nonpoint Source Pollution and Treatment Measure Evaluation for the Morro Bay Watershed 1992-2002 Final Report*. Prepared for U.S. Environmental Protection Agency, by Central Coast Regional Water Quality Control Board and California Polytechnic State University – San Luis Obispo, CA. Accessed February 10, 2016. http://www.elkhornsloughctp.org/uploads/files/1149032089MorroBayEstuaryProgram.pdf.

Chiles, J-P. and P. Delfiner. 1999. *Geostatistics, Modeling Spatial Uncertainty*. John Wiley, New York.

Clausen, J.C. 2007. *Jordan Cove Watershed Project Final Report*. University of Connecticut, College of Agriculture and Natural Resources, Department of Natural Resources Management and Engineering. Accessed January 8, 2016. http://jordancove.uconn.edu/jordan_cove/publications/final_report.pdf.

Coffey, S.W., J. Spooner, and M.D. Smolen. 1993. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring*. North Carolina State University, Department of Biological and Agricultural Engineering, NCSU Water Quality Group, Raleigh, NC.

Cunningham, W.L., and C.W. Schalk, comps. 2011. *Groundwater Technical Procedures of the U.S. Geological Survey*. Techniques and Methods 1–A1. U.S. Geological Survey, Reston, VA. Accessed January 26, 2016. http://pubs.usgs.gov/tm/1a1/pdf/tm1-a1.pdf.

Davenport, T.E. 1984. Comprehensive monitoring and evaluation of the Blue Creek watershed. *Lake and Reservoir Management* 1(1):570-573.

DE CIB (Delaware Center for the Inland Bays). n.d. *Wetlands and Waterways of the Inland Bays Watershed*. Delaware Center for the Inland Bays, Rehoboth Beach, DE.

Dietz, M. 2006. *Automated Introduction to the Paired Watershed Approach for Jordan Cove, Jordan Cove Urban Watershed Project.* University of Connecticut, College of Agriculture and Natural Resources, Storrs, CT. Accessed January 26, 2016. http://jordancove.uconn.edu/jordan_cove/study_design.html.

Domagalski, J.L., S. Ator, R. Coupe, K .McCarthy, D. Lampe, M. Sandstrom and N. Baker. 2007. Comparative study of transport processes of nitrogen, phosphorus, and herbicides to streams in five agricultural basins, USA. *Journal of Environmental Quality* 37(3):1158-1169.

Donald, R., G. Southam, R. Foust, R. Rottach, and B.D. Burch. 1998. *The Oak Creek Canyon Section 319(h) National Monitoring Project Final Report January, 1994-June, 1998*. Prepared for U.S. Environmental Protection Agency, by Arizona Department of Environmental Quality. Northern Arizona University, Flagstaff, AZ.

Endale, D.M., D.S. Fisher, L.B. Owens, M.B. Jenkins, H.H. Schomberg, C.L. Tebes-Stevens, and J.V. Bonta. 2011. Runoff water quality during drought in a zero-order Georgia piedmont pasture: nitrogen and total organic carbon. *Journal of Environmental Quality* 40:969–979.

EPA (Environmental Protection Agency) Victoria. 2006. *Hydrogeological Assessment (Groundwater Quality) Guidelines*. Publication 668, ISBN 0 7306 7658 7. Environmental Protection Agency Victoria, Southbank, Victoria, Australia. Accessed January 26, 2016. http://www.epa.vic.gov.au/~/media/Publications/668.pdf.

Exner, M.E., H. Perea-Estrada, and R.F. Spalding. 2010. Long-term response of groundwater nitrate concentrations to management regulations in Nebraska's Central Platte Valley. *The Scientific World Journal* 10:286-297.

FACWI (Federal Advisory Committee on Water Information). 2013. *A National Framework for Ground-Water Monitoring in the United States*. The Subcommittee on Ground Water of the Federal Advisory Committee on Water Information. Accessed January 26, 2016. http://acwi.gov/sogw/ngwmn_framework_report_july2013.pdf.

Gaines, S. and S. Airame. 2010. *Upwelling*. Ocean Explorer. U.S. Department of Commerce, National Oceanic and Atmospheric Administration. Accessed January 26, 2016. http://oceanexplorer.noaa.gov/explorations/02quest/background/upwelling/upwelling.html.

Gaydos, J. 2009. Puget Sound Natural Resources Overview. In *Puget Sound and a National Ocean Policy,* The Ocean Foundation, Seattle, WA August 24, 2009. Accessed January 26, 2016. http://www.seadocsociety.org/files/pdfs/Gaydos-Presentation-Seattle-Ocean-Foundation-Mtg-Aug-09.pdf.

Gilliom, R.J., W.M. Alley, and M.E. Gurtz. 1995. *Design of the National Water-Quality Assessment Program: Occurrence and Distribution of Water-Quality Conditions*. Circular 1112. U.S. Geological Survey. Accessed January 26, 2016. http://pubs.usgs.gov/circ/circ1112/.

GLERL (Great Lakes Environmental Research Laboratory). 2011. *Great Lakes Coastal Forecasting System*. National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory, Ann Arbor, MI. Accessed January 26, 2016. http://www.glerl.noaa.gov/res/glcfs/.

GLWWA (Grand Lake/Wabash Watershed Alliance). 2009. *Grand Lake St. Marys/Wabash River Watershed Action Plan*. Grand Lake/Wabash Watershed Alliance, Celina, OH. Accessed April 22, 2016. https://water.ohiodnr.gov/portals/soilwater/downloads/wap/Wabash.pdf.

Goodman, J., D. German, and J. Bischoff. 1996. *Ground Water Monitoring – a Guide to Monitoring for Agricultural Nonpoint Source Pollution Projects*. South Dakota Department of Environment and Natural Resources, Pierre, SD. Accessed April 22, 2016. https://archive.org/details/CAT10819301.

Haffner, G.D., M.L. Yallop, P.D.N. Hebert, and M. Griffiths. 1984. Ecological significance of upwelling events in Lake Ontario. *Journal of Great Lakes Research* 10(1):28-37.

Harmel, R.D., K.W. King, B.E. Haggard, D.G. Wren, and J.M. Sheridan. 2006. Practical guidance for discharge and water quality collection in small watersheds. *Transactions of the American Society of Agricultural and Biological Engineers* 49(4):937-948.

Hewlett, J.D. and L. Pienaar. 1973. Design and Analysis of the Catchment Experiment, In *Proceedings of a Symposium on Use of Small Watersheds in Determining Effects of Forest Land Use on Water Quality*, ed. E. H. White, University of Kentucky, Lexington, KY, May 22-23, 1973. Accessed January 26, 2016. http://coweeta.uga.edu/publications/hewlett_73_catchment.pdf.

Johnson, H.O., S.C. Gupta, A.V. Vecchia, and F. Zvomuya. 2009. Assessment of water quality trends in the Minnesota River using non-parametric and parametric methods. *Journal of Environmental Quality* 38:1018–1030.

Klatt, J.G., A.P. Mallarino, J.A. Downing, J.A. Kopaska and D.J. Wittry. 2003. Soil phosphorus, management practices, and their relationship to phosphorus delivery in the Iowa Clear Lake agricultural watershed. *Journal of Environmental Quality* 32(6):2140-2149.

Lampert, W. 1989. The adaptive significance of diel vertical migration of zooplankton. *Functional Ecology* 3:21-27.

Lee, S-J., M.L. Serre, and G. Christakos. 2005. Bayesian maximum entropy mapping of arsenic using pH-arsenic empirical laws. *ESE Research Notes* 2(1):8-15. University of North Carolina, Department of Environmental Sciences and Engineering, Chapel Hill.

Lentz, R.D. and G. A. Lehrsch. 2010. Nutrients in runoff from a furrow-irrigated field after incorporating inorganic fertilizer or manure. *Journal of Environmental Quality* 39(4):1402-1415.

Lewis, J. 2006. *Fixed and Mixed-Effects Models for Multi-Watershed Experiments*. Pacific Southwest Research Station, U.S. Forest Service, Arcata, CA. Accessed March 24, 2016. http://www.fs.fed.us/psw/publications/4351/Lewis06.pdf.

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.

Meals, D.W. 2001. *Lake Champlain Basin Agricultural Watersheds Section 319 National Monitoring Program Project, Final Project Report: May, 1994-September, 2000*. Vermont Department of Environmental Conservation, Waterbury, VT.

Meals, D.W. and S.A. Dressing. 2008. *Surface Water Flow Measurement for Water Quality Monitoring Projects*. Tech Notes #9, September 2013. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W., J. Spooner, S.A. Dressing, and J.B. Harcum. 2011. *Statistical Analysis for Monotonic Trends*, Tech Notes #6, September 2011. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Montgomery, D.R. and J.M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. *Geological Society of America Bulletin* 109: 596-611

MRGOMWNTF (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force). 2008. *Gulf Hypoxia Action Plan 2008*. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force. Accessed January 26, 2016. http://www.epa.gov/ms-htf/gulf-hypoxia-action-plan-2008.

NCSU (North Carolina State University) and Harbridge House. 1983. *The Model Implementation Program: Lessons Learned From Agricultural Water Quality Projects*. National Water Quality Evaluation Project at North Carolina State University and Harbridge House, Inc. Prepared for USDA and USEPA, Washington, DC.

Needelman, B.A., W.J. Gburek, G.W. Petersen, A.N. Sharpley, and P.J.A. Kleinman. 2004. Surface runoff along two agricultural hillslopes with contrasting soils. *Soil Science Society of America Journal* 68:914–923.

NOAA (National Oceanic and Atmospheric Administration). 1990. *Estuaries of the United States: Vital Statistics of a National Resource Base*. National Oceanic and Atmospheric Administration, Silver Spring, MD.

OCWC (Oak Creek Watershed Council). 2012. Improvement Plan for the Oak Creek Watershed, Arizona. Prepared for Arizona Department of Environmental Quality, by Oak Creek Watershed Council, Sedona, AZ. Accessed April 22, 2016.

OEPA (Ohio Environmental Protection Agency). 2013. *Statewide Biological and Water Quality Monitoring and Assessment*. Ohio Environmental Protection Agency, Division of Surface Water, Columbus, OH. Accessed April 22, 2016. http://www.epa.state.oh.us/dsw/bioassess/ohstrat.aspx.

Ohrel, R.L. Jr. and K.M. Register. 2006. *Volunteer Estuary Monitoring - a Methods Manual, Second Edition*. The Ocean Conservancy, Washington, DC. Accessed January 27, 2016.

Plattner, S., D.M. Mason, G.A. Leshkevich, D.J. Schwab, and E.S. Rutherford. 2006. Classifying and forecasting coastal upwellings in Lake Michigan using satellite derived temperature images and buoy data. *Journal of Great Lakes Research* 32:63–76.

Poff, B. and A. Tecle. 2002. Bacteriological Water Quality Trend Analysis in Oak Creek Canyon, Arizona. In *Ground Water/Surface Water Interactions*, ed. J.F. Kenny, AWRA (American Water Resources Association) Specialty Summer Conference, Keystone, CO, July 1-3, 2002, pp. 431-436. Accessed April 22, 2016.

Rich. T. 2001. *Results of Monitoring from 1989 through 1997 for the Statewide Ground Water Quality Monitoring Network*. Open-File Report 89-UR. Prepared for South Dakota Department of Environmental and Natural Resources, by University of South Dakota, Akeley-Lawrence Science Center Science Center, Vermillion, SD. Accessed April 22, 2016. http://www.sdgs.usd.edu/pubs/pdf/UR-89%20-%20text.pdf.

Rosgen, D.L. and H.L. Silvey. 1996. *Applied River Morphology*. Wildland Hydrology Books, Fort Collins, CO.

Scalf, M.R., J.F. McNabb, W.J. Dunlap, R.L. Cosby, and J.S. Fryberger. 1981. *Manual of Ground-Water Quality Sampling Procedures*. U.S. Environmental Protection Agency, Office of Research and Development, Robert S. Kerr Environmental Research Laboratory, Ada, OK.

Schilling, K.E. 2002. Chemical transport from paired agricultural and restored prairie watersheds. *Journal of Environmental Quality* 31(4):1184-1193.

Schilling, K.E. and C.F. Wolter. 2001. Contribution of baseflow to nonpoint source pollution loads in an agricultural watershed. *Ground Water* 39(1):49-58.

SDDENR (South Dakota Department of Environment and Natural Resources). 2001. *South Dakota's Ground Water Quality Monitoring Network Project Completion Report: Section 319 Nonpoint Source Pollution Program*. Prepared for U.S. Environmental Protection Agency, by South Dakota Department of Environment and Natural Resources, Division of Financial and Technical, Assistance, Geological Survey Program. Accessed April 22, 2016. http://denr.sd.gov/dfta/wp/Documents/GroundWaterMonitoring.pdf.

Sekely, A.C., D.J. Mulla, and D.W. Bauer. 2002. Streambank slumping and its contribution to the phosphorus and suspended sediment loads of the Blue Earth River, Minnesota. *Journal of Soil and Water Conservation* 57(5):243-250.

Shirmohammadi, A., K.S. Yoon, and W.L. Magette. 1997. Water quality in mixed land-use watershed – Piedmont Region in Maryland. *Transactions of the American Society of Agricultural Engineers* 40(6):1563-1572.

Shirmohamadi, A. and H. Montas. 2004. *Modeling and Monitoring the Hydrologic and Water Quality Response of the Mixed Land Use Basin FFY 03 Annual Report*. University of Maryland, Department of Biological Resources Engineering, College Park, MD.

SIO (Scripps Institution of Oceanography). 2003. *Waves, Nearshore Currents, and Tides*. Coastal Morphology Group, Scripps Institution of Oceanography and the Kavli Institute. Accessed January 28, 2016. http://coastalchange.ucsd.edu/st3_basics/waves.html.

Slovinsky, P.A.. 2008. *Coastal Circulation in the Vicinity of the Kennebunk River and Goochs Beach*. Maine Geological Survey, Augusta, ME. Accessed April 22, 2016. http://www.maine.gov/dacf/mgs/explore/marine/sites/may08.pdf.

Sommer, T.R., W.C. Harrell, and T.J. Swift. 2008. Extreme hydrologic banding in a large-river floodplain, California, U.S.A., *Hydrobiologia* 598:409-415.

Spooner, J., L.A. Szpir, D.E. Line, D. Meals, G.L Grabow, D.L. Osmond and C. Smith. 2011. *2011 Summary Report: Section 319 National Monitoring Program Projects, National Nonpoint Source Watershed Project Studies*. North Carolina State University, Biological and Agricultural Engineering Department, NCSU Water Quality Group, Raleigh, NC. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/319monitoring/toc.html.

STAC (Scientific and Technical Advisory Committee). 2005. Understanding "Lag Times" Affecting the Improvement of Water Quality in the Chesapeake Bay. STAC Publ. 05–001. Chesapeake Bay Program, Edgewater, MD. Accessed January 27, 2016.

Stich, H. and W. Lampert. 1981. Predator evasion as an explanation of diurnal vertical migration by zooplankton. *Nature* 293(1):396-398.

Striffler, W.D. 1965. The selection of experimental watersheds and methods in disturbed forest areas. In *IASH Symposium of Budapest 1965 - Representative and Experimental Basins*, Publication No. 66, pp. 464-473, International Association of Hydrological Sciences. Accessed April 22, 2016. http://hydrologie.org/redbooks/a066/iahs_066_0464.pdf.

Stuntebeck, T.D., Komiskey, M.J., Owens, D.W., and Hall, D.W. 2008. *Methods of Data Collection, Sample Processing, and Data Analysis for Edge-of-Field, Streamgaging, Subsurface-Tile, and Meteorological Stations at Discovery Farms and Pioneer Farm in Wisconsin, 2001–7*. Open-File Report 2008–1015. U.S. Geological Survey, Reston, VA. Accessed April 22, 2016. http://pubs.usgs.gov/of/2008/1015/pdf/ofr2008-1015.pdf.

Suppnick, J. 1999. *Water Chemistry Trend Monitoring in Sycamore Creek and Haines Drain, Ingham County, Michigan 1990-1997*. Staff report MI/DEQ/SWQD-99-085. Michigan Department of Environmental Quality, Surface Water Quality Division. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Taylor, C.B. and H.B. Stewart. 1959. Summer upwelling along the east coast of Florida. *Journal of Geophysical Research* 64(1): 33-40.

Taylor, C.J. and W.M. Alley. 2001. *Ground-Water-Level Monitoring and the Importance of Long-Term Water-Level Data*. Circular 1217, U.S. Geological Survey, Denver. CO. Accessed January 28, 2016. http://pubs.usgs.gov/circ/circ1217/pdf/circular1217.pdf.

Tetra Tech. n.d. *Section 319 National Monitoring Program Projects Lessons Learned*. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Tiner, R.W. 2002. *In Search of Swampland – A Wetland Sourcebook and Field Guide*. Rutgers University Press, New Brunswick.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2003. *National Water Quality Handbook*. 450–VI–NWQH. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC. Accessed January 29, 2016. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044775.pdf.

USEPA (U.S. Environmental Protection Agency). 1983. *Results of the Nationwide Urban Runoff Program: Volume 1 - Final Report*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 28, 2016. http://www.epa.gov/npdes/pubs/sw_nurp_vol_1_finalreport.pdf.

USEPA (U.S. Environmental Protection Agency). 1990. *The Lake and Reservoir Restoration Guidance Manual*. EPA-440/4-90-006. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1993a. *Evaluation of the Experimental Rural Clean Water Program*. EPA-841-R-93-005. U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1993b. *Paired Watershed Study Design*. EPA 841-F-93-009. U.S. Environmental Protection Agency, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997. *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls*. EPA 841-B-96-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-guidance-determining-effectiveness-nonpoint.

USEPA (U.S. Environmental Protection Agency). 1998. *Nearshore Waters and Your Coastal Watershed*. EPA 842-F-98-007. U.S. Environmental Protection Agency, Washington, DC. Accessed January 28, 2016.

USEPA (U.S. Environmental Protection Agency). 2001. *EPA Requirements for Quality Assurance Project Plans, EPA QA/R-5*. EPA 240/B-01/003. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002a. *Methods for Evaluating Wetland Condition: #16 Vegetation-Based Indicators of Wetland Nutrient Enrichment*. EPA-822-R-02-024. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 15, 2016. http://www.epa.gov/sites/production/files/documents/wetlands_16indicators.pdf.

USEPA (U.S. Environmental Protection Agency). 2002b. *Methods for Evaluating Wetland Condition: Study Design for Monitoring Wetlands*. EPA-822-R-02-015.U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2003a. *Elements of a State Water Monitoring and Assessment Program*. EPA 841-B-03-003. U.S. Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans and Watersheds, Assessment and Watershed Protection Division, Washington, DC. Accessed January 28, 2016.

USEPA (U.S. Environmental Protection Agency). 2003b. *Guidance for 2004 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d) and 305(b) of the Clean Water Act, July 21, 2003*. United States Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans, and Watersheds, Assessment and Watershed Protection Division, Watershed Branch, Washington, DC. Accessed January 28, 2016.

USEPA (U.S. Environmental Protection Agency). 2004. *Review of Rapid Assessment Methods for Assessing Wetland Condition*. EPA/620/R-04/009. U.S. Environmental Protection Agency, National Health and Environmental Effects Laboratory, Corvallis, OR.

USEPA (U.S. Environmental Protection Agency). 2006a. *Application of Elements of a State Water Monitoring and Assessment Program For Wetlands*. U.S. Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans, and Watersheds, Wetlands Division, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2006b. Guidelines for the Award of Monitoring Initiative Funds under Section 106 Grants to States, Interstate Agencies, and Tribes. U.S. Environmental Protection Agency, Washington, DC. *Federal Register*, March 29, 2006, 71:15718-15724. Accessed January 28, 2016. https://federalregister.gov/a/E6-4585.

USEPA (U.S. Environmental Protection Agency). 2008a. Amendment to the Guidelines for the Award ofMonitoring Initiative Funds under Section 106 Grants to States, Interstate Agencies, and Tribes. U.S. Environmental Protection Agency, Washington, DC. *Federal Register*, July 17, 2008, 73:41069-41076. Accessed January 28, 2016. https://federalregister.gov/a/E8-16385.

USEPA (U.S. Environmental Protection Agency). 2008b. *Handbook for Developing Watershed Plans to Restore and Protect Our Waters.* EPA 841-B-08-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 10, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/handbook-developing-watershed-plans-restore-and-protect.

USEPA (U.S. Environmental Protection Agency). 2011. *Aquatic Resource Monitoring - Frequently Asked Questions - Survey Sampling*. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division, Corvallis, OR. Accessed January 28, 2016. http://archive.epa.gov/nheerl/arm/web/html/designfaqs.html.

USEPA (U.S. Environmental Protection Agency). 2012a. *EPA Program Guidance*. U.S. Environmental Protection Agency, Washington, DC. Accessed January 28, 2016. http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/guidance.cfm#program.

USEPA (U.S. Environmental Protection Agency). 2012b. *Starting Out in Volunteer Monitoring*. EPA 941-F002-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed January 28, 2016.

Wagner, R.J., R.W. Boulger, Jr., C.J. Oblinger, and B.A. Smith. 2006. *Guidelines and Standard Procedures for Continuous Water-Quality Monitors—Station Operation, Record Computation, and Data Reporting*. Techniques and Methods 1–D3. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.usgs.gov/tm/2006/tm1D3/.

WAL (Wisconsin Association of Lakes). 2009. *Lake Types: How Does Water Get into Your Lake?* Wisconsin Association of Lakes, Madison, WI. Accessed January 28, 2016. http://www.wisconsinlakes.org/index.php/the-science-of-lakes/21-lake-types.

Welch, E.B. and G.D. Cooke. 1999. Effectiveness and longevity of phosphorus inactivation with alum. *Journal of Lake and Reservoir Management* 15:5-27.

Wetzel, R.G. 1975. *Limnology*. W.B. Saunders Company, Philadelphia, PA.

Wilde, F.D. 2005. Preparations for Water Sampling. Book 9, Chapter A1 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.water.usgs.gov/twri9A1/.

Wilde, F.D., ed. 2006. Collection of Water Samples (Version 2.0). Book 9, Chapter A4 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.water.usgs.gov/twri9A4/.

Wilson, M.L., C.J. Rosen, and J.F. Moncrief. 2010. Effects of polymer-coated urea on nitrate leaching and nitrogen uptake by potato. *Journal of Environmental Quality* 39: 2: 492-499.

Winter, T.C., J.W. Harvey, O.L. Franke, and W.M. Alley. 1998. *Ground Water and Surface Water a Single Resource*. Circular 1139. U.S. Geological Survey, Denver. Accessed January 28, 2016. http://pubs.usgs.gov/circ/circ1139/index.html.

Zaret, T.M. and J.S. Suffern. 1976. Vertical migration in zooplankton as a predator avoidance mechanism. *Limnology and Oceanography* 21(6):804-813.

This page intentionally left blank.

3 Monitoring Plan Details

By D.W. Meals, S.A. Dressing, J. Spooner, and J.B. Harcum

In chapter 2 we discussed monitoring objectives, the fundamentals of good monitoring, the selection of an appropriate geographic scale for monitoring, and the selection of a basic monitoring design. In this chapter we discuss the nuts and bolts of monitoring, beginning with the selection of variables and concluding with data reporting and presentation. Because the emphasis of this guidance is placed on monitoring for watershed-level problem assessment, load estimation, trend analysis, and the effectiveness of BMPs or watershed projects, the details that follow in this and subsequent chapters will be centered on these objectives.

3.1 Variable Selection

Monitoring variables are often grouped into three general categories:

- Physical (e.g., flow, temperature, or suspended sediment)
- Chemical (e.g., DO, P, atrazine)
- Biological (e.g., *E. coli* bacteria, benthic macroinvertebrates, fish)

It is usually most appropriate for projects to monitor a mix of variables, although some projects may focus in one specific area such as physical measurements. Variables are often interrelated across these categories. For example, DO concentrations and temperature influence the fish assemblage present at a site.

Selection of the appropriate variables to monitor is a crucial task. A monitoring program cannot afford to measure every single variable nor should a project attempt to do so because some variables contribute to achieving project goals more than others. It is wasteful to measure characteristics that are unimportant or irrelevant to project objectives, and it is equally problematic to fail to measure key variables. In general, it is better to monitor a minimum set of variables well than a large number of variables poorly (e.g., minimal sampling frequency and/or duration).

3.1.1 General Considerations

The selection of which variables to measure in a monitoring program requires consideration of several important factors. It is important to resist the temptation to measure more variables than are needed for the project or to adopt a generic list of traditionally monitored water quality variables. The final design of a monitoring program often represents a compromise based on balancing information requirements, budget, personnel, and other constraints. Excess resources spent on analyzing unnecessary variables may force a reduction in the number of sampling stations, the sampling frequency, or the duration of monitoring, which can threaten program effectiveness.

The following sections discuss important factors to be considered when selecting variables to monitor. Variables commonly measured in watershed nonpoint source monitoring efforts are also discussed.

3.1.2 Selection Factors

Several factors should be considered when selecting variables to measure in a monitoring project. These factors, discussed below, are:

- Program objectives
- Waterbody use
- Water resource type
- Use impairment
- Pollutant sources
- Expected response to treatment
- Difficulty or cost of analysis
- Logistical constraints
- Need for covariates
- Priorities

3.1.2.1 Program Objectives

The overarching principle of monitoring variables selection is that the variables should be tied directly to the program objectives with due consideration of the other factors described in this section. In many cases, the stated program objective will clearly indicate the appropriate variable(s) to monitor. For example, an objective to document the effectiveness of BMPs on *E. coli* levels at a public beach clearly calls for measurement of *E. coli* bacteria. An objective to reduce TP loading to a lake would suggest measuring TP (perhaps not dissolved P) concentration and measuring flow because both concentration and flow data are required to calculate load (see section 3.8 and section 7.9). An objective to restore a fishery might require, at a minimum, monitoring the fish population as well as chemical (e.g., DO, ammonium) and physical (e.g., temperature, substrate) variables that support acceptable fish habitat.

It is more challenging to select monitoring variables when program objectives are less specific. For monitoring aimed at assessing water quality standards compliance or TMDL implementation, the selected variables should focus on what is required to assess water quality standards violations or TMDL achievement. For monitoring objectives that involve watershed reconnaissance or characterization, other factors such as the nature of the impairment, type of water resource, or likely pollutant sources must be considered.

3.1.2.2 Waterbody Use

Variable selection may be driven by a waterbody's designated use. Designated uses are one of three elements contained in water quality standards. The other elements are water quality criteria to protect those uses and determine if they are being attained, and antidegradation policies to help protect high quality water bodies (USEPA 2011c). States and tribes designate water bodies for specific uses based on their goals and expectations for their waters. Typical designated uses include:

- Protection and propagation of fish, shellfish, and wildlife
- Recreation
- Public water supply

- Agricultural, industrial, navigational, and other purposes.

Numeric and narrative water quality criteria are set to protect each designated use by describing the chemical, physical and biological conditions necessary for safe use of waters by humans and aquatic life. These criteria should be used to help guide variable selection and other monitoring details (e.g., sampling period and frequency) where use attainment or protection is the primary monitoring concern. Failure to meet some or all of the applicable water quality criteria can result in less than full support of designated uses.

For example, monitoring of a waterbody used for recreation might emphasize sediment, nutrient, or bacteria variables because these help define the aesthetic and health-related character of the waterbody. However, variables for monitoring irrigation water might include total dissolved solids and salinity variables and exclude less relevant biological variables. General applicability of water monitoring variable groups to selected designated uses is shown in Table 3-1.

3.1.2.3 Waterbody Use Impairment

Monitoring of waterbodies with documented use impairments can differ substantially from monitoring to assess use attainment or protection. For example, the impairment could be the result of a single pollutant (e.g., violation of a turbidity criterion) or failure to achieve one portion of a narrative criterion (e.g., fish assemblage), rather than a failure to meet all applicable criteria. In these situations, monitoring can be focused on the specific variables violating criteria instead of all potential variables indicated by the applicable water quality standard. While the variable list associated with criteria may be narrowed, additional variables should be considered to address the causes of the violation(s). For example, turbidity problems could be caused by streambank erosion, high phytoplankton production, or wash from impervious surfaces. Fish assemblage could be impacted by a number of factors such as lack of suitable flow or cover, water quality, or physical obstructions. For projects with an objective to relate water quality changes to pollution control efforts, it is essential to track variables associated with the causes of identified water quality problems.

3.1.2.4 Type of Water Resource Sampled

Variables monitored should be suitable for the type of waterbody under study. Appropriate variables often differ significantly between surface and ground water and between streams and lakes. Examples of variable groups that can be applicable to different water resource types are shown in Table 3-2.

3.1.2.5 Pollutant Sources

Variables monitored should reflect the nonpoint sources known or suspected to be present in the watershed. Crop agriculture, for example, is likely to influence suspended sediment, turbidity, nutrients and pesticides measured in water. The presence of intensive livestock agriculture in a watershed would justify measuring biochemical oxygen demand (BOD), nutrients and indicator bacteria. Urban stormwater sources are likely to influence variables such as discharge, temperature, turbidity, metals and indicator bacteria. Examples of variable groups that can be responsive to different nonpoint source activities are shown in Table 3-3.

Table 3-1. Monitoring variable groups by direct relationship to selected designated water use (adapted from USDA-NRCS 2003)

Variable	Designated Use				
	Aquatic life support	Contact recreation	Aesthetics	Irrigation	Drinking water supply
Physical					
Discharge	X				
Dissolved oxygen (DO)	X		X		X
Salinity	X			X	X
Secchi disk transparency	X	X	X		
Specific conductance	X			X	X
Suspended sediment	X	X	X		X
Temperature	X				
Total dissolved solids (TDS)	X			X	X
Turbidity	X	X	X		X
Chemical					
BOD	X		X		
Inorganics (Cl, F)	X			X	X
Metals (As, Cd, Cr, Cu, Fe, Hg, Pb, Zn)	X	X		X	X
Nutrients (N, P) – dissolved	X		X	X	
Nutrients (N, P) – total/particulate	X		X		
pH	X			X	X
Biological					
Benthic macroinvertebrates	X				
Chlorophyll a	X	X	X		X
Fish	X				
Indicator bacteria (fecal coliform, *E. coli*)		X			X
Macrophytes	X		X		
Pathogens (*Giardia, Cryptosporidium*)		X			X
Plankton (algae)	X		X		X

3.1.2.6 Response to Treatment

In a monitoring program designed to evaluate water quality response to management measure implementation, it is critical that monitored variables focus on dimensions of water quality expected to change in response to treatment. For example, an agricultural watershed uses conservation tillage as the principal management measure implemented to address an erosion problem. The water quality monitoring program should measure flow, peak flow, suspended sediment, and turbidity as variables likely to respond to widespread changes in tillage practices. It would be less appropriate to monitor for *E. coli*, even if *E. coli* standards are also violated in the watershed, unless land application of organic wastes in the watershed occurs in the watershed.

Table 3-2. Monitoring variables by selected water resource types (adapted from USDA-NRCS 2003)

Variable	Lake	Stream	Wetland	Ground Water
Discharge	X	X	X	
Dissolved oxygen	X	X	X	X
Habitat	X	X	X	
Riffle/pool ratio		X		
Salinity	X	X	X	X
Secchi disk transparency	X			
Specific conductance	X	X	X	X
Substrate characteristics	X	X	X	
Suspended sediment	X	X	X	
Temperature	X	X	X	
Total dissolved solids	X	X	X	X
Turbidity	X	X	X	
BOD	X	X	X	
Inorganics (Cl, F)		X	X	X
Metals (As, Cd, Cr, Cu, Fe, Hg, Pb, Zn)	X	X	X	X
Nutrients (N, P) – dissolved	X	X	X	X
Nutrients (N, P) – total/particulate	X	X	X	
pH	X	X	X	X
Benthic macroinvertebrates	X	X	X	
Chlorophyll a	X	X		
Fish	X	X	X	
Indicator bacteria (fecal coliform, *E. coli*)	X	X	X	X
Macrophytes	X	X	X	
Pathogens (*Giardia, Cryptosporidium*)	X	X	X	X
Plankton (algae)	X	X	X	

Research has shown that some BMPs can have unintended side effects. For example, increasing conservation tillage may result in increased herbicide use or increased concentrations and delivery of soluble nutrients. While conservation tillage has been shown to greatly reduce sediment bound P, P can become concentrated at the soil surface because of the lack of mixing by tillage, resulting in significant losses of soluble P in runoff (Beegle 1996). In these situations, it is advisable to monitor either TP or both particulate and dissolved P to ensure that BMP effectiveness is accurately assessed. Decisions on whether to track these variables, including adding subsurface monitoring sites, should be made at the beginning of a monitoring program.

Table 3-3. Monitoring variable groups by selected nonpoint source activities (adapted from USDA-NRCS 2003)

Variable	Nonpoint Source Activity				
	Crop Agriculture	Livestock Agriculture	Construction	Mining	Urban Stormwater
Physical					
Discharge	X	X	X	X	X
Dissolved oxygen	X	X		X	X
Salinity	X	X			
Secchi disk transparency	X	X	X		X
Specific conductance				X	X
Suspended sediment	X	X	X	X	X
Temperature			X	X	X
Total dissolved solids	X	X	X	X	X
Turbidity	X	X	X	X	X
Chemical					
BOD	X	X			X
Inorganics (Cl, F)				X	X
Metals (As, Cd, Cr, Cu, Fe, Hg, Pb, Zn)				X	X
Nutrients (N, P) – dissolved	X	X	X		X
Nutrients (N, P) – total/particulate	X	X	X		X
pH				X	
Biological					
Benthic macroinvertebrates	X	X	X	X	X
Chlorophyll a	X	X	X	X	X
Fish	X		X	X	X
Indicator bacteria (fecal coliform, *E. coli*)		X			X
Macrophytes	X	X		X	X
Pathogens (*Giardia*, *Cryptosporidium*)		X			X
Plankton (algae)	X	X		X	X

3.1.2.7 Difficulty or Cost of Analysis

The difficulty and cost of analysis must be considered in the selection of variables to monitor. While other factors like program objectives and pollutant sources should be more important criteria in the selection process, cost of analysis often drives choices among suitable variables because of budget constraints. Analytical costs will vary by region of the country and by laboratory. In-house laboratories, such as a university or a state agency, may have lower unit costs than an independent contract laboratory.

Some representative analytical costs are shown in Table 3-4. For several monitoring objectives, alternative monitoring variables that are lower cost may be available. For example turbidity analysis is half the cost of suspended sediment; a total dissolved solids measurement is about twice the cost of a laboratory analysis of specific conductance. Field measurement of conductivity is even cheaper if the equipment is available. These pairs of variables are likely to be highly correlated, making the lower cost alternative possibly the best choice (see section 3.1.3.3 for a discussion of surrogates). However, this will not always be the case and cost alone should not be a primary criterion for variable selection. For example, a lower-cost analysis for NO_3-N ($17) measures an entirely different form of nitrogen from TKN.

Table 3-4. Representative laboratory analytical costs for selected water quality variables. Costs will vary by region and by laboratory (Dressing 2014)

Variable	Cost per analysis ($)
NO_3-N	17
TKN	35
TN	20
Soluble reactive P	15
Total P	22
Turbidity	8
Suspended sediment	16
Specific conductance	8
Total dissolved solids	15
Pesticide scan	135
COD	25
Oil and grease	45
Lead (ICP)	15
Invertebrates	150

It should also be noted that many variables can be analyzed by different methods that have both different costs and different levels of sensitivity. For example, a lead analysis by inductively coupled plasma (ICP) has a cost of $15/analysis using EPA method 200.9 (Barnstable County 2016) and a method detection limit of 0.7 µg/L (Creed et al. 1994). Compare this to a lead analysis by EPA method 200.5 with a method detection limit of 1.1 µg/L (Martin 2003) at a cost of $29/analysis (PSU 2016). Project objectives, data quality objectives and pollutant sources would factor into the trade-off between cost and sensitivity. Specific analytical methods can be further investigated in the *National Environmental Methods Index* (NEMI) at www.nemi.gov.

Finally, it should be noted that analytical costs, while potentially high, are often considerably lower than other categories of project costs, particularly personnel costs (see chapter 9). While cost alone is an important consideration, it cannot be the primary driver of variable selection. If monitoring of the appropriate variables cannot be correctly performed, money spent on monitoring is wasted.

3.1.2.8 Method Comparability

Advances in sampling and analytical methods are common. While these advances are welcomed on the one hand by reducing interference and improving reliability and accuracy, they can introduce challenges during the course of the project or when trying to design a new project that takes advantage of existing data. For example, it is wrong to compare historical turbidity data determined by the Jackson Candle method (units: Jackson Turbidity Unit or JTU) with turbidity data collected from a calibrated nephelometer (units: Nephelometric Turbidity Units or NTU). This caution extends to practically every phase of the monitoring program, from field sampling, sample preservation, and laboratory procedures. Ensuring that data from multiple methods can be compared is critical. One approach is to perform a comparability study by implementing both methods with laboratory splits and comparing the resulting paired data. Depending on the results, it is prudent for projects with limited duration to continue with an older method rather than updating to a new method.

3.1.2.9 Logistical Constraints

Logistical issues like refrigeration availability at a sampling station or travel time between field sites and the laboratory may constrain selection of monitoring variables. Most water quality variables have specified permissible holding times and holding conditions. These parameters determine the length of time a sample can be stored after collection and prior to analysis without significantly affecting the analytical results. Maximum holding times and storage conditions have been established by the U.S. EPA (40 CFR 136.3, USEPA 2008b). Examples of these specifications are shown in Table 3-5.

Holding times and conditions will influence the choice of analytical variables. Unless samples can be delivered to the laboratory within six hours, *E. coli* analysis may be impractical. The demand for immediate filtration of samples for orthophosphate analysis may restrict that analysis to grab samples, while samples for TP can be held for 28 days. Samples for metals analysis can be held for up to six months before analysis, offering flexibility in analytical schedules and laboratory selection.

Table 3-5. EPA-recommended preservation conditions and hold times for selected water quality variables (40 CFR 136.3 and NEMI 2006)

Variable	Preservation	Maximum Holding Time From Sample Collection
pH	None	15 minutes
Ammonia	Cool, ≤6 ºC, H_2SO_4 to pH<2	28 days
Nitrate	Cool, ≤6 ºC	48 hours
Orthophosphate	Filter immediately, Cool, ≤6 ºC	48 hours
Total Phosphorus	Cool, ≤6 ºC, H_2SO_4 to pH<2	28 days
Total Dissolved Solids	Cool, ≤6 ºC	7 days
Specific Conductance	Cool, ≤6 ºC	28 days
Turbidity	Cool, ≤6 ºC	48 hours
Total Suspended Solids	Cool, ≤6 ºC	7 days
Pesticides	Amber glass bottle, sealed, Cool, 4 ºC	4 to 7 days depending on method
COD	Cool, 4 ºC, H_2SO_4 to pH<2	28 days
Oil and Grease	Cool, 4 ºC, H_2SO_4 to pH<2	28 days
Soluble metals (except Hg, B)	HNO3 to pH<2	6 months
E. coli	Cool, ≤10 ºC	6 hours

All of these constraints will drive station location, field schedules and staff requirements in a monitoring project. For example, in the St. Albans Bay Rural Clean Water project, samples from four tributary monitoring stations were analyzed for both orthophosphate and TP. This work required sample collection by a field technician two to three times each week in order to collect and retrieve samples and deliver them to the laboratory 28 mi (45 km) away (Vermont RCWP Coordinating Committee 1991). In contrast, the Lake Champlain Basin Agricultural Watersheds NNPSMP project collected weekly composite samples for P analysis that were analyzed for TP only, requiring a single trip by a field technician each week to retrieve samples and deliver them to the laboratory (Meals and Hopkins 2002). In both examples, power from the electrical grid was available to run the refrigerated samplers required to maintain sample temperatures at ≤6 ºC. Without power, there would be additional logistical challenges to keeping samples cold with ice or visiting stations more frequently.

3.1.2.10 Need for Covariates

It is important to consider monitoring variables not directly required by project objectives or pollutant sources but that may be important in understanding or explaining the behavior of other critical variables. Such explanatory variables that vary in concert with critical project variables are called covariates. Some covariates are obvious. For nonpoint source issues, precipitation and other weather variables are usually important covariates (see section 2.2.1). Even where load measurement is not required, flow (or stage) should always be measured, for example, as a key covariate in explaining observed patterns of suspended sediment or particulate P that are delivered predominantly in surface runoff in high-flow events. A monitoring program for a lake impaired by eutrophication may benefit from measurement of temperature, chlorophyll α, and algae, even if the focus is on reducing nutrient loads. In cases where paired watersheds are expected to have somewhat dissimilar hydrologic responses to precipitation events, it may be helpful to monitor additional variables such as instantaneous peak flow rate and average flow rate for inclusion in data analysis approaches (see section 7.8.2.2).

3.1.2.11 Set Priorities

Because numerous potential water quality variables exist and because selection criteria may conflict or overlap, it is useful to take a deliberate approach to setting priorities when designing a monitoring program. There are several ways to begin this approach. The USDA National Handbook of Water Quality Monitoring (USDA-NRCS 2003) recommends formulating a written justification for each candidate variable. If the justification is weak, the variable may be of low priority and might not be essential. A ranking system may be useful, where a minimum set of essential variables are identified (e.g., flow and TP for a TMDL aimed at a eutrophic lake), followed by a set of additional, justifiable variables to be monitored if other constraints allow (e.g., orthophosphate, nitrogen, Secchi disk transparency, and chlorophyll *a*). Finally, systematic evaluation of correlations among candidate variables may suggest that one variable (e.g., turbidity) is highly correlated to another (e.g., TSS) so that both need not be measured. Examination of such correlations may also show that some variables do not have direct covariates (e.g., NO_3-N) and should be given priority. Because some relationships between variables (e.g., turbidity and suspended sediment) can change as a result of watershed plan implementation (e.g., turbidity correlation with suspended sediment increases as nutrient levels and biological component of turbidity decrease), it may be appropriate to monitor both variables.

3.1.3 Physical and Chemical Water Quality Data

3.1.3.1 Measuring Surface Water Flow

Measuring surface water flow is an important component of many NPS water quality monitoring projects. Flooding, stream geomorphology, and aquatic life support are directly influenced by streamflow. Runoff and streamflow drive the generation, transport, and delivery of many NPS pollutants. Pollutant load calculations require knowledge of water flow (see section 3.8 and section 7.9).

Surface water *flow* is simply the continuous movement of water in runoff or open channels. This flow is often quantified as *discharge*, the rate of flow or the volume of water that passes through a channel cross section during a specific period of time. Discharge can be reported as total volume (e.g., acre-foot [ac/ft] or millions of gallons) or as a rate such as cubic feet per second (ft^3/s or cfs) or cubic meters per second (m^3/s). The depth of flowing water (m or ft) is commonly measured as *stage*, the elevation of the water surface relative to an arbitrary fixed point. Stage is itself important. Peak stage may exceed the capacity of stream channels, culverts, or other structures. Very low stage may stress aquatic life.

Flow data can be used for a variety of purposes, including problem assessment, watershed project planning, assessment of treatment needs, targeting source areas, design of management measures, and project evaluation. The selection of appropriate flow variables depends on the specific purpose and situation. Two common uses of flow data by watershed monitoring projects are pollutant load calculation (see section 7.9) and model calibration. Pollutant loads are critical elements of TMDL development and implementation. A pollutant load reduction is often one of the principal measures of success in NPS watershed projects. Discharge data are essential for the estimation of loads of sediment or chemical pollutants exported from a river or stream.

Evaluation of specific BMPs or watershed-scale BMP implementation often requires measurement of both pollutant concentration and flow. Many BMPs, particularly stormwater practices in urban settings, are designed to reduce total flow, peak flow, and/or flow velocity, as well as pollutant concentrations. The degree to which these practices achieve pollutant load reductions due to changes in flow versus changes in pollutant concentration varies. Careful consideration of the expected impacts of specific BMPs or combinations of BMPs should help guide decisions regarding flow variables to be monitored.

Basic principles of discharge measurement. Discharge is typically calculated as the product of *velocity* and *cross-sectional area* (Figure 3-1). Surface water *velocity* is the direction and speed with which the water is moving, measured in feet per second (ft/s) or meters per second (m/s). The cross-sectional area of an open channel is the area (ft^2 or m^2) of a slice in the water column made perpendicular to the flow direction.

Determination of discharge (usually symbolized as Q) thus requires two measurements: the cross-sectional area of the water in the channel (A, e.g., in m^2) and the area-weighted average velocity of moving water (V, e.g., in m/s). The product of these two measurements gives discharge in volume per unit time:

$$Q = V \times A$$

For example,

$$1.25 \frac{m}{s} \times 36m^2 = 45 \frac{m^3}{s}$$

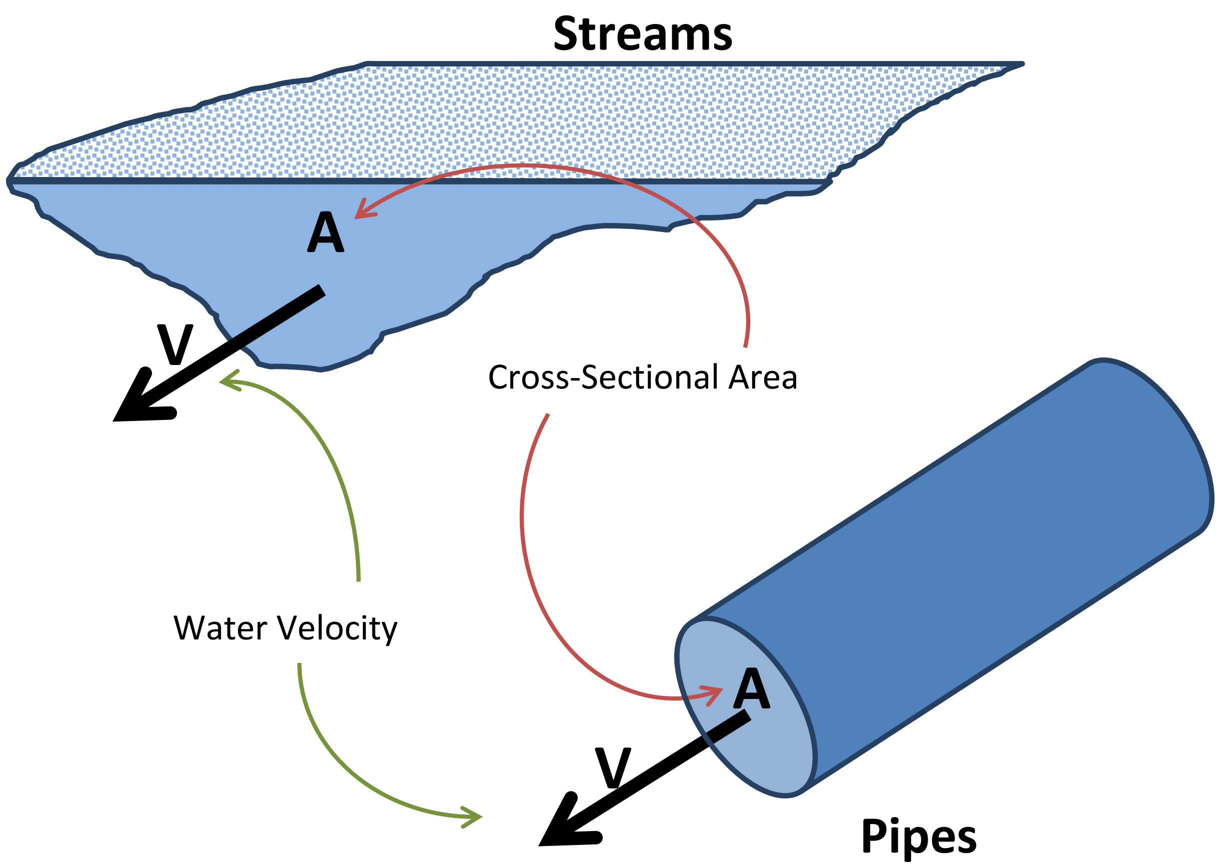


Figure 3-1. Cross-sectional area and water velocity for streams and pipes

It is important to recognize that the *velocity* of moving water varies both across a stream channel and from the surface to the bottom of the stream because of friction and irregularities in cross section and alignment – hence the use of average velocity in the above equation (see section 2.2.1.4.1). Friction caused by the rough channel surfaces slows the water near the bottom and sides of a channel so that the fastest water is usually near the center of the channel and near the surface. On a river bend, the water on the outside of the bend moves faster than the water on the inside of the bend, as it has to cover more distance in the same time frame. Clearly, more than a single measurement is needed to accurately characterize the velocity of water moving down the stream, particularly when the stream channel is irregular.

Flow measurement in water quality monitoring projects can take several forms, from a single measurement of peak stage during a high-flow event to continuous recording of stream discharge. Various approaches to measuring flow are described below.

Peak stage measurement. How high the water reaches during a storm event or flood, also known as peak stage, is often crucial information. In urban watershed projects where reduction of peak stormwater flows is a major goal, tracking peak stream stage (and precipitation) during storm events before and after watershed treatment can be a simple and inexpensive surrogate for monitoring actual streamflow. Peak stage is important to determine for stream restoration projects where high flows shape the physical habitat of the stream. Peak stage is also essential to determine in flood planning, especially for flood frequency statistics, floodplain management, and design/protection of structures.

Peak stage can be observed by several informal means such as high water marks and debris lines on buildings or vegetation. More precise records of peak stage can be obtained using specialized crest gages (Figure 3-2). Information about crest gages is available at http://pubs.usgs.gov/fs/2005/3136/fs2005-3136-text.htm.

Figure 3-2. Traditional crest-stage gage

Instantaneous flow measurements. It is often necessary to estimate or measure discharge at a particular site at a particular time, either to document flow under certain conditions or to develop a data base for further analysis. There are several ways to determine instantaneous discharge, varying in accuracy and in applicability by the size of the stream.

- **Manning's Equation.** Discharge may be computed based on a slope-area method using the Manning equation:

$$Q = \left(\frac{1.486}{n}\right) AR^{\frac{2}{3}}S^{\frac{1}{2}}$$

Where:
Q = discharge in ft^3/s
A = mean area of the channel cross section in ft^2
R = mean hydraulic radius of the channel in ft
S = slope of the water surface in ft/ft
n = roughness factor depending on the character of the channel lining
1.486 = conversion factor in $ft^{1/3}/s$

The n factor can be estimated from tabular values and depends on the character of the channel, varying between 0.01 for smooth concrete to 0.10 for weedy streams with deep pools. The proper selection of a roughness factor is difficult in many cases and discharge determined by this method is only approximate.

- **Volumetric measurement.** For very small flows, e.g., low-flows in ditches or small streams or discharge from drain outlets, the most accurate method of discharge measurement is to simply measure the time required to fill a container of known volume. In some circumstances, it may be necessary to use sandbags to temporarily channel flow to a practical collection point.

- **Dilution methods.** Dilution methods of discharge measurement consist of adding a concentrated tracer solution (salt or dye) of known strength to the stream and by chemical analysis determining its dilution after it has flowed far enough to mix completely with the stream and produce a uniform final concentration in the stream. Discharge is calculated as:

  $$Q = q * (C_1 - C_2)/(C_2 - C_o)$$

  Where:
  Q = stream discharge
  q = tracer injection rate
  C_1 = tracer concentration in injection
  C_2 = final concentration of tracer in the stream
  C_o = background tracer concentration in the stream

  The particular tracer selected should be conservative (i.e., slow to decay and not taken up by sediments or living organisms in the stream) and should be easily measured in the laboratory or field. Salt (NaCl) and rhodamine dye are commonly used tracers. Rhodamine dye can be analyzed in the field by fluorescence.

  When using tracers it is important to inventory all downstream uses of the water and check for notification requirements. Downstream users should be given advance notice of the study, including use of clear signage and other methods of communication.

- **Weirs and flumes.** For long-term projects, discharge can be measured using a weir or a flume, structures that water flows through or over that have a known relationship between stage and flow. If such a device is used, discharge measurement can be as simple as observing the stage of water just upstream of the device and consulting a table or using a simple equation to calculate discharge.

  Weirs are essentially dams built across an open channel over which water flows through a specially shaped opening or edge. Weirs are classified according to the shape of their opening – e.g., a 90º V-notch weir has a notch shaped like an inverted right triangle, whereas a rectangular weir has a rectangular notch. Figure 3-3 shows a 120º V-notch weir in operation. Each type of weir has an associated equation for determining the discharge rate, based on the depth (stage) of water in the pool formed upstream of the weir (see Rantz et al. 1982 for examples). In practice, weirs can range from small wood or metal plates temporarily mounted across small ditches or streams to more permanent installations involving concrete walls and other structures. Note that erecting any obstruction in a stream will create a pool upstream and care must be taken to avoid creating the potential for flooding during high flows.

  Flumes are specially shaped open channel flow sections that restrict the channel area, resulting in increased velocity and a change in water level as water flows through a flume. The discharge through a flume is determined by measuring the stage in the flume at a specific point, depending on the type of flume (see Rantz et al. 1982 for examples). In general, flumes are used to measure discharge where weirs are not feasible; flumes are often used to measure field runoff where flows during storm events can be collected and channeled through the device. Commonly used flumes include the Parshall (Figure 3-4) and Palmer-Bowlus (Figure 3-5). The H-flume is a special flume developed for agricultural field research that can measure discharge over a wide range with good accuracy. Figure 3-6 shows an H-flume in operation in a field runoff monitoring project. Flumes come in a wide range of sizes denoting the maximum depth of flow they can accommodate and can be purchased as prefabricated units or built on-site. While flow control structures such as weirs and flumes can be pre-calibrated, the accuracy of discharge measurements can be compromised by faulty installation (Harmel et al. 2006, Komiskey et al. 2013).

Figure 3-3. 120º V-notch weir, Englesby Brook, Burlington, VT

Figure 3-4. Field application of small Parshall flume

Figure 3-5. Palmer-Bowlus flume

Figure 3-6. 2-foot (0.6 m) H-flume in place for edge-of-field monitoring, East Montpelier, VT

- **Area-velocity technique.** The most common method of measuring discharge in open channels is by measuring the cross-sectional area and the water velocity, as generally described earlier (Figure 3-7). Discharge in a small, wadeable stream can be measured by the following process:
    - **Select location.** Choose a straight reach, reasonably free of large rocks or obstructions, with a relatively flat streambed, away from the influence of abrupt changes in channel width.
    - **Establish cross-section.** Determine the width of the stream and string a cable or measuring tape across the stream at a right-angle to the flow. Divide the width into 20 to 25 segments (streams less than 10 ft [3 meters (m)]) wide may not allow as many segments) using tape or string to mark the center of each segment on the cable (Figure 3-8). Typically the stream is divided into enough segments so that each one has no more than 10 percent of the total streamflow.
    - **Measure depth of each segment.** At each mark across the stream, measure the depth from the water surface to the bottom with a graduated rod or stick (Figure 3-9).
    - **Measure water velocity.** At each mark, measure the velocity of the water (see below). Where depth is less than 2.5 feet (ft) (0.8 m), a single velocity measurement at 0.6 of the total depth below the water surface gives a reasonable estimate of the average velocity with respect to depth. For depths of 2.5 ft or more, the average of velocity measurements taken at 0.2 and 0.8 of depth is preferred.
    - **Calculate discharge for each segment.** For each segment, stream discharge is the product of width of the segment and the measured depth (giving area) multiplied by the velocity for that segment.
    - **Sum discharges.** Total stream discharge is the sum of all segment discharges.

Figure 3-7. Measuring stream discharge (USGS)

While wading is the preferred method for accurate discharge measurement, there are safety considerations that limit the flows at which wading can be accomplished. The USGS has a rule of thumb that prohibits wading if the product of depth (in ft) and velocity (in ft/s) exceeds 8 anywhere in the cross-section. Discharge measurement in larger rivers or at high flows follows the same principles of area and velocity but requires specialized techniques. These include suspension of equipment from bridges (Figure 3-10), cranes (Figure 3-11), or cableways, use of weighted sounding lines, and the use of heavy equipment for velocity measurement (Turnipseed and Sauer 2010).

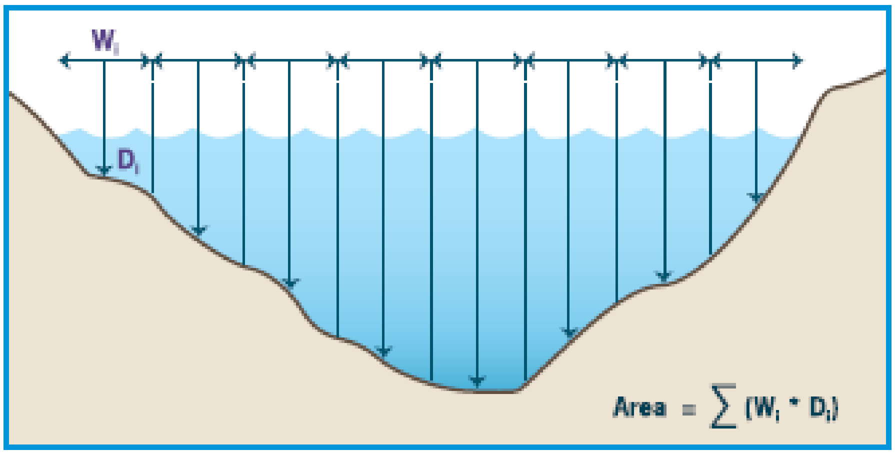


Figure 3-8. Delineation of stream-width segments for discharge measurement

Figure 3-9. Measuring the cross-section profile of a stream channel

Figure 3-10. Measuring discharge from bridge using an ADCP (acoustic Doppler current profiler) unit (USGS 2007)

Figure 3-11. Measuring discharge from a bridge using a current meter and crane (USGS n.d.)

Accurate velocity measurement is a critical component of the area-velocity technique. A variety of instruments are available to measure water velocity, from traditional mechanical current meters to electronic sensors (Turnipseed and Sauer 2010). Velocity measurement technology is evolving. For example, acoustic Doppler technology can measure velocity distributions within the flow, eliminating the need for wading or suspending instruments into the water (Fulton and Ostrowski 2008).

Continuous flow measurements. A single instantaneous measurement of stream discharge provides limited value because it provides information about only a single point in time. It is usually necessary to monitor discharge continuously when a project attempts to measure pollutant load over time or assess relationships between stream discharge and pollutant concentrations or aquatic life.

Continuous discharge measurement in open channels usually requires that the stage-discharge relationship is known, either through the installation of a weir or flume or through development of a stream rating. A stream rating is an equation determined for a specific site that relates discharge to stage based on a linear regression of a series of concurrent measurements of stage and discharge (e.g., by the area-velocity technique). Stage can be measured by a staff gage, a rigid metal plate graduated in meters or feet attached to a secure backing, linked through survey to a fixed elevation and located in a part of the stream where water is present even at low flows (Figure 3-12). Stage can also be read by measuring the distance from a fixed overhead point to the water surface (e.g., using a weighted wire or tape lowered from a bridge beam or using an ultrasonic sensor).

Figure 3-12. Staff gauge in stream

The rating equation should be based on measurements taken over a full range of streamflow conditions. It is usually unacceptable to extrapolate the rating equation beyond the range of observations that it is based on. As shown in the stream rating curve in Figure 3-13, stage-discharge relationships usually have a log-log form. With a valid stream rating, discharge can be determined simply from a stage observation plugged into the equation or read from a table. For more information on stage-discharge ratings, see http://training.usgs.gov/TEL/Nolan/SWProcedures/Index.html. Note that stream rating curves should be checked periodically, especially after major high-flow events. Rating curves frequently shift due to changes in streambed slope, channel roughness, and filling, scouring, or reshaping of streambanks.

Harmel et al. (2006) recommended against using Manning's equation in lieu of direct streamflow measurements to establish a stage-discharge relationship because it results in unacceptable uncertainty. In their analysis of various methods to estimate discharge they found that streamflow estimation with Manning's equation with a stage-discharge relationship for an unstable, mobile bed and a shifting channel resulted in a probable error range of ±42 percent. This compares with a range of 6 percent to 19 percent for typical scenarios using other methods.

Once the stream rating has been developed, continuous discharge measurement becomes an exercise in continuously measuring stream stage. Continuous measurement of stage is also used to record discharge through weirs or flumes where the rating is already known. Depending on the installation, this continual measurement can be accomplished in a number of ways.

A stilling well is a vertical tube or pipe hydraulically connected to the channel such that the level of water in the stilling well matches that in the channel, but the transient variations due to waves or turbulence are damped out (Figure 3-14). Stilling wells can range from an 8-inch-diameter (in) (20 centimeters [cm]) pipe connected to the side of a flume to a 3-ft-diameter (0.9 m) pipe placed in the ground and connected by pipes to a stream. Several devices exist to measure and record stage in a stilling well. Traditionally, this method was conducted using a float attached to a pulley that rose and fell with the water level in the well and moved a pen on a clock-drive chart recorder (Figure 3-15). There are modern versions that use electric chart drives or digital recording systems.

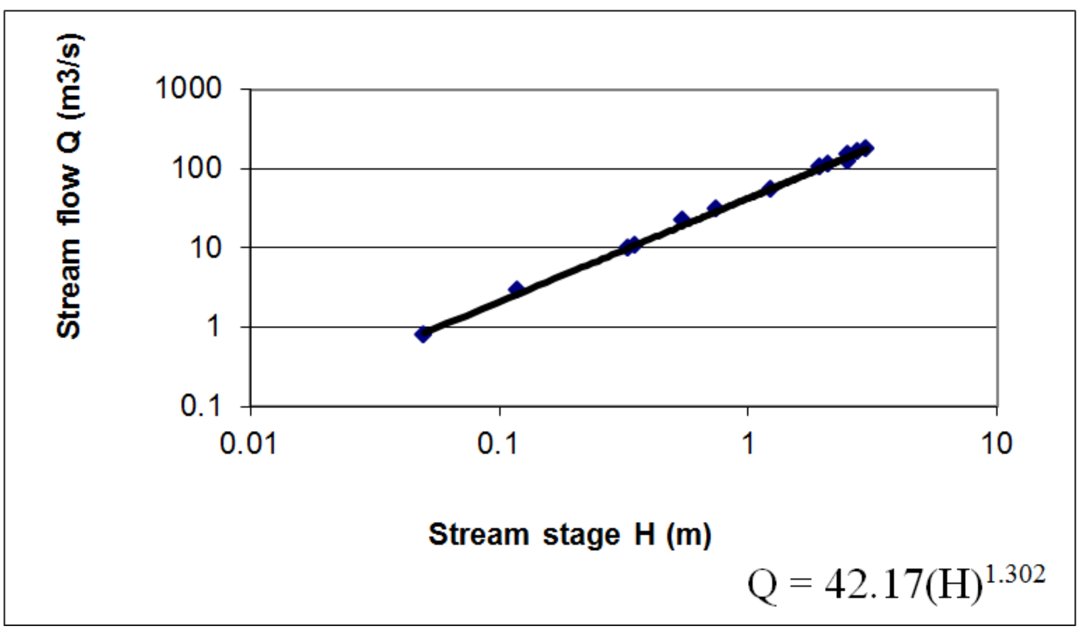


Figure 3-13. Example of a stage-discharge rating for a stream

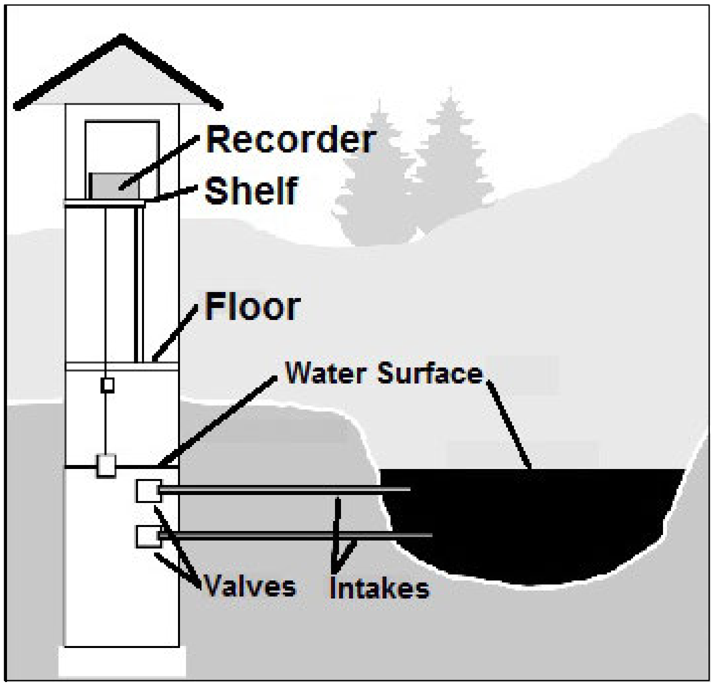


Figure 3-14. Stilling well design schematic (Wahl et al. 1995)

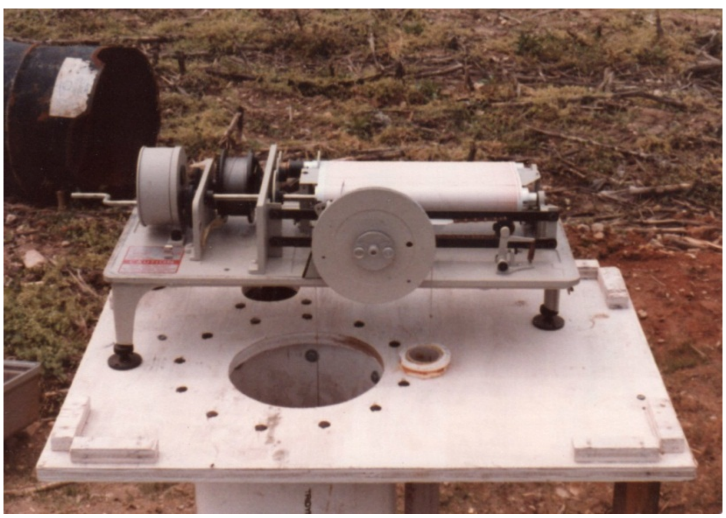


Figure 3-15. Traditional clock-drive chart recorder at a stilling well

Other approaches to measuring and recording level, either in stilling wells or directly in the channel include:

- **Bubblers.** Air or an inert gas is forced through a small diameter bubble line submerged in the flow channel. The water level is measured by determining the pressure needed to force air bubbles out of the line.
- **Pressure transducers.** A probe fixed to the bottom of the channel senses the pressure of the overlying water.
- **Ultrasonic sensors.** The sensor is mounted above the flow stream, and transmits a sound pulse that is reflected by the surface of the flow. The elapsed time between sending a pulse and receiving an echo determines the level in the channel.

Output from level recording sensors can either be recorded directly into a data logger (an electronic device connected to an instrument or sensor that records data over time) for later processing or into a specialized flow meter. There are several manufacturers of such meters; the meters often include the facility to calculate and record discharge and summary statistics, record other data such as precipitation, and interact with other devices such as automated water samplers.

Additional information on flow measurement can be obtained from:

- USDI Bureau of Land Reclamation. *Water Measurement Manual*. http://www.usbr.gov/tsc/techreferences/mands/wmm/WMM_3rd_2001.pdf.
- USGS Measurement and Computation of Streamflow. http://pubs.usgs.gov/wsp/wsp2175/

Streamflow measurement in a natural channel can be challenging for the novice, and mistakes made by technicians can greatly increase measurement uncertainty beyond the ranges reported by Harmel et al. (2006). This can result in highly unreliable stage-discharge relationships, inaccurate estimates of pollutant load, and spurious relationships between flow and other measured parameters. For these reasons, cost savings, and convenience, many monitoring teams seek flow data from USGS wherever possible. Real time daily stream flow data from USGS stations are available at http://waterdata.usgs.gov/usa/nwis/rt.

3.1.3.2 Commonly Measured Physical and Chemical Water Quality Constituents

Selected physical and chemical characteristics and constituents commonly measured in NPS monitoring programs are listed in Table 3-6. This is by no means an exhaustive list. These and other water quality variables are discussed in detail in the following sources:

- *Standard Methods for the Examination of Water and Wastewater* (Rice et al., 2012)
- U.S. EPA Clean Water Act Analytical Methods (http://water.epa.gov/scitech/methods/cwa/index.cfm)
- National Environmental Methods Index (www.nemi.gov/)

There are several complex issues associated with chemistry and analysis of some constituents that should be clarified. Below are a few brief discussions of some of the most important issues that those creating monitoring systems for NPS might encounter.

The traditional measurement of particulate matter suspended in water has been TSS, measured by filtering a subsample of water through a glass fiber filter and weighing the dried residue captured on the filter. In the last decade, research has reported a significant bias in the TSS analysis (Gray et al. 2000). The TSS analysis typically involves subsampling an aliquot from a bulk sample by pipette or pouring from an open container. This method often results in a significant underestimate of heavier particles (i.e., sand) in the sample and thus an underestimate of the total amount of suspended material in the original water. In contrast, the suspended sediment concentration (SSC) analysis entails measurement of the entire mass of sediment and the net weight for the entire sample, capturing all the particles in the original sample. An extensive comparative analysis (Gray et al. 2000) concluded that the TSS method frequently underestimates suspended sediment concentration and is fundamentally unreliable for the analysis of natural water samples. In contrast, the SSC method produces relatively reliable results for samples of natural water, regardless of the amount or percentage of sand-size material in the samples. SSC and TSS data collected from natural water are not comparable and should not be used interchangeably. NPS monitoring projects should monitor SSC, not TSS, to conduct accurate monitoring of suspended sediment loads. However, if comparability with past monitoring is required, it still may be necessary to measure TSS.

Table 3-6. Selected physical and chemical water quality variables commonly measured in NPS watershed monitoring programs

Variable	Abbreviation	Units	Definition	Notes
Physical Characteristics				
Salinity	-	g/kg mg/L	A measure of the total level of salts such as chlorides, sulfates, and bicarbonates in water.	Affects suitability of water (especially groundwater) for drinking, irrigation, and industrial use.
Secchi disk transparency	-	m	A measurement of water transparency in lakes using a black and white disc lowered into the water; the secchi depth is noted as the depth at which the pattern on the disk is no longer visible.	A common, inexpensive measurement of turbidity and an indicator of trophic status of lakes.
Specific conductance	COND	mS/m µmhos/cm	A measure of the ability of water to pass an electrical current; affected by the presence of inorganic dissolved solids.	Indirect measure of dissolved solids in water, highly correlated with salinity.
Total dissolved solids	TDS	mg/L	The sum of all dissolved matter (e.g., Ca, Cl, NO_3, P, Fe, S, and other ions) in a sample.	Indirect indicator of salinity; affects suitability of water for drinking, irrigation, industrial use.
Total suspended solids	TSS	mg/L	A measure of the weight of all particulate matter suspended in water obtained by separating particles from an aliquot of a water sample using filtration.	Affects water clarity, aquatic life support, suitability for drinking water and/or irrigation; indicates sediment from field and/or streambank erosion; particles carry other pollutants, e.g., P, metals, toxicants. It is a measure of wastewater treatment efficiency.
Suspended sediment concentration	SSC	mg/L	A measure of the weight of all suspended sediment in water obtained by separating particles from the entire water sample by filtration.	Related to TSS, but considered more representative of full range of particle sizes present in water because the entire sample, not a subsample, is filtered.
Temperature	T	° C	A measure of the thermal energy content of water.	Rates of biological and chemical processes depend on temperature. Solubility of oxygen is determined by temperature. Aquatic organisms from microbes to fish depend on certain temperature ranges for their optimal health, reproduction, and survival.
Turbidity	-	NTU	A measure of water clarity, i.e., how much suspended particulate material in water decreases the passage of light.	Indirect measure of suspended solids in water; particles may include soil particles, algae, plankton, microbes, and other substances.
Volatile suspended solids	VSS	mg/L	A measure of the organic portion of TSS.	Indicate what portion of TSS is organic in origin such as algal cells or organic wastes.
Nonvolatile suspended solids	NVSS	mg/L	A measure of the inorganic portion of TSS, usually calculated as the difference between TSS and VSS.	Indicate what portion of TSS is comprised of inorganic materials such as soil particles.
Chemical Characteristics				
Biochemical Oxygen Demand	BOD	mg/L	The amount of dissolved oxygen consumed by microorganisms in water in the decomposition of organic matter.	Indirect measure of organic pollutant levels. Usually referenced by oxygen consumed over specified time, e.g., 5-day BOD (BOD_5).
Dissolved Oxygen	DO	mg/L	Oxygen dissolved in water.	Supports aquatic life; influences form and availability of other pollutants.

Variable	Abbreviation	Units	Definition	Notes
Metals	(various)	mg/L or µg/L	Metals are trace elements having atomic weight from 60 – 200 (e.g., As, Cd, Cr, Cu, Hg, Ni, Pb, Zn). Metals exist in surface waters in colloidal, particulate, and dissolved phases; dissolved concentrations are generally low.	Behavior and toxicity varies by element, but metals generally exert chronic and/or acute health effects on aquatic organisms and humans. Presence of elevated concentrations may indicate influence of industrial waste, landfill leachate, or urban stormwater runoff.
Nitrogen – Ammonia	NH_3-N	mg/L	Unionized form of N produced by microbial mineralization of organic N.	Important plant nutrient; may contribute to eutrophication. Toxic to fish at high levels Results reported for ammonia N typically include both NH_3-N and NH_4^+-N forms.
Nitrogen – Ammonium	NH_4^+-N	mg/L	Ionized form of N produced by microbial mineralization of organic N.	Important plant nutrient. Under typical conditions, most ammonia in surface waters occurs as ammonium.
Nitrogen – Nitrite	NO_2-N	mg/L	A partially-oxidized form of N that is a short-lived product of mineralization and nitrification of N from organic materials, usually rapidly further oxidized to NO_3-N.	Nitrites have similar behavior and toxicity to nitrates; significant levels are rarely found in surface waters as they are rapidly converted to NO_3-N in aerobic environments.
Nitrogen – Nitrate	NO_3-N	mg/L	An oxidized form of N that is a common component of inorganic fertilizer; also produced by the mineralization and nitrification of N from organic materials.	Nitrates are highly soluble and mobile in surface and ground water; excess amounts can promote eutrophication and pose a health threat to humans and animals in drinking water.
Nitrogen – Nitrite + Nitrate	NO_2-N+NO_3-N	mg/L	Sum of nitrite and nitrate N in a sample.	Nitrite and nitrate are often analyzed together, depending on laboratory method
Nitrogen – Organic	-	mg/L	Nitrogen in a complex organic form (e.g., proteins) prior to mineralization to ammonia.	The presence of organic N indicates recent presence of organic wastes.
Nitrogen – Total Kjeldahl	TKN	mg/L	TKN is the sum of organic N and ammonia-N.	TKN includes all forms of N except NO_2-N and NO_3-N
Nitrogen - Total	TN	mg/L	Total N is the sum of all forms of N in a water sample.	TN can be determined directly through chemical analysis or calculated as the sum of TKN+ NO_2-N+NO_3-N
Phosphorus – Orthophosphate	OP PO_4-P	mg/L	The simplest and most stable of inorganic P compounds, H_3PO_4.	Ortho P (also referred to as "reactive" P) is a plant nutrient that may contribute to eutrophication.
Phosphorus – Soluble Reactive	SRP	mg/L	A dissolved form of P operationally defined as the P that reacts with specific reagents in a laboratory analysis.	Related and functionally similar to ortho phosphate, usually measured on a filtered sample.
Phosphorus – Total	TP	mg/L	Total P is the sum of all forms of P in a water sample, as determined by chemical digestion to a dissolved form.	Generally includes both particulate and dissolved P, unless operationally separated into dissolved and particulate forms in the laboratory.
pH	-	-	A measure of the acidity or basicity of water, expressed as the negative log of the H^+ ion concentration.	Affects chemical form of some pollutants, may have direct effects on aquatic life. Indicator of mine drainage.

Nitrogen (N) undergoes a complex cycle in the environment that includes both air and water pathways, mediated by microorganisms. A simplified N cycle is illustrated in Figure 3-16. Forms of N commonly measured as chemical water quality variables (Table 3-6) track the aqueous components of this cycle well. It should be noted that the term "ammonia" commonly refers to two chemical species that are in equilibrium in water (NH_3, un-ionized and NH_4^+, ionized). Water quality analyses for ammonia usually measure and report total ammonia (NH_3 plus NH_4^+). The toxicity of ammonia is primarily attributable to the un-ionized form (NH_3), as opposed to the ionized form (NH_4^+) (NCSU 2003). Ambient conditions of pH determine the net toxicity of total ammonia in water; in general, more un-ionized NH_3 and therefore greater toxicity exist at higher (alkaline) pH. NPS monitoring projects concerned with nitrogen should monitor total N, either as a discrete analysis or by measuring TKN and NO_2+NO_3 and summing the two for an estimate of total N unless there is a compelling reason to select different N variables.

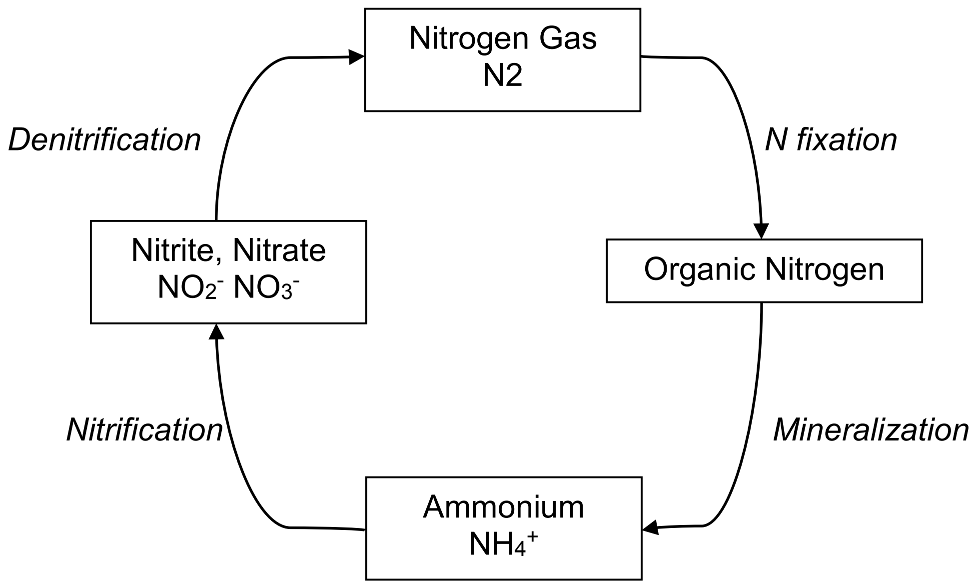


Figure 3-16. Simplified version of the nitrogen cycle

Phosphorus (P) undergoes a somewhat simpler cycle (USEPA 2012), lacking the atmospheric component, but the analytical scheme does not correspond perfectly to that cycle. In water quality monitoring, P is reported largely on an operational basis corresponding to sampling and laboratory procedures, rather than to specific points on a biogeochemical cycle.

P in freshwater systems exists in either a particulate phase or a dissolved phase. Particulate matter includes living and dead plankton, precipitates of P, P adsorbed to particulates, and amorphous P. The dissolved phase includes inorganic P (generally in the soluble orthophosphate form), organic P excreted by organisms, and macromolecular colloidal P. Most of these forms, however, are rarely analyzed specifically.

For some purposes, a water sample may be split between dissolved and particulate forms by filtration prior to further analysis. Thus, it is important to specify and know whether a specific P analysis is done for the dissolved phase, the particulate phase, or on the total sample. Ortho phosphorus is frequently analyzed as the primary dissolved form of P and is readily available to algae and aquatic plants. Most of the P discharged by wastewater treatment facilities is in the dissolved form. Another P fraction is also sometimes defined operationally as "reactive P" because it reacts with certain reagents in chemical analysis to form a color, resulting in reporting as Soluble Reactive P (SRP), which is related to but not exactly equivalent to the ortho-P analysis[1]. To gauge the potential impact of a P discharge on eutrophication, “bioavailable P” is sometimes evaluated by measuring a sample’s potential to support algae growth in a bioassay. Bioavailable P does not usually correspond exactly to a form of P directly measurable in chemical analysis.

Because the organic and inorganic particulate and soluble forms of P undergo continuous transformations (e.g., through uptake and release by algae and other plants or by chemical sorption and desorption on soils, suspended sediment, and other particulate material), many monitoring programs measure TP rather than individual forms to determine the amount of nutrient that can potentially support the growth of aquatic plants and contribute to eutrophication. The TP analysis uses digestion by acid and strong chemicals to convert all P in a sample to a soluble reactive form that can be easily measured in the laboratory (USEPA 2012). Several different digestion procedures are available and a monitoring program should be sure to specify the appropriate method for their situation.

3.1.3.3 Surrogates

In some cases, it may be preferable to use surrogate variables to represent other variables that may be mentioned specifically in project objectives but are difficult or expensive to measure. In some cases, it is necessary to use surrogates because a desired response variable is a complex composite of many individual factors. If, for example, the objective is to monitor the condition of salmon spawning areas, surrogate measures are necessary because the quality of spawning areas responds to many influences. Good surrogate variables would be stream bank undercut, substrate embeddedness, and vegetative overhang (Platts et al. 1983).

Two important criteria must be met by surrogate measures:

- A strong and consistent relationship must exist between the surrogate and the primary variable(s) of interest. Such a relationship can be established by simple linear regression using a local data set.
- A scientific basis is needed to assert that the surrogate and primary variable(s) will respond similarly to environmental management (e.g., BMP implementation) and change (i.e., the relationship remains the same). This assertion should be confirmed with data collected after such management or change.

While some surrogate relationships are widely appropriate in principle, the specifics of the relationship vary from site to site and, in most cases, should be based on locally derived data. For example, the relationship between turbidity and TP is usually highly specific to an individual watershed and should be used only in the system where the relationship can be documented. It is very important that the physical, chemical, or biological relationships between candidate surrogates and primary variables are considered in some depth to ensure that plausible relationships exist. For example, while erosion and sedimentation rates are often related in principle, using measured or estimated field erosion rates as a surrogate for watershed sediment load, for example, is likely to give poor results because the relationship between the two variables (i.e., the sediment delivery ratio) is not known.

---

[1] The term "orthophosphate" is a chemistry-based term that refers to the phosphate molecule all by itself (USEPA 2012). "Reactive phosphorus" is a corresponding method-based term that describes what you are actually measuring when you perform the test for orthophosphate. Because the lab procedure is not perfect, you get mostly orthophosphate but you also get a small fraction of some other forms.

Cost and ease of analysis are the primary reasons why specific conductance (fast and easy to measure with an electronic instrument) is often used as a surrogate for total dissolved solids (TDS) that requires measuring, drying and repeated weighing of a sample in a laboratory. In addition, specific conductance can be expected to respond to environmental management in the same way as TDS in many cases. Improved irrigation management, for example, might be expected to reduce levels of both actual TDS and specific conductance generated by the dissolved ions. Specific conductance can be expected to reflect the effect of irrigation management on TDS.

Indicator bacteria like *E. coli* are commonly used to indicate the likely presence of true pathogens in water because indicators are relatively fast and inexpensive to measure compared to pathogens. A good application of *E. coli* as a surrogate would be a study evaluating the effects of fencing livestock from streams because reductions in direct manure deposition to the stream would be expected to reduce both *E. coli* bacteria and manure-borne pathogens.

Turbidity is fast and easy to measure directly in the field and can be recorded continuously by field instruments. It is often highly correlated with TSS or SSC and can be used as a surrogate for these more expensive analyses when such correlations are established with local data. For example, if turbidity data will be used to predict or estimate TSS concentrations or loads (e.g., through a regression equation), the specific parameters of the equation must be documented in the local system because soils and suspended sediment vary widely among watersheds. In addition, turbidity can be a poor surrogate for SSC if the particles causing turbidity are not consistently related to those comprising SSC. Management changes that reduce SSC through trapping of larger sediment fractions, for example, may change the relationship between SSC and turbidity, which is more strongly linked to finer sediment. When turbidity is used as a surrogate for either TSS or SSC it is recommended that the relationship between the surrogate and primary variable is checked throughout the monitoring period to determine if changes have occurred.

3.1.4 Biological Data

Biological data, including aquatic organisms, habitat, and pathogens, are often central to NPS monitoring efforts. Selected biological characteristics commonly measured in NPS monitoring programs are listed in Table 3-7. This is not an exhaustive list. These and other water quality variables are discussed in detail in chapter 4 and in the following sources:

- Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish (Barbour et al. 1999).
- The Qualitative Habitat Evaluation Index (QHEI): Rationale, Methods, and Application (Rankin 1989).
- Methods for Assessing Habitat in Flowing Waters: Using the Qualitative Habitat Evaluation Index (QHEI) (MBI 2006).

Aquatic organisms are particularly useful because they integrate the exposure to various NPS pollutant stressors over time. Measures of biological communities can integrate the effects of different pollutant stressors like excess nutrients, toxic chemicals, increased temperature, and riparian degradation. They provide an aggregate measure of the impact of stressors from the watershed. When the objectives of a NPS watershed project focus on biological response (e.g., restoration of fish in a stream) or when treatment in the watershed focuses on in-stream practices like habitat restoration, biological monitoring is essential.

Table 3-7. Selected biological water quality variables commonly measured in NPS watershed monitoring programs

Variable	Units	Definition	Notes
Habitat variables			
Bottom substrate	Qualitative score	Percent rubble, gravel; presence of undercut banks, woody debris	Quality and diversity of substrate influences suitability for fish reproduction and habitat quality for benthic invertebrates.
Embeddedness	Qualitative score	Percent gravel, cobble, and boulder particles surrounded by fine sediment	Substrate condition influences suitability for fish reproduction and habitat quality for benthic invertebrates.
Flow velocity	cm/s	Range of current velocity	Prevailing current velocity influences suitability for stream biota.
Channel alteration	Qualitative score	Channelization, presence of point bars, silt deposition	Altered channels may reduce habitat diversity; sediment deposition can render substrate unsuitable for fish or invertebrate communities.
Pool/riffle ratio	Qualitative score	Variety of pool/riffle environments	A diversity or lack of pool and riffle environments influences suitability of a stream environment for fish and other biota.
Qualitative Habitat Evaluation Index (QHEI)[1]	Numerical score	Multiple metric index of habitat variables including substrate, cover, channel quality, riparian condition, bank erosion, pool/riffle distribution, drainage area, and gradient	The QHEI is composed of an array of metrics that describe attributes of physical habitat that may be important in explaining the presence, absence, and composition of fish communities in a stream. A significant correlation between QHEI and IBI has been documented in Ohio.
Microorganisms			
Indicator bacteria	#/100 ml cfu/100 ml MPN/100 ml	Bacteria of fecal origin whose presence is indicative of the probability of existence of true pathogens, e.g., fecal coliform, *E. coli*, *enterococci*	Use of indicator bacteria is based on rapid, inexpensive analysis, presumed association with true pathogens, and some epidemiological evidence of gastrointestinal disease.
Pathogens	#/L MPN/100 ml	Waterborne microorganisms that cause disease in humans or animals, including bacterial pathogens like *E. coli* O157:H7 and *Salmonella* and protozoans like *Giardia* and *Cryptosporidium*	Rarely analyzed as a routine because of expense and required expertise.
Microbial Source Tracking	-	Use of DNA, antibiotic resistance, or other techniques to attribute bacteria found in water to specific host group, e.g., human, cow, waterfowl	Increasingly used in situations of significant microbiological impairment where multiple sources are possible and specific cause(s) of impairment is unknown.
Plants			
Chlorophyll α	mg/L	Measurement of chlorophyll α pigment extracted from algae collected in a water sample	Used as an indicator of biological productivity or trophic state of lakes and as a surrogate for algal biomass. Often correlated with other measures of lake eutrophication such as P load and secchi disk transparency.
Algae	-	Identification and classification of algae taxa found in a sample of lake water	Presence and/or dominance of certain algal taxa may indicate trophic status of lakes, e.g., presence of *cyanobacteria* (blue-green algae) often indicate eutrophication due to excess P concentrations.
Macrophytes	(various)	Identification, classification, of macrophyte taxa and measurement of extent and abundance in a lake or stream	Presence of extensive growth of some species is considered a nuisance, especially invasive species; extent and abundance of other species is considered ecologically desirable.

Variable	Units	Definition	Notes
Benthic Macroinvertebrates			
# of organisms	#/m^2	Number of organisms found per unit area	Provides crude estimate of biomass for comparison between sites or over time.
Taxa richness	# of families	Number of families present	Reflects general health of community; generally increases with improving water quality, habitat diversity, and habitat suitability.
Biotic Index	Numerical score	Index based on tolerance of taxonomic groups (e.g., family) to organic pollution	Indicates general impacts of organic pollution on invertebrate community; values of the BI increase with decreasing water quality.
EPT Index	# of taxa	Number of distinct taxa within the *Ephemeroptera*, *Plecoptera*, and *Trichoptera* groups	Summarizes taxa richness within the insect groups that are generally considered to be pollution sensitive.
Functional feeding groups	(various)	Classification of organisms by feeding style, e.g., shredders, scrapers, filter-feeders, predators	Certain feeding groups indicate certain impairment types, e.g., shredders are sensitive to riparian zone impacts that change the inputs of coarse particulate organic matter to a stream.
Fish			
# of individuals	#/m^2	Number of individuals found per unit area	Reflects crude estimate of fish population size and biomass for comparison between sites or over time.
# of species	# of species	Number of different species present	Reflects general health of fish community; generally increases with improving water quality, habitat diversity, and habitat suitability. Presence/absence of particular species can be associated with water quality or particular stressors.
Index of Biotic Integrity	Numerical score	Integrated index of multiple metrics of species richness and composition, trophic composition, and fish abundance and condition	Individual component metrics can be used. IBI is adaptable and often modified on a regional basis.

[1] Developed by Ohio EPA (Rankin 1989)

Biological monitoring of macroinvertebrates, fish, and other aquatic biota must consider more dimensions than is the case for most physical and chemical monitoring. For example, the presence or absence of certain species or assemblages is not simply an indicator of ambient water quality or water quality impairment. The biotic community present at a particular location is always a reflection of the available habitat required to support those life forms. Data on aquatic biota cannot be interpreted without reference to the habitat at a particular site and is the reason that several habitat metrics are listed in Table 3-7. The nature of aquatic communities is also strongly determined by ecoregion. A warm-water river in Iowa is not capable of supporting the same biotic community as a Rocky Mountain stream in Montana. This is not necessarily because of a water quality impairment but because climate, watershed, soils, vegetation, and other factors differ between the two ecoregions. In NPS watershed monitoring projects, it is common practice to monitor biological variables at impaired or treated sites and at reference sites within the ecoregion indicating the best biological condition that can be expected in the subject watershed. For this reason, data on biological variables are often presented in comparison with data on the same variables collected at one or more reference sites. See chapter 4 for additional details on biological monitoring.

The use of indicator bacteria in biological monitoring is an evolving issue (Meals et al. 2013). Organisms like fecal coliform and *E. coli* are not themselves pathogenic but are assumed to have a significant association with the presence of true pathogens. Empirical evidence has suggested a statistical probability of increased incidence of gastrointestinal disease at some threshold of indicator bacteria count (Dufour 1984). However, the adequacy of the association between indicator bacteria and true pathogenic microorganisms has been increasingly challenged in recent years (Harwood et al. 2005). Indicator organisms have been found in high numbers where few pathogens were detected and pathogens have been documented when a waterbody meets water quality standards for indicator bacteria. True pathogens like *Cryptosporidium* have been shown to survive considerably longer than *E. coli* in animal waste spread on agricultural land (Hutchison et al. 2005). Furthermore, the traditional presumption that indicator bacteria indicate recent fecal pollution is increasingly in doubt as fecal coliform and *E. coli* have been shown to survive for long periods and even reproduce in aquatic sediments, beach sands, and urban storm drains (Jiang et al. 2007, Yamahara 2009). However, other research continues to support an association between both *E. coli* and *enterococci* and the incidence of gastrointestinal disease (Arnone and Walling 2007), so the matter is far from resolved.

Indicator bacteria will likely continue to be widely used monitoring variables in the future. Water quality standards for shellfishing continue to be based on fecal coliform counts. TMDLs for bacteria are nearly always focused on fecal coliform, *E. coli* or some other indicator organism. As microbial source tracking becomes more widely cost-effective, that technology may become more important than simply measuring indicator bacteria counts at a sampling station (USEPA 2005a, 2011b). Furthermore, when waterborne disease outbreaks are an immediate concern, evaluation of true pathogens could be warranted.

See chapter 4 for a detailed discussion of biological monitoring approaches.

3.1.5 Weather Data

Weather is an essential variable set for NPS monitoring projects. Precipitation drives NPS pollutant generation and delivery and patterns of wet/dry weather, seasonality, and extremes are major influences on NPS loads. Actual weather data during a watershed project are needed to place the monitoring period in context with long-term average conditions. Weather is often a critical covariate in NPS projects, as unusually dry or wet weather may exaggerate or mask response to treatment. Precipitation variables like total rainfall, rainfall intensity, storm duration, and storm interval are often key design components in urban stormwater/LID practices. Temperature may be an important response variable in restoration of stream habitat and for implementation of urban stormwater BMPs. Finally, good weather data are usually key drivers for modeling and the extent and quality of precipitation data often determines the success of model calibration.

Variable selection is largely driven by specific project needs. In most cases, at least daily precipitation totals are needed. Data on storm intensity, duration, and frequency may also be needed where pollutant delivery is highly episodic and monitoring is focused on storm events. Air temperature (daily minimum, maximum, and mean) data may be needed because temperature drives evapotranspiration. In northern regions air temperature determines the form of precipitation as rain or snow and controls snowmelt. The majority of the annual NPS pollutant load in northern regions may be delivered by winter and spring snowmelt events (Hanson et al. 2000, Panuska et al. 2008). Other weather variables may be required by specific project objectives. For example, monitoring of stream fishery status after restoration of forested riparian buffers may benefit from data on solar radiation to correlate to shading and stream water temperature (Whitney 2007). A study of bacteria survival and transport in field runoff might need to monitor solar radiation, relative humidity, wind velocity, and soil temperature in addition to basic precipitation and air temperature as variables that affect microorganism survival after manure application.

Monitoring personnel should query local sources of weather data to determine the need for additional weather stations. A source of information for this step is:

- NOAA Earth System Research Laboratory. http://www.esrl.noaa.gov/psd/data/faq/
    - Provides information and links for locating climate and weather data and information.

Sources of current and historical weather data include:

- NOAA National Weather Service Internet Weather Source. http://weather.noaa.gov/
    - Provides weather conditions for the past 24 hours, forecasts, watches, and warnings. Data are easily copied and pasted into a spreadsheet.
- NOAA National Centers for Environmental Information Climate Data Online. https://www.ncdc.noaa.gov/cdo-web/
    - Provides for historical data retrievals and download to a comma delimited file.
- NOAA Meteorological Assimilation Data Ingest System (MADIS). https://madis.ncep.noaa.gov/index.shtml
    - MADIS is a meteorological observational database and data delivery system that provides observations that cover the globe. Data are available from July 2001 to the present.
- Weather Underground. http://www.wunderground.com/
    - Provides current and historical data that can be downloaded to a comma delimited file. Weather data come from more than 180,000 weather stations across the country.

3.1.6 Watershed Characterization

In designing any watershed monitoring program, it is essential to characterize the watershed to identify causes and sources of NPS pollution, understand how water and pollutants are transported through the watershed, and determine where and how to implement a monitoring program. In any specific project, data on particular watershed characteristics like geology or impervious cover may be needed, but in nearly all NPS projects, data on topography, soils, surface and subsurface drainage, hydrology (e.g., NHDPlus), and land use/land cover will be necessary. These data are often collected as part of the watershed project planning process described in detail by U.S. EPA (2008a).

3.1.6.1 Topographic Data

Topographic data may be needed to determine water flow paths, including mapping subcatchments, and to identify areas of steep slope, critical elevation, or particular aspect. Application of simulation models like SWAT (Soil and Water Assessment Tool) and AGNPS (AGricultural Non-Point Source Pollution Model) requires detailed topographic data. The main sources of topographic data in the recent past were published topographic maps. Today topographic data are readily available as Digital Elevation Models (DEMs) derived from remote sensing and assembled in a geographic information system (GIS). DEMs are commonly available from state or local agencies and, once imported into a GIS, can be readily manipulated to generate derived data on drainage area boundaries, hydrography, elevation, slope, and aspect.

A major consideration in DEM data for monitoring programs is resolution. Standard DEMs generally offer 30-meter resolution (i.e., vertical accuracy of ±30 meters), with 10-meter resolution possible in some cases, providing an improved representation of landscape features. Recent advances in using

LiDAR (*Light Detection and Ranging*, a remote sensing system using aircraft-mounted lasers) can provide DEMs with a resolution of 1 meter or better. High-resolution DEMs can be useful in locating and mapping very small-scale landscape features such as drainage ditches, swales, and ephemeral gullies, all of which can be important in understanding runoff and pollutant transport and identification of critical source areas to design land treatment.

USGS provides information on several sources of free geospatial data at:

http://education.usgs.gov/lessons/geospatialwebsites.html

3.1.6.2 Soil Characteristics

Data on soil physical characteristics and soil chemistry may be required for some NPS monitoring projects. Physical characteristics like hydrologic soil group strongly influence where surface runoff commonly occurs. Soil type and factors (e.g. soil erodibility) influence erosion and soil loss and are sometimes used as parameters to identify critical source areas of NPS pollutants in a watershed. Soil and vadose zone variables like permeability, hydrologic conductivity, or depth to water table may be important to determine in ground water monitoring efforts. Soil chemistry data (e.g., soil test P, organic matter, cation exchange capacity) may be essential to identify important source areas and understand pollutant transport over and through watershed soils. Testing for soil P levels can be helpful at the beginning of a project to ensure that paired watersheds, for example, are suitably matched (Bishop et al. 2005).

Data on soil characteristics may be available from specific studies in local areas or can be obtained from national databases such as the USDA *State Soil Geographic* (STATSGO) and *Soil Survey Geographic* (SSURGO) (http://soils.usda.gov/survey/geography/).

3.1.6.3 Land Use/Land Cover

Land use/land cover data includes information on the natural and cultural character of the land surface (e.g., forest, grassland, wetland, water, pavement) and on the activities taking place on the land (e.g., crop agriculture, pasture, residential, commercial, highways). Because NPS pollution is predominantly a function of land use, detailed knowledge of land uses and their spatial distribution is critical in developing a watershed monitoring program.

Land use/land cover data are usually derived from remote sensing data, either aerial photography or satellite imagery. Specific classification of land use/land cover types vary according to project objectives. For an urban stormwater/LID monitoring effort, data on many classes of developed land may be needed, as well as aggregate variables like impervious cover. In urban watersheds, the percent of direct and indirect impervious cover and other metrics of urban land use have been clearly documented as a determinant of many dimensions of stream impairment (Paul and Meyer 2001, Roy et al. 2003). In contrast, an agricultural NPS monitoring effort may need detailed information on many agricultural land uses like corn, soybeans, hay, pasture, farmstead but may lump urban land uses into a single broad category. A common land use/land cover classification scheme is shown in Table 3-8.

Table 3-8. Anderson Level II land use and land cover classification system for use with remote sensor data (Anderson et al. 1976)

1 Urban or Built-up Land		**6 Wetland**	
	11 Residential		61 Forested Wetland
	12 Commercial and Services		62 Nonforested Wetland
	13 Industrial	**7 Barren Land**	
	14 Transportation, Communications, and Utilities		71 Dry Salt Flats
	15 Industrial and Commercial Complexes		72 Beaches
	16 Mixed Urban or Built-up Land		73 Sandy Areas other than Beaches
	17 Other Urban or Built-up Land		74 Bare Exposed Rock
2 Agricultural Land			75 Strip Mines, Quarries, and Gravel Pits
	21 Cropland and Pasture		76 Transitional Areas
	22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas		77 Mixed Barren Land
	23 Confined Feeding Operations	**8 Tundra**	
3 Rangeland			81 Shrub and Brush Tundra
	31 Herbaceous Rangeland		82 Herbaceous Tundra
	32 Shrub and Brush Rangeland		83 Bare Ground Tundra
	33 Mixed Rangeland		84 Wet Tundra
4 Forest Land			85 Mixed Tundra
	41 Deciduous Forest Land	**9 Perennial Snow or Ice**	
	42 Evergreen Forest Land		91 Perennial Snowfields
	43 Mixed Forest Land		92 Glaciers
5 Water			
	51 Streams and Canals		
	52 Lakes		
	53 Reservoirs		
	54 Bays and Estuaries		

The land use/land cover variables of interest for watershed characterization are mainly static but are spatially referenced. A single map of current watershed land use/land cover may suffice for designing a water quality monitoring program; an annual update may be useful to relate to observed trends in water quality over time. Such data are distinct from land use activity data needed on a fine scale to relate to observed water quality at a site level, which include a critical temporal element. This kind of land use data monitoring is discussed later in section 3.7.

3.2 Sample Type Selection

3.2.1 General Considerations

The goal of collecting water samples is to obtain information representative of the target population for the monitoring effort. If monitoring is directed only at storm flows, the goal is to collect samples representative of storm flow conditions. If base flows are of greatest importance, then samples need to represent base flow conditions. For pollutant load estimation, it is most important that samples represent flow conditions that generate the greatest share of the pollutant load most strongly related to the identified problem. When monitoring is directed at specific conditions that threaten or harm aquatic life, sampling may need to favor extreme conditions such as low flow or high temperature. Sample type choices can be a major determinant of the success or failure of a monitoring program.

As described in chapter 2, water quality varies both temporally and spatially. The extent that water quality spatial variability is addressed in a monitoring program is determined by the station location and the sample type. Station location determines where on the landscape a particular sample is taken, whereas sample type determines the spatial representation of each sample taken at that location. Similarly, sampling frequency and duration combine with sample type to determine the extent of temporal variability of water quality captured by the monitoring program. Sampling duration defines the timeframe for sampling, and sampling frequency determines how many times samples are collected during that timeframe. Sample type determines the degree to which temporal variability is captured within each sampling event.

There are generally four types of water quality samples (USDA-NRCS 2003):

- **Grab.** A discrete sample taken at a specific point and time.
- **Composite.** A series of grab samples collected at different times and mixed together.
    - Time-weighted: A fixed volume of sample collected at prescribed time intervals and then mixed together.
    - Flow-weighted: A series of samples each taken after a specified volume of flow has passed the monitoring station and then mixed together.
- **Integrated.** Multi-point sampling to account for spatial variations in water quality within a water body.
- **Continuous.** Truly continuous or very frequent sequential measurements using electrometric probes.

Each sample type has advantages and disadvantages and is discussed in the remainder of the next section. Ultimately, the selection of the appropriate sample type is determined by study objectives, variable(s) sampled, and whether concentration or mass is of interest (USDA-NRCS 2003). Integrated samples are generally preferred when suspended sediment is measured, and grab samples are preferred for bacteria. Generally appropriate sample type selection as a function of monitoring objective is illustrated in Table 3-9.

Table 3-9. Sample type as a function of monitoring objective (adapted from USDA-NRCS 2003)

	Sample Type				
		Composite			
Objective	Grab	Time-Weighted	Flow-Weighted	Integrated	Continuous
Problem Identification & Assessment	X	X	X	X	X
NPS Load Allocation			X		
Point Source Wasteload Allocation		X	X		
Trend Analysis	X	X	X	X	
Assess Watershed Project Effectiveness		X	X		
Assess BMP Effectiveness		X	X		
Assess Permit Compliance	X	X	X		

Objective	Sample Type				
	Grab	Composite		Integrated	Continuous
		Time-Weighted	Flow-Weighted		
Validate or Calibrate Models		X	X	X	
Conduct Research		X	X	X	X

3.2.2 Types

3.2.2.1 Grab

Grab samples are discrete samples taken from a specific point and time (USDA-NRCS 2003). For this reason, grab samples provide the narrowest representation of the spatial and temporal variability of water quality conditions. Grab samples are usually obtained manually with plastic or glass bottles or jars but can also be taken with automatic samplers. Grab sampling typically occurs in wadeable streams or from boats on lakes, but sampling can also be taken from bridges during high flows for larger streams and rivers. It is important to document both when and where grab samples are taken. Location can be recorded by recording depth and position along the width of the stream or depth and coordinates on a lake.

The specific method used to collect grab samples can have a significant influence on the content of the sample. Wilde et al. (2014) define samples for which the velocities of the stream and water entering the sampler intake are the same and different as isokinetic and nonisokinetic, respectively. Because the suspension of particulate materials depends largely on stream velocity, an isokinetic sample may therefore have a different and more accurate sediment concentration compared to a nonisokinetic sample. Isokinetic, depth-integrated samplers are described in section 3.2.2.3. Nonisokinetic samplers include the hand-held bottle, the weighted-bottle sampler, the BOD sampler, and the so-called "thief samplers" such as the Kemmerer and Van Dorn samplers that are often used for lake sampling at specific depths (Wilde et al. 2014).

3.2.2.2 Composite

Composite samples are generally considered a series of simple grab samples taken over time and lumped together (USDA-NRCS 2003). Isokinetic, depth-integrated samples collected to produce a discharge-weighted sample may also be included in this grouping (Wilde 2006). Composite samples are usually collected with automatic samplers (see section 3.6.2.4), but passive samplers (Bonilla et al. 2006) and labor-intensive manual methods can also be used. Composite samples derived from simple grab samples are taken from a single location and do not address the spatial variability of water quality conditions. When automatic samplers with fixed-depth intake(s) are used, the sample is considered by USGS to be a point-integrated sample (Wilde et al. 2014).

Sample preservation is always a concern but is of particular importance when automatic samplers are used for composite sampling. Analyte loss can occur between sample collection and laboratory analysis because of physical, chemical, and biological processes that result in chemical precipitation, adsorption, oxidation, reduction, ion exchange, degassing, or degradation (Wilde et al. 2009). Acidification and/or refrigeration is required for many monitoring variables.

The trigger for collecting samples distinguishes time-weighted from flow-weighted composite sampling. Time-weighted composite samples are derived from samples collected at pre-determined intervals such as hourly or daily samples taken and composited in a single container (Stuntebeck et al. 2008). Because flow is not considered (but could be measured) in the sampling scheme, time-weighted composites are generally inappropriate for load estimation (see section 7.9) in nonpoint source applications. Where flow is constant, however, time-weighted composites would be useful for load estimation. If flow is measured in a time-weighted sampling scheme where samples are collected in multiple bottles, it is possible to make up a flow-weighted composite samples from individual discrete samples by adding amounts of individual samples in proportion to the flow that occurred over the collection interval (Stuntebeck et al, 2008). Peak pollutant concentrations may be missed in a time-weighted sampling design, however, resulting in low estimates of pollutant load.

Flow-weighted or flow-proportional samples are better for capturing the influence of both peak concentrations and peak flows, resulting in more accurate estimates of pollutant loads (see section 7.9). Collecting flow-weighted samples requires an established stage-discharge relationship, prediction of flow conditions during the period between sample collections, continuous flow measurement, and instantaneous and continuous calculation of flow volume that has passed the sampling station. Any fouling of the stage measurement by backflow, icing or other causes will result in incorrect flow volume calculations and the collection of non-representative samples. Remote access to the monitoring station provides some capability to address these potential problems. Flow-weighted composite sampling has as many applications as time-weighted composite sampling, with the additional advantage of being useful for pollutant load estimation (Table 3-9). The cost for flow-weighted sampling will exceed that of time-weighted sampling that does not include flow measurement. Both composite sampling types offer reduced laboratory costs per unit of temporal information gained when compared to grab sampling over the same time period because fewer samples are analyzed. Compositing results in information loss, however, as the individual samples are averaged either by time or flow. This information loss corresponds with reduced sample-to-sample variability which can be helpful in efforts to evaluate BMP and project effectiveness. For the same number of samples, composite sampling also offers the advantage of fewer trips to the field compared to grab sampling, reducing labor costs (see chapter 9 for a discussion of monitoring costs).

Advances in remote access and control of automatic sampling equipment have made it possible to adjust the sampling program based on current knowledge of weather conditions and discharge (Stuntebeck et al. 2008). This technology provides considerable flexibility for the researcher, including the ability to change flow-weighted sampling if flow conditions differ markedly from those assumed when the sampler was programmed.

3.2.2.3 Integrated

Grab samples can be integrated over depth and/or width. At flowing-water sites, USGS collects an isokinetic, depth-integrated, discharge-weighted sample as standard procedure (Wilde 2006). However, such a sample would not integrate temporal variations (USDA-NRCS 2003). Depth integration in lake sampling can be achieved by mixing grab samples taken from each lake stratum, by obtaining a simultaneous sample of the entire water column with a hose, or by automatic devices that collect at different depths over time (USDA-NRCS 2003).

Isokinetic, depth-integrating methods are designed to produce a discharge-weighted (velocity-weighted) sample (Wilde 2006). Using this method, each unit of stream discharge is equally represented in the sample, either by dividing the stream cross section into intervals of equal width (EWI) or equal discharge (EDI) (Wilde 2006). With the EWI method, depth integrated samples are collected at equally spaced intervals at the cross section and then composited (USDA-NRCS 2003). Under the EDI method, knowledge of stream discharge is used to divide the cross section into equal discharge subsections for sampling. In theory, the two methods will produce composite samples with identical constituent concentrations. The instantaneous load could be determined by multiplying the analyte concentration by the measured instantaneous stream discharge (Wilde 2006). If nonisokinetic sampling methods are used, the method will not result in a discharge-weighted sample unless the stream is completely mixed laterally and vertically.

An isokinetic, depth-integrating sampler is designed to accumulate a representative water sample continuously and isokinetically from a vertical section of a stream while transiting the vertical at a uniform rate (Wilde et al. 2014). Isokinetic, depth-integrating samplers are categorized into either hand-held samplers or cable-and-reel samplers. The USGS provides details on how to use these devices for both isokinetic and nonisokinetic sampling (Wilde et al. 2014).

Integrated samples may be the best approach for situations where water quality is known to be spatially variable, e.g., vertical integration for lake sampling, or horizontal integration for river sampling. Given that the temporal variability of lake conditions is generally not as great as that in streams, integrated grab samples may be the most useful sample type for lakes. Grab samples at various lake depths, however, may provide necessary information that integrated samples "average out," so both types of samples could be appropriate depending upon the monitoring objectives. A combination of seasonal, integrated and simple grab samples taken at representative depths could be the best approach for problem assessment and trend analysis for lakes and other still water bodies. Composite or continuous sampling under these conditions would be likely to generate datasets with substantial serial correlation issues at a cost far greater than simple or integrated grab sampling.

The best approach for lake sampling will depend on project monitoring objectives and lake characteristics. Because sampling throughout the entire water column is not always necessary to characterize conditions of interest, integrated sampling can be unimportant. For example, when monitoring a vertically stratified lake for nutrient problems, it may be most desirable to collect surface grab samples for chlorophyll *a* and use meters to develop depth profiles of temperature, pH, conductivity, and DO. Nutrients could be monitored with surface grab samples only unless project objectives dictated that bottom samples were also necessary. Pairing the chlorophyll *a* and nutrient data from grab samples taken at various surface locations would be appropriate for analysis in most cases.

3.2.2.4 Continuous

Continuous sampling is not usually used in nonpoint source pollution studies, but the USGS uses continuous water-quality monitors in its national assessment of surface waters (Wagner et al. 2006). A commonly used configuration for USGS data collection is the four-parameter monitoring system, which collects temperature, specific conductance, dissolved oxygen, and pH data. Devices currently on the market have sensors for DO, conductivity, pH, turbidity, depth, chlorophyll α, blue-green algae, ammonia, NO_3, Cl^{-1}, total dissolved gas, temperature, and other parameters. Sondes are available that can measure 15 parameters simultaneously. Some instruments can store measurements to internal or external memory in a format compatible with a hand-held display, personal digital assistant (PDA), or laptop computer (Gibs et al. 2007).

Continuous sampling can be performed during short-term or long-term periods depending upon the monitoring objectives. Like grab and composite sampling, continuous monitoring provides no information about the spatial aspects of water quality conditions. Continuous sampling also has the potential to create information overload if carried out during a long period, with the potential consequence of expensive data reduction requirements, including addressing the problem of autocorrelation.

Other challenges associated with continuous sampling include the need for careful field observation, cleaning, and calibration of the sensors (Wagner et al. 2006). Despite manufacturer claims, even “self-cleaning” sensors require cleaning. Most electrodes are temperature dependent and many cannot be placed in areas of high stream velocity (USDA-NRCS 2003), but flow-through systems can be designed to address the stream velocity issue (Wagner et al. 2006). An advantage to continuous sampling is the ability to track the duration of values exceeding thresholds, in particular, those with significant diurnal variability.

With the exception of flow, continuous sampling is not frequently used in nonpoint source monitoring. It may be useful for variables such as temperature or dissolved oxygen, which should be measured *in situ* and for which minimum and maximum daily values are critical concerns. Continuous monitoring cost considerations include the cost of sondes and sensors, labor associated with keeping the sensors clean and operative, and costs associated with reducing the datasets for statistical analysis. Problem assessment and research are two areas for which continuous measurement could be highly appropriate. Continuous sampling could also track the exposure of aquatic organisms to harmful levels of temperature or DO, providing a very useful tool for trend analysis or an assessment of BMP or watershed project effectiveness.

3.3 Station Location

Monitoring station locations must be determined at two distinct scales. At the macro-scale, sampling locations must be determined by monitoring objectives, experimental design and resource type. The micro-scale issues of site access and physical configuration will drive the final selection of station locations.

3.3.1 Macro-scale

At the watershed or macro-scale, monitoring design (see section 2.4) will control station location. A single-watershed or trend design will require a station to be located at a watershed outlet where collected data represent water quality from the entire drainage area. An above-below or input-output design calls for two or more stations bracketing a treated area or an individual BMP to compare concentrations or loads entering and leaving the area. A synoptic or reconnaissance design will need numerous stations, located at areas that can isolate particular drainage areas or NPS pollutant source areas (Figure 3-17). Ground water monitoring for flow and/or mass determination will require an extensive network of monitoring wells to determine flow into and out of the area and to map hydrogeologic properties of the aquifer.

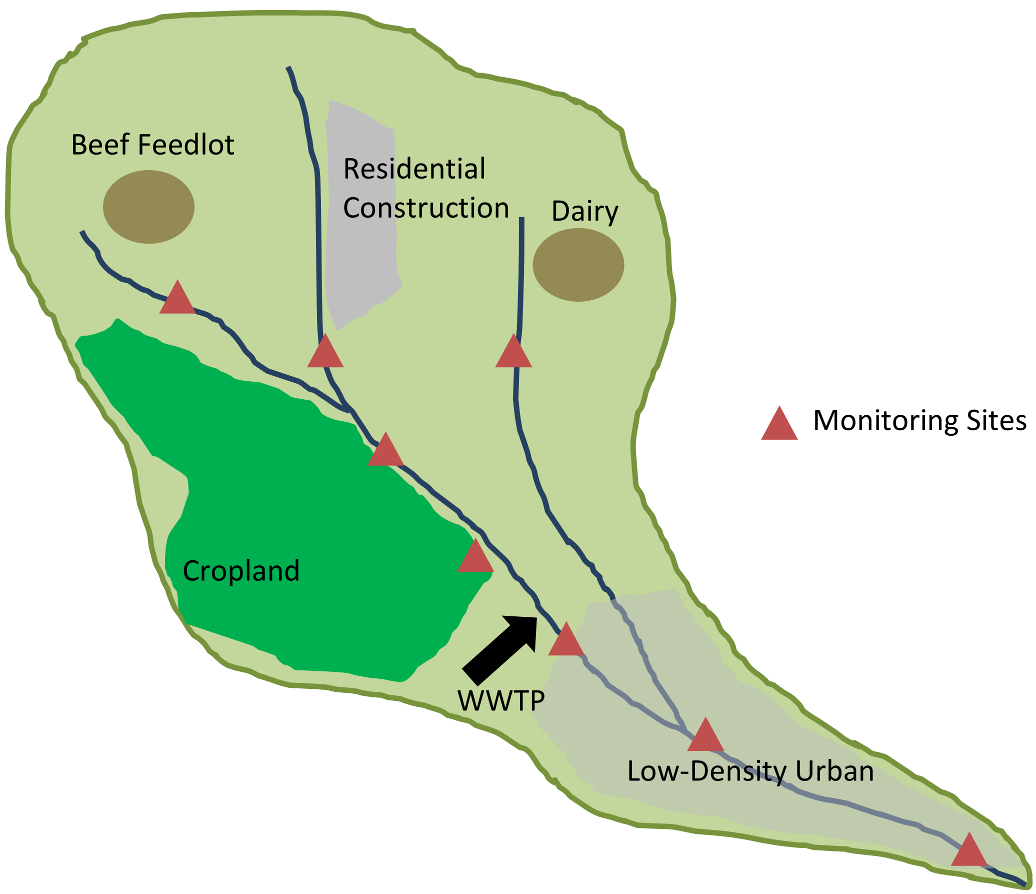


Figure 3-17. Possible sampling locations for a synoptic survey

Water body type is another macro-scale factor in station location. On stream or river networks, station locations might be selected to either capture or avoid the effects of tributary streams, to isolate sub-catchments, or to focus on areas of particular characteristics, e.g., high-quality regional biological reference sites. In lakes and reservoirs, monitoring stations at each major tributary discharge may be required to effectively measure load for a TMDL. In the lake itself, lake morphology, vertical stratification, and currents may require samples in several lake regions and/or at several depths in order to adequately represent water quality (Figure 3-18). Lake sampling designs and factors to consider when selecting sampling locations are described in detail by Nevers and Whitman (n.d.), and U.S. EPA (1998) provides guidance on sampling designs and locations for bioassessments.

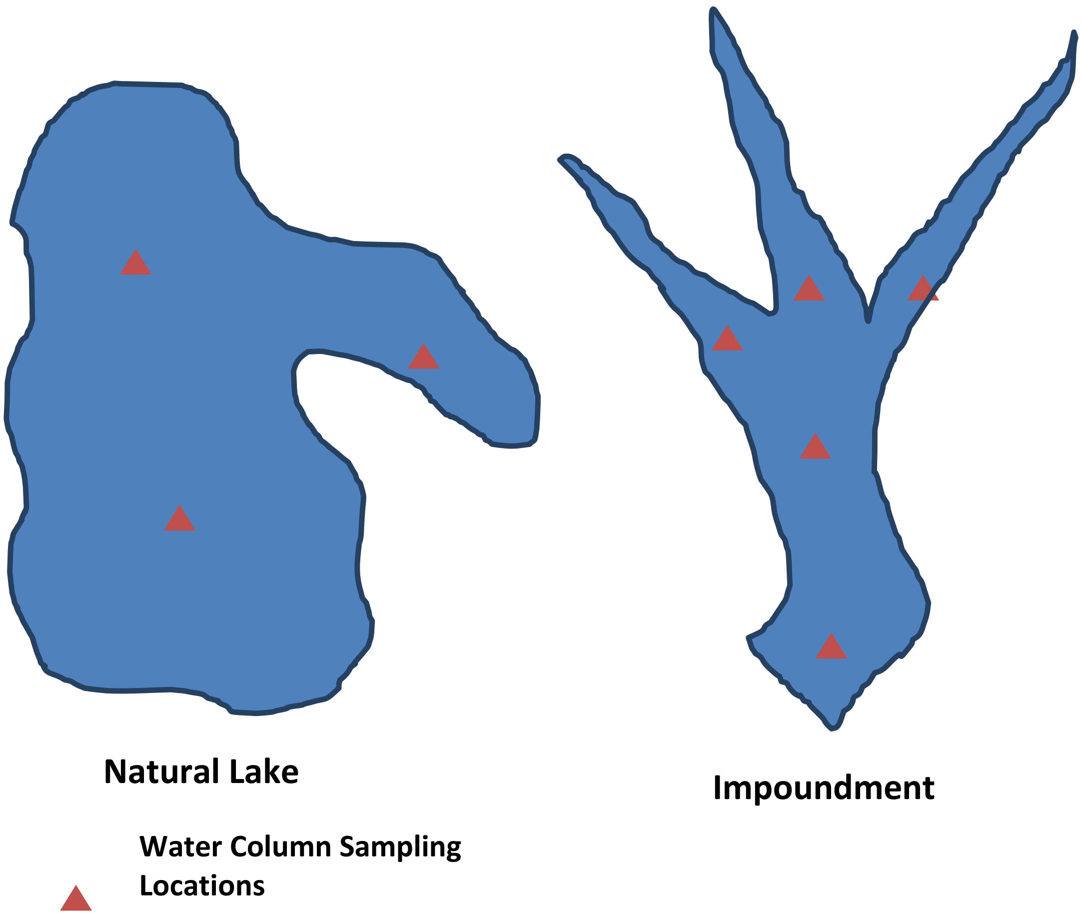


Figure 3-18. Potential lake monitoring locations

In its 2012 National Lakes Assessment, U.S. EPA randomly selected 904 natural lakes, ponds, and reservoirs across the lower 48 states using a probability based survey design (USEPA 2011a). To be included in the survey, these lakes (excluding the Great Lakes and the Great Salt Lake) had to be at least one meter deep and greater than 2.5 ac (1 ha) in size. In addition to these 904 sites, some sites were resampled for quality assurance purposes, and reference sites representing least-disturbed conditions were also sampled. A variety of field measurements were taken at “index sites” which are either the deepest point in a natural lake or the middle of a reservoir (USEPA 2011a). If the deepest point exceeded 50 m in depth, the index site was set as close to the middle of the lake as field staff could go without exceeding 50 m in depth. In addition, conditions of the littoral zone and shoreline were documented from stations around the lake.

The location of monitoring stations in ground water systems is determined by aquifer type and vertical, horizontal, and longitudinal variability in both water quality and water quantity (Figure 3-19). Both USDA-NRCS (2003) and Lapham et al. (1997) provide additional information on well selection for ground water monitoring.

In some cases, it may be appealing to adopt sampling stations that were part of a past monitoring network or are active in another project or program. Piggy-backing on past or existing monitoring stations may offer advantages of an historical data record or significant cost-savings. A prime example is co-locating with an operating USGS station. High-quality continuous flow data (sometimes in real-time through a website) is a major benefit for a monitoring program because flow data are challenging and expensive to acquire. However, adopting sites from past or other monitoring programs must be carefully evaluated before decision making. Such stations may not be located optimally for the current monitoring program's objectives, and data collected for other purposes, objectives and schedules, or by other methods, may not be useful for current needs.

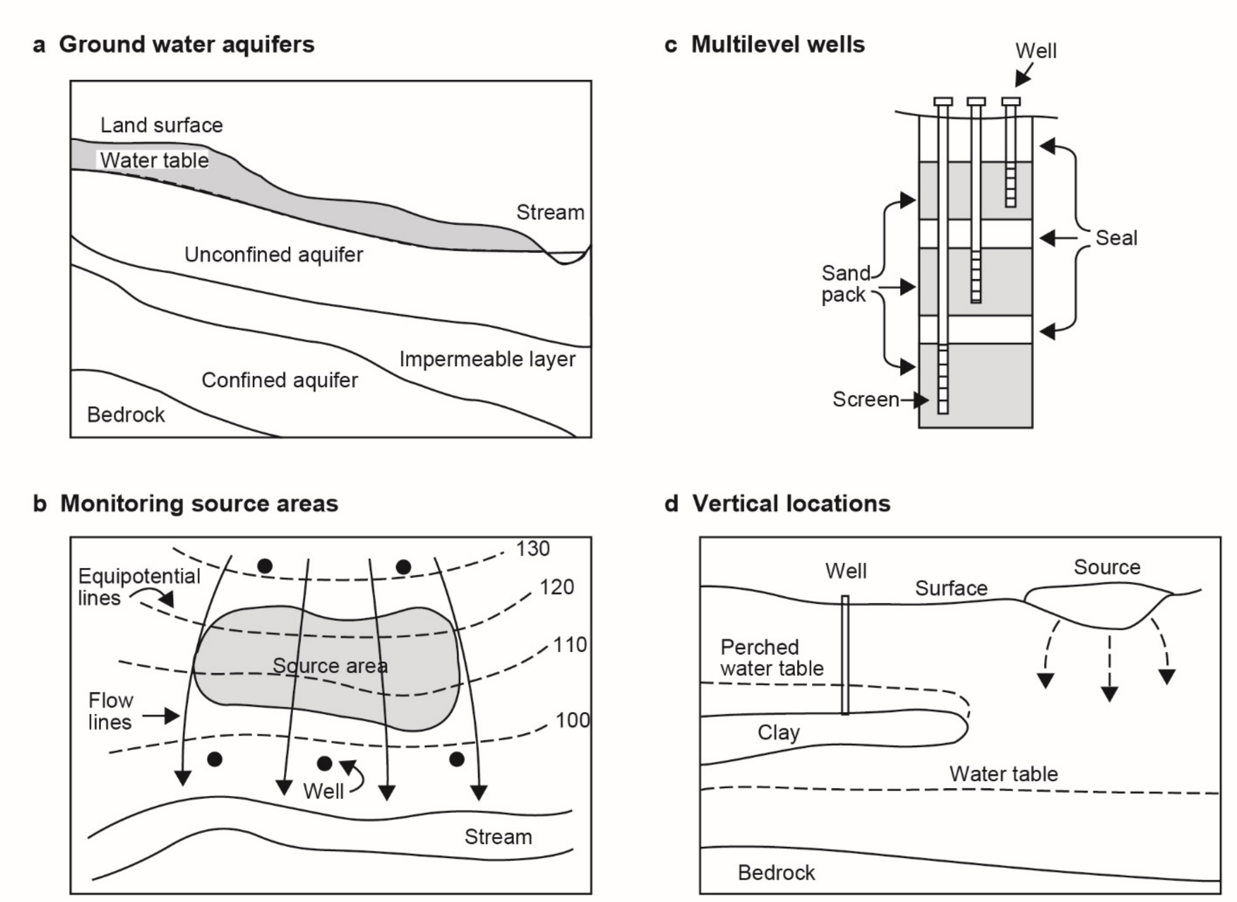


Figure 3-19. Possible groundwater monitoring locations (after USDA-NRCS 2003)

3.3.2 Micro-scale

Some general considerations apply to choosing the location of sampling stations at the local scale, and some specific factors apply to locations for flow measurement and biological monitoring.

3.3.2.1 General Considerations

> **Micro-Scale Site Location Considerations**
>
> - Representativeness
> - Easy access
> - Safety
> - Power
> - Permission
> - Security

Stations must be located so samples and other data can be collected that are representative of the conditions being monitored according to project objectives. In practice, this means that stream stations should be located on relatively straight runs, away from obvious eddies or backwaters, far enough from major obstructions that prevent adequate mixing, and far enough downstream of tributary or other inputs to ensure complete mixing before samples are collected. Lake stations should be located far enough into open water to avoid obvious near-shore influences and outside of confined embayments unless near-shore or embayment conditions are of primary interest. In lakes of complex morphometry, multiple sampling stations may be required to collect representative data. Ground water sampling wells should be arrayed and installed in locations (both horizontally and vertically) that represent the resource of interest, e.g., a known contaminant plume or a regional aquifer system.

Many relevant general considerations for local-scale station location relate to practical matters of logistics (see section 2.2.3.1). Access, in terms of both travel from a base to the site and foot access to the stream and/or station facilities, is critical. Considering the safety of field staff, especially in harsh seasons or inclement weather, is vital. The availability of power and communication links may be essential to some station types. Security, from both human interference and natural threats like flooding, is important as is land ownership. In some cases, stations can be located in the highway right-of-way or on a bridge structure, avoiding the need for negotiations with private landowners (although permission/approval from the state or local transportation agency is usually required). In some cases, permission from or lease or rental agreements with property owners may be required. Finally, if buildings, electrical power, or other physical structures are to be installed, local land use permits may be required.

3.3.2.2 Locations for Flow Measurement

There are some special considerations for locating stream stations at which flow will be measured in open channels.

- Select a straight reach, reasonably free of large rocks or obstructions, with a relatively flat streambed, away from the influence of abrupt changes in channel width.
- Avoid culverts, waterfalls, and bridges where obstructions or degraded structures may cause hydraulic anomalies that interfere with a stable stage-discharge relationship.
- Seek an area with a stable cross-section and avoid areas subject to frequent deposition of sand or gravel bars or severe bank erosion.
- Look for an area where depth and velocity measurements can be conducted safely at low flows.
- Look for an area where a bridge crossing or walkway allows safe velocity measurements at high flows.
- Look for areas where stage can be measured and/or recorded continuously, e.g., protected area for a staff gage.

Where flow is to be measured at the edge of a field or elsewhere using a weir or a flume, look for sites where flow can be collected and/or diverted into the device, where ponding caused by a weir will not cause problems, and where concentrated discharge from a flume can be safely conveyed away downstream. See section 3.1.3.1 for additional information on flow measurement.

3.3.2.3 Locations for Biological Monitoring

Rapid Bioassessment Protocols (Barbour et al. 1999) lists several important considerations for locating biomonitoring sites.

- Ensure a generally comparable habitat at each station. Otherwise, differences in biology attributable to local habitat alone will be difficult to separate from differences or changes in response to water quality degradation due to NPS pollution.
- Locally modified sites, such as small impoundments and bridge areas, should be avoided unless project objectives are to assess their effects.
- Sampling near the mouths of tributaries entering large waterbodies should be avoided because these areas will have habitat more typical of the larger waterbody.
- Biological monitoring programs generally require a reference site to provide data on the best attainable biological conditions in a local or regional system of comparable habitat.

See chapter 4 for additional information on locating biological monitoring sites.

3.4 Sampling Frequency and Duration

The questions of how often to collect samples (the sampling frequency or interval between samples) and how long to conduct a sampling program are critical and without simple or stock answers. The choice of sampling frequency depends on program objectives, type of water body involved, variables measured, and available budget.

3.4.1 General Considerations

In general, sampling frequency must be relatively high (e.g., daily to weekly) for monitoring to evaluate effectiveness of a single BMP or to document the mechanisms controlling water quality at a particular site. Automatic samplers with flow meters that can collect composite, flow-weighted, samples over storm events and collect weekly or biweekly samples enable effective sampling for concentration and load data to evaluate BMP effectiveness. They also reduce the high cost of retrieving and analyzing samples collected more frequently. A program with an objective of detecting a long-term trend or evaluating watershed program effectiveness can accept longer intervals (e.g., weekly to monthly) between samples. Considerations specific to monitoring for load estimation are discussed in section 3.8.

Sampling frequency must also be determined based on the type of waterbody being monitored, and in particular the variability of water quality in the waterbody. Greater variability requires higher sampling frequency to obtain a reasonable picture of water quality. For example, water quality in edge-of-field runoff from cropland is likely to be considerably more variable and require considerably more frequent sampling than water quality in a large lake or a regional aquifer. Water quality in intermittent streams is usually more variable than in large river systems. A general guide to the relationship between system variability and sampling interval is illustrated in Figure 3-20.

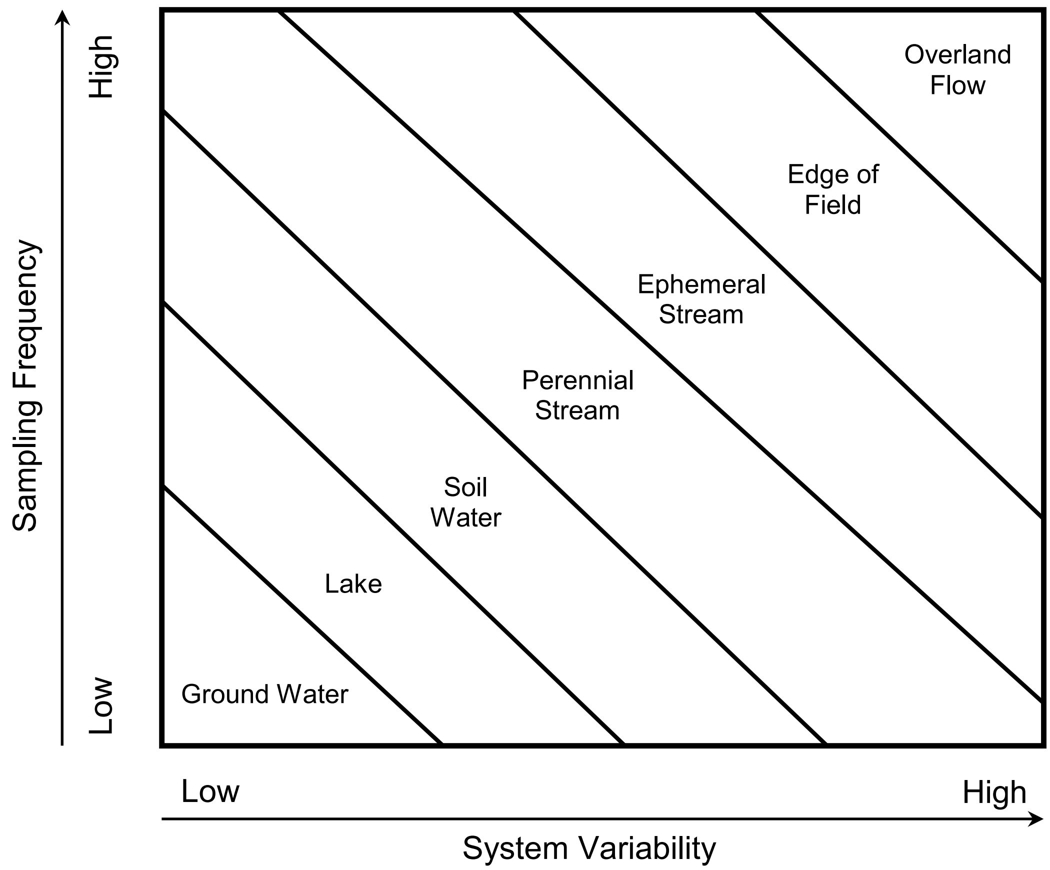


Figure 3-20. Schematic of sampling frequency as a function of system type (after USDA-NRCS 2003)

Project budgets, staff availability, and laboratory capability typically put limitations on sampling frequency, but financial resources should not be the primary basis for decisions on sampling frequency. A sampling program that cannot achieve desired objectives because of inadequate sampling frequency is not cost-effective. Where resources are limited, consider reducing the list of variables analyzed or even the number of stations before cutting back on the sampling frequency that is required to meet project objectives. Use of less expensive surrogate variables, simplifying field instrumentation, and the use of composite sampling programs are all ways to reduce costs while maintaining the critical sampling frequency.

Calculation of appropriate sampling frequency varies with the statistical objective for the monitoring data and sampling regime. Following are examples of how sampling frequency in the context of simple random sampling can be calculated for estimating the mean and for detecting trends.

3.4.1.1 Estimating the Mean

A common objective for monitoring data is to be able to estimate the mean value of a water quality variable with a specific level of confidence in the estimate. The equation for calculating the sample size (Reckhow and Chapra 1983, USDA-NRCS 2003) is:

$$n = \frac{t^2 s^2}{d^2}$$

where:
n = the calculated sample size
t = Student's t at n-1 degrees of freedom and a specified confidence level
s = estimate of the sample standard deviation
d = acceptable difference of the estimate from the estimate of the true mean, or ½ of the confidence interval from the mean

The t value is taken from a table of Student's t at the desired confidence level (*P*) (typically 0.05 or 0.10). In general, a two-tailed t-test should be used because we are usually interested in error on both sides of the mean. The estimate of the population standard deviation is best obtained from baseline data from the monitored water body; if such data are lacking, an estimate from a comparable nearby system can be used. The acceptable confidence interval from the true mean can be expressed as a percent of the mean. The actual calculation may be an iterative process because the value of t may change with the particular value of n chosen. See file nmean.pdf for an example.

3.4.1.2 Detecting a Step or Linear Trend

Another objective for monitoring data might be to detect a change or trend in the value of a water quality variable with a specific level of confidence (see section 7.8.2.4 for a discussion of trend analysis techniques).

Commonly in watershed studies, there are two types of change in the water quality variable studied:

- a step change that compares the pre- and post- water quality mean values
- a linear (gradual, consistent) trend over time

To determine sample size to detect a step change (e.g., comparing the change in baseline mean due to implementation of land management changes), the detectable change must first be calculated based upon the standard deviation of the difference between the pre- and post- means with an anticipated number of samples. See section 3.4.2.3 for an example calculation to determine the detectable step change with a given sample size. With an iterative process of trying different pre- and post- sample sizes, a sample size to detect a step change difference of acceptable magnitude can be estimated. See file ntrend.pdf for an example.

As with documenting a step change between pre- and post- BMP periods, monitoring for trend detection must be sensitive enough to detect the level of water quality change likely to occur in response to management changes. For a linear trend, this monitoring is based upon the confidence interval on the standard deviation of the slope. The standard deviation of the slope (S_{b1}) is a function of both the square root of the MSE (which is the standard deviation of the water quality data with any linear trend removed), as well as the spread of the X's (in this case, length of monitoring:

$$S_{b1} = \frac{\sqrt{\mathrm{MSE}}}{\sqrt{\sum(\mathrm{Xi} - \overline{\mathrm{X}})^2}}$$

Where: MSE = standard deviation of the water quality data with any linear trend removed
X_i = X value at time i, $\overline{X}$ is the average X values.

Typically, for watershed studies, X is expressed as a 'DATE' value which represents 1 day. The slope is therefore expressed as change per day. To express as a change per year over N number of years, the slope/per day would be multiplied by 365 days/year and N number of years.

Therefore, for a linear regression of water quality values vs. time, one-half of the confidence interval on the slope is:

½ confidence interval = (N) * $t_{(n*N-2)df}$ * 365 * s_{b1} [same as Minimal Detectable Change]

Where: $t_{(n*N-2)df}$ = One-sided Student's t-statistic (α=.05)
N = Number of monitoring years
n = Number of samples per year
df = degrees of freedom
365 = Correction factor to put the slope on an annual basis when DATE is entered as a Date (day) variable, e.g., the slope is in units per day. If DATE values were 1-12 for months and the slope was expressed 'per month' then this value would be "12."

The sample size could therefore be calculated interactively by trying various sample frequencies and durations until the watershed monitoring would be able to detect the amount of change anticipated by BMP implementation.

If pre-BMP data exist, the sample variance can be used to estimate MSE (or capture the MSE by running the sampled data through a linear regression computer program. Table 3-10 gives sample size for common sample intervals and durations. Table 3-11 provides example values of $\sum(Xi - \bar{X})^2$ for biweekly sampling that were generated using P concentration data from a long-term NPS monitoring project. This information is used in the linear trend example in file ntrend.pdf. Note that the required sample duration will increase when corrected for autocorrelation (See section 3.4.2).

See Spooner et al. (2011) for more details on calculating the minimum detectable change (MDC) for linear trends. See file ntrend.pdf for an example (hyperlink to be added).

Table 3-10. Number of total samples per indicated sample frequency and number of years

	Total number of samples, n		
Number of years, N	Weekly	Biweekly	Monthly
1	52	26	12
2	104	52	24
3	156	78	36
4	208	104	48
5	260	130	60
6	312	156	72
7	364	182	84
8	416	208	96
9	168	234	108
10	520	260	120

Table 3-11. Values of $\sqrt{\sum(\mathrm{Xi} - \bar{\mathrm{X}})^2}$ for biweekly sampling for selected monitoring durations, assuming X_i is measured as a 'Date' or daily variable

Number of years, N	$\sqrt{\sum(\mathrm{Xi} - \bar{\mathrm{X}})^2}$
2	1,472
4	4,224
8	15,955

3.4.2 Minimum Detectable Change (MDC) Analysis

3.4.2.1 Definition and Overview

The MDC is the minimum change in a pollutant concentration (or load) during a given time period required for the change to be considered statistically significant. Most of the material presented is taken from Spooner et al. (2011) where the reader will find a more detailed discussion, relevant equations, and illustrative examples.

The calculation of MDC has several practical uses, including determining appropriate sampling frequencies (discussed here) and assessing whether a BMP implementation plan will be sufficient for creating change that is measurable with the planned monitoring design (see section 7.6.3). The same basic equations are used for both applications with the specific equations depending primarily on whether a gradual (linear) or step trend is anticipated. The reader is referred to Meals et al. (2011b) for a discussion of these types of trends. In simple terms, one can estimate the required sampling frequency based on the anticipated change in pollutant concentration or load, or turn the analysis around and estimate the change in pollutant concentration or load that is needed for detection with a monitoring design at a specified sampling frequency. The basic steps for conducting MDC analysis and consideration of matters such as the availability of representative data, the distribution of available data, independence of data values, the need for data transformation, and level of statistical significance are touched upon lightly here, but described and illustrated in detail in Spooner et al. (2011).

Sampling frequency determination is very closely related to MDC calculations. Sample size determination is usually performed by fixing a significance level, power of the test, the minimum change one wants to detect, the duration of monitoring, and the type of statistical test. MDC is calculated similarly, except that the sample size (i.e., number of samples), significance level, and power are fixed and the minimum detectable change is computed. In short, MDC is the amount of change you can detect given the sample variability.

3.4.2.2 Steps to Calculate the MDC

The calculation of MDC or the water quality concentration change required to detect significant trends requires several steps described by Spooner et al. (1987 and 1988) for a power of 50 percent. This general procedure varies slightly based upon:

- Whether the appropriate statistical model assumes a step or linear trend.
- Whether the data used are on the original scale (e.g., mg/L or kg) or log transformed.
- Incorporation of time series to adjust for autocorrelation.
- Addition of explanatory variables such as streamflow or season.
- Whether an alternative power is selected.

The following assumptions are made in the calculation of MDC.

- Historical sample measurements are representative of the temporal and spatial variation of the past and future conditions.
- Variability due to sampling, transport or laboratory error is negligible compared to variability over time.

3.4.2.2.1 Step 1. Define the Monitoring Goal and Choose the Appropriate Statistical Trend Test Approach.

One goal may be to detect a statistically significant linear trend in the annual mean (geometric mean is using log transformed data) pollutant concentrations that may be related to land treatment changes. A linear regression model using log-transformed data would be appropriate. An alternative goal to detect a statistically significant change in the post-BMP period as compared to a pre-BMP period would require a step change statistical test such as the *t*-test or ANCOVA.

3.4.2.2.2 Step 2. Exploratory Data Analyses.

The water quality data sets are examined to verify distributional assumptions required for parametric statistical procedures. Specific attention is given to the statistics on normality, skewness, and kurtosis. Preliminary data inspections are used to determine if the residuals follow a normal distribution with constant variance, both of which are required for the parametric analyses to be used. Both the original and logarithmic transformed values are tested. See section 7.10 for a list of available software packages. Options for exploratory data analysis (EDA) include Minitab Basic Statistics (Minitab 2016) and the SAS procedure PROC UNIVARIATE (SAS Institute 2012).

3.4.2.2.3 Step 3. Data Transformations.

Water quality data often follow log-normal distributions. In these cases, use the base 10 logarithmic transformation for the dependent variables (e.g., TP) to minimize the violation of the assumptions of normality and constant variance. Explanatory variables in statistical trend models do not have any distributional requirements because it is only the distribution of the residuals that is crucial. However, if they do exhibit log normal distribution, exploratory variables (e.g., upstream concentrations, flow) are also log-transformed which usually helps with the distribution of the residuals. When log transformation is required for the dependent variables, the log-transformed data are used in all MDC calculations leading to Step 7.

3.4.2.2.4 Step 4. Test for Autocorrelation.

Perform tests for autocorrelation on the water quality time series. An autoregressive, lag-1 (AR(1)) structure in biweekly or weekly samples is common. The tests usually assume samples are collected with equal time intervals. Methods to test for autocorrelation are described in detail in section 7.3.6.

3.4.2.2.5 Step 5. Calculate the Estimated Standard Error.

The variability observed in the historic or pre-BMP water quality monitoring data is used to calculate the MDC estimate. The estimated standard error is obtained from using the same statistical model selected in Step 1.

For a linear trend, use regression models with a linear trend, time series errors, and other optional explanatory variables to obtain an estimate of the standard deviation on the slope over time. If adjusting for autocorrelation, use a software procedure such as SAS's PROC AUTOREG to get the correct standard error on the slope. Alternatively, you can use the standard error adjustment for autocorrelated data given below in this step. For a step trend, use a *t*-test or ANCOVA with appropriate time series and explanatory variables to estimate the standard deviation of the difference between the mean values of the pre-BMP vs. post-BMP data ($s_{(\bar{x}_{pre}-\bar{x}_{post})}$). In practice, an estimate is obtained by using the following formula:

$$s_{(\bar{x}_{pre}-\bar{x}_{post})} = \sqrt{\frac{\text{MSE}}{n_{pre}} + \frac{\text{MSE}}{n_{post}}}$$

Where: $n_{pre} + n_{post}$ = the combined number of samples in the pre- and post-BMP periods

$s_{(\bar{x}_{pre}+\bar{x}_{post})}$= estimated standard error of the difference between the mean values in the pre- and the post- BMP periods.

MSE= s_p^2 = Estimate of the pooled Mean Square Error (MSE) or, equivalently, weighted average ("pooled") of the variances within each period. The MSE estimate is obtained from the output of a statistical analysis using a *t*-test or ANCOVA with appropriate time series and explanatory variables. If post-BMP data are not available, no autocorrelation is present, and no explanatory variables are appropriate (i.e., the simplest case), MSE or s_p^2 can be estimated by the variance (square of the standard deviation) of pre-BMP data.

The standard error on the trend estimate for simple trend models (e.g., step, linear, or ramp trends) with AR(1) error terms is **larger** than that (incorrectly) calculated by software procedures that do not include a correction for autocorrelation. The following adjustment can be applied to obtain the correct standard error for weekly or biweekly water quality data (Matalas, 1967; see Spooner et al. 2011 for additional details):

$$std.dev._{corrected} = std.dev._{uncorrected}\sqrt{\frac{1+\rho}{1-\rho}}$$

Where: $std.dev_{corrected}$ = true standard deviation of the trend (slope or difference between 2 means) estimate

$std.dev_{uncorrected}$ = incorrect standard deviation of the trend estimate calculated without regard to autocorrelation

ρ = autocorrelation coefficient for autoregressive lag 1, AR(1)

3.4.2.2.6 Step 6. Calculate the MDC.

For a power of 50 percent, the MDC is essentially one-half of the confidence interval for the slope of a linear regression trend or for the step trend difference between the mean values of the pre-and post-BMP periods. For a linear trend, the MDC is equal to one-half of the confidence interval on the slope obtained by multiplying the estimate standard deviation of the slope by the *t*-statistic, the total monitoring timeframe, and a correction factor for the additional planned monitoring years (see Spooner et al. 2011 for formulas). For a step trend, the MDC is one-half of the confidence interval to detect a change between the mean values in the pre- vs. post- BMP periods.

3.4.2.2.7 Step 7. Express MDC as a Percent Change.

If the data analyzed were not log-transformed, this is just the MDC divided by the average values in the pre-BMP period expressed as a percentage. If the data were log-transformed, a simple calculation can be performed to express the MDC as a percent decrease in the geometric mean concentration relative to the initial geometric mean concentration or load. The calculation is (see details and examples below):

$$\mathrm{MDC\%} = \left(1 - 10^{-\mathrm{MDC}'}\right) \times 100$$

where MDC′ is the MDC on the log scale and MDC% is a percentage.

3.4.2.3 Examples

The simplest example of an MDC calculation assumes a step trend, no autocorrelation, no covariates or explanatory variables, and Y values on the original scale (i.e., not transformed); see Spooner et al. (2011) for examples of linear trends with autocorrelation and covariates, as well as a paired watershed study or above/below-before/after studies. In this simple example, the planned comparison would be to detect a significant change in the average values between the pre- and post-BMP periods. The pre- and post-periods can have different sample sizes but should have the same sample frequency. Note: in this simplified example, the MDC would be equivalent to the Least Significant Difference (LSD) and would be calculated with a power of 50 percent as:

$$\mathrm{MDC} = t_{(\mathrm{n}_{pre}+\mathrm{n}_{post}-2)} * \mathrm{s}_{(\mathrm{X}_{pre}+\mathrm{X}_{post})}$$

Or, equivalently:

$$\mathrm{MDC} = t_{(\mathrm{n_{pre}}+\mathrm{n_{post}}-2)} \sqrt{\frac{\mathrm{MSE}}{\mathrm{n}_{pre}} + \frac{\mathrm{MSE}}{\mathrm{n}_{post}}}$$

Where: $t_{(\mathrm{n}_{pre}+\mathrm{n}_{post}-2)}$ = one-sided[2] Student's *t*-value with (n_{pre} + n_{post} -2) degrees of freedom.

n_{pre} + n_{post} = the combined number of samples in the pre- and post-BMP periods

$\mathrm{s}_{(\bar{x}_{pre}+\bar{x}_{post})}$= estimated standard error of the difference between the mean values in the pre- and the post- BMP periods.

MSE= s_p^2 = Estimate of the pooled Mean Square Error (MSE)

[2] The choice of one- or two-sided *t*-statistic is based upon the question being asked. Typically, the question is whether there has been a statistically significant decrease in pollutant loads or concentrations and a one-sided *t*-statistic would be appropriate. A two-sided *t*-statistic would be appropriate if the question being evaluated is whether a change in pollutant loads or concentrations has occurred. The value of the *t*-statistic for a two-sided test is larger, resulting in a larger MDC value.

Calculation Example #1 (post-BMP data not available): It is assumed that there will be two years of pre-BMP monitoring following by five years of post-BMP monitoring. For this example calculation, we assume bi-weekly sampling to avoid serious autocorrelation concerns and the need for adjustment. Example #2 illustrates an approach to address autocorrelation associated with weekly sampling.

n_{pre} = 26 samples/yr × 2 yr = 52 in the pre-BMP period
n_{post} = 26 samples/yr × 5 yr = 130 in the post-BMP period
Mean X = 36.9 mg/l, mean of the 52 samples in the pre-BMP period
s_{p} = 21.2 mg/L = standard deviation of the 52 pre-BMP samples
MSE= $s_p^2 = 449.44$
$t_{(n_{pre}+n_{post}-2)} = t_{180} = 1.6534$ (one-sided)

The MDC would be:

$$\text{MDC} = t_{(n_{pre}+n_{post}-2)}\sqrt{\frac{\text{MSE}}{n_{pre}}+\frac{\text{MSE}}{n_{post}}}$$

$$\text{MDC} = 1.6534\sqrt{\frac{449}{52}+\frac{449}{130}}$$

$$\text{MDC} = 5.7 \text{ mg/l}$$

Percent change required = 100 × (5.7/36.9) = 15%

So, in this example, sampling bi-weekly before (2 years) and after (5 years) BMP implementation would require a 15 percent change in concentration to be detectable at the 95 percent confidence level and a 50 percent power. If a smaller change was anticipated, then sampling frequency (or duration) would need to be increased to adjust for autocorrelation. If a decrease of more than 15 percent was expected, then sampling frequency could be decreased and the MDC recalculated to determine if the reduced sampling frequency would be adequate to meet project goals. Because MDC analysis is used to "estimate" detectable change it is recommended that estimated sampling frequency needs are assumed to be higher than calculated to reduce the risk of failure.

Calculation Example #2 (post-BMP data not available, similar data distribution as in example #1): It is assumed that there will be two years of pre-BMP monitoring following by five years of post-BMP monitoring. For this example

Autocorrelation

Essentially means that subsequent samples are influenced by previous samples. These subsequent samples contain less new information than would otherwise be obtained from a completely independent additional sample (i.e., there is information overlap). The result is that autocorrelation reduces the *effective* sample size compared to the situation with no autocorrelation.

Rho**(ρ)** is the **coefficient of autocorrelation**, and basically describes the relationship between the current and its past values.

- Rho increases as the strength of the relationship between current and past samples increases.
- Larger rho means that each collected sample has less new information (i.e., effective sample size is reduced).
- So, the relative improvement in estimates of a mean or a minimum detectable change decreases as sample size increases.
- Rho is used to adjust the standard deviation for inclusion in the step-change MDC calculations demonstrated in this section.

calculation, we assume weekly sampling and address autocorrelation by assuming an autocorrelation coefficient of ρ=0.3 (common for NPS projects with weekly sampling). A corrected standard deviation is calculated as (see Spooner et al. 2011 for additional details):

$$pooled\ std.dev._{corrected} = 21.2 \times \sqrt{\frac{1+\rho}{1-\rho}} = 21.2\ x \sqrt{\frac{1+0.3}{1-0.3}} = 28.9$$

Therefore:

n_{pre} = 52 samples/yr × 2 yr = 104 in the pre-BMP period
n_{post} = 52 samples/yr × 5 yr = 260 in the post-BMP period
Mean X = 36.9 mg/l, mean of the 52 samples in the pre-BMP period
s_p = 28.9 mg/L = corrected standard deviation of the 52 pre-BMP samples
MSE= s_p^2 = 834.67
$t_{(n_{pre}+n_{post}-2)} = t_{362}$ = 1.6491 (one-sided)

The MDC would be:

$$\text{MDC} = t_{(\text{n}_{\text{pre}}+\text{n}_{\text{post}}-2)} \sqrt{\frac{\text{MSE}}{\text{n}_{pre}} + \frac{\text{MSE}}{\text{n}_{post}}}$$

$$\text{MDC} = 1.6491 \sqrt{\frac{835}{104} + \frac{835}{260}}$$

$$\text{MDC} = 5.5\ \text{mg/l}$$

Percent change required = 100 × (5.5/36.9) = 15%

So, in this example, sampling weekly before (2 years) and after (5 years) BMP implementation would also require at least a 15 percent change in concentration to be detectable at the 95 percent confidence level and a 50 percent power. In essence, autocorrelation results in diminishing returns for higher sample frequencies. However, it should be noted that even biweekly sample frequency such as used example #1 also have autocorrelation, just a lesser amount (e.g., ρ=0.1) which would have resulted in a MDC estimate for biweekly sampling in example #1 of 17.2 percent.

3.4.2.4 Factors Affecting the Magnitude of the MDC

Up to this point the discussion of MDC analysis has been based on simplifying assumptions. The reality, however, is that the true MDC value for a specific significance level varies as a function of pollutant variability, sampling frequency, length of monitoring time, other factors (e.g., potential explanatory variables such as season, meteorological, and hydrologic variables), the magnitude and structure of the autocorrelation (see Calculation Example #2 above), and the statistical techniques used to analyze the data. Variations in water quality measurements are due to several factors including:

- A change in land treatment or land use resulting in decreased (hopefully) concentrations and/or loadings to receiving waters (determining the amount of water quality change is usually a key objective of a watershed project).
- Sampling and analytical error.
- Monitoring design (e.g., sampling frequency, sampling location, variables measured).
- Changes in meteorological and hydrologic conditions.
- Seasonality.
- Changes in input to and exports from the system. For example, changes in upstream concentrations can affect the downstream water quality.

The bottom line is that the magnitude of MDC is often larger than expected but can be reduced by:

- Accounting for changes in discharge, precipitation, ground water table depth or other applicable hydrologic/meteorological explanatory variable(s).
- Accounting for changes in incoming pollutant concentrations upstream of the BMP implementation subwatershed (i.e., upstream concentrations).
- Increasing the length of the monitoring period.
- Increasing the sample frequency.
- Applying the statistical trend technique that best matches the implementation of BMPs and other land use changes.

Figure 3-21 through Figure 3-24 illustrate how MDC varies with sampling frequency/duration, confidence level (expressed as percent), coefficient of variation (CV), and autocorrelation coefficient (ρ), respectively using a 50 percent power. These examples all assume a step trend and no covariates or explanatory variables and use the basic equation found in section 3.4.2.3. The CV is used in lieu of standard deviation because it has broader applicability (CV=std.dev./mean). Figure 3-22 to Figure 3-24 assume a seven-year monitoring program (two pre-BMP and five post-BMP) with the same sampling frequency each year. Data are assumed to follow a normal distribution and pre- and post-BMP CVs are assumed to be the same. In Figure 3-21 through Figure 3-23, the values of ρ were assumed to be 0.1 and 0.3 for sampling frequencies of 26 and 52 times per year, respectively. No autocorrelation was assumed for less frequent sampling.

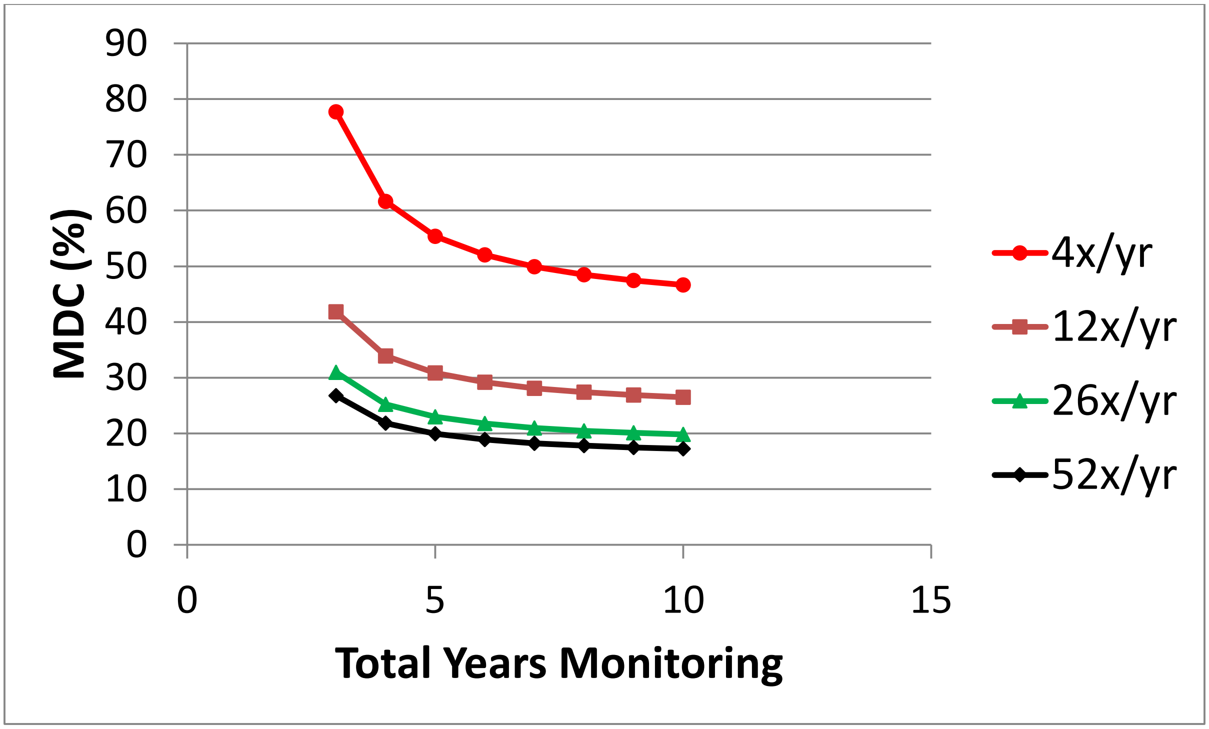


Figure 3-21. MDC versus frequency and years of monitoring. Assumes ρ=0.1 for 26x/yr and 0.3 for 52x/yr, CV=0.7, and 95% confidence level.

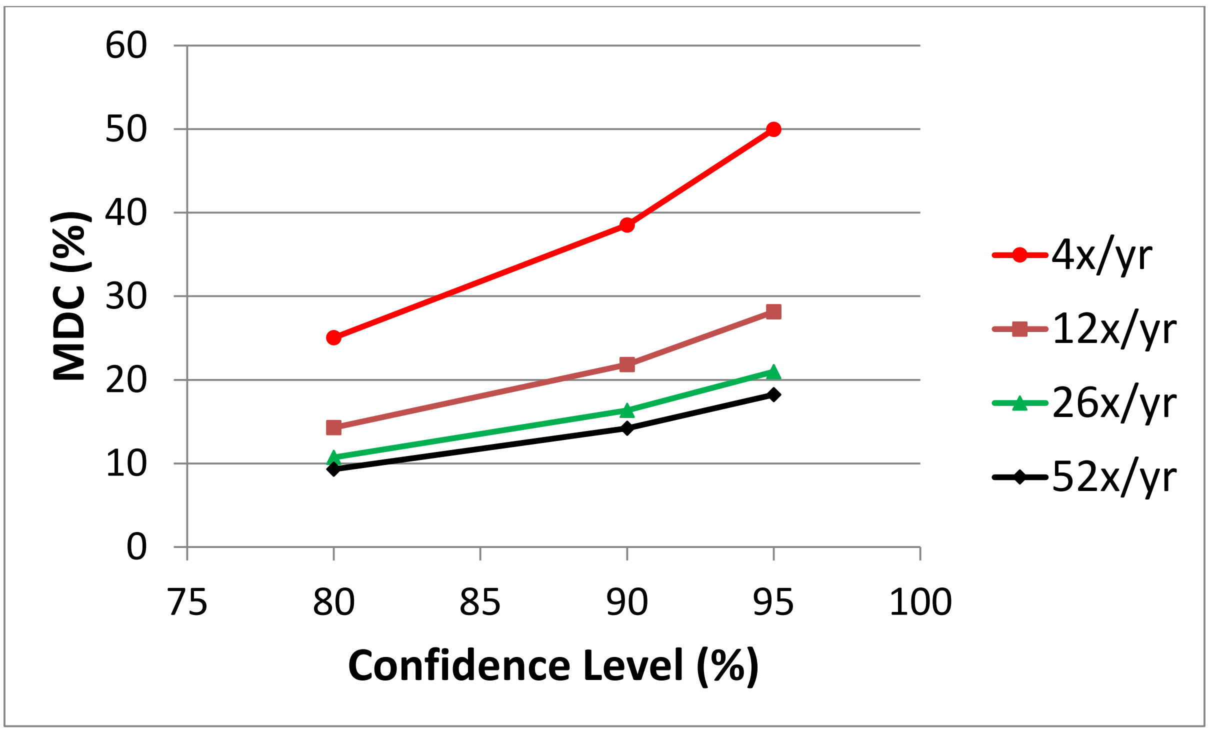


Figure 3-22. MDC versus confidence level. Assumes ρ=0.1 for 26x/yr and 0.3 for 52x/yr, 7 years of monitoring, and CV=0.7.

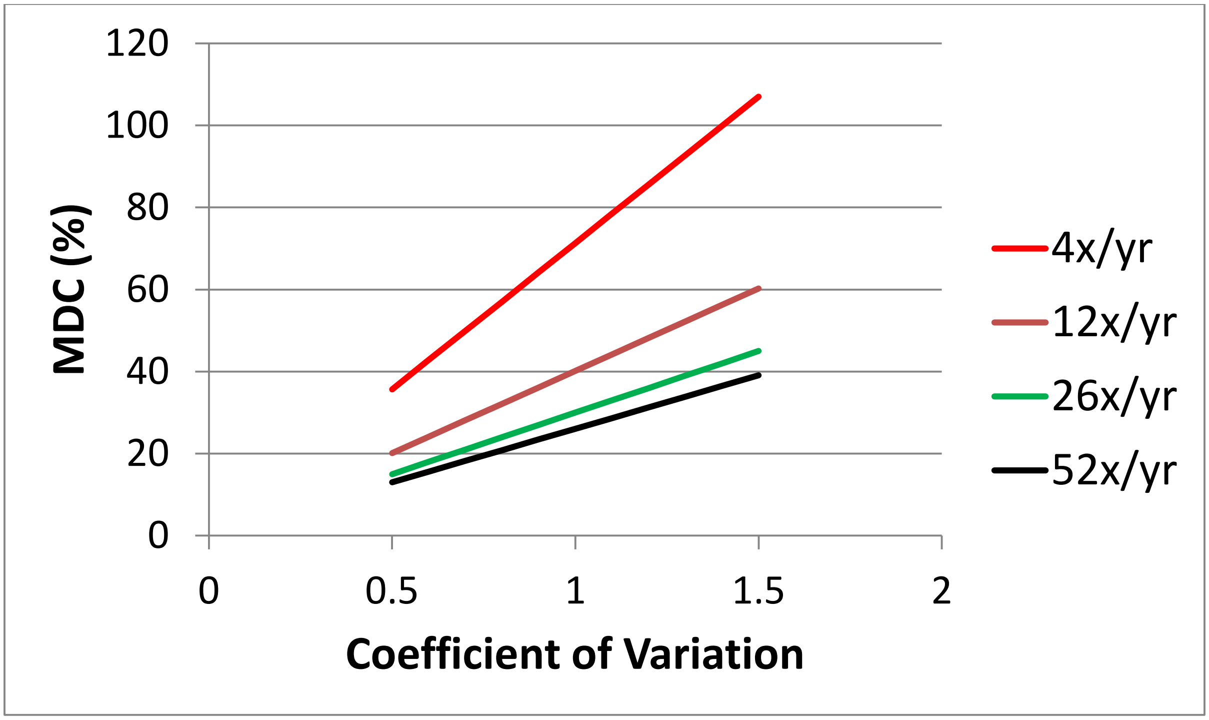


Figure 3-23. MDC versus coefficient of variation. CV calculated using unadjusted std. dev. Assumes ρ=0.1 for 26x/yr and 0.3 for 52x/yr, 7 years of monitoring, and 95% confidence level.

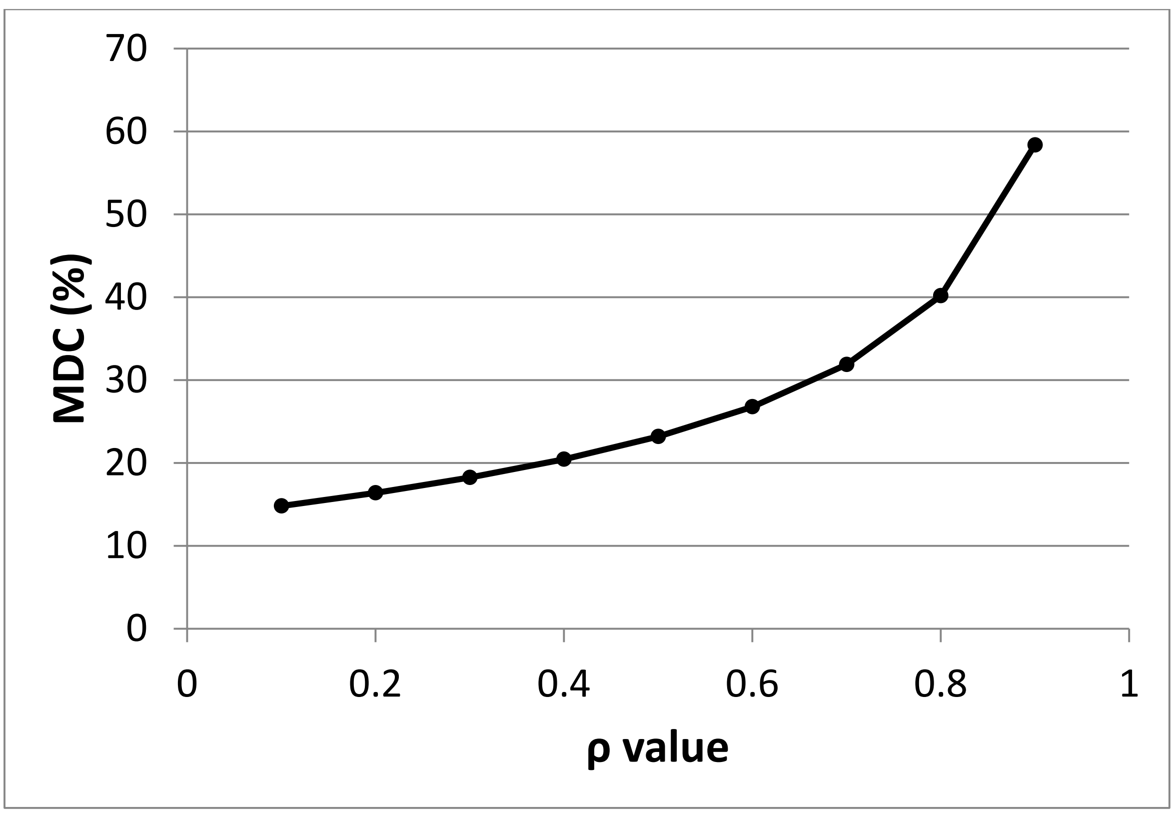


Figure 3-24. MDC versus coefficient of autocorrelation (ρ). Assumes 7 years of monitoring, 52x/yr, CV=0.7, and 95% confidence level. MDC = 13% if no autocorrelation is assumed.

Figure 3-21 shows that the change in MDC is less pronounced with increasing duration for designs with more frequent sampling. For example, MDC drops from 78 percent to about 62 percent when monitoring is extended from three to four years with quarterly sampling; the corresponding change for weekly sampling is only 5 percent. In addition, the change in MDC is minor after seven years for monthly or more frequent sampling. Figure 3-22 illustrates the benefits of considering the statistical confidence needed in changes that might be documented. Sampling 26 times/year over seven years, the MDC drops from 21 to 11 percent when the confidence level is changed from 95 to 80 percent, respectively. In some cases management decisions can be based on less than 95 percent confidence. These changes are more pronounced at lower sampling frequencies. Figure 3-23 illustrates the importance of having a good estimate of variance when calculating MDC. An assumption that the CV=0.5 when it is actually 1.5 could result in a monitoring plan designed to detect an MDC of 15 percent at 26 samples per year when the actual MDC is 45 percent. Finally, Figure 3-24 illustrates the impact of autocorrelation on MDC estimates. In this example (52 samples/year for seven years), the MDC increases with increasing autocorrelation, with an MDC of 20 percent at ρ=0.4 and an MDC of 32 percent at ρ=0.7. Testing for autocorrelation is an important element of using existing data to aid in monitoring plan development, particularly when anticipated sampling frequencies exceed about 25 or more per year.

The reader is referred to Spooner et al. (2011) for additional details on estimation of MDC.

3.4.3 Sampling Duration

How long should a monitoring program be conducted? The answer is essentially: as long as needed to achieve the objectives or document a change. Following are basic guidelines for ensuring that a planned monitoring program has a reasonable chance of success.

- **Capture at least one full cycle of natural or cultural variability.** Especially for NPS situations, monitoring should be conducted long enough to capture the full range of expected variability from weather, seasons and cultural factors such as cropping patterns or construction management. Similarly, if the first year of monitoring is done in a notable drought period, it would be wise to extend monitoring to capture a more representative set of weather conditions.
- **Use statistical tests to evaluate the adequacy of a monitoring period.** Data from some monitoring designs can be tested statistically to determine if an adequate database exists. For example, data from a paired-watershed design (see section 2.4.2.8) can be tested to determine if acceptable calibration has been achieved and if treatment can begin (USEPA 1993b). Pre-treatment data from a before/after design can be evaluated for MDC to help determine if it is likely that enough data exist to document an expected change.
- **Consider lag time.** Lag time between land treatment and water quality response is a common phenomenon (see section 6.2). Knowledge of key lag time factors can help determine the required duration of a monitoring program. For example, if groundwater travel time from an agricultural field through a riparian forest buffer to a stream is known to be five to 10 years, it is reasonable to expect to continue monitoring at least that long. Similarly, a lake with a flushing rate of 1.5 years may respond much more quickly to changes in pollutant inputs and a shorter monitoring program could suffice.

3.5 Monitoring Station Construction and Operation

This section discusses the design and operation of physical facilities involved in fixed monitoring stations. The type of station required depends on both project objectives and the nature of the resource being monitored. Not all monitoring designs require fixed station facilities, e.g., synoptic/grab sampling, lake monitoring, biological monitoring. When physical facilities are required, several important principles apply, regardless of station type.

- **Select monitoring sites according to specific criteria based on program objectives and needs** (see section 3.3).
- **Design the station to collect representative samples from the target population under foreseeable circumstances.** Make certain that measurements and samples are taken from areas that represent the resource or problem of interest, e.g., from the main flow of a stream, not an eddy; from a well-mixed area below a discharge; from the geologic formation transmitting subsurface flow. In situations where vertical or horizontal variability exists, depth-integrated samples or several discrete samples may be required. Physical facilities should allow access and sample collection during anticipated high flows, harsh climates, or inclement weather.
- **Strive for simplicity.** While sophisticated technology offers many capabilities and advantages, power failures and unexpected errors may occur and cause problems in complex designs. When possible, the simple alternative may well be the best choice. A passive crest gage may provide necessary information on peak stream stage more reliably than an electronic sensor. In addition, monitoring systems with data loggers and real-time internet uplinks may function well most of the time, but there is often no substitute for a regular visit by a field technician to maintain equipment and to record key data and observations.
- **Include redundancy.** When possible, provide a backup means of collecting essential samples or data. This may mean including a passive sampling device like a US U-59 single stage sampler (Wilde et al. 2014) as a backup to an autosampler. A flow totalizer on a flow meter provides data on total event discharge in case a data logger fails or a file is corrupted and the continuous stage and flow data are lost.
- **Provide security.** Monitoring instruments and equipment need to be protected both from the elements and from potential vandalism. Field technicians need safe access and protection from inclement weather and other hazards. The integrity of samples and accumulated data should be protected so that adequate chain of custody is maintained.

The following sections discuss important aspects of monitoring station design for several common applications including streams and rivers, edge of field, and individual structures or BMPs. The following are examples of comprehensive references that provide additional detail on these and other matters of monitoring station design.

- USDA *Field Manual for Research in Agricultural Hydrology* (Brakensiek et al. 1979)
- USDA-NRCS *National Handbook of Water Quality Monitoring* (USDA-NRCS 2003)
- USGS *National Field Manual for the Collection of Water Quality Data* (USGS variously dated)

When selecting specific instrumentation and equipment for monitoring stations, review manufacturer information for features and specifications to be sure that equipment can do the jobs required.

3.5.1 Grab Sampling

Even though monitoring programs based exclusively on grab sampling may not require “stations” with physical facilities, grab sampling stations must be located and identified so that samples can be repeatedly collected from the same location. Such locations may be fairly obvious such as road crossings on streams or pipes delivering flow to or from a stormwater treatment system. These sampling locations can simply be recorded on a map or in a standard operating procedure. In lakes, however, repeated navigation to a specific location will likely require use of a global positioning system (GPS) device. Determination of sampling depth will also be required at some lake stations, using a weighted line or an electronic depth sounder. There are numerous devices available to collect grab samples. The choice will depend on water resource characteristics, the type of sample desired (e.g., surface vs. depth-integrated), and on the variable(s) to be monitored (see section 3.6.2.1).

3.5.2 Perennial Streams and Rivers

Long-term stations to continuously record streamflow and collect periodic water samples require structures and facilities to house monitoring equipment. Specific considerations for flow measurement have been discussed previously (see sections 3.1.3.1 and 3.3.2.2). Stations for continuous flow measurement require a staff gage and a means of continuously recording stage, e.g., using a stilling well with a float or bubbler or directly in the channel using a bubbler, pressure transducer, or ultrasonic device. The traditional float gage in a stilling well is highly reliable and is protected from turbulence, ice and debris in the stream channel. Advantages of bubblers, transducers or ultrasonic devices are they can be placed directly in the stream channel, data can be logged electronically, and flow data can be linked to an autosampler. A diagram of a stream station with an in-stream pressure transducer and staff gages is shown in Figure 3-25 (Freeman et al. 2004).

Water samples at continuous monitoring stations are typically collected by autosamplers. Autosamplers commonly pump samples from the stream through plastic tubing and collect the water in one or more bottles. Modern autosamplers are sophisticated instruments that can collect timed samples of specific volume based on their own internal programs or collect storm-event or flow-proportional samples when linked to a flow recorder or other triggering device (see section 3.6.2.4). One common issue associated with pumping autosamplers is the nature and placement of the intake. Sampler intake is usually fixed at some point in the stream and may not collect a sample representative of vertical or horizontal variability. Some depth-integrated intake devices have been proposed and tested with success (Eads and Thomas 1983), but some of these devices can require frequent maintenance and can be impractical in northern climates where ice is a problem. Selbig and Bannerman (2011), however, demonstrated the idea of vertical stratification of solids in storm sewer runoff using a fully-automated, depth-integrated sample arm (DISA) for collecting integrated samples within pipes (Figure 3-26). Subsequent laboratory testing showed that the DISA was better able to characterize suspended-sediment concentration and particle size distribution compared to fixed-point methods (Selbig et al. 2012).

Some variables (like temperature, turbidity, specific conductance, and dissolved oxygen) can be monitored *in situ* without collecting actual water samples using sensors deployed directly in the stream. Installation and operation of such sensors for continuous monitoring requires consideration of site-specific characteristics related to exposure of the sensors to the water, mounting platforms, protection from fouling and impact from debris, calibration, and maintenance. Consult manufacturer recommendations and additional resources for specific guidance on sensors (e.g., Miles 2009, USEPA 2005b).

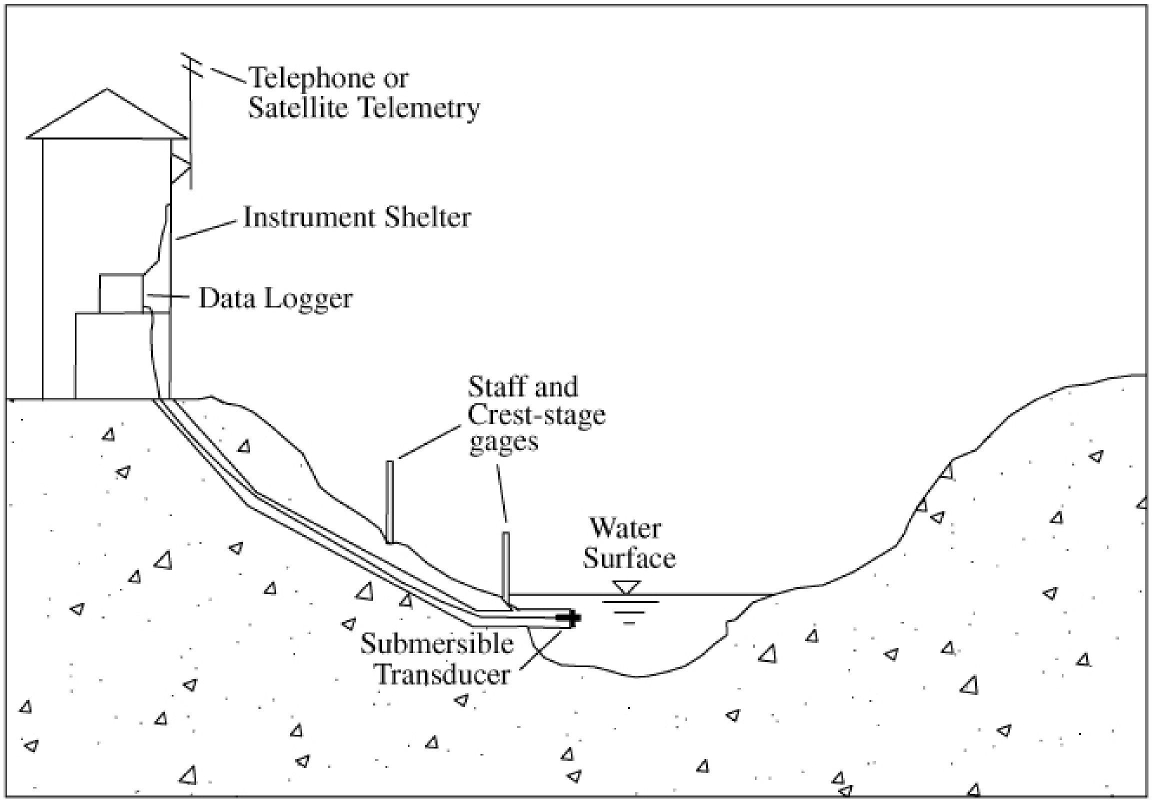


Figure 3-25. Monitoring station with submersible transducer in stream (Freeman et al. 2004)

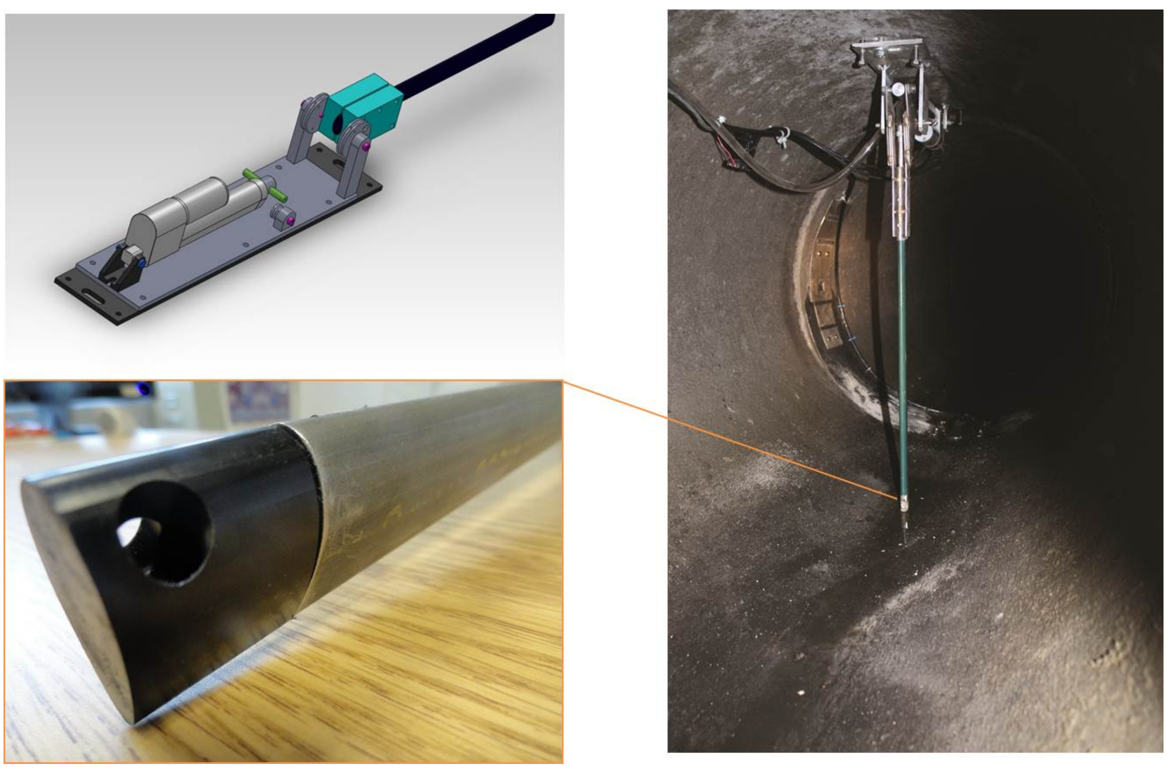


Figure 3-26. Drawing and field installation of depth-integrated sample arm for automatic samplers (photo by R.T. Bannerman, Wisconsin DNR)

The major advantage of autosamplers and recording sensors is that they can operate unattended for extended periods. Autosamplers, for example, can remain dormant for weeks and triggered by precipitation or rising flow independent of personnel action. This is particularly important when monitoring transient storm events is an objective. However, such equipment is expensive and requires regular maintenance and calibration.

Modern monitoring instruments can be linked together with a data logger (either a separate unit or part of either the flow meter or autosampler) for sampling control and data storage. Where resources are available, stations can be equipped to communicate through cell phone systems or Internet in real time. In such cases data can be downloaded and commands for sampling or recording data can be sent remotely. If this kind of system is used, issues of communication linkage such as line-of-sight for radio or connectivity for cell phones must be considered during station design.

Unless a completely passive, mechanical system is devised, most water quality monitoring stations will require electrical power. Power can be provided with deep-cycle automotive or marine batteries, but servicing and recharging batteries may be problematic and battery power may be inadequate for running refrigeration or heating. For long-term application, it is desirable to obtain AC power from either the electrical grid or a properly designed solar charging system. It should be cautioned, however, that electronic monitoring instruments are often vulnerable to voltage spikes that may occur, especially in rural areas, and computer-type power surge protectors should be used to prevent instrument damage.

Finally, it should be noted that stream monitoring stations face a number of challenges in northern climates. Ice in the stream channel can disrupt a stage-discharge rating (see section 3.1.3.1) and disable or destroy sampling lines or instruments located in the stream. Winter weather may require robust shelter and prolonged low temperatures may require heat from heating tape or propane heaters to prevent samples and equipment from freezing. Conversely, stations in hot climates may require special cooling and/or ventilation for proper operation. Such requirements must be considered in designing monitoring stations.

3.5.3 Edge of Field

"Edge of field" generally describes a situation where flow is intermittent and may or may not move through defined channels. For the purposes of this manual, this includes monitoring in waterways or points of concentrated flow at the edges of agricultural fields or in intermittent streams in any location associated with field drainage. Edge of field monitoring stations share many common requirements with stations on perennial streams, i.e., the need to measure flow (when it occurs), the need to collect representative water samples and other data, the need for power, and challenges of extreme weather. Edge-of-field stations face several additional challenges including:

- **Lack of a defined drainage channel,** requiring measures such as wingwalls or berms to direct flow into and/or out of the station.
- **Intermittent flow,** requiring that monitoring equipment be prepared for activation (e.g., by precipitation or flow) at any time.
- **Unpredictable timing and magnitude of flow,** requiring wide tolerances in flow and sampling capacity.
- **Remote location,** usually lacking easy access and power from the grid.

Stuntebeck et al. (2008) provides a comprehensive discussion of how these challenges were met in edge-of-field monitoring stations at the Discovery and Pioneer Farms in Wisconsin. Typical edge-of-field stations included these elements:

- **Enclosures** consisting of custom-made, aluminum, clam-style structure to house equipment designed to measure stage, collect water samples, and provide two-way telecommunication.
- **Stage and discharge equipment** including
  - Wingwalls and berms to collect overland flow.
  - A flume for discharge measurement.
  - A discharge outlet to prevent erosion and ensure proper flume operation.
  - A bubbler gage, pressure transducer, or acoustic sensor for water level recording.
  - A crest gage as a backup and calibration check for recorded stage data.
- **Sampling equipment** including an autosampler and sample intake line protected from freezing by using a down-gradient slope, heat tape, and foam insulation.
- **Data logging and control instruments.**
- **Communications** including radio modem and datalogging communications software.
- **Power** including solar-charged DC batteries for electronics operation and an AC generator for heating and sample refrigeration.
- **Digital time-lapse camera** to periodically record field conditions.

Finally, it should be noted that edge-of-field stations typically require more maintenance than continuous stream stations. Edge-of-field stations may have to remain dormant but ready for activation over extended periods between events, and regular maintenance visits are required even when inactive. This is particularly true in northern climates where removal of ice and snow in preparation for monitoring critical winter thaw or spring runoff events is especially labor-intensive.

Figure 3-27 shows examples of edge of field monitoring stations.

3.5.4 Structures/BMPs

Monitoring stations for specific BMPs or stormwater treatment structures are similar in many respects to edge-of-field stations, but require some additional considerations because of site characteristics and constraints.

Many individual BMP monitoring efforts have similar requirements for flow measurement, water sampling, data logging, communications, and security as other station types, but are often constrained by physical characteristics. Monitoring inflow and outflow from a constructed wetland is generally comparable to monitoring flow in an intermittent stream. Runoff from a parking lot entering an infiltration BMP, however, may be very difficult to quantify and sample, and outflow from the BMP may be carried in an underground pipe. Some specialized equipment for such monitoring has been developed, including passive runoff samplers (Figure 3-28) and flume inserts for pipes with integrated stage sensors (Figure 3-29). In a review of passive samplers for urban catchment studies, Brodie and Porter (2004) classified them based on the main hydraulic principle applied in their design: gravity flow, siphon flow, rotational flow, flow splitting, and direct sieving. In two Wisconsin studies, Parker and Busch (2013) demonstrated the capabilities and limitations of a crown divisor sampler in the laboratory and at the edge of a small field, while Graczyk et al (2000) compared siphon samplers to automatic samplers in a stream setting. In urban settings, much of the monitoring equipment may need to fit into a catch basin or storm sewer access point. Station enclosures and security in urban areas may present additional challenges.

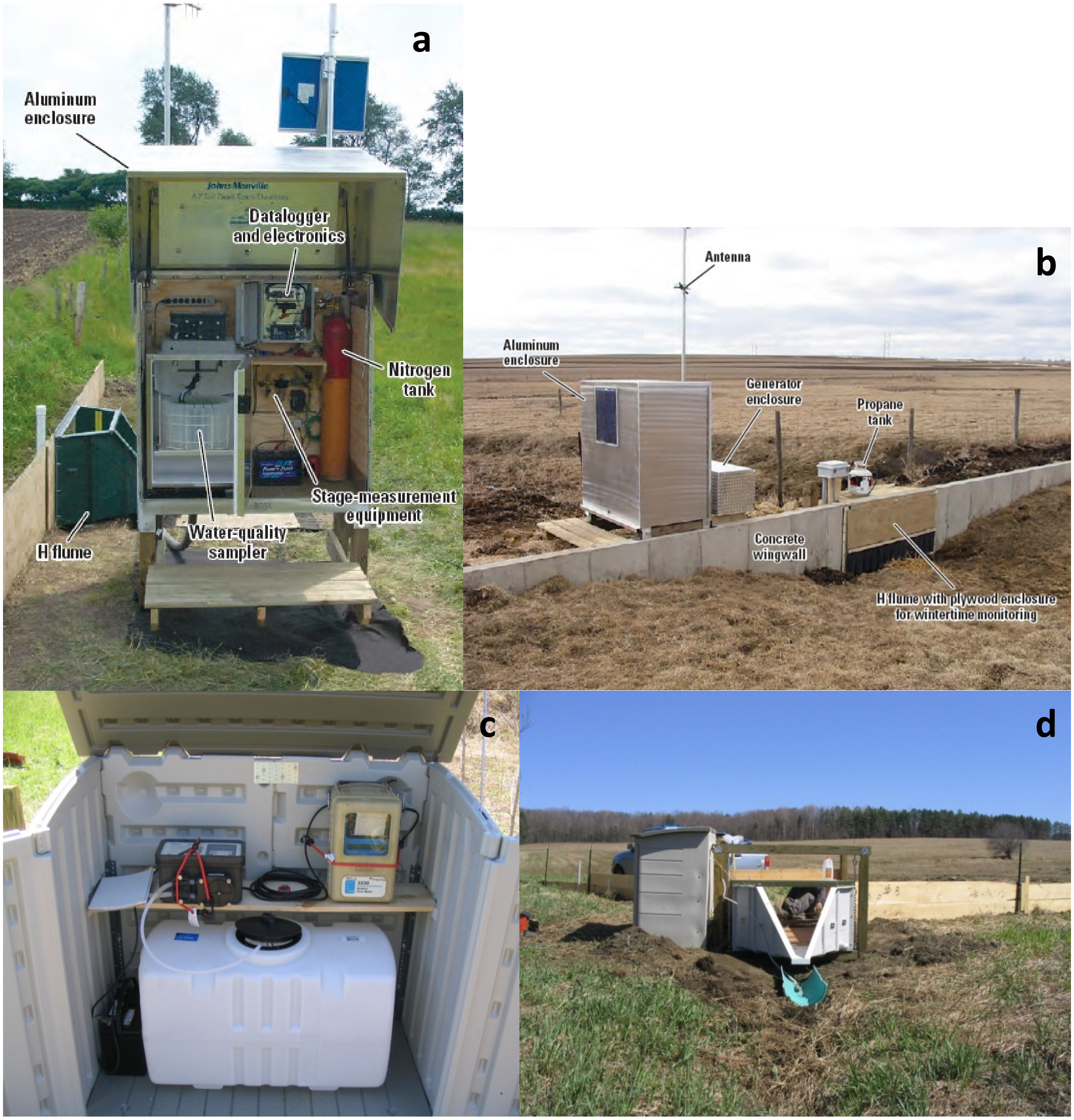


Figure 3-27. Edge-of-field monitoring stations. *a, b,* Wisconsin Discovery and Pioneer Farms (Stuntebeck et al. 2008); *c, d*, Vermont (Meals et al. 2011a).

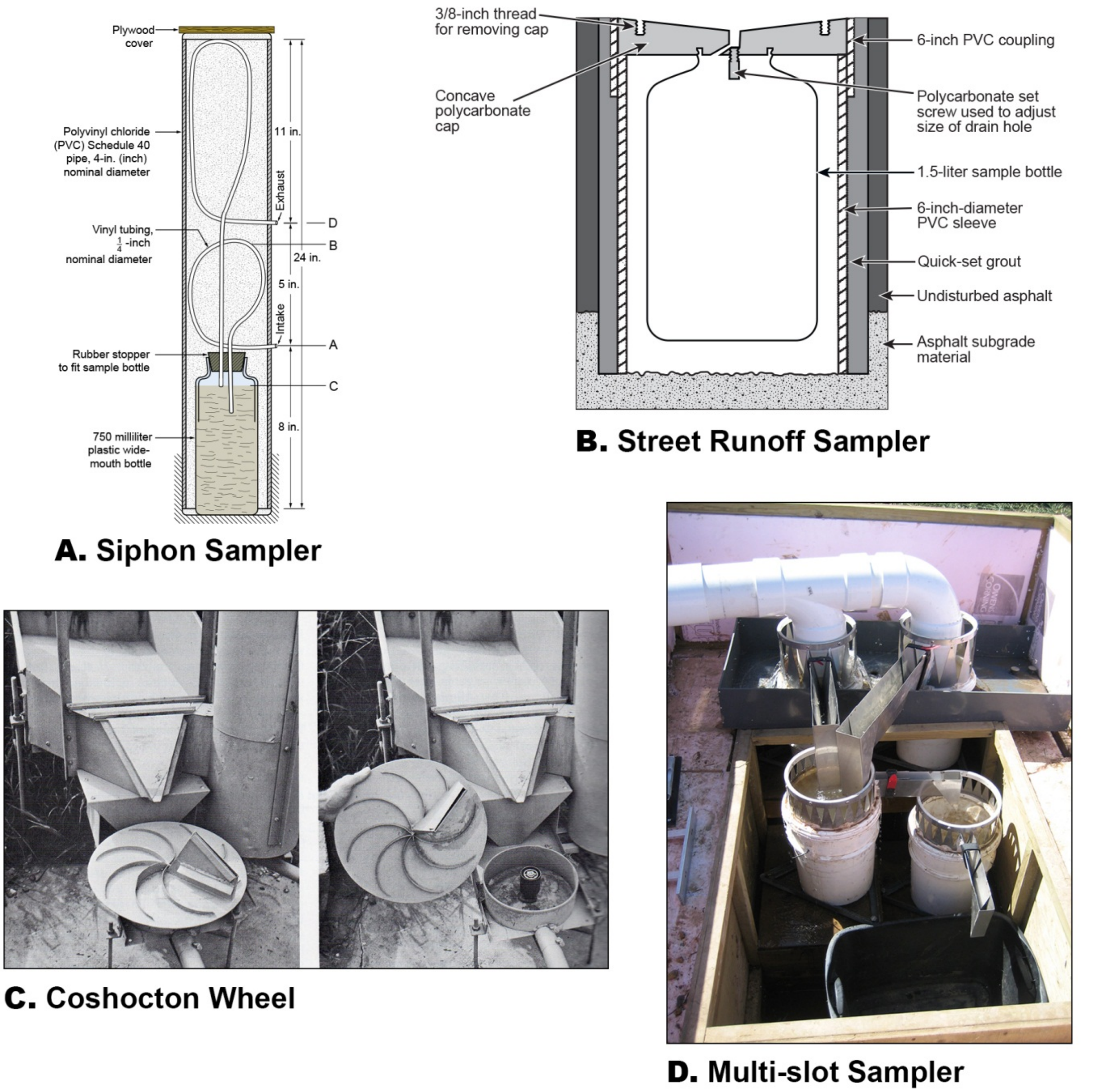


Figure 3-28. Examples of passive runoff samplers that can be used for edge-of-field or BMP studies (A-Graczyk et al. 2000, B-Waschbusch et al. 1999, C-Brakensiek et al. 1979, and D-Parker and Busch 2013; photo D by P. Parker, University of Wisconsin-Platteville)

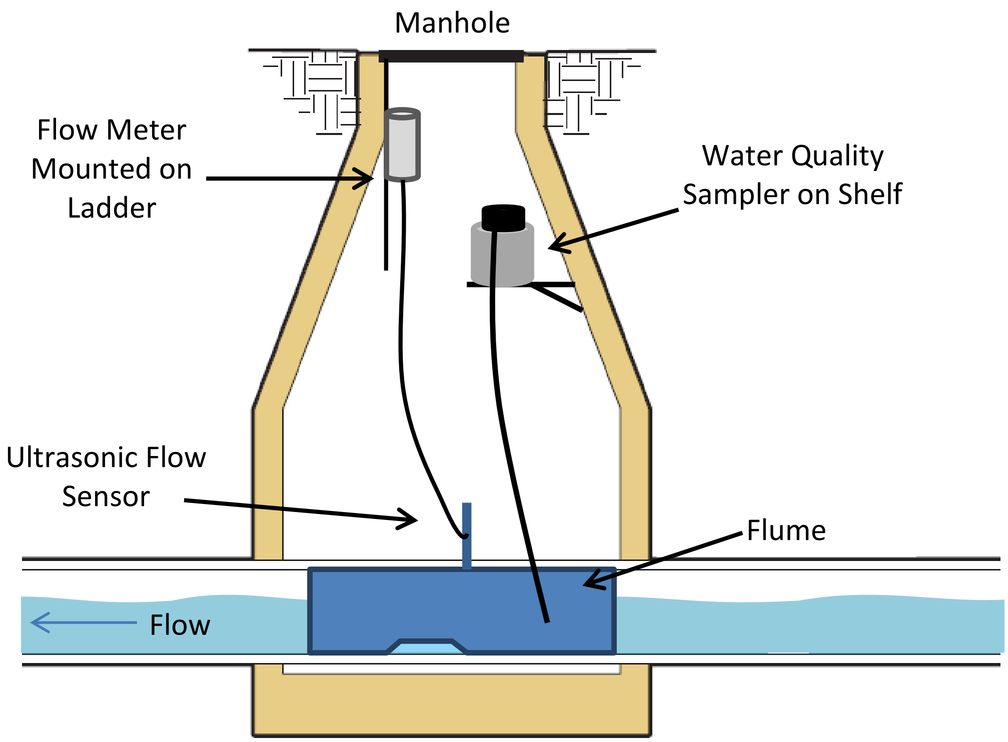
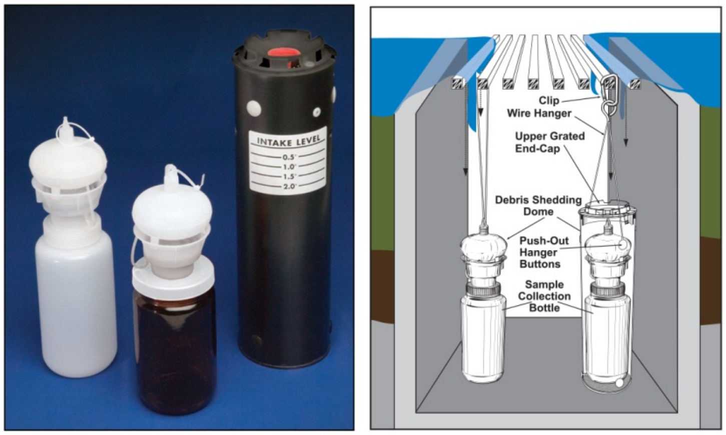


Figure 3-29. Flow measurement and water quality sampling in stormwater pipes

In urban runoff monitoring, the first flush phenomenon (the initial surface runoff from a rainstorm carrying high levels of pollutants that accumulated on impervious surfaces during dry weather) requires special consideration because pollutant loads during the first part of an event may be much larger than those in the later flows. Several approaches have evolved to monitor this phenomenon. Low-cost passive first-flush samplers are available that capture early surface runoff, then close when filled (Figure 3-30). Waschbusch et al. (1999) used a range of passive samplers to monitor street runoff, driveway runoff, lawn runoff (Figure 3-31), roof runoff, and parking lot runoff. Some modern autosamplers offer special settings for activation of intensive sampling programs at certain flow levels, then scale back sampling frequency later in the event (Figure 3-32).



A. Nalgene® first-flush sampler. Installed below grate (at right).

B. Edge-of-road sampler.

C. GKY first-flush sampler.

Figure 3-30. Examples of first-flush runoff samplers (A-Nalgene 2007, B-Barrett 2005, C-GKY 2014)

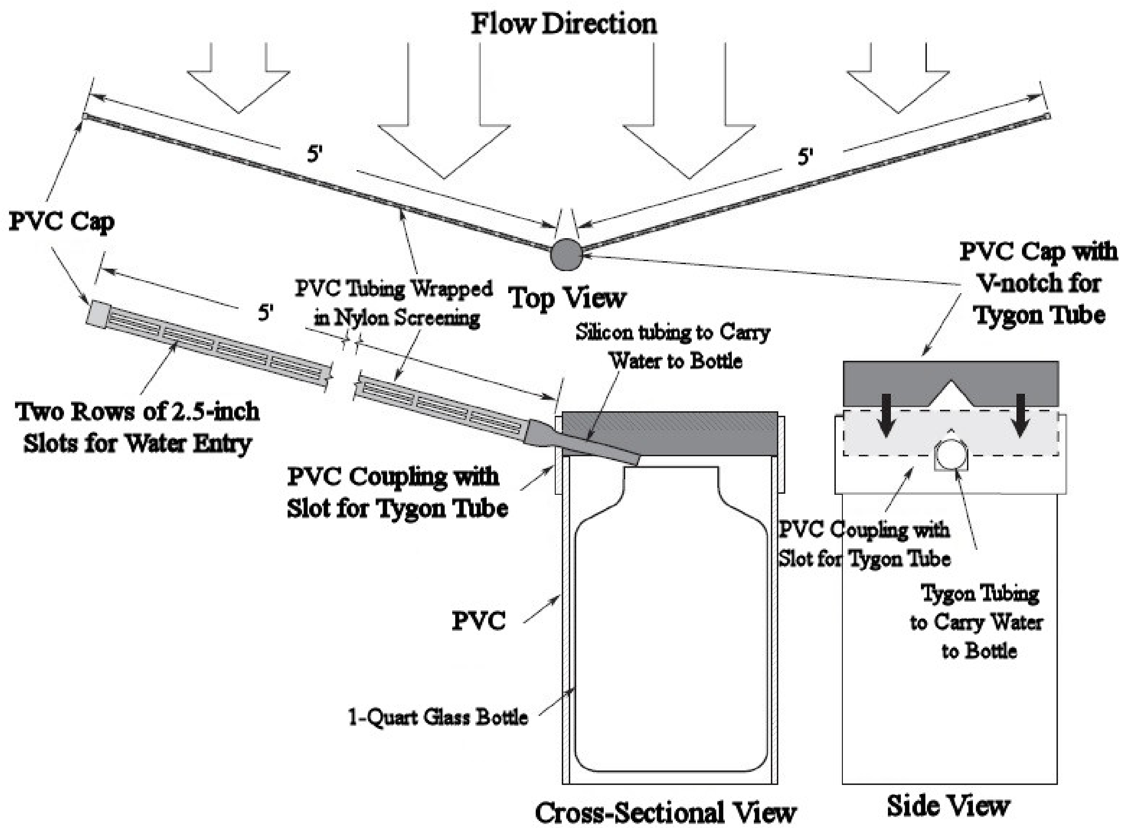


Figure 3-31. Passive sampling setup for lawn runoff (after Waschbusch et al. 1999)

Sigma 900 MAX Portable Standard Sampler

Isco 6712 Portable Sampler

Figure 3-32. Examples of automatic samplers with capabilities for variable sampling frequencies (Hach® 2013a, Teledyne Isco 2013a)

Many BMP monitoring efforts follow an input/output design, where water quality (i.e., concentration or load) is measured entering and again leaving the structure to assess pollutant reduction performance. Such cases not only require two monitoring stations but also require that the stations be coordinated so that water actually treated by the BMP is sampled properly. If sampling is conducted simultaneously at the entrance and exit of a stormwater BMP, for example, the outflow sample may represent "old" water pushed out of the BMP by "new" inflow, rather than new inflow after treatment by the BMP. Similarly, water quality measured simultaneously upstream and downstream of a feedlot may not reflect the influence of the feedlot, at least early in a storm event. Time of travel or residence time in the BMP must be considered in setting up monitoring stations. This can be accomplished by linking the above and below stations to better coordinate downstream with upstream sampling. Stuntebeck (1995), for example, modified the basic above/below design in a Wisconsin barnyard runoff study by setting the samplers to be activated by precipitation and programming them to collect time-integrated samples for an initial period. This modification allowed for sampling of barnyard runoff in the receiving stream before streamwater level increases could be sensed, thereby effectively isolating the barnyard runoff from nonpoint-pollution sources upstream. Secondly, this approach allowed sampling during small storms in which local inputs from the barnyard were apparent, but little storm runoff from the upstream areas of the watershed were observed. A second modification took advantage of the close proximity of the two stations to create a direct electronic connection between the stations for collection of concurrent samples.

3.5.5 Meteorology

Meteorological data, particularly precipitation data, are nearly always relevant to NPS monitoring projects (see section 3.1.5). The nature and extent of meteorological monitoring will vary according to monitoring objectives. Precipitation data are useful in driving event sampling and for documenting rainfall conditions relative to long-term averages. Particular monitoring objectives may require monitoring of other meteorological variables. A study of indicator bacteria runoff from agricultural fields, for example, may call for monitoring of weather conditions that influence bacteria survival in the field, such as air temperature, soil temperature, solar radiation, relative humidity, and wind velocity.

Guidance for meteorological monitoring is given in the USDA Agricultural Handbook 224 (Brakensiek et al. 1979) and in the National Weather Service Observing Handbook No. 2 (NWS 1989). Probably the most important criterion for precipitation measurement is location. For BMP or field monitoring efforts, a single meteorological station may be sufficient. For larger watershed monitoring, multiple stations are usually necessary to account for variation of weather with elevation, and other geographic factors. Multiple precipitation stations are especially important in monitoring efforts designed to provide data for model application. Successful application of watershed models such as SWAT is highly dependent on accurate precipitation data (Gassman et al. 2007). Precipitation monitoring stations must be located so that there are no obstructions within 45º of the lip of the gage (USDA-NRCS 2003). A more restrictive general rule, illustrated in Figure 3-33, indicates that an obstruction should not be closer to the gage than two to four times the obstruction's height above the gage (Brakensiek et al. 1979).

A variety of instrumentation is available for meteorological monitoring, including many electronic instruments that record directly into dataloggers. Tipping bucket rain gages measure both total accumulated rainfall and rainfall rate and can be connected to other monitoring instruments to log data and/or trigger sample collection. For winter operation, tipping bucket gages must be heated electrically. A weighing bucket precipitation gage can measure both rain and snow if it is charged with anti-freeze in the winter. It is generally a good idea to provide a manual (non-recording) rain gage on the station site as a backup and calibration check for the recording instrument.

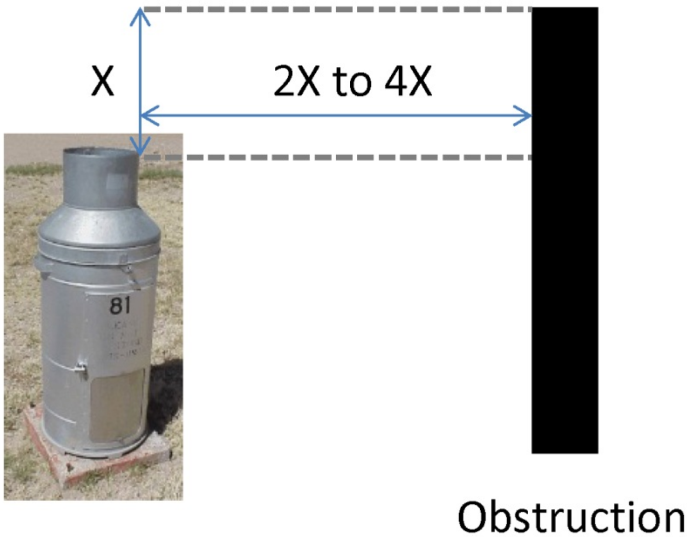


Figure 3-33. Precipitation gage placement relative to obstructions

An example of a meteorological station measuring precipitation, air temperature, solar radiation, relative humidity, and wind velocity is shown in Figure 3-34.

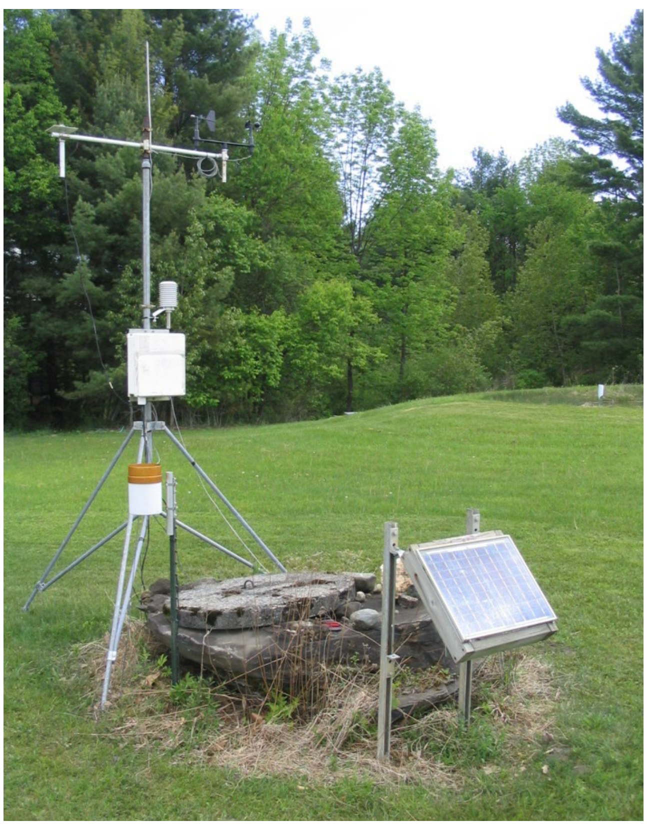


Figure 3-34. Photograph of a meteorological monitoring station (Meals et al. 2011a)

3.6 Sample Collection and Analysis Methods

Collection and analysis of samples, and obtaining measurements and other data from monitoring stations is an exacting task that requires training, appropriate equipment, careful adherence to standard procedures, and detailed record-keeping. This guidance discusses basic principles and important rules of thumb. Other sources such as the USGS National Field Manual for the Collection of Water Quality Data provide specific information and procedures.

3.6.1 General Considerations

This section presents some general aspects of sample collection and is primarily focused on preparation to collect specific types of samples. A preliminary step in determining sample collection and analysis methods for a new monitoring project is to examine how sampling was performed under other past or current monitoring efforts in the area or in other locations you may be interested in. As noted in the discussion of trend monitoring (section 2.4.2.4) changes in methods over time can doom the analysis, so it can be very important to align your methods with those used in the past. Unless there is a compelling reason to use different sample collection and analysis methods from those used to generate past data, it may be best to simply use the same methods to increase the likelihood of data compatibility.

3.6.1.1 Documentation and Records

Because field personnel may rotate assignments in a monitoring project, it is critical that field procedures be documented clearly to ensure consistency, both day-to-day and over the long term. Preparation of field manuals and written standard operating procedures (SOPs) will help supplement the basic training that will be required for field personnel. Field personnel should also keep meticulous sample collection records to support and explain the data being collected. These records should include a logbook of calibration and maintenance records for field instruments and notes concerning variations from SOPs, errors, extreme events and field conditions.

3.6.1.2 Preparation for Sampling

Preparation for a sampling trip includes activities such as cleaning, calibrating, and testing field instruments and sampling equipment as well as making certain that all needed supplies and equipment are assembled. The USGS recommends that a formal checklist be filled out in preparation for each sampling trip to make sure that nothing essential is forgotten (Wilde variously dated).

3.6.1.3 Cleaning

Sample containers must be clean to avoid contamination and preserve sample integrity (Wilde 2004). Most water quality variables have specific requirements for the type and composition of sample container and the cleaning process appropriate for that constituent (see section 3.6.3.2 for references to sources of information on analytic methods). Field personnel must ensure that sample containers they take are prepared for use. In the field, most polyethylene sample bottles and those glass sample bottles that are designated for analysis of inorganic constituents should be field rinsed with the same water that will ultimately fill the sample bottle. Specific field rinsing procedures recommended by USGS are described in Table 5-2 of Wilde et al. (2009).

3.6.1.4 Safety

Field personnel are subject to the basic safety policies and regulations of their employer. In addition, field work for water quality monitoring presents special hazards and considerations that should be addressed. Some important safety protocols include:

- Field personnel should not work alone, should have capacity for communication, and should leave contact and itinerary information with their base.
- Pay attention to inclement weather, especially when sampling from boats in open water or sampling in flashy urban streams. Seek shelter or head back to shore if threatening conditions approach.
- When wading to collect samples or make measurements, wear a personal flotation device (PFD) and do not attempt to wade a stream where the depth exceeds 4 ft or where the product of depth (in ft) times velocity (in ft/s) equals or exceeds 8 anywhere in the cross section. This guidance is based on a study that tested the stability of human subjects over a velocity range of 1.2-10 ft/s and a depth range of 1.6-4 ft (Abt et al. 1989).
- When electrofishing (see section 3.6.2.6 and chapter 4), always work in teams of two properly trained technicians and use proper protective equipment.
- Follow standard safety procedures around mechanical equipment and hazardous chemicals.
- Use caution and extra protection when working with water known or suspected to contain high levels of pathogens.

These and other important procedures are documented in detail in chapter 9 of the USGS National Field Manual (Lane and Fay 1997).

3.6.2 Field Procedures

General procedures are discussed below for different types of sampling. The reader is encouraged to consult other resources (e.g., Barbour et al. 1999; USGS variously dated) for more detailed information on specific sampling procedures. A detailed discussion of sample types can be found in section 3.2.

3.6.2.1 Field Measurements

Collection of data on some water quality characteristics must be based on field measurements, rather than samples collected for later analysis in a laboratory. Variables such as water temperature and dissolved oxygen concentration must be measured directly in the waterbody (Figure 3-35). Other properties such as pH, specific conductance, and turbidity can be measured either *in situ* or immediately on the site using a sample taken from the source, depending on the specific instruments involved.

An *in situ* measurement is made by immersing one or more instrument sensors directly into the waterbody. In flowing water, a single sampling point in a well-mixed area is generally used to represent an entire cross-section, often after a preliminary investigation of variability has been made from repeated measurements at points along the cross-section. In lakes or other still water, field measurements may be made at multiple locations and depths, depending on monitoring objectives and the variability of the waterbody. It is important to record the results of individual measurements from the field, not averaged values. Field measurements in ground water generally require purging the monitoring well of standing water before taking measurements so that the measurements accurately represent the properties of the water in the geologic formation at the time of collection. Following purging, field measurements are performed either above ground by pumping water from the well or downhole, using submersible sensors.

Figure 3-35. Measuring dissolved oxygen, specific conductance, pH, and water temperature using a hand-held probe

Detailed procedures for making field measurements are presented in chapter 6 of the USGS National Field Manual (Wilde variously dated).

3.6.2.2 Grab Sampling

There are a variety of devices available to collect grab samples from waterbodies for different purposes (Wilde et al. 2014).

- **Isokinetic depth-integrated samplers** are designed to accumulate a representative water sample continuously and isokinetically (water approaching and entering the sampler intake does not change in velocity) from a vertical section of a stream while transiting the vertical at a uniform rate. Isokinetic samplers may be hand-held or used with cable systems. Such devices are often used for suspended sediment sampling because maintaining constant velocity facilitates the collection of a sample that is representative of all suspended matter moving in the water column. Some examples of isokinetic samplers are shown in Figure 3-36.

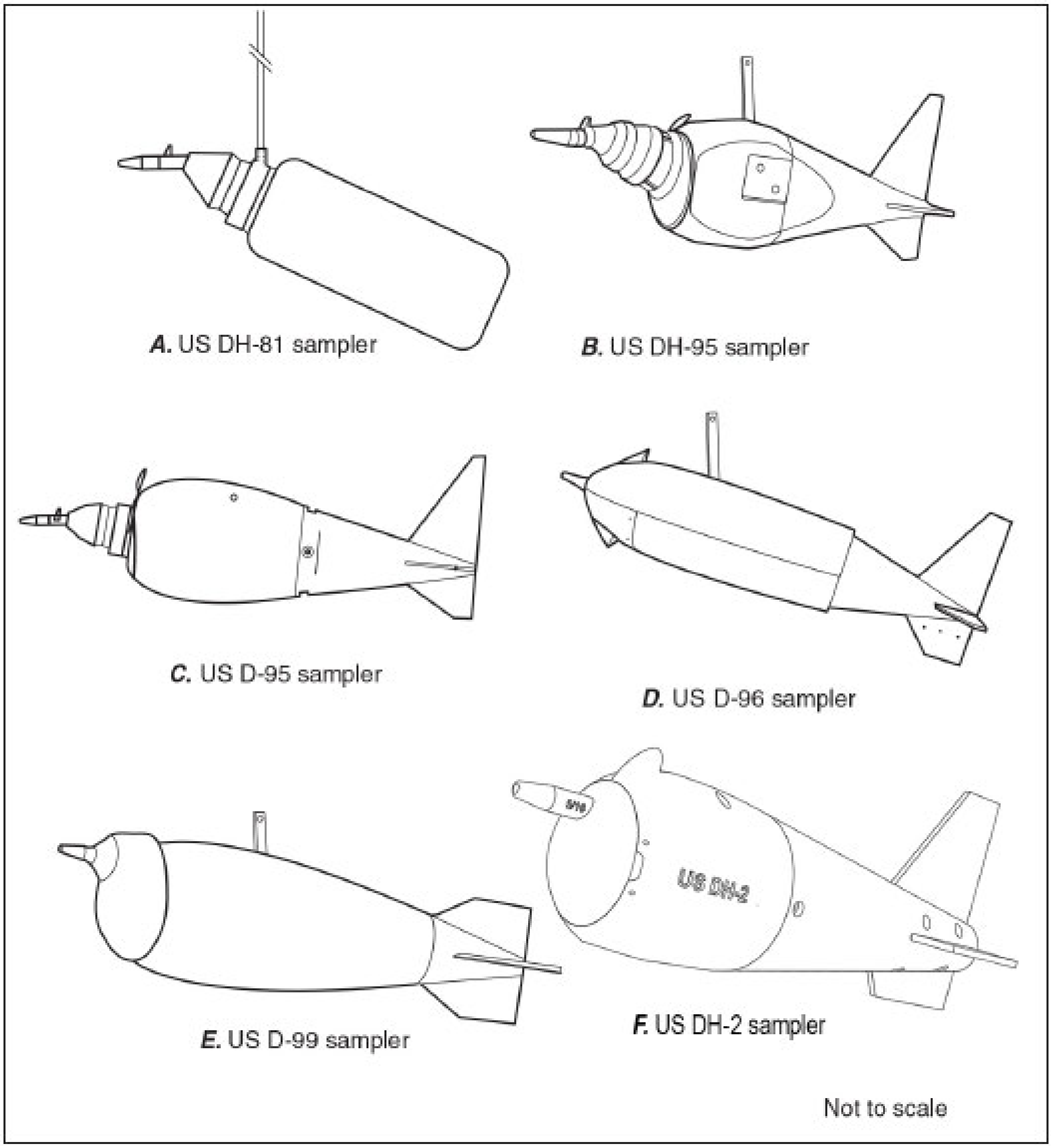
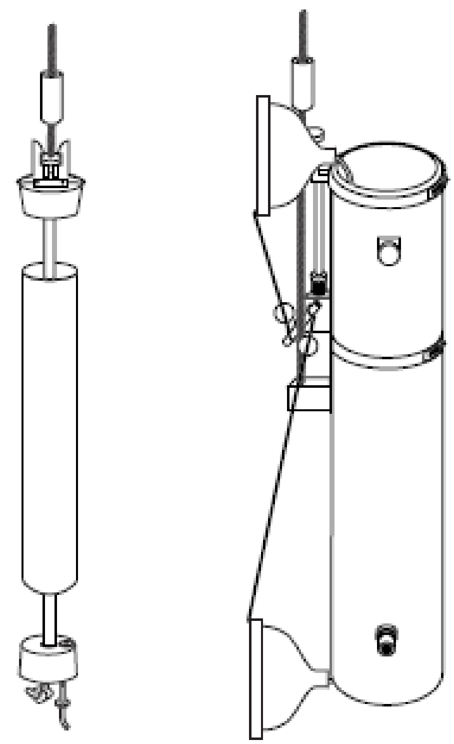


Figure 3-36. Examples of isokinetic depth-integrating samplers (Wilde et al. 2014)

- **Nonisokinetic samplers** are sampling devices in which the sample enters the device at a velocity that differs from ambient. Nonisokinetic samplers include ordinary hand-held open-mouth bottles, weighted bottles on cables, and specialized BOD and volatile organic compound (VOC) samplers for collecting non-aerated samples.
- **Depth-specific samplers** (also called "thief samplers") are used to collect discrete samples from lakes, estuaries and other deep water at a known depth. Common samplers of this type (another form of nonisokinetic sampler) include the Kemmerer and Van Dorn samplers (Figure 3-37).



A. Kemmerer sampler B. Van Dorn sampler

Figure 3-37. Depth-specific samplers for lake sampling (Wilde et al. 2014)

3.6.2.3 Passive Sampling

Passive samplers are devices to collect unattended grab samples without reliance on external power or electronic activation. They offer the convenience of unattended operation, however in most cases the exact time and circumstance of sampling is unknown unless other data are taken at the same time. Some passive samplers are also limited to collecting samples from the rising limb of the hydrograph, so resulting data may be biased compared to samples collected during the full event. Examples of passive samplers include:

- **Runoff samplers** are used to collect overland flow from urban or rural areas. A first-flush sampler is often a bottle buried so that its mouth is flush with the ground (see Figure 3-30). When the bottle is filled, a check-valve closes, preventing subsequent flow from entering. Another type of runoff sampler/flow splitter collects overland flow and splits off a subsample into a down-slope container. Examples are shown in Figure 3-38.
- **Single-stage samplers** (Figure 3-39) are designed to collect unattended samples for suspended sediment or other constituents from streams during storm events. Multiple units can be mounted above each other to collect samples from different elevations or times as stream stage increases.

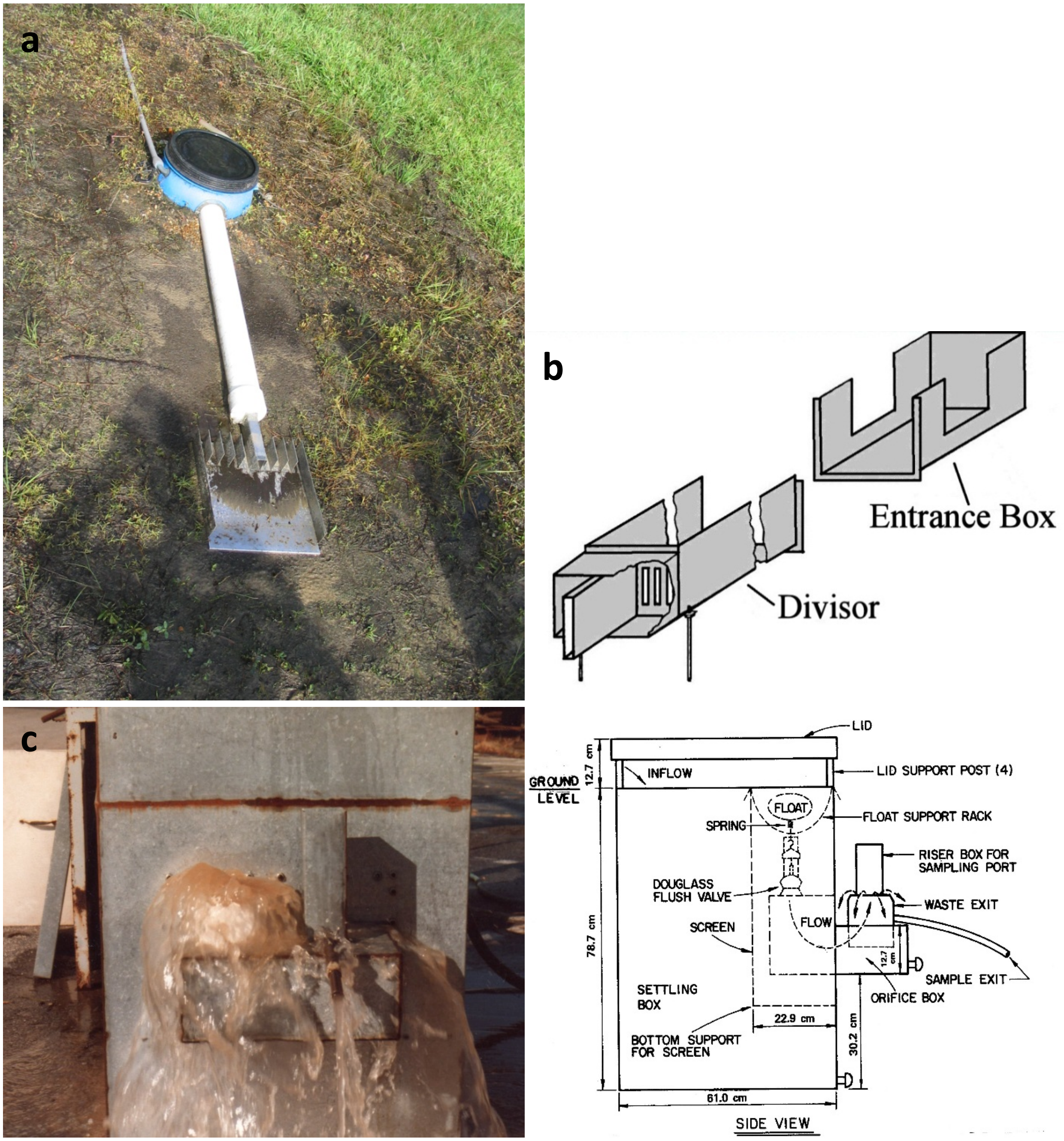


Figure 3-38. Examples of passive samplers. *a,* Passive runoff sampler/flow splitter, University of Georgia, Tifton, GA (photo by D.W. Meals); *b*, Multi-slot divisor (after Brakensiek et al. 1979); *c*, Water and sediment sampler (Dressing et al. 1987, photo by S.A. Dressing).

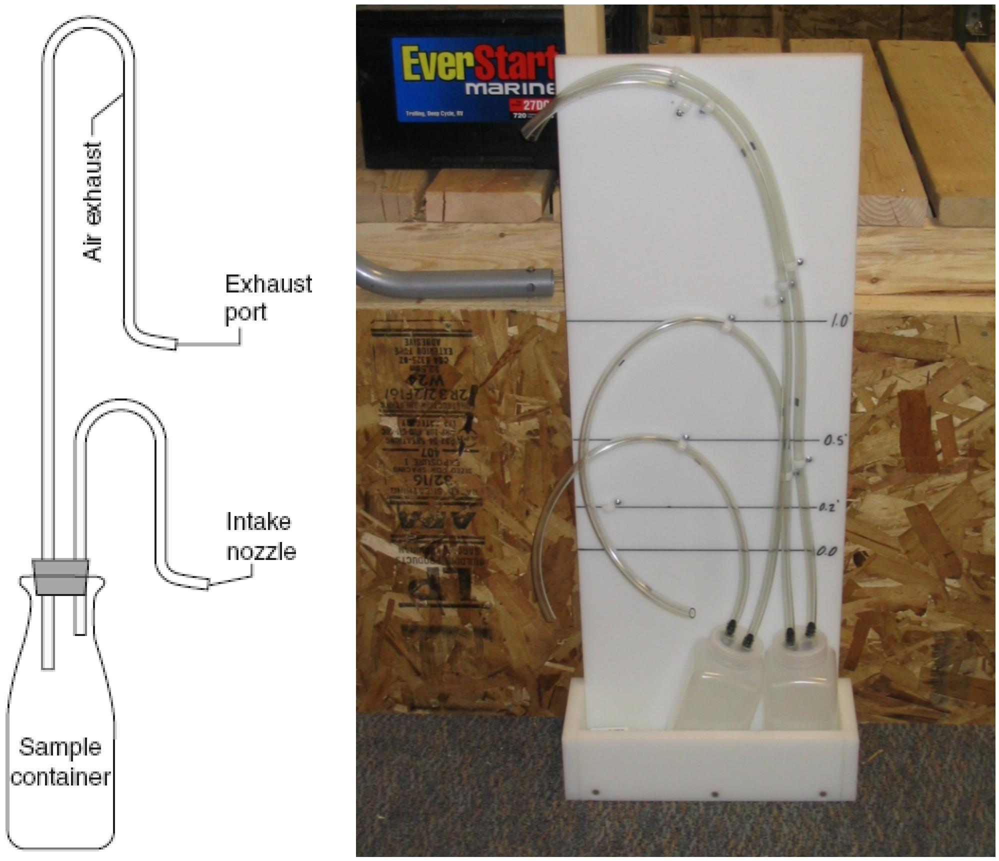


Figure 3-39. Single-stage passive sampler (diagram: Wilde et al. 2014, photo by D.W. Meals)

- **Tipping-bucket samplers** are mechanical devices that capture water flowing from a pipe or other concentrated discharge in one of two pans that tip back and forth on an axis as one pan fills and the other discharges to a large pan. Slots or a funnel can passively convey a sample to a collection bottle; the resulting sample is a flow-proportional composite. A tipping bucket sampler has the additional feature that total discharge can be measured by counting the number of tips with a mechanical counter. An example of design and application of a tipping-bucket system for sampling field runoff and suspended sediment with a pipe collector is given by Kahn and Ong (1997).
- **Coshocton wheel** samplers are rotating wheels driven by the force of water discharging from a pipe or flume (see Figure 3-28). A standing slot collects a sample each time it rotates under the discharge. Coshocton wheels collect a sample volume proportional to the total discharge (usually 1 percent of the discharge) and therefore can provide an estimate of total event discharge.
- **Lysimeters** are devices buried in the ground to sample soil water moving through the vadose zone, the area between the ground surface and water table (Figure 3-40). Lysimeters may be entirely passive (“zero-tension lysimeters”) collecting gravitational water in funnels, pans or troughs. Alternatively, tension lysimeters extract a sample of soil water by applying suction through porous plates or cups.

Figure 3-40. Lysimeters before and after installation (photos by R. Traver, Villanova University)

3.6.2.4 Autosampling

Autosamplers generally consist of an intake line submerged in the waterbody or the flow through a pipe or flume, a peristaltic or submersible pump that pumps water to the sampler, one or more bottles to contain collected samplers, and electronic controls to initiate sample collection and record data. Some autosamplers may be refrigerated to preserve samples for extended periods. Some may be designed specifically to fit into storm drains and catch basins. Most operate with either DC or AC power. Examples of autosamplers are shown in Figure 3-41.

Figure 3-41. Examples of portable and refrigerated autosamplers (Hach® 2013b, Teledyne Isco 2013b)

Autosamplers can be set to take time-based samples either continuously, i.e. collect a sample every eight hours, or as initiated by an external trigger such as detection of rainfall or rising stream stage. Some samplers can be set in variable time programs, e.g., to collect samples every 15 minutes during the early part of a storm event, then take hourly samples as the event subsides. When connected to a flow meter, autosamplers can take flow-proportional samples, collecting a subsample for every m^3 that passes the station during a set time period or during a discrete storm event. Flow-proportional sampling may be the most appropriate way of sampling for many NPS pollutants, where high concentrations are associated with high flows and where events that could be missed by timed sampling carry the bulk of the pollutant load (see section 3.2.2.2).

Most autosamplers can collect discrete samples in individual bottles so that a picture of constituent concentration variation across a time period or storm event (i.e., a chemograph) can be plotted and the relationships among time, flow and concentration evaluated. Autosamplers can also combine individual samples into a single larger container to yield a composite sample that represents an extended time period (see section 3.2.2.2). Collecting composite samples can reduce analytical costs by sending a single sample (representing the time period or the storm event) to the laboratory. A flow-proportional sample provides an event mean concentration (EMC) with a single analysis and facilitates load estimation by providing a single EMC result that can be multiplied by the total period or event flow for a load estimate (see section 3.8 and section 7.9).

The flexibility, capacity for self-contained unattended operation, and potential linkage to flow data are major advantages of autosamplers. There are also a few disadvantages with autosamplers. First, autosampler intakes are generally fixed in one position in a waterbody and may therefore not be fully representative of variability, especially where strong vertical or horizontal gradients exist. Second, the size of the intake line and the velocity achieved by the autosampler pump, as well as the position in the streamflow, may prevent the collection of a representative sample, especially of suspended sediment and particulate-bound pollutants. Third, monitoring for some pollutants like volatile organics or pathogens, may be challenging because of special limitations for materials contacting the sample and requirements for sterilization between sample intake events. Finally, because samples are taken at intervals, regardless of whether an autosampler collects on a time- or flow-based program, the possibility always exists that a transient pulse of a pollutant (e.g., from a spill or first-flush) may pass by unsampled. This of course is also a risk in manual sampling.

Autosamplers must be maintained properly to ensure that sample collection is reliable and performed in accordance with programming instructions. Routine maintenance, sample volume calibration, and probe calibration procedures specified in user manuals should be strictly followed.

3.6.2.5 Benthic Macroinvertebrate Sampling

Sampling of benthic macroinvertebrates from aquatic substrates like stream bottoms and lake beds must consider not only how to physically collect samples but also the diversity of stream habitats that influence the numbers and types of organisms to be sampled. Different types of assemblages of macroinvertebrates inhabit different aquatic habitats (Hawkins et al. 1993). While a monitoring program need not necessarily sample all these habitat types, the habitats sampled should be based on monitoring objectives and on regional stream or lake characteristics. Two distinct types of stream habitats are generally sampled: riffles (shallow areas of fast-moving water, generally with a stony or gravelly bottom) and pools (areas of deeper, slow-flowing water, generally with a softer sediment substrate) (Figure 3-42). In lakes, near-shore areas offer different substrates and habitats from those in deeper lake regions that might lack light, vegetation, and oxygen. Different groups of organisms tend to occupy these habitats, and different approaches for sampling them are required.

The Rapid Bioassessment Protocols (RBPs) recommended by U.S. EPA (Barbour et al. 1999) specify many of the parameters of benthic macroinvertebrate sampling. These issues are discussed in greater detail in chapter 4 of this guidance. In general, benthic macroinvertebrates can be collected actively or passively. In rivers and streams, active collection is often accomplished by disturbing the streambed and capturing the dislodged organisms in a net as the current carries them downstream (Figure 3-43). Kick-seines, D-frame nets, and Surber square-foot samplers are common devices used (Figure 3-44). Regardless of the specific device, it is important to quantify both the area of the streambed disturbed and the time/effort of sampling so that results can be quantified (e.g., organisms/m^2), repeated and compared among different sampling events over time. In lakes, active sampling in shallow areas can be done by similar methods. Grab samplers such as the petite ponar (Figure 3-44) or larger dredges are used for taking sediment samples from hard bottoms such as sand, gravel).

Passive sampling for benthic macroinvertebrates often uses artificial substrates like the Hester-Dendy plate sampler or rock baskets (Figure 3-44) that are anchored in the waterbody. Organisms colonize the devices and then the devices are retrieved to collect and enumerate the organisms.

All of these techniques have advantages and disadvantages that are discussed in chapter 4.

Figure 3-42. Preparing to take samples in a low-gradient stream

Figure 3-43. Using a D-frame net to sample a gravel bottom stream for benthic macroinvertebrates

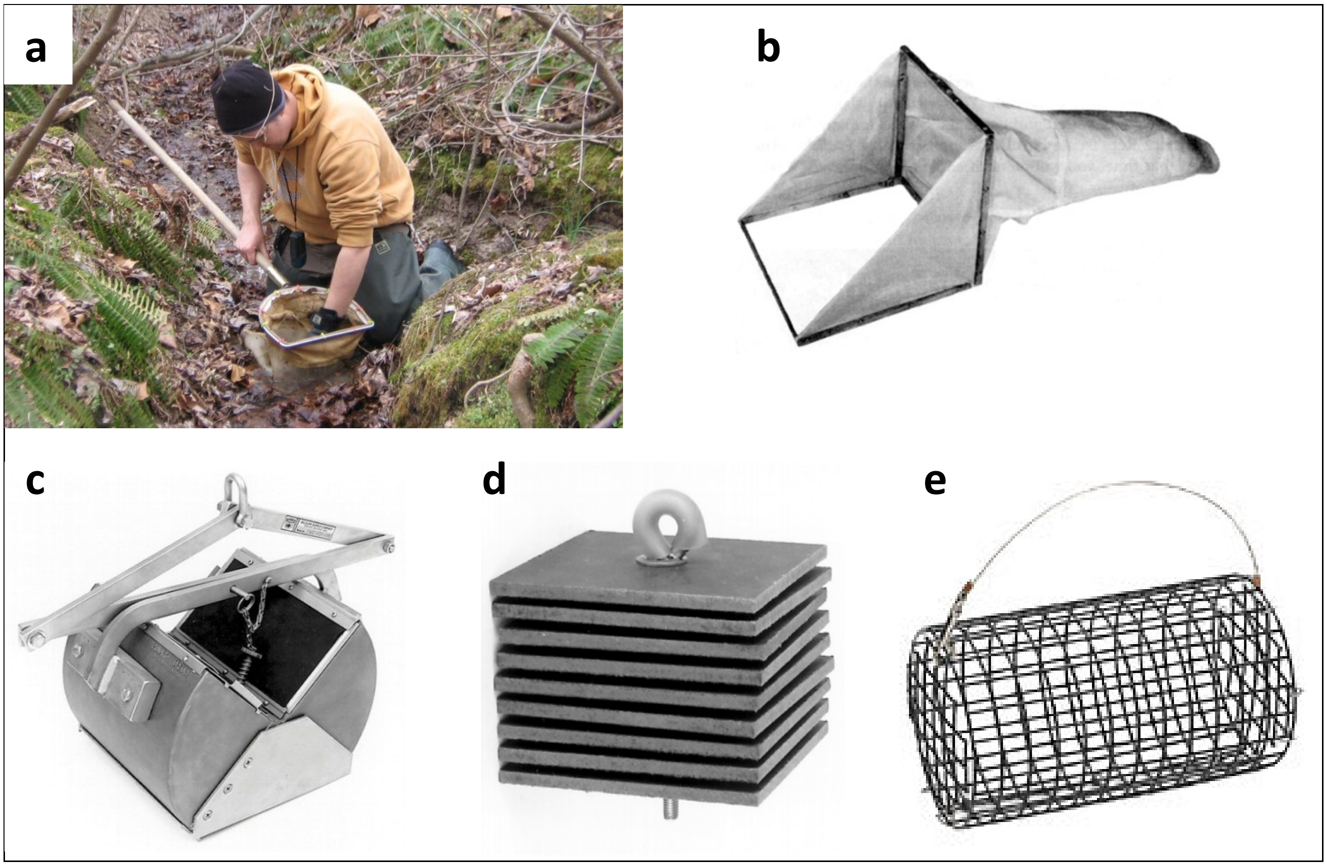


Figure 3-44. Sampling devices for biological and habitat variables. *a*, D-frame net; *b*, Surber sampler (Rickly 2016); *c*, Ponar dredge (Rickly 2016); *d*, Hester-Dendy artificial substrate (Rickly 2016); *e*, Rock basket artificial substrate (Ben Meadows 2016).

3.6.2.6 Fish Sampling

As with benthic macroinvertebrates, distinct fish assemblages are found in different habitat types. For fish, characteristics like water temperature, flow velocity, dissolved oxygen levels, cover and shade, in addition to substrate type, are important habitat characteristics. In general, biomonitoring efforts should sample fish habitats based on project objectives and resource characteristics. Major habitat types like riffles, pools and runs (stream reaches between riffles and pools) should normally be sampled. Habitats and the size of sampling areas should be consistent between sampling events to allow long-term comparisons.

Fish are most commonly sampled by electrofishing, where a portable generator system introduces an electric current into the water, temporarily stunning fish within a certain range (Figure 3-45). In practice, the ends of a sampling reach (approximately 30 m in length) are closed off with nets and a sampling crew walks through the reach. One person runs the shocker, while the others retrieve stunned fish into buckets. When collection is complete, the fish are counted and identified (usually to species), then returned to the stream (Figure 3-46). The process may be repeated at several different sites of similar habitat to ensure a representative sampling has been achieved. Other approaches to fish sampling include use of seines, gill nets, traps, or underwater observation. For a discussion of the advantages and limitations of different fish sampling gear, see Klemm et al. (1992). Ohio EPA (OEPA 1987) discusses electrofishing techniques for bioassessment.

Figure 3-45. Backpack electrofishing (USEPA)

Figure 3-46. Field processing of fish sample: taxonomic identification and data recording

3.6.2.7 Aquatic Plant Sampling

Aquatic plants sampled for water quality monitoring include algae (small free-floating plants), periphyton (the community of algae, microbes, and detritus attached to submerged surfaces), and macrophytes (large, plants rooted in aquatic sediments). Many of these plants are good indicators of nutrient enrichment and ecosystem condition. Algae are usually evaluated in lakes or other bodies of standing water and are sampled using a plankton net towed through the water column (Figure 3-47). Collected organisms are identified and counted under a microscope. As a surrogate for algal biomass, chemical analysis of a water sample for chlorophyll *a* may be performed. Periphyton biomass is usually measured in streams, either by scraping known areas of rock surfaces or by use of artificial substrates (typically glass microscope slides) placed in the stream and retrieved after a specified period. Aquatic macrophytes, often monitored in near-shore areas of lakes or in large rivers, may be surveyed to assess species composition, quantified in small plots by counting individual plants or harvesting vegetation, or mapped by remote sensing to document areal extent of growth.

Figure 3-47. Plankton nets (NOAA 2014)

3.6.2.8 Bacteria/Pathogen Sampling

Collection of water samples for monitoring indicator bacteria, pathogens, or other microorganisms is usually conducted by grab sampling. Samples for fecal coliform and *E. coli* bacteria analysis typically require small volumes (e.g., 100 milliliter [ml]). Samples for detection and enumeration of protozoan pathogens like *Giardia* and *Cryptosporidium* may require up to 20 Liter (L) of sample. Sterile sample containers such as autoclaved polyethylene containers or pre-sterilized single-use bags or bottles are required. Sample collection should be done by clean technique, with samples allowed to contact only sterile surfaces; field personnel should wear gloves when collecting grab samples, both to protect themselves from water-borne pathogens and to prevent sample contamination. Samples for bacteria and/or pathogens might require more rapid delivery to the laboratory than samples from physical and chemical analysis (see section 3.6.3).

3.6.2.9 Habitat Sampling

Assessment of aquatic habitat may be essential to interpretation of data collected from monitoring of benthic invertebrates and fish. Habitat characteristics might also be an important response variable to land treatment or stream restoration efforts. Habitat quality may be measured in three dimensions: habitat structure, flow regime, and energy source. Habitat structure includes physical characteristics of stream environment such as channel morphology, gradient, instream cover (boulders and woody debris), substrate types, riparian condition, and bank stability. Flow regime is defined by velocity and volume of water moving through a stream, both on the average and during extreme events (wet or dry). Energy enters stream systems through nutrients from runoff or ground water, as leaves and other debris falling into streams, or from photosynthesis by aquatic plants and algae.

Some important metrics of habitat sampling were shown in Table 3-7 in section 3.1.4. Many habitat characteristics are quantified by direct measurement in representative stream reaches, e.g., by surveying, substrate sampling, and soil/geophysical measurements. Sets of habitat measurements are often incorporated into indices that facilitate comparison between sites and between sampling times. For example, the Qualitative Habitat Evaluation Index (QHEI) used by Ohio EPA (Rankin 1989) includes measurements of:

- **Substrate:** type and quality
- **Instream cover:** type and amount
- **Channel morphology:** sinuosity, development, channelization, stability
- **Riparian zone:** width, quality, bank erosion
- **Pool quality:** maximum depth, current, morphology
- **Riffle quality:** depth, substrate stability, substrate embeddedness
- **Map gradient**

Habitat assessment is discussed further in chapter 4 of this guidance. The reader is referred to additional resources for more information on habitat sampling:

- Rapid Bioassessment Protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish (Barbour et al. 1999)
- The Qualitative Habitat Evaluation Index (QHEI): rationale, methods, and application (Rankin 1989)
- Methods for assessing habitat in flowing waters: using the Qualitative Habitat Evaluation Index (QHEI) (OEPA 2006.).

3.6.2.10 Specialized Sampling

Specialized sampling techniques may be required for unusual or emerging pollutants. For example, microbial source tracking analyzes DNA to attribute indicator bacteria to specific host sources (USEPA 2011b, Meals et al. 2013). This method requires water sampling and might also involve collection of fecal material from human and animal sources in the watershed.

Urban stormwater monitoring may test for optical brighteners (fluorescent whitening agents added to laundry detergent) in stormwater as indicators of wastewater or septic effluent contamination. Because these chemicals are absorbed by fabric, cotton pads are deployed in streams for several days, then collected and tested for fluorescence with a UV source (Gilpin et al. 2002).

Sentinel chambers, dialysis membrane diffusion samplers, polar organic chemical integrative samplers (POCIS), and other passive sampling devices have been used to passively sample low-concentration pollutants like volatile organic compounds, estrogen analogs, endocrine disruptors, and other emerging pollutants in a variety of settings (Vrana et al. 2005, Liscio et al. 2009, Kuster et al. 2010).

3.6.3 From Field to Laboratory

There are several important steps to consider between sample collection and analysis including sample processing, sample preservation and transport, sample custody tracking, and performance audits. Quality assurance and quality control procedures are described in detail in chapter 8.

3.6.3.1 Sample Processing

Sample processing refers to the measures taken to prepare and preserve a water sample at or after collection, and before it is delivered to the laboratory for analysis. The goals of sample processing are to prepare samples for appropriate analysis (e.g., dissolved vs. TP), prevent contamination and cross-contamination, and preserve sample integrity until analysis. The USGS National Field Manual includes detailed sample processing procedures for many specific analytes, and recommends the following order of sample processing: organic fraction, organic C, inorganic constituents, nutrients, radiochemicals, isotopes, and then microorganisms (Wilde et al. 2009).

Samples requiring filtration (e.g., dissolved P, dissolved organic C) must be filtered during or immediately after collection (Wilde et al. 2009). Surface water samples may be composited or subsampled in the field using an appropriate device, such as a churn or cone splitter (Figure 3-48). Ground water samples are not composited but are pumped either directly through a splitter or through a filtration assembly into sample bottles unless a bailer or other downhole sampler is used to collect the sample.

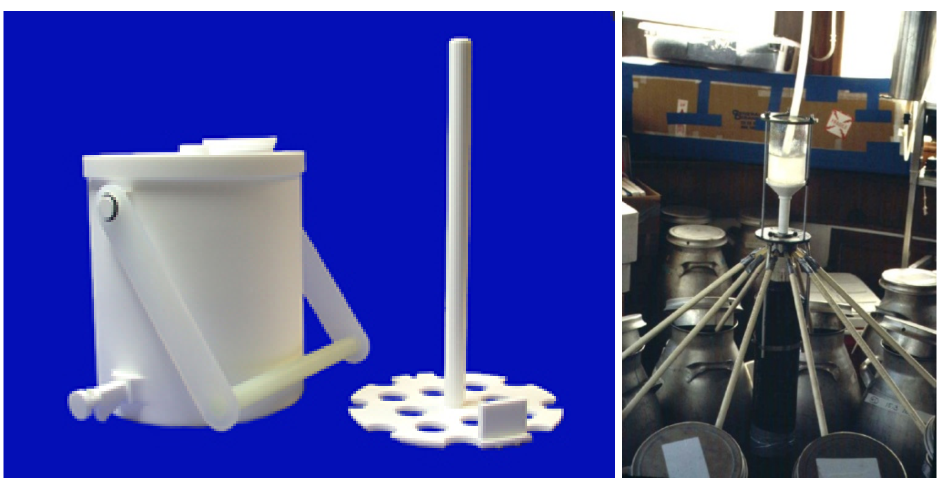


Figure 3-48. Churn and cone splitters (FISP 2014)

3.6.3.2 Sample Preservation and Transport

Water samples to be analyzed for most water quality variables have specified permissible holding time and holding conditions that determine the length of time a sample can be held between collection and analysis without significantly affecting the analytical results. Maximum holding times and storage conditions have been established by the EPA (40 CFR 136.3, USEPA 2008b) and are shown in Table 3-12. Storage and preservation for most analytes involve cooling to below 6 ºC; chemical preservatives such as nitric acid (HNO_3) or sulfuric acid (H_2SO_4) may also be used, depending on the analyte (Wilde et al. 2009).

Samples should be packaged and transported to the laboratory for analysis as soon as possible. The shorter the time between sample collection and analysis, the more reliable the analytical results will be. If samples must be shipped to a laboratory, check to insure that sample containers are sealed, labeled, and packed to prevent breakage. It is necessary to follow receiving laboratory protocols for labeling, documenting, and packaging samples.

Table 3-12. Required containers, preservation techniques, and holding times

Parameter number/name	Container [1]	Preservation [2,3]	Maximum holding time [4]
Table IA—Bacterial Tests:			
1-5. Coliform, total, fecal, and ***E. coli***	PA, G	Cool, <10 °C, 0.0008% $Na_2S_2O_3$ [5]	8 hours.[22,23]
6. Fecal streptococci	PA, G	Cool, <10 °C, 0.0008% $Na_2S_2O_3$ [5]	8 hours.[22]
7. Enterococci	PA, G	Cool, <10 °C, 0.0008% $Na_2S_2O_3$ [5]	8 hours.[22]
8.*Salmonella*	PA, G	Cool, <10 °C, 0.0008% $Na_2S_2O_3$ [5]	8 hours.[22]
Table IA—Aquatic Toxicity Tests:			
9-12. Toxicity, acute and chronic	P, FP, G	Cool, ≤6 °C [16]	36 hours.
Table IB—Inorganic Tests:			
1. Acidity	P, FP, G	Cool, ≤6 °C [18]	14 days.
2. Alkalinity	P, FP, G	Cool, ≤6 °C [18]	14 days.
4. Ammonia	P, FP, G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
9. Biochemical oxygen demand	P, FP, G	Cool, ≤6 °C [18]	48 hours.
10. Boron	P, FP, or Quartz	HNO_3 to pH <2	6 months.
11. Bromide	P, FP, G	None required	28 days.
14. Biochemical oxygen demand, carbonaceous	P, FP G	Cool, ≤6 °C [18]	48 hours.
15. Chemical oxygen demand	P, FP, G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
16. Chloride	P, FP, G	None required	28 days.
17. Chlorine, total residual	P, G	None required	Analyze within 15 minutes.
21. Color	P, FP, G	Cool, ≤6 °C [18]	48 hours.
23-24. Cyanide, total or available (or CATC) and free	P, FP, G	Cool, ≤6 °C [18], NaOH to pH >10 [5 6], reducing agent if oxidizer present	14 days.
25. Fluoride	P	None required	28 days.
27. Hardness	P, FP, G	HNO_3 or H_2SO_4 to pH <2	6 months.
28. Hydrogen ion (pH)	P, FP, G	None required	Analyze within 15 minutes.
31, 43. Kjeldahl and organic N	P, FP, G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
Table IB—Metals: [7]			
18. Chromium VI	P, FP, G	Cool, ≤6 °C [18], pH = 9.3-9.7 [20]	28 days.
35. Mercury (CVAA)	P, FP, G	HNO_3 to pH <2	28 days.
35. Mercury (CVAFS)	FP, G; and FP-lined cap [17]	5 ml/L 12N HCl or 5 ml/L BrCl [17]	90 days.[17]

Parameter number/name	Container [1]	Preservation [2,3]	Maximum holding time [4]
3, 5-8, 12, 13, 19, 20, 22, 26, 29, 30, 32-34, 36, 37, 45, 47, 51, 52, 58-60, 62, 63, 70-72, 74, 75. Metals, except boron, chromium VI, and mercury	P, FP, G	HNO_3 to pH <2, or at least 24 hours prior to analysis [19]	6 months.
38. Nitrate	P, FP, G	Cool, ≤6 °C [18]	48 hours.
39. Nitrate-nitrite	P, FP, G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
40. Nitrite	P, FP, G	Cool, ≤6 °C [18]	48 hours.
41. Oil and grease	G	Cool to ≤6 °C [18], HCl or H_2SO_4 to pH <2	28 days.
42. Organic Carbon	P, FP, G	Cool to ≤6 °C [18], HCl, H_2SO_4, or H_3PO_4 to pH <2	28 days.
44. Orthophosphate	P, FP, G	Cool, to ≤6 °C [18,24]	Filter within 15 minutes; Analyze within 48 hours.
46. Oxygen, Dissolved Probe	G, Bottle and top	None required	Analyze within 15 minutes.
47. Winkler	G, Bottle and top	Fix on site and store in dark	8 hours.
48. Phenols	G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
49. Phosphorous (elemental)	G	Cool, ≤6 °C [18]	48 hours.
50. Phosphorous, total	P, FP, G	Cool, ≤6 °C [18], H_2SO_4 to pH <2	28 days.
53. Residue, total	P, FP, G	Cool, ≤6 °C [18]	7 days.
54. Residue, Filterable	P, FP, G	Cool, ≤6 °C [18]	7 days.
55. Residue, Nonfilterable (TSS)	P, FP, G	Cool, ≤6 °C [18]	7 days.
56. Residue, Settleable	P, FP, G	Cool, ≤6 °C [18]	48 hours.
57. Residue, Volatile	P, FP, G	Cool, ≤6 °C [18]	7 days.
61. Silica	P or Quartz	Cool, ≤6 °C [18]	28 days.
64. Specific conductance	P, FP, G	Cool, ≤6 °C [18]	28 days.
65. Sulfate	P, FP, G	Cool, ≤6 °C [18]	28 days.
66. Sulfide	P, FP, G	Cool, ≤6 °C [18], add zinc acetate plus sodium hydroxide to pH >9	7 days.
67. Sulfite	P, FP, G	None required	Analyze within 15 minutes.
68. Surfactants	P, FP, G	Cool, ≤6 °C [18]	48 hours.
69. Temperature	P, FP, G	None required	Analyze.
73. Turbidity	P, FP, G	Cool, ≤6 °C [18]	48 hours.
Table IC—Organic Tests: [8]			
13, 18-20, 22, 24-28, 34-37, 39-43, 45-47, 56, 76, 104, 105, 108-111, 113. Purgeable Halocarbons	G, FP-lined septum	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$ [5]*	14 days.
6, 57, 106. Purgeable aromatic hydrocarbons	G, FP-lined septum	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$ [5], HCl to pH 2 [9]*	14 days.[9]
3, 4. Acrolein and acrylonitrile	G, FP-lined septum	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$, pH to 4-5[10]*	14 days.[10]
23, 30, 44, 49, 53, 77, 80, 81, 98, 100, 112. Phenols [11]	G, FP-lined cap	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$*	7 days until extraction, 40 days after extraction.
7, 38. Benzidines [11] [12]	G, FP-lined cap	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$ [5]*	7 days until extraction.[13]

Parameter number/name	Container [1]	Preservation [2,3]	Maximum holding time [4]
14, 17, 48, 50-52. Phthalate esters [11]	G, FP-lined cap	*Cool, ≤6 °C [18]*	7 days until extraction, 40 days after extraction.
82-84. Nitrosamines [11 14]	G, FP-lined cap	*Cool, ≤6 °C[18], store in dark, 0.008% $Na_2S_2O_3$[5]*	7 days until extraction, 40 days after extraction.
88-94. PCBs [11]	G, FP-lined cap	*Cool, ≤6 °C [18]*	1 year until extraction, 1 year after extraction.
54, 55, 75, 79. Nitroaromatics and isophorone [11]	G, FP-lined cap	*Cool, ≤6 °C [18], store in dark, 0.008% $Na_2S_2O_3$[5]*	7 days until extraction, 40 days after extraction.
1, 2, 5, 8-12, 32, 33, 58, 59, 74, 78, 99, 101. Polynuclear aromatic hydrocarbons [11]	G, FP-lined cap	*Cool, ≤6 °C [18], store in dark, 0.008% $Na_2S_2O_3$[5]*	7 days until extraction, 40 days after extraction.
15, 16, 21, 31, 87. Haloethers [11]	G, FP-lined cap	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$[5]*	7 days until extraction, 40 days after extraction.
29, 35-37, 63-65, 107. Chlorinated hydrocarbons [11]	G, FP-lined cap	*Cool, ≤6 °C [18]*	7 days until extraction, 40 days after extraction.
60-62, 66-72, 85, 86, 95-97, 102, 103. CDDs/CDFs [11]			
Aqueous Samples: Field and Lab Preservation	G	*Cool, ≤6 °C [18], 0.008% $Na_2S_2O_3$[5], pH <9*	1 year.
Solids and Mixed-Phase Samples: Field Preservation	G	*Cool, ≤6 °C [18]*	7 days.
Tissue Samples: Field Preservation	G	*Cool, ≤6 °C [18]*	24 hours.
Solids, Mixed-Phase, and Tissue Samples: Lab Preservation	G	*Freeze, ≤ −10 °C*	1 year.
114-118. Alkylated phenols	G	*Cool, <6 °C, H_2SO_4 to pH <2*	28 days until extraction, 40 days after extraction.
119. Adsorbable Organic Halides (AOX)	G	*Cool, <6 °C, 0.008% $Na_2S_2O_3$ HNO_3 to pH <2*	Hold ***at least*** 3 days, but not more than 6 months.
120. Chlorinated Phenolics		*Cool, <6 °C, 0.008% $Na_2S_2O_3$ H_2SO_4 to pH <2*	30 days until acetylation, 30 days after acetylation.
Table ID—Pesticides Tests:			
1-70. Pesticides [11]	G, FP-lined cap	*Cool, ≤6 °C [18], pH 5-9 [15]*	7 days until extraction, 40 days after extraction.
Table IE—Radiological Tests:			
1-5. Alpha, beta, and radium	P, FP, G	*HNO_3 to pH <2*	6 months.
Table IH—Bacterial Tests:			
1. ***E. coli***	PA, G	*Cool, <10 °C, 0.0008% $Na_2S_2O_3$[5]*	8 hours.[22]
2. Enterococci	PA, G	*Cool, <10 °C, 0.0008% $Na_2S_2O_3$[5]*	8 hours.[22]
Table IH—Protozoan Tests:			

Parameter number/name		Container [1]	Preservation [2,3]	Maximum holding time [4]
	8.***Cryptosporidium***	LDPE; field filtration	*1-10 °C*	96 hours.[21]
	9.***Giardia***	LDPE; field filtration	*1-10 °C*	96 hours.[21]

[1] "P" is for polyethylene; "FP" is fluoropolymer (polytetrafluoroethylene (PTFE); Teflon®), or other fluoropolymer, unless stated otherwise in this Table II; "G" is glass; "PA" is any plastic that is made of a sterilizable material (polypropylene or other autoclavable plastic); "LDPE" is low density polyethylene.

[2] Except where noted in this Table II and the method for the parameter, preserve each grab sample within 15 minutes of collection. For a composite sample collected with an automated sample (e.g., using a 24-hour composite sample; see 40 CFR 122.21(g)(7)(i) or 40 CFR Part 403, Appendix E), refrigerate the sample at ≤ 6 °C during collection unless specified otherwise in this Table II or in the method(s). For a composite sample to be split into separate aliquots for preservation and/or analysis, maintain the sample at ≤ 6 °C, unless specified otherwise in this Table II or in the method(s), until collection, splitting, and preservation is completed. Add the preservative to the sample container prior to sample collection when the preservative will not compromise the integrity of a grab sample, a composite sample, or aliquot split from a composite sample within 15 minutes of collection. If a composite measurement is required but a composite sample would compromise sample integrity, individual grab samples must be collected at prescribed time intervals (e.g., 4 samples over the course of a day, at 6-hour intervals). Grab samples must be analyzed separately and the concentrations averaged. Alternatively, grab samples may be collected in the field and composited in the laboratory if the compositing procedure produces results equivalent to results produced by arithmetic averaging of results of analysis of individual grab samples. For examples of laboratory compositing procedures, see EPA Method 1664 Rev. A (oil and grease) and the procedures at 40 CFR 141.34(f)(14)(iv) and (v) (volatile organics).

[3] When any sample is to be shipped by common carrier or sent via the U.S. Postal Service, it must comply with the Department of Transportation Hazardous Materials Regulations (49 CFR part 172). The person offering such material for transportation is responsible for ensuring such compliance. For the preservation requirement of Table II, the Office of Hazardous Materials, Materials Transportation Bureau, Department of Transportation has determined that the Hazardous Materials Regulations do not apply to the following materials: Hydrochloric acid (HCl) in water solutions at concentrations of 0.04% by weight or less (pH about 1.96 or greater; Nitric acid (HNO3) in water solutions at concentrations of 0.15% by weight or less (pH about 1.62 or greater); Sulfuric acid (H2SO4) in water solutions at concentrations of 0.35% by weight or less (pH about 1.15 or greater); and Sodium hydroxide (NaOH) in water solutions at concentrations of 0.080% by weight or less (pH about 12.30 or less).

[4] Samples should be analyzed as soon as possible after collection. The times listed are the maximum times that samples may be held before the start of analysis and still be considered valid. Samples may be held for longer periods only if the permittee or monitoring laboratory has data on file to show that, for the specific types of samples under study, the analytes are stable for the longer time, and has received a variance from the Regional Administrator under Sec. 136.3(e). For a grab sample, the holding time begins at the time of collection. For a composite sample collected with an automated sampler (e.g., using a 24-hour composite sampler; see 40 CFR 122.21(g)(7)(i) or 40 CFR part 403, Appendix E), the holding time begins at the time of the end of collection of the composite sample. For a set of grab samples composited in the field or laboratory, the holding time begins at the time of collection of the last grab sample in the set. Some samples may not be stable for the maximum time period given in the table. A permittee or monitoring laboratory is obligated to hold the sample for a shorter time if it knows that a shorter time is necessary to maintain sample stability. See 136.3(e) for details. The date and time of collection of an individual grab sample is the date and time at which the sample is collected. For a set of grab samples to be composited, and that are all collected on the same calendar date, the date of collection is the date on which the samples are collected. For a set of grab samples to be composited, and that are collected across two calendar dates, the date of collection is the dates of the two days; e.g., November 14-15. For a composite sample collected automatically on a given date, the date of collection is the date on which the sample is collected. For a composite sample collected automatically, and that is collected across two calendar dates, the date of collection is the dates of the two days; e.g., November 14-15. For static-renewal toxicity tests, each grab or composite sample may also be used to prepare test solutions for renewal at 24 h, 48 h, and/or 72 h after first use, if stored at 0-6 °C, with minimum head space.

[5] ASTM D7365-09a specifies treatment options for samples containing oxidants (e.g.,chlorine). Also, Section 9060A of Standard Methods for the Examination of Water and Wastewater (20th and 21st editions) addresses dechlorination procedures.

[6] Sampling, preservation and mitigating interferences in water samples for analysis of cyanide are described in ASTM D7365-09a. There may be interferences that are not mitigated by the analytical test methods or D7365-09a. Any technique for removal or suppression of interference may be employed, provided the laboratory demonstrates that it more accurately measures cyanide through quality control measures described in the analytical test method. Any removal or suppression technique not described in D7365-09a or the analytical test method must be documented along with supporting data.

[7] For dissolved metals, filter grab samples within 15 minutes of collection and before adding preservatives. For a composite sample collected with an automated sampler (e.g., using a 24-hour composite sampler; see 40 CFR 122.21(g)(7)(i) or 40 CFR Part 403, Appendix E), filter the sample within 15 minutes after completion of collection and before adding preservatives. If it is known or suspected that dissolved sample integrity will be compromised during collection of a composite sample collected automatically over time (e.g., by interchange of a metal between dissolved and suspended forms), collect and filter grab samples to be composited (footnote 2) in place of a composite sample collected automatically.

[8] Guidance applies to samples to be analyzed by GC, LC, or GC/MS for specific compounds.

[9] If the sample is not adjusted to pH 2, then the sample must be analyzed within seven days of sampling.

[10] The pH adjustment is not required if acrolein will not be measured. Samples for acrolein receiving no pH adjustment must be analyzed within 3 days of sampling.

[11] When the extractable analytes of concern fall within a single chemical category, the specified preservative and maximum holding times should be observed for optimum safeguard of sample integrity (i.e.,use all necessary preservatives and hold for the shortest time listed). When the analytes of concern fall within two or more chemical categories, the sample may be preserved by cooling to ≤ 6 °C, reducing residual chlorine with 0.008% sodium

thiosulfate, storing in the dark, and adjusting the pH to 6-9; samples preserved in this manner may be held for seven days before extraction and for forty days after extraction. Exceptions to this optional preservation and holding time procedure are noted in footnote 5 (regarding the requirement for thiosulfate reduction), and footnotes 12, 13 (regarding the analysis of benzidine).

[12] If 1,2-diphenylhydrazine is likely to be present, adjust the pH of the sample to 4.0 ± 0.2 to prevent rearrangement to benzidine.

[13] Extracts may be stored up to 30 days at < 0 °C.

[14] For the analysis of diphenylnitrosamine, add 0.008% Na2S2O3and adjust pH to 7-10 with NaOH within 24 hours of sampling.

[15] The pH adjustment may be performed upon receipt at the laboratory and may be omitted if the samples are extracted within 72 hours of collection. For the analysis of aldrin, add 0.008% Na2S2O3.

[16] Place sufficient ice with the samples in the shipping container to ensure that ice is still present when the samples arrive at the laboratory. However, even if ice is present when the samples arrive, immediately measure the temperature of the samples and confirm that the preservation temperature maximum has not been exceeded. In the isolated cases where it can be documented that this holding temperature cannot be met, the permittee can be given the option of on-site testing or can request a variance. The request for a variance should include supportive data which show that the toxicity of the effluent samples is not reduced because of the increased holding temperature. Aqueous samples must not be frozen. Hand-delivered samples used on the day of collection do not need to be cooled to 0 to 6 °C prior to test initiation.

[17] Samples collected for the determination of trace level mercury (<100 ng/L) using EPA Method 1631 must be collected in tightly-capped fluoropolymer or glass bottles and preserved with BrCl or HCl solution within 48 hours of sample collection. The time to preservation may be extended to 28 days if a sample is oxidized in the sample bottle. A sample collected for dissolved trace level mercury should be filtered in the laboratory within 24 hours of the time of collection. However, if circumstances preclude overnight shipment, the sample should be filtered in a designated clean area in the field in accordance with procedures given in Method 1669. If sample integrity will not be maintained by shipment to and filtration in the laboratory, the sample must be filtered in a designated clean area in the field within the time period necessary to maintain sample integrity. A sample that has been collected for determination of total or dissolved trace level mercury must be analyzed within 90 days of sample collection.

[18] Aqueous samples must be preserved at ≤ 6 °C, and should not be frozen unless data demonstrating that sample freezing does not adversely impact sample integrity is maintained on file and accepted as valid by the regulatory authority. Also, for purposes of NPDES monitoring, the specification of "≤ °C" is used in place of the "4 °C" and "< 4 °C" sample temperature requirements listed in some methods. It is not necessary to measure the sample temperature to three significant figures (1/100th of 1 degree); rather, three significant figures are specified so that rounding down to 6 °C may not be used to meet the ≤6 °C requirement. The preservation temperature does not apply to samples that are analyzed immediately (less than 15 minutes).

[19] An aqueous sample may be collected and shipped without acid preservation. However, acid must be added at least 24 hours before analysis to dissolve any metals that adsorb to the container walls. If the sample must be analyzed within 24 hours of collection, add the acid immediately (see footnote 2). Soil and sediment samples do not need to be preserved with acid. The allowances in this footnote supersede the preservation and holding time requirements in the approved metals methods.

[20] To achieve the 28-day holding time, use the ammonium sulfate buffer solution specified in EPA Method 218.6. The allowance in this footnote supersedes preservation and holding time requirements in the approved hexavalent chromium methods, unless this supersession would compromise the measurement, in which case requirements in the method must be followed.

[21] Holding time is calculated from time of sample collection to elution for samples shipped to the laboratory in bulk and calculated from the time of sample filtration to elution for samples filtered in the field.

[22] Sample analysis should begin as soon as possible after receipt; sample incubation must be started no later than 8 hours from time of collection.

[23] For fecal coliform samples for sewage sludge (biosolids) only, the holding time is extended to 24 hours for the following sample types using either EPA Method 1680 (LTB-EC) or 1681 (A-1): Class A composted, Class B aerobically digested, and Class B anaerobically digested.

[24] The immediate filtration requirement in orthophosphate measurement is to assess the dissolved or bio-available form of orthophosphorus (i.e.,that which passes through a 0.45-micron filter), hence the requirement to filter the sample immediately upon collection (i.e.,within 15 minutes of collection).

[38 FR 28758, Oct. 16, 1973]

Source: Electronic Code of Federal Regulations, U.S. Government Printing Office (http://www.ecfr.gov)

Title 40: Protection of Environment

PART 136—GUIDELINES ESTABLISHING TEST PROCEDURES FOR THE ANALYSIS OF POLLUTANTS , § 136.3 Identification of test procedures. (Accessed January 29, 2016).

3.6.3.3 Sample Custody

The location and status of collected samples must be tracked at all points between the source waterbody and the final data report (see chapter 8). The purposes of tracking sample custody are to prevent loss of samples and/or data, document the conditions under which the samples were held between collection and analysis, and preserve sample and data security and integrity. The principal goal is to be able to track each individual analytical result back through all the steps between collection and analysis should any questions arise concerning analytical results. Records of sample custody are important in all monitoring programs, but are especially critical where data may be used for regulatory or litigation purposes.

Sample custody starts with a consistent numbering and labeling system that uniquely identifies each sample with respect to source, monitoring program, date and time of collection, responsible person(s), and desired analysis. Custody is usually tracked through forms and other records that are signed and dated by each individual in the chain. For example, in addition to field logs and notes, field personnel will generally fill out a form upon delivery of samples to the laboratory documenting sample identification numbers, program name, date and time of collection, date and time of delivery, and name of delivery person. Laboratory staff will incorporate sample identification numbers into their own custody and data tracking system.

3.6.3.4 Performance Audits

Regular field operations performance audits should be part of the overall quality assurance/quality control process embodied in the QAPP (see chapter 8). These performance audits might include actions such as:

- **Sample container and equipment blanks:** distilled/deionized water is processed through sampling equipment and sample containers to rule out contamination.
- **Trip blanks:** distilled/deionized water is transported from the laboratory through the field sampling process to document any potential contamination during travel and transport.
- **Field duplicates:** two grab samples are collected in quick succession to assess repeatability of sampling.
- **Field splits:** a collected sample is split into two subsamples to assess analytical performance by the laboratory or to make comparisons between labs.

3.6.4 Laboratory Considerations

Water quality samples collected from field sites are generally analyzed in a laboratory. While field test kits are widely available and commonly used in volunteer/citizen monitoring, the accuracy and precision generally required in NPS monitoring programs, especially those evaluating the effects of treatment or the achievement of TMDL objectives, demand formal laboratory analysis. Laboratories used for NPS monitoring projects may include those operated by state agencies, universities, and private companies.

Specific analytical methods exist for all the water quality variables discussed in this guidance. For all monitoring efforts, analyses should be conducted by accepted laboratory methods. These methods are too numerous to explore in this guidance. There are several resources available to learn about and select appropriate analytical methods, including:

- U.S. EPA Approved Clean Water Act Methods http://www.epa.gov/cwa-methods
- Standard Methods for the Examination of Water and Wastewater, 22nd Edition, American Public Health Association, American Water Works Association, and Water Environment Foundation, http://www.standardmethods.org/ (Rice et al. 2012)
- National Environmental Methods Index (NEMI) https://www.nemi.gov/home/

Select a laboratory to analyze monitoring samples with care. While there is no national certification program for water quality laboratories, most states operate their own certification or registration programs. U.S. EPA operates a Drinking Water Laboratory Certification Program in partnership with EPA regions and states in which laboratories must be certified to analyze drinking water samples for compliance monitoring. Certified laboratories must successfully analyze proficiency testing samples annually, use approved methods, and successfully pass periodic on-site audits. Such certified laboratories may also perform analyses on non-drinking water samples.

When selecting a laboratory, look for one that is certified either by a state program or under the EPA Drinking Water program, one that uses approved methods for analysis, and one that participates in regional comparative proficiency testing programs, if available. In general, it is easier to locate a laboratory to conduct physical and chemical analyses than one to perform analysis of benthic macroinvertebrates, fish, and other aquatic biota. State environmental or natural resource agency biomonitoring programs or university laboratories may be the best bet for bioassessment sample processing. Any laboratory selected, however, should be able to provide documentation of methods and QA/QC protocols used, as well as provide assurance that samples will be handled and processed expeditiously. In making arrangements with the selected laboratory, consider the lab's data approval and reporting system, particularly the likely delays between sample delivery and final data reporting. Long delays in data reporting will inhibit the feedback between land treatment and water quality monitoring that is critically important in watershed project management. Finally, while most water quality laboratories are equipped to analyze water samples for common indicator bacteria like *E. coli*, analysis for pathogens like *E. coli* O157:H7 or *Cryptosporidium* requires considerable expertise generally found only in state health department or private consultant laboratories.

3.7 Land Use and Land Treatment Monitoring

3.7.1 General Considerations

As discussed in section 2.2.1, NPS pollution is generated by activities on the land that vary in location, intensity, and duration. For all monitoring objectives addressed in this guidance (see section 2.1), it will be important to track both land use and land treatment. Note that for the purposes of this guidance, the term "land use" refers not only to the general category of land use or cover (e.g., residential, row crop) but also to land management or source activities (e.g., street sweeping, agrichemical applications, tillage). Similarly, in many cases, the term "land treatment" refers not just to the existence of a specific treatment or BMP (e.g., sediment basin, reduced tillage) but also to the management of the BMP (e.g., sediment basin clean-out, tillage dates, or nutrient application rate, timing, and method). Land use/treatment monitoring encompasses both land use and land treatment.

In general, linking land treatment to water quality response requires both land use/treatment and water quality monitoring. Specific needs for land use/treatment monitoring may differ by monitoring type. For example, assessment monitoring often includes complete spatial coverage of source activities, but temporal variability is not generally addressed because of the short timeframe for problem assessment. Modeling is often used to address the long-term temporal aspects of source activities, including land use changes like conversion of agricultural land to residential use. Evaluating the land uses of a watershed is an important step in understanding watershed condition and source dynamics. Additional details regarding the role of land use in watershed assessment can be found in U.S. EPA's Watershed Planning Handbook (USEPA 2008a).

Understanding of pollutant loading patterns requires information on both the spatial and temporal variability of source activities, particularly when load and wasteload allocations are developed as part of a TMDL. The size of the margin of safety in a TMDL is often directly related to the level of uncertainty associated with the variability of nonpoint source loads (see USEPA 2008a for a discussion of margin of safety [MOS]).

It is necessary to track land use/treatment when planning to attribute water quality trends to activities on the land (see section 2.4.2.4). Because monitoring for trend analysis can continue for decades, costs need to be factored carefully into decisions about the scope, level of detail, and frequency of land use/treatment monitoring that will be done.

For individual BMP effectiveness monitoring, it is important to document:

- Design specifications of the practice evaluated;
- Degree to which the practice was implemented, maintained, and operated according to specifications;
- Management activities conducted under the scope of the practice; and
- Any situations where the BMP operated under conditions outside of the design range.

For example, it is important to flag any monitoring data collected when the design capacity of a stormwater runoff device is exceeded because performance will often suffer. These same considerations apply to all BMPs to be evaluated at the watershed scale, with the additional proviso that both the spatial distribution and interrelationships of BMPs should be addressed.

Existing guidance provides recommendations for tracking the implementation of agricultural, silvicultural, and urban BMPs (USEPA 1997b, USEPA 1997c, USEPA 2001b). This guidance addresses data sources, methods of data collection, temporal and spatial scales of land use/treatment monitoring, monitoring variables, and sampling frequency.

3.7.2 Basic Methods

3.7.2.1 Direct Observation

Personal observations may be the best way to track land use/treatment for plot and field studies. Studies at this scale are frequently visited for equipment monitoring and sample collection, so a good record of source activities can be obtained. It is recommended that a form be developed and used to ensure that tracking is complete and consistent over time (USDA-NRCS 2003). Examples of such forms are shown in Figure 3-49. Advantages of this method include the ability to schedule visits and the fact that the observer controls the quality of data collected.

Agronomic Data Form

Site name				

MANURE APPLICATION				
Date	Field # (map)	Amt applied (spreader #, loads)	Date incorporated	Comments

Date	Field # (map)	Crop or stocking rate	Activity (till, plant, harvest, etc.)	Comments

Figure 3-49. Examples of agricultural activity data recording forms

Other forms of direct observation include windshield surveys such as those performed by the Conservation Technology Information Center (www.ctic.purdue.edu/CRM/) (CTIC 2016). For some applications, photography can be an important tool. At an edge-of-field monitoring station, an automated digital camera can be installed to take periodic photographs looking up into the drainage area to record crop growth or other visible information. A detailed discussion of the use of photo points for monitoring is presented in chapter 5.

Disadvantages of direct observation methods include the potential for bias due to the observer's lack of understanding of management activities, scheduling that misses important events, and the inability to assess rate or quantity information based only on observation (USDA-NRCS 2003).

3.7.2.2 Log Books

Log books can be given to land owners and managers to record activities relevant to the monitoring study (USDA-NRCS 2003). An advantage of this method is that the same individual who is responsible for the activity does the reporting. However, it is difficult to guarantee compliance or consistent reporting across individuals.

3.7.2.3 Interviews

For interviews, as for log books, reporting is performed by the individual responsible for the activity. When conducted in person, interviews also offer the opportunity to gather additional information of importance to the study. Disadvantages include the potential for less than complete reporting of information by the person interviewed, as well as potentially inadequate or uneven interview skills by those conducting the interviews (USDA-NRCS 2003).

A combination of the log book and interview approach may work well in small watersheds with a relatively small number of participants. A Vermont project (Meals 2001) successfully used a combination of log books distributed to watershed farmers with an annual interview to collect the logbook and record other information. Interviews were conducted by a local crop consultant who was known and trusted in the region.

3.7.2.4 Agency reporting.

USDA maintains data on conservation practices implemented with USDA cost-share funds or technical assistance. The utility of this information is limited for watershed projects, however, because Section 1619 of the Food, Conservation, and Energy Act of 2008 (section 1619) provides that USDA, or any contractor or cooperator of USDA, may not generally disclose farm-specific information. Exceptions to this prohibition include the disclosure of such information with consent of the producer or owner of the land and statistical or aggregate summaries of the data by which specific farms are not identifiable. Publicly-available data are typically aggregated at the county level and some implementation is not reported due to confidentiality restrictions. In addition, cumulative implementation is difficult to ascertain because maintenance and operation of practices is not tracked. Note also that the information in the system is verified and finalized annually, so data within a current year may be incomplete or inaccurate.

State-level information on USDA conservation programs can be obtained through the RCA Report – Interactive Data Viewer. This information may be useful during the project planning phase to determine the level of program activity and degree to which specific practices are implemented in the state. Farm-specific data, however, would need to be obtained directly from the producer or owner of the land or through a section 1619 agreement with USDA. Hively et al. (2013) describe in detail several section 1619 agreements established within the Chesapeake Bay watershed.

There are also several survey-based inventories of land use information, including USDA's National Resources Inventory (NRI) and the Census of Agriculture (USDA-NASS 2012). Because of confidentiality requirements, the Census of Agriculture does not disclose information on animal populations, crop acreage, or the like for counties with fewer than four individual producers. Data for such non-disclosed counties may need to be estimated, using a variety of approaches (see section 3.7.6).

Other specialized land use datasets include NOAA's Coastal Change Analysis Program's (C-CAP) nationally standardized database of land cover and land change information for the coastal regions of the U.S. Various historical GIS datasets are also available, including the National Land Cover Data and USGS's Land Use and Land Cover data (USEPA 2008a). GIS data for mapping human population are provided by the U.S. Census Bureau through the TIGER (Topologically Integrated Geographic Encoding and Referencing) program. TIGER data consist of man-made features (such as roads and railroads) and political boundaries. Population data from the 2010 Census can be linked to the TIGER data to map population numbers and density for small (census blocks) and large areas (counties and states). In addition, a number of states and counties also have statewide or local land use and land cover information available.

3.7.2.5 Remote Sensing

The basic categories of remote sensing are described in existing guidance (USEPA 2008a). Aerial imagery includes images and data collected from an aircraft and involves placing a sensor or camera on a fixed-wing or rotary aircraft. Space-based imagery includes images and data collected from space-borne satellites that orbit the earth. A wide range of remote sensing datasets are available for free or at low cost, including data products at the USGS's National Map Viewer and Download Platform or Earth Resources Observation and Science (EROS) data center. Other datasets include Landsat data, elevation, greenness, "Nighttime Lights," and coastal and Great Lakes Shorelines (USEPA 2008a). In some regions, FSA conducts annual low-altitude aerial photography to assess compliance with crop insurance programs. If this photography can be accessed with appropriate permissions, it can provide an annual record of crops grown, changes in field boundaries, land development, and other features.

Commercial web-based resources such as Bing Maps and Google Earth can be useful tools for land use monitoring. Although the date of the imagery in these or other resources may not exactly match what is required for a specific project, features such as roads, farmsteads, rivers, and lakes are readily apparent and general land use types (e.g., urban, agriculture, or forest) can be identified and mapped in preparation for acquisition of more current detailed data.

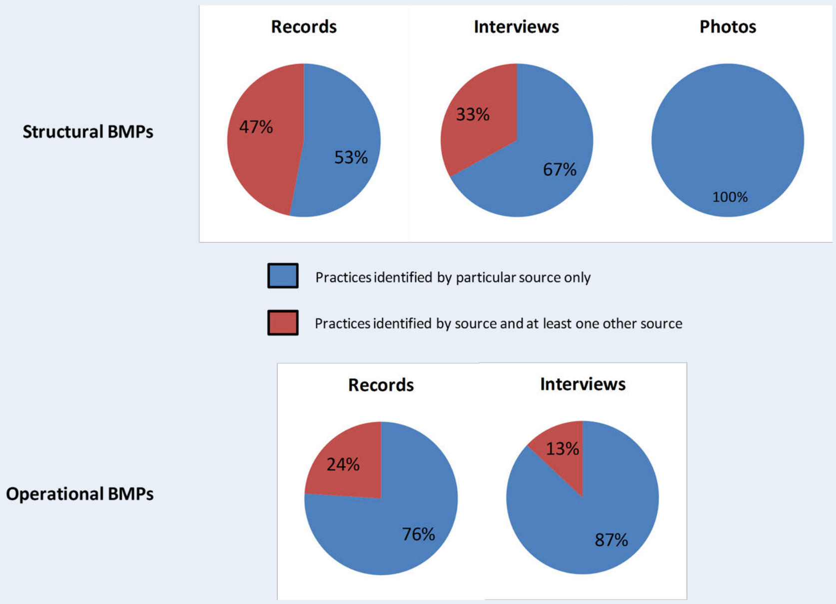
Remote sensing can be useful for tracking practices and land management that are monitored visually. For example, cover crops are easily identified with remote sensing, but whether the cover crops have been fertilized is not easily identifiable. McCarty et al. (2008) used remote sensing technologies to scale point measurements of BMP effectiveness from field to subwatershed and watershed scales, demonstrating that optical satellite (SPOT-5) data and ground-level measurements can be effective for monitoring nutrient uptake by winter cover crops in fields with a wide range of management practices. Hively et al. (2009a and 2009b) combined cost-share program enrollment data with satellite imagery and on-farm sampling to evaluate cover crop N uptake on 136 fields within the Choptank River watershed in Maryland. Annual cost-share program enrollment records were used to locate cover crop fields and provide agronomic management information for each field. Satellite imagery from December and March was used to measure pre-winter and spring cover crop biomass, respectively. Data collected simultaneously from fields were used to convert satellite reflectance measurements to estimates of biomass and nutrient uptake, thus providing a means to estimate aboveground biomass and N uptake estimates for all fields enrolled in the cover crop program.

Locating Best Management Practices by Three Methods
Eagle Creek Watershed, IN — NIFA-CEAP Watershed Project

Objective: **To assess the effects of BMPs on water quality, researchers needed to identify all BMPs implemented in an agricultural watershed since 1995 under a variety of state and federal programs**

Three approaches based on different data sources were used:

- Examination of government records (NRCS, FSA, Indiana Dept. of Env. Mgt.)
- Interviews with producers (structural and operational, polygon and line format)
- Analysis of aerial photography (structural only)



Observations:

1. NRCS data required processing to eliminate double-counting because each point potentially represented multiple practices. After eliminating all the double-counting, 107 structural practices were reduced to 48 standard practices and 299 operational BMPs to 84 distinct practices.
2. Remote sensing picked up only 27 structural practices and no operational practices
3. Producer interviews detected 47 structural practices and 185 operational practices
4. Using all three sources of information, 94 structural practices and 215 operational practices were identified.
5. 53% of the structural practices were identified by government records, while 67% were identified through producer interviews.
6. Operational practices were identified in government records 76% of the time relative to 87% from producer surveys.
7. Researchers found that:
   - Government records identified the majority of BMPs, but were incomplete and difficult to obtain
   - Interviews were information-rich but time-consuming to conduct
   - Photos were effective to confirm and supplement records and interviews
   - Combined data collection techniques provided a clearer picture of conservation practices in the watershed compared to any single approach.

(Grady et al. 2013)

3.7.3 Temporal and Spatial Scale

Land use/treatment monitoring should address the entire area contributing to flow at the water quality sampling point. Depending on the specific study area and monitoring design, some parts of a larger area may be emphasized more than other parts. For example, land nearest to the sampling point can sometimes have a major effect on the measured water quality, so these areas must be monitored carefully. Thus, the spatial coverage of land use monitoring may range from a single field (or portion of a field) up to an entire river basin.

In designing a land use/treatment monitoring system, it is logical to begin with the assumption that the temporal scale of land use/treatment monitoring should match that of the water quality monitoring when the data are to be combined for analyses. Data from weekly composite water quality samples, for example, would be associated with weekly measures of source activity. However, this design should be tempered by understanding the inherent variability of what is being measured (see section 3.7.5). Some metrics of land use and land treatment do not in fact vary on a weekly time scale. It would be wasted effort, for example, to determine and record the crop present in an agricultural field each week during a single growing season or note that a residential subdivision is composed of moderate density detached homes. On the other hand, some highly transient land management activities are very critical to water quality. Manure application on cropland, tillage operations, and street sweeping are examples, and weekly records of such phenomena would be important. Still other land management activities may be important to identify exactly in time and magnitude, for example in relation to a storm event. Herbicide losses from cropland, for example, are strongly influenced by proximity of application to the first few runoff events; pollution potential of pasture runoff may be influenced by the number of grazing animals around the time of major runoff events.

A multi-level land use/treatment monitoring approach can address these multiple temporal concerns:

- **Characterization:** an initial snapshot of land use/land cover, focusing on relatively static parameters (at least relative to the project period) such as water bodies, highways, impervious cover, and broad patterns of urban, agricultural, and forest land uses;
- **Annual:** an annual survey for annually-varying features such as crop type;
- **Weekly:** weekly observations or log entries to identify specific dates/times of critical activities like manure or herbicide applications, tillage, construction, and street sweeping; and
- **Quantitative:** data collection on rates and quantities (e.g., nutrient or herbicide application rates, number of animals on pasture, logging truck traffic).

The guiding principle of timing is to collect land use/treatment data at a fine enough time resolution to be able to (at least potentially) explain water quality observations (e.g., a spike in P concentration) as they occur.

It is important to note that associations between land use/treatment observations and water quality patterns can be confounded by the timing of the source activities (USDA-NRCS 2003). For example, road salt is applied under icing conditions, while wash off tends to occur during periods of thawing or rainfall. Matching weekly water quality and land use/treatment in this case could result in associating high salinity levels with periods of no road salt application. As another example, nutrient concentrations peak during wet periods, but manure is not usually applied when fields are muddy. Using weekly data, high nutrient concentrations would be associated with periods of no manure application. An understanding of pollutant pathways and lag time (section 6.2) and some creative data exploration are often needed to effectively pair land use/treatment observations with water quality data, but this becomes more difficult moving from the BMP level to the watershed scale. Such issues may be addressed by pairing annual water quality data with annual land use/treatment data (Meals 1992); although fine-scale relationships may be lost by this data aggregation, broad patterns of the influence of land use on water quality may be established.

3.7.4 Monitoring Variables

The appropriate set of land use/treatment variables for any monitoring plan will depend on the monitoring objectives, monitoring design and characteristics of the watershed or site to be monitored. The set of land use/treatment variables needed for problem assessment is usually broad (USEPA 2008a), whereas the set of variables for BMP effectiveness monitoring is tailored to the BMP and the conditions under which it is being evaluated.

Table 2-2 in section 2.2.2 illustrates an important first step in selecting land use/treatment variables appropriate for the monitoring plan. The next step involves selecting the specific water quality variables and matching those with specific land use/treatment variables for which a relationship is likely. Table 3-13 shows examples of pairing water quality and land use/treatment variables.

Table 3-13. Relationship of water quality and land use/land treatment variables

Source	Water Quality Monitoring Variable	~Weekly Land Use/Treatment Monitoring Variables	~Annual Land Use/Treatment Monitoring Variables
Cropland Erosion	Suspended Sediment	• Date of tillage operations; • Tillage equipment used; • Crop canopy development; • Cover crop density	• Acreage (and percentage) of land under reduced tillage; • Acreage (and percentage) served by terrace systems; • Acreage (and percentage) of land converted to permanent cover; • Linear feet (and percentage of linear feet) of watercourse protected with riparian buffers
Agricultural Cropland	Total Nitrogen	• Manure and/or fertilizer application rates; • Manure and/or fertilizer forms; • Date of manure and/or fertilizer application; • Manure and/or fertilizer application methods	• Number (and percentage) and acreage (and percentage) of farms implementing comprehensive nutrient management plans; • Annual fertilizer and manure N applications per acre; • Legume acreage; • N fertilizer sales
Urban	Stream Flow	• Operation and maintenance of stormwater system; • Functioning of stormwater diversions or treatment devices	• Percent impervious cover; • Acreage (and percentage) served by water detention/retention; • Number and area of rain gardens or other infiltration practices

"~Weekly" variables are those that must be monitored frequently to record the exact date or quantity associated with the metric. "~Annual" variables can be determined less frequently as they generally remain constant within a crop year.

3.7.5 Sampling Frequency

As discussed briefly in section 3.7.3, land use/treatment data can be either static or dynamic (USDA-NRCS 2003). Static land use/treatment data such as soil type and slope do not generally change with time, but dynamic land use/treatment data can vary with time. Examples of dynamic land use/treatment data include the number of animals, crop rotations, cover crops, undisturbed area, nutrient and pesticide applications, road salting, and irrigation schedules.

Sampling frequency will vary depending on the study design and source activity. For BMP effectiveness studies at the plot or field scale, observations should be made each time the site is visited (USDA-NRCS 2003). It is possible to easily observe the entire study area at these scales, but observations made at monitoring stations for larger-scale projects, although important to do, will not cover the entire study area. The frequency for sampling dynamic data will vary depending on the type and magnitude of the variable's impact on measured water quality. For example, construction activities occur on a daily basis at any given construction site, but there are construction phases that are more important than others (e.g., site clearing) and therefore warrant closer attention. The availability of records should also be considered when determining sampling frequency. Producers under many nutrient management plans, for example, must keep field-by-field records of manure and chemical nutrient applications, so sampling can theoretically be done on an annual basis assuming that the records are clear and accurate.

3.7.6 Challenges

There are many challenges associated with tracking land use/treatment, including gaining access to locations for direct observation or communication with landowners or managers. Obtaining cooperation on field logs also represents a major hurdle in many cases, especially when confidential business information is involved. At the watershed scale, the task of checking all source activities of potential interest can be difficult logistically, labor intensive, and complicated in a mixed use watershed where different areas of expertise may be needed to track a wide range of source activities.

Data confidentiality can present major challenges to monitoring land management in a watershed project. Confidentiality applies at many levels, from individual landowners participating in USDA cost-share programs through their local NRCS district to county or watershed-level data reported in the Census of Agriculture. In small projects, a good way to overcome this obstacle is to obtain permission from the landowner; with such permission, NRCS and FSA records of BMP implementation will be accessible. In some field-scale projects, it may be possible to have the cooperating landowner(s) sign a release at the beginning of the project to allow access to their records, including nutrient management plans, participation in cost-share programs, BMP installation, etc.

Dealing with larger scale agency data is more problematic. As noted previously (section 3.7.2.4), data reported by the Census of Agriculture are not disclosed if a limited number of producers are present in a county or watershed. This data gap presents a challenge to determining basic characteristics of a county or watershed such as cropland acres or animal populations. There are, however, some helpful approaches to estimate the undisclosed data. For example, if dairy cow numbers are not disclosed for a county of interest, it is possible to add up the numbers for reported counties, subtract that sum from the state total to arrive at a number for the "remainder" dairy cows. If data from more than one county are non-disclosed, the "remainder" animals can be apportioned by county area, cropland acres, or other reported variable. Although such procedures are cumbersome and add uncertainty, they often represent the best or only source of data for a project area.

Estimation of non-disclosed Census of Agriculture data Nutrient Use Geographic Information System (NuGIS) International Plant Nutrition Institute

Objective: **A major national study of fertilizer nutrient balance by county needed to derive estimates of fertilizer nutrients applied and removed in harvested crops for each U.S. county**

Standard procedures for estimating data missing due to non-disclosure in Census of Agriculture were developed:

- *When Census of Ag Production data for a commodity were not disclosed for some counties in a state, subtracting the sum of disclosed production for a commodity from the state total production for that same commodity yielded a remainder – the 'State Production Remainder' – that represents the sum of production in non-disclosed counties for that commodity. We apportion the State Production Remainder for this commodity to each county in a state with non-disclosed production for this commodity, based on each county's harvested acres of this commodity as reported in the Census of Ag or as estimated.*
- *For each commodity, the amount of State Production Remainder that is apportioned to each county with a non-disclosed production value was calculated using a 'Production to Harvested Acres coefficient'; this could also be thought of as an estimated yield. This coefficient was calculated, for each commodity, in each state, by dividing the (State Production Remainder) by the (Sum of Harvested Acres in counties with non-disclosed Production). The county crop production was then calculated using:*

  (County Total Cropland Acres) X (Harvested Acres to Total Cropland Acres coefficient).

Example:

State Production Remainder for Corn =
(State total production of corn) - (sum of corn production in counties with data disclosed)

2 million bu corn – 1 million bu corn = 1 million bu corn

State Production to Harvested Acres coefficient for corn =
(State Production Remainder for Corn) /
(Sum of Harvested Acres of Corn in counties with non-disclosed production of corn) =

(1 million bu) / (5,000 Harvested Acres of Corn) = 200 bu corn / harvested acre of corn

Estimated Production of Corn for County A =
(Harvested Acres of Corn in County A) X (Production to Harvested Acres Coefficient for corn) =
(3,000 Harvested acres) X (200 bu corn / harvested acre) = 600,000 bu of corn production

(IPNI 2010)

3.8 Special considerations for pollutant load estimation

Because of the central role pollutant loads and load reduction targets play in many watershed projects, especially those with TMDLs, the accuracy of load estimates is very important to all project stakeholders. Further, the potentially high relative cost of monitoring for load estimation (see chapter 9) places a premium on cost-effectiveness. This section combines many of the observations made in this chapter about monitoring for load estimation in one place to provide basic guidelines and considerations for this special type of monitoring. Richards (1998) provides a comprehensive discussion of pollutant load estimation techniques and is the source of much of the information presented here.

Pollutant flux (see Box) varies tremendously with both flow and pollutant concentration. Because we cannot measure flux directly or continuously, we usually compute unit loads (e.g., daily or monthly) as the product of discharge and pollutant concentration, then sum these unit loads to produce an estimate of annual load.

> **Basic Pollutant Load Terms**
>
> **Flux** – instantaneous loading rate (e.g., kg/sec)
>
> **Flow rate** – instantaneous rate of water passage (e.g., L/sec)
>
> **Discharge** – quantity of water passing a specified point (e.g., m3)
>
> **Load** – mass of substance passing a specified point (e.g., metric tons).

The following steps are recommended to plan a monitoring effort for load estimation:

1. Determine whether the project goals require knowledge of load, or if goals can be met using concentration data alone. In many cases, especially when trend detection is the goal, concentration data may be easier to work with and be more accurate than crudely estimated load data. However, some concurrent hydrologic/meteorologic data (flow, stage height, rainfall, etc.) are often needed for some aspect of any watershed study.

2. If load estimates are required, determine the accuracy and precision needed based on the uses to which they will be put. This is especially critical when the purpose of monitoring is to look for a change in load. It is foolish to attempt to document a 25 percent load reduction from a watershed program with a monitoring design that gives load estimates ± 50 percent of the true load (see Spooner et al. 2011).

3. Decide which approach will be used to calculate the loads based on known or expected attributes of the data. This decision will also lead to choices on monitoring equipment (e.g., whether an automatic sampler will be used). See section 7.9.2 for a discussion of approaches to load estimation and see below for a discussion of sampling equipment.

4. Use the precision goals from Step 2 to calculate the sampling requirements for the monitoring program. Sampling requirements include both the total number of samples and the distribution of the samples with respect to some auxiliary variable such as flow or season. See section 3.4 and below for information on sampling frequency and distribution.

5. Calculate the loads based on the samples obtained after the first full year of monitoring, and compare the precision estimates (of both flow measurement and the sampling program) with the initial goals of the program. Adjust the sampling program if the estimated precision deviates substantially from the goals. See *Interval Estimation* (p. 4-18 of the 1997 guidance [USEPA 1997a]) or Spooner et al. (2011) and section 3.4.2 for information relevant to this step.

3.8.1 Sample Type and Sampling Equipment

The basic approaches for load estimation described in section 7.9 of this guidance are numeric integration, regression, and ratio methods. With numeric integration, the goal is to collect representative concentration samples for each sampling interval which is typically defined either by the calendar (e.g., daily, weekly) or by the volume of flow that passes by the sampling point. In other words, there are no data gaps. For both the regression and ratio methods, it is assumed that a strong relationship exists between concentration and flow and that there will be sampling intervals for which only flow is measured (i.e., no concentration samples taken). With the regression approach, the missing concentration values are then estimated from the relationship of flow and concentration (when concentration samples were taken). The ratio approach assumes that flow is measured for each sampling interval and that daily loads are calculated for those days when concentration samples are taken. A flow ratio (annual flow/flow for days with concentration samples) is then used in combination with a bias correction factor (to account for correlation between discharge and load) to estimate annual pollutant load. Both the regression and ratio methods can be performed using annual or seasonal relationships. These relationships may change over time, particularly in cases where BMPs are implemented, so it is important that the relationships are re-examined at least annually.

Autosamplers are typically required for numeric or composite integration because of the large number of concentration samples needed. Grab sampling is typical for both the regression and ratio methods, but autosamplers can be used. Continuous or near-continuous flow measurement is required for all three methods but in some cases flow data are obtained from others (e.g., USGS).

Section 3.2 describes many options for sample type, the simplest of which is a grab sample. The specific type of sample appropriate for each project will depend on the details of the load estimation objective. For example, it may be desirable to track the variability of both concentration and load during the sampling interval. In this case, multiple discrete samples over time would be preferred over composite samples; the cost for sample analysis, however, would increase considerably. Where fluctuations within the sampling interval are not of interest, composite samples would be recommended. Flow-proportional samples are recommended for load estimation in these cases.

3.8.2 Sampling Frequency and Timing

Sample type is an essential consideration involved in sampling for good load estimation but sampling frequency and sample distribution over time are equally important. The selection of sampling frequency required for accurate estimation of pollutant loads is more challenging than for concentration because load is a product of concentration and flow, both of which usually vary significantly. Furthermore, in NPS situations, because the majority of the annual pollutant load often occurs in a few major events, the choice of *when* to sample is also critical.

Ideally, the most accurate approach to estimating pollutant load would be to sample very frequently and capture all the variability. Flow is relatively straightforward to measure continuously (see Meals and Dressing 2008 and section 3.1.3.1), but concentration is expensive to measure and in most cases impossible to measure continuously. It is therefore critically important to choose a sampling interval that will yield a suitable characterization of concentration. Strategies for determining sampling frequency and timing for accurate load estimation are described below; see Richards (1998) for additional information.

Sampling frequency determines the number of unit load estimates that can be computed and summed for an estimate of total load. Using more unit loads increases the probability of capturing variability across the year and not missing an important event; in general, the accuracy and precision of a load estimate increase as sampling frequency increases. For example, the top panel in Figure 3-50 shows load estimated from weekly sampling superimposed on idealized daily load data. The bottom panel shows results plotted from monthly and quarterly sampling on top of the same daily load data. The weekly data appear to capture much of the variation of the daily series, but the monthly series does much more poorly. Quarterly sampling clearly misses many important peaks and overstates periods of low flux.

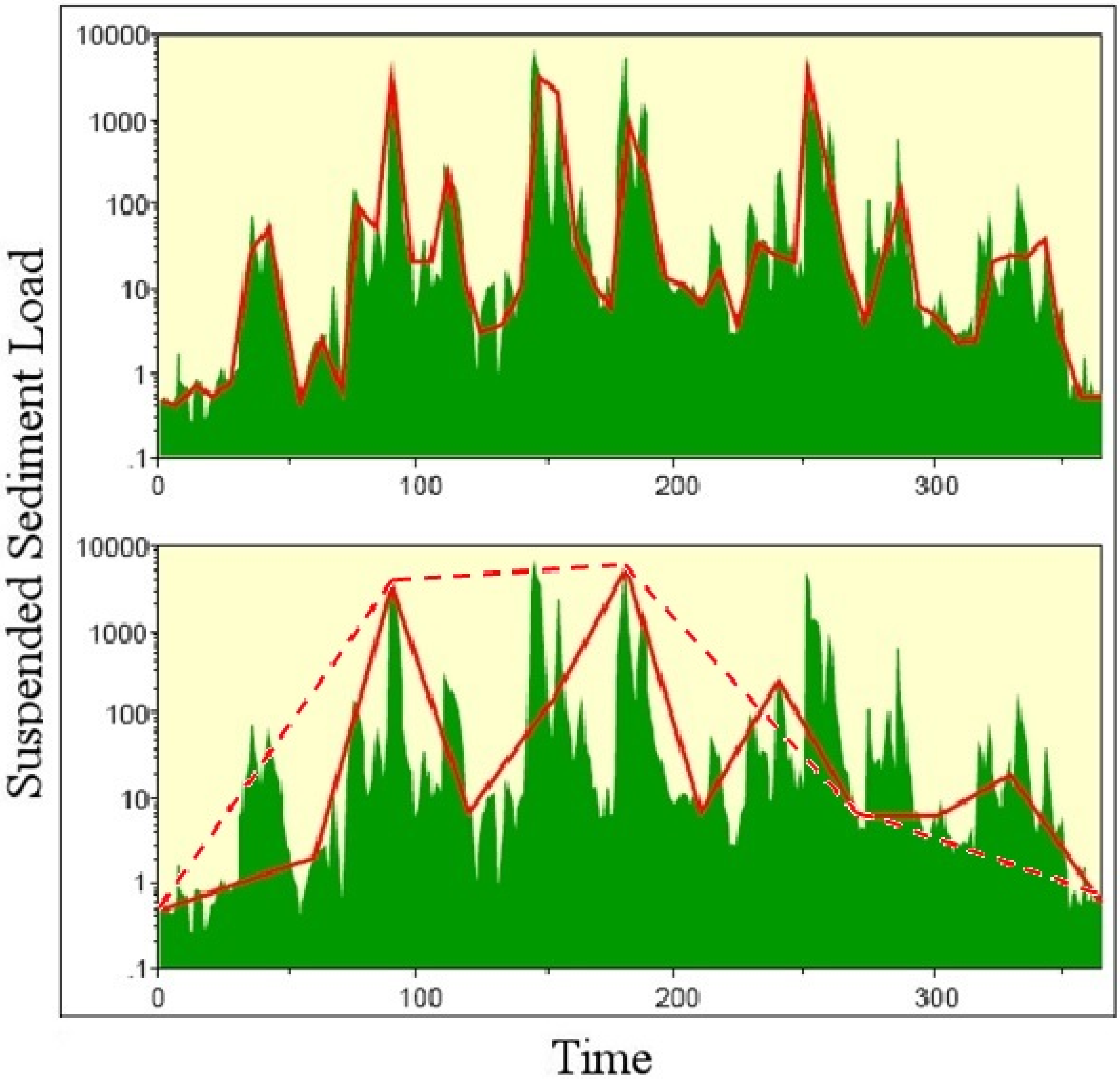


Figure 3-50. Weekly (top panel) and monthly and quarterly (bottom panel, solid and dashed lines, respectively) load time series superimposed on idealized daily load time series (adapted from Richards 1998)

There is a practical limit to the benefits of increasing sampling frequency, however, due to the fact that water quality data tend to be autocorrelated (see section 7.3.6). The concentration or flux at a certain point today is related to the concentration or flux at the same point yesterday and, perhaps to a lesser extent, to the concentration or flux at that spot last week. Because of this autocorrelation, beyond some point, increasing sampling frequency will accomplish little in the way of generating new information. This is usually not a problem for monitoring programs but can be a concern when electronic sensors are used to collect data nearly continuously.

The choice of *when* to collect concentration samples is critical. Most NPS water quality data have a strong seasonal component as well as a strong association with other variable factors such as precipitation, streamflow, or watershed management activities such as tillage or fertilizer application. Selecting when to collect samples for concentration determination is essentially equivalent to selecting when the unit loads that go into an annual load estimate are determined. That choice must consider the fundamental characteristics of the system being monitored. In northern climates, spring snowmelt is often the dominant export event of the year; sampling during that period may need to be more intensive than during midsummer in order to capture the most important peak flows and concentrations. In southern regions, intensive summer storms often generate the majority of annual pollutant load; intensive summer monitoring may be required to obtain good load estimates. For many agricultural pesticides, sampling may need to be focused on the brief period immediately after application when most losses tend to occur. In arid areas, it may be more appropriate to collect storm composites, focusing sampling efforts on the normal wet periods. Regardless of the approach chosen, it is essential that loads are calculated after the first year in accordance with Step 5 above to determine if precision needs are met.

For both the regression and ratio approaches, determination of sampling frequency may assume a normal distribution for concentration and random sampling. Several formulas are available to calculate the number of samples (random or within strata) required to obtain a load estimate of acceptable accuracy based on known variance of the system (see chapter 2 of the 1997 guidance). Stratification may improve the precision and accuracy of the load estimate by allocating more of the sampling effort to the aspects which are of greatest interest or which are most difficult to characterize because of great variability such as high flow seasons.

3.8.3 Planning and Cost Considerations

As described here, the sampling regime needed for load estimation must be established in the initial monitoring design, based on quantitative statements of the precision required for the load estimate. In many cases, the decision to calculate loads is sometimes made after the data are collected, often using data collected for other purposes. At that point, little can be done to compensate for a data set that contains too few observations of concentration, discharge, or both, collected using an inappropriate sampling design. Many programs choose monthly or quarterly sampling with no better rationale than convenience and tradition. A simulation study for some Great Lakes tributaries revealed that data from a monthly sampling program, combined with a simple load estimation procedure, gave load estimates which were biased low by 35 percent or more half of the time (Richards and Holloway 1987).

Monitoring programs often struggle with a conflict between the number of observations a program can afford and the number needed to obtain an accurate and reliable load estimate. Most use flow as a means to estimate the best intervals between concentration observations. For example, planning to collect samples every x thousand ft^3 of discharge would automatically emphasize high flux conditions while economizing on sampling during baseflow conditions.

It is possible, however, that funding or other limitations may prevent a monitoring program from collecting the data required for acceptable load estimation. In such a case, the question must be asked: is a biased, highly uncertain load estimate preferable to no load estimate at all? Sometimes the correct answer will be no.

3.9 Data Management

3.9.1 General considerations

Data management can be defined as the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver, and enhance the value of data and information assets (Mosley et al. 2009). Small, short-term monitoring projects can often set up and operate their own effective data management system using basic tools like spreadsheets and paper files. Depending on the magnitude and duration of the monitoring project, it may be advisable to go beyond immediate local data storage and reporting practices and participate in and comply with ongoing USEPA data management programs (e.g., USEPA 2010). Regardless of the magnitude of the monitoring effort, data management must be part of initial project planning.

Data management planning should be an integral part of developing a monitoring plan as reflected by its inclusion as a Group B element in QAPPs (USEPA 2001a). The aspects of data management to be described in a QAPP include the path of the data from their generation to their final use or storage, the standard record-keeping procedures, document control system, and the approach used for data storage and retrieval on electronic media. In addition, the control mechanism for detecting and correcting errors and for preventing loss of data during data reduction, data reporting, and data entry to forms, reports, and databases are to be described in the QAPP. Examples of any forms or checklists to be used are also required, as are descriptions of all data handling equipment and procedures to process, compile, and analyze the data. This includes procedures for addressing data generated as part of the project as well as secondary data from other sources. Required computer hardware and software and any specific performance requirements for the hardware/software configuration used are to be described. Data analysis software options are described in chapter 7.

3.9.2 Data acquisition

Sections 2.1 through 3.7 and chapters 4-5 address experimental design, sample collection, and sample analysis methods for a wide range of nonpoint source monitoring projects. The data generated by these monitoring projects must be collected (data acquisition) and transferred to the data management system for storage and analysis.

Field and laboratory procedures may include the use of field books or data entry sheets to record observations and measurements and either paper or electronic data report forms. The transcription of data reported in these fashions into a database is a potential source of typographic errors, switched digits, and other errors in data entry. It is crucial that all data be error-checked after entry into electronic forms, but before analysis and reporting. Finding errors in a dataset after analysis and reporting is underway can be very frustrating.

Newer methods of data acquisition include the use of data loggers (either external loggers that record multiple data streams or loggers directly built into sensor devices), laptops, tablets, and smartphones to allow direct acquisition, transmission, and entry of data to electronic media. An advantage of using data loggers is that manual data entry and the associated transcription errors are avoided (USDA-NRCS 2003). Remote access allows direct transfer of field data from a data logger to the main data storage site. One disadvantage of data loggers is that their storage capacity is limited; once full, new data may not be recorded or older data may be overwritten and thus lost. It is strongly recommended that monitoring protocols include prompt and routine downloads of data from field data loggers.

Not all data are generated directly by the project. Element B-9 of a QAPP addresses data obtained from non-measurement sources such as computer databases, programs, literature files, and historical data bases (USEPA 2001a). Whenever data are obtained from other sources, it is important to determine the sufficiency of the data for project purposes (USEPA 2008a). One of the challenges of using GIS data, for example, is the need to ground-truth and fill gaps in the data layers (USDA-NRCS 2003). Johnson and Zelt (2005) present a method for filling in a data gap of spatial scale in woodland LULC (land-use/land-cover) between the land-cover data available from the 30-m 1990s National Land Cover Dataset (NLCD) and the reach-level data available from the prescribed National Water Quality Assessment (NAWQA) habitat assessment.

Data provided by others may have been collected at different locations, by different methods, or to serve different objectives from those of the current project, so it is important to carefully review the data and methods used for its collection. This situation is a common occurrence in the watershed project planning phase during which projects often must use whatever data are available to characterize problems and suggest actions to solve those problems. The QAPP should include acceptance criteria for the use of such data in the project, as well as any data use limitations (USEPA 2001a).

3.9.3 Data storage

Data storage includes both manual and computerized technologies (USDA-NRCS 2003). All field and laboratory notebooks must be fully documented and stored safely, and all data contained in the notebooks should be backed up in paper or digital form, perhaps as scanned images.

A data inventory is important for monitoring projects, particularly those focused on problem assessment. Information on ways to organize and manage a data inventory is provided in existing guidance (USEPA 2008a). Naming and labeling conventions should be established, and metadata (e.g., where, how, why, when and what was monitored) should be included with all datasets.

Spreadsheets might be adequate for data generated by small projects, but a relational database is usually preferable for more complex projects involving many sites or variables (USEPA 2008a). A relational database houses data, metadata (information about the data), and other ancillary information in a series of relational tables including station information, sample information, analyses, methods used, and quality control information.

All computerized data and electronic project files should be backed up using one of many options, including USB flash drives, external hard drives, CDs, remote servers via File Transfer Protocol (FTP), and commercial data storage systems available on-line. All media have their advantages and disadvantages. As technology changes, computerized data should be copied to the latest media using the latest software. For archival purposes, data storage as paper printouts may be a preferred choice; consider that 1985 data archived on 5.25-in floppy disks would be next to useless today. Daily backup of computerized data and electronic project files is recommended. Where practical, backups should be stored offsite for protection against theft, fire or water damage. Today, with the proliferation of relatively inexpensive and free options for data backups, there is little excuse for losing data due to computer failure once the data has returned from the field or laboratory.

3.10 Data Reporting and Presentation

3.10.1 General considerations

Data reporting and presentation occur at multiple levels in many forms to address a wide range of audiences and purposes. Communication with stakeholders is often best done on a frequent, informal basis, whereas communication with outside audiences is more commonly accomplished via presentations at professional meetings or publications of project findings.

Funding agencies generally include reporting requirements in their grants or contracts. Some include requirements to upload the data to repositories such as EPA's STORET (http://www.epa.gov/storet/). States receiving section 319 grants are required to use GRTS (Grants Reporting and Tracking System) to report specific nationally mandated data elements (USEPA 2013).

3.10.2 Communicating with Stakeholders

Project managers should schedule regular meetings with stakeholders to present available data and discuss both successes and failures. Project staff will often find that stakeholders have information, ideas, and resources they need to improve the project or make their objectives easier to accomplish. Quarterly meetings are recommended, so those collecting and analyzing the data should examine the data frequently to be familiar with the current status of the project and to identify and fix problems. The USGS, for example, recommends that field and laboratory results be examined as soon as possible, preferably before the next sample-collection field trip (Wilde 2005). Results indicating potential bias in the data may trigger needed changes in equipment, equipment-cleaning procedures, or field methods used.

Communicating with groups of individuals with varied levels of understanding and different learning styles requires a multimedia approach that includes written materials, audio-visual presentations, and face-to-face communication. Simple quarterly reports with easily interpreted graphs, summary tables, and maps will enhance the communication. Reports should highlight observed patterns and both raw data and metadata should be attached for those in the audience with more advanced understanding of project data. A particularly powerful tool for presenting information to any audience is a Geographic Information System (GIS) that can be used to create watershed maps and display a variety of spatial information (USEPA 2008a). Users can display selected data and a combination of spatial coverages tailored to the specific audience and venue.

3.10.3 Final reports

Final reports are an essential element of all monitoring projects, but experiences of the Rural Clean Water Program (USEPA 1993a) and similar watershed programs show that project budgets frequently do not provide sufficient resources for final data analysis and reporting. One way to address this problem is to require quarterly reports and meetings as described in section 3.10.2. A major hurdle associated with final reports is the task of pulling together all project data and performing the final analyses. This burden is reduced substantially if reports and analyses have been generated on a regular basis since the beginning of the project.

The basic elements of a project report are the title, abstract, introduction, body, summary and conclusions, references, forward, preface, appendices, glossary, tables, and illustrations (USGS 2008). The introduction should include the purpose and scope of the report, and will usually include background information pertinent to the study. The body of the report includes the purpose of the study, data summaries, and the analyses and interpretation of the data. The summary and conclusions pull together the major results and conclusions described in the body. A concise Executive Summary is useful as a pull-out section to distribute project results to a wide audience.

State and federal agencies have their own guidelines and reporting requirements. Professional publications and journals specify reporting requirements at their websites.

3.11 References

Abt, S.R., Wittler, R.J., Taylor, A., and D.J. Love, 1989. Human stability in a high flood hazard zone. *Water Resources Bulletin* 25(4) 881-889.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Geological Survey Professional Paper 964. U.S. Department of the Interior, Geological Survey. Accessed April 22, 2016. https://pubs.er.usgs.gov/publication/pp964.

Arnone, R.D. and J.P. Walling. 2007. Waterborne pathogens in urban watersheds. *Journal of Water and Health* 5(1):149-162.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*. 2nd ed. EPA/841-D-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 9, 2016. http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm.

Barnstable County. 2016. *Comprehensive Laboratory Pricing*. Water Quality Laboratory, Barnstable County Department of Health and Environment, Barnstable, MA. Accessed February 2, 2016.

Barrett, M. 2005. *Stormwater Quality Documentation of Roadside Shoulders Borrow Ditches: a Summary*. Project Summary Report 0-4605-S. The University Of Texas, Center for Transportation Research, Austin, TX. Accessed February 5, 2016. http://library.ctr.utexas.edu/hostedpdfs/txdot/psr/0-4605-s.pdf.

Beegle. D. 1996. *Nutrient Management in Conservation Tillage Systems*. Conservation Tillage Series Number 4. The Pennsylvania State University, College of Agricultural Sciences, University Park, PA. Accessed February 2, 2016. http://pubs.cas.psu.edu/freepubs/pdfs/uc127.pdf.

Ben Meadows. 2016. Ben Meadows Company, Janesville, WI. Accessed February 8, 2016. http://www.benmeadows.com/.

Bishop, P.L., W.D. Hively, J.R. Stedinger, M.R. Rafferty, J.L. Lojpersberger, and J.A. Bloomfield. 2005. Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. *Journal of Environmental Quality* 34:1087-1101.

Bonilla, C.A., D.G. Kroll, J.M. Norman, D.C. Yoder, C.C. Molling, P.S. Miller, J.C. Panuska, J.B. Topel, P.L. Wakeman, and K.G. Karthikeyan. 2006. Instrumentation for measuring runoff, sediment, and chemical losses from agricultural fields. *Journal of Environmental Quality* 35:216-223.

Brakensiek, D.L., H.B. Osborn, and W.J. Rawls. 1979. *Field Manual for Research in Agricultural Hydrology*. Agriculture Handbook No. 224, U.S. Department of Agriculture, Science and Education Administration, Washington, DC.

Brodie, I. and M. Porter. 2004. Use of Passive Stormwater Samplers in Water Sensitive Urban Design. In *Proceedings of WSUD 2004 International Conference on Water Sensitive Urban Design: Cities As Catchments*, Adelaide, South Australia, November 21-25, 2004. Accessed February 5, 2016. https://eprints.usq.edu.au/3579/1/Brodie_Porter.pdf.

Creed, J.T., T.D. Martin, and J.W. O'Dell. 1994. *Method 200.9: Determination of Trace Elements by Stabilized Temperature Graphite Furnace Atomic Absorption Revision 2.2*. U.S. Environmental Protection Agency, Office Of Research And Development, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio. Accessed February 2, 2016. http://www.caslab.com/EPA-Methods/PDF/EPA-Method-2009.pdf.

CTIC (Conservation Technology Information Center). 2016. *Crop Residue Management Survey*. Conservation Technology Information Center, West Lafayette, IN. Accessed January 29, 2016. http://www.ctic.purdue.edu/crm.

Dressing, S.A. 2014. *Monitoring Cost Estimation Spreadsheet – Simplified Version*. Tetra Tech, Inc., Fairfax, VA. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Dressing, S.A., J. Spooner, J.M. Kreglow, E.O. Beasley, and P.W. Westerman. 1987. Water and sediment sampler for plot and field studies. *Journal of Environmental Quality* 16(1):59-64.

Dufour, A.P. 1984. *Health Effects Criteria for Fresh Recreational Waters*. EPA-600/1-84-004, U.S. Environmental Protection Agency, Cincinnati, OH.

Eads, R.E. and R.B. Thomas. 1983. Evaluation of a depth proportional intake device for automatic pumping samplers. *Water Resources Bulletin* 19(2): 289-292. Accessed February 5, 2016. http://www.fs.fed.us/psw/publications/eads/Eads83.pdf.

FISP (Federal Interagency Sedimentation Project). 2014. Federal Interagency Sedimentation Project. Accessed August 26, 2014. http://water.usgs.gov/fisp/.

Freeman, L.A., M.C. Carpenter, D.O. Rosenberry, J.P. Rousseau, R. Unger, and J.S. McLean. 2004. Use of Submersible Pressure Transducers in Water-Resources Investigations. Book 8, Chapter A3 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://pubs.usgs.gov/twri/twri8a3/pdf/twri8-a3.pdf.

Fulton, J. and J. Ostrowski. 2008. Measuring real-time streamflow using emerging technologies: radar, hydroacoustics, and the probability concept. *Journal of Hydrology* 357:1–10. Accessed February 5, 2016. http://hydroacoustics.usgs.gov/publications/fulton-ostrowski08.pdf.

Gassman, P.W., M.R. Reyes, C.H. Green, and J.G. Arnold. 2007. The Soil and Water Assessment Tool: historical development, applications, and future research directions. *Transactions of the American Society of Agricultural and Biological Engineers.* 50(4):1211-1250. Accessed February 5, 2016. http://www.card.iastate.edu/environment/items/asabe_swat.pdf.

Gibs, J., F.D. Wilde, and H.A. Heckathorn. 2007. Use of Multiparameter Instruments for Routine Field Measurements (Ver. 1.1). Book 9, Chapter A6 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://water.usgs.gov/owq/FieldManual/Chapter6/6.8_contents.html.

Gilpin, B. J., J. E. Gregor, and M. G. Savill. 2002. Identification of the source of fecal pollution in contaminated rivers. *Water Science and Technology* 46:9-15.

GKY (GKY and Associates). 2014. First-Flush Sampler. GKY and Associates, Inc., Chantilly, VA.

Graczyk, D.J., D.M. Robertson, W.J. Rose, and J.J. Steuer. 2000. *Comparison of Water-Quality Samples Collected by Siphon Samplers and Automatic Samplers in Wisconsin*. Fact Sheet FS-067-00. U.S. Geological Survey, Middleton, WI. Accessed July 23, 2013. http://wi.water.usgs.gov/pubs/FS-067-00/FS-067-00.pdf.

Grady, C., A.P. Reimer, J.R. Frankenberger, and L.S. Prokopy. 2013. Locating existing best management practices within a watershed: The value of multiple methods. *Journal of American Water Resources Association* 49(4):883-895.

Gray, J.R., G. D. Glysson, L.M. Turcios, and G.E. Schwarz. 2000. *Comparability of Suspended-Sediment Concentration and Total Suspended Solids Data*. Water-Resources Investigations Report 00-4191. U. S. Geological Survey, Reston, VA.

Hach®. 2013a. 900MAX Portable Standard Sampler with (24) 1 Liter Bottle Kit. Hach Company, Loveland, CO. Accessed February 5, 2016. http://www.hach.com/product.detail-overview.jsa?id=7640267036.

Hach®. 2013b. 900 MAX Refrigerated Sampleter, 115V with 5.5 Gallon Bottle Kit. Hach Company, Loveland, CO. Accessed February 5, 2016. http://www.hach.com/900-max-refrigerated-sampler-115v-with-5-5-gallon-bottle-kit/product?id=7640263242&callback=qs.

Hansen, N. C., Gupta, S. C. and Moncrief, J. F. 2000. Snowmelt runoff, sediment, and phosphorus losses under three different tillage systems. *Soil Tillage Research* 57(1-2):93-100.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Transactions of the American Society of Agricultural and Biological Engineers* 49(3): 689−701.

Harwood, V.J., A.D. Levine, T.M. Scott, V. Chivukula, J. Lukasik, S.R. Farrah, and J.B. Rose. 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Applied Environmental Microbiology* 71(6):3163-3170.

Hawkins, C.P., J.L. Kershner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman, and M.K. Young. 1993. A hierarchical approach to classifying stream habitat features. *Fisheries* 18(6):3-12.

Hively, W.D., M. Lang, G.W. McCarty, J. Keppler, A. Sadeghi, and L.L. McConnell. 2009a. Using satellite remote sensing to estimate winter cover crop nutrient uptake efficiency. *Journal of Soil and Water Conservation* 64(5):303-313.

Hively, W.D., G.W. McCarty, and J. Keppler. 2009b. Federal-state partnership yields success in remote sensing analysis of conservation practice effectiveness: Results from the Choptank River Conservation Effects Assessment Project. *Journal of Soil and Water Conservation* 64(5):154A.

Hively, W.D., O.H. Devereux, and P. Claggett. 2013. *Integrating Federal and State Data Records to Report Progress in Establishing Agricultural Conservation Practices on Chesapeake Bay Farms*.

Open-File Report 2013–1287. U.S. Geological Survey, Reston, VA. Accessed February 1, 2016. http://pubs.usgs.gov/of/2013/1287/pdf/of2013-1287.pdf.

Hutchison, M.L., L.D. Walters, S.M. Avery, F. Munro, and A. Moore. 2005. Analyses of livestock production, waste storage, and pathogen levels and prevalences in farm manures. *Applied Environmental Microbiolog.* 71(3): 1231–1236.

IPNI (International Plant Nutrition Institute). 2010. *A Preliminary Nutrient Use Geographic Information System (NuGIS) for the U.S.* IPNI Publication No. 30-3270. International Plant Nutrition Institute, Norcross, GA.

Jiang, S.C., W. Chu, B.H. Olson, J-W. He, S. Choi, J. Zhang, J.Y Le, and P.B. Gedalanga. 2007. Microbial source tracking in a small southern California urban watershed indicates wild animals and growth as the source of fecal bacteria. *Applied Microbiology and Biotechnology* 76(5):927-934.

Johnson, M.R. and R.B. Zelt. 2005. *Protocols for Mapping and Characterizing Land Use/Land Cover in Riparian Zones*. Open-File Report 2005–1302. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://pubs.usgs.gov/of/2005/1302/.

Kahn, A.H., and C.K. Ong. 1997. Design and calibration of tipping bucket system for field runoff and sediment quantification. *Journal of Soil and Water Conservation* 52(6):437-443.

Klemm, D.J., Q.J. Stober, and J.M. Lazorschak. 1992. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. EPA/600/R-92/111. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Komiskey, M.J., T.D. Stuntebeck, A.L. Cox, and D.R. Frame. 2013. *Implications of Flume Slope on Discharge Estimates from 0.762-Meter H Flumes Used in Edge-of-Field Monitoring*. Open-File Report 2013–1082, U.S. Geological Survey, Reston, VA. Accessed July 11, 2013. http://pubs.usgs.gov/of/2013/1082/pdf/ofr2013-1082_web.pdf.

Kuster, M., A. De la Cal, E. Eljarrat, M.J. Lopez de Alda, and D. Barcelo. 2010. Evaluation of two aquatic passive sampling configurations for their suitability in the analysis of estrogens in water . *Talanta* 83(2):493-499.

Lane, S.L. and R.G. Fay. 1997. Safety in Field Activities. Book 9, Chapter A9 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://water.usgs.gov/owq/FieldManual/Chap9/content.html.

Lapham, W.W., F.D. Wilde, and M.T. Koterba. 1997. *Guidelines and Standard Procedures for Studies of Ground-Water Quality: Selection and Installation of Wells, and Supporting Documentation*. Water-Resources Investigations Report 96-4233. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://pubs.usgs.gov/wri/wri964233/pdf/wri964233.pdf.

Liscio, C., E. Magi, M.DiCarro, M.J.-F. Suter, and E.L. M. Vermeirssen. 2009. Combining passive samplers and biomonitors to evaluate endocrine disrupting compounds in a wastewater treatment plant by LC/MS/MS and bioassay analyses. *Environmental Pollution* 157(10):2716-2721.

Martin, T.D. 2003. *Method 200.5: Determination of Trace Elements in Drinking Water by Axially Viewed Inductively Coupled Plasma - Atomic Emission Spectrometry Revision 4.2*. U.S.

Environmental Protection Agency, Office Of Research And Development, National Exposure Research Laboratory, Cincinnati, Ohio. Accessed February 5, 2016. https://www.nemi.gov/methods/method_summary/9798/.

Matalas, N.C. 1967. Time series analysis. *Water Resources Research* 3(3):817–829.

MBI (Midwest Biodiversity Institute). 2006. *Methods for Assessing Habitat inIn Flowing Waters: Using the Qualitative Habitat Evaluation Index*. Technical Bulletin EAS/2006-06-1. Ohio Environmental Protection Agency, Midwest Biodiversity Institute for Division of Surface Water, Groveport, OH. Accessed February 5, 2016. http://www.epa.state.oh.us/portals/35/documents/qheimanualjune2006.pdf.

McCarty, G.W., L.L. McConnell, C.J. Hapeman, A. Sadeghi, C. Graff, W.D. Hively, M.W. Lang, T.R. Fisher, T. Jordan, C.P. Rice, E.E. Codling, D. Whitall, A. Lynn, J. Keppler, and M.L. Fogel. 2008. Water quality and conservation practice effects in the Choptank River watershed. *Journal of Soil and Water Conservation* 63(6):461-474.

Meals, D.W. 1992. Relating Land Use and Water Quality in the St. Albans Bay Watershed, Vermont. In *Seminar Publication: The National Rural Clean Water Program Symposium*. EPA/625/R-92/006, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, pp. 131-143.

Meals, D.W. 2001. *Lake Champlain Basin Agricultural Watersheds Section 319 National Monitoring Program Project, Final Project Report: May, 1994-September, 2000*. Vermont Department of Environmental Conservation, Waterbury, VT.

Meals, D.W. and R.B. Hopkins. 2002. Phosphorus reductions following riparian restoration in two agricultural watersheds in Vermont, USA. *Water Science & Technology* 45(9):51–60.

Meals, D.W. and S.A. Dressing. 2008. *Surface Water Flow Measurement for Water Quality Monitoring Projects*. Tech Notes #9, September 2013. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W., D.C. Braun, and J.A. Hanzas. 2011a. *Assessment of Dairy Manure Management Practices to Reduce Pathogen Runoff Losses in Agricultural Watersheds*. Final Report. Prepared for U.S. Department of Agriculture, National Institute of Food and Agriculture, by Stone Environmental, Inc., Montpelier, VT.

Meals, D.W., J. Spooner, S.A. Dressing, and J.B. Harcum. 2011b. *Statistical Analysis for Monotonic Trends*, Tech Notes #6, September 2011. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W., J.B. Harcum, and S.A. Dressing. 2013. *Monitoring for Microbial Pathogens and Indicators*. Tech Notes #9, September 2013. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Miles, E.J. 2009. *Guidelines: Shallow Water Quality Monitoring: Continuous Monitoring Station: Selection, Assembly & Construction*. Special Report in Applied Marine Science and Ocean Engineering No. 412. Virginia Institute of Marine Science, Chesapeake Bay National Estuarine Research Reserve in Virginia. Accessed February 5, 2016. http://www.truecostofdata.com/cms/uploads/Guidelines-Water-Quality-Monitoring-Continuous-Monitoring-Stations.pdf.

Minitab. 2016. *Summary Statistics in Minitab*. Minitab Inc., State College, PA. Accessed February 5, 2016. http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/summary-statistics/summary-statistics-in-minitab/.

Mosley, M., M.H. Brackett, S. Earley, and D. Henderson. 2009. *DAMA Guide To The Data Management Body Of Knowledge (DAMA-DMBOK Guide)*. Technics Publications, Bradley Beach, NJ.

Nalgene. 2007. Storm Water Sampler by Nalgene®. Thermo-Fisher Scientific, Rochester, NY. Accessed February 5, 2016. http://www.coleparmer.com/assets/literature/99159-XXBrochure.pdf.

NCSU (North Carolina State University). 2003. *Watershedss - a Decision Support System for Nonpoint Source Pollution Control*. North Carolina State University, NCSU Water Quality Group, Raleigh, NC. Accessed February 5, 2016. http://www.water.ncsu.edu/watershedss/info/nh3.html.

NEMI (National Environmental Methods Index). 2006. *National Environmental Methods Index*. National Water Quality Monitoring Council. Accessed February 5, 2016. www.nemi.gov.

Nevers, M.B. and R.L. Whitman. n.d. *Lake Monitoring Field Manual*. U.S. Geological Survey, Lake Michigan Ecological Research Station, Porter, IN. Accessed February 5, 2016.

NOAA (National Oceanic and Atmospheric Administration). 2014. *Ocean Explorer*. National Oceanic and Atmospheric Administration. Accessed February 5, 2016. http://oceanexplorer.noaa.gov/.

NWS (National Weather Service). 1989. *Cooperative Station Observations*. National Weather Service Observing Handbook No. 2. National Weather Service, Office of Systems Operations, Observing Systems Branch, Silver Spring, MD. Accessed April 24, 2016. http://www.nws.noaa.gov/os/coop/Publications/coophandbook2.pdf.

OEPA (Ohio Environmental Protection Agency). 1987. *Biological Criteria for the Protection of Aquatic Life: Volume III. Standardized Biological Field Sampling and Laboratory Methods for Assessing Fish and Macroinvertebrate Communities*. Revised 1989 and 2015. Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, OH. Accessed February 9, 2016. http://www.epa.state.oh.us/portals/35/documents/BioCrit15_Vol3.pdf.

OEPA (Ohio Environmental Protection Agency). 2006. *Methods for Assessing Habitat in Flowing Waters: Using the Qualitative Habitat Evaluation Index (QHEI)*. Technical Bulletin EAS/2006-06-1. Prepared for Ohio Environmental Protection Agency, by Midwest Biodiversity Institute Groveport, OH. Accessed February 5, 2016. http://www.epa.state.oh.us/portals/35/documents/qheimanualjune2006.pdf.

Panuska, J. C., Karthikeyan, K. G. and Norman, J. M. 2008. Sediment and phosphorus losses in snowmelt and rainfall runoff from three corn management systems. *Transactions of the American Society of Agriculture and Biological Engineers* 51(1):95-105.

Parker, P.J. and D.L. Busch. 2013. Field-scale evaluation of a multislot passive sampler. *Journal of Soil and Water Conservation* 68:83-88.

Paul M.J. and J.L. Meyer. 2001. Streams in the urban landscape. *Annual Review of Ecology and Systematics* 32:333–365.

Platts, W.S., W.F. Megahan, and G.W. Minshall. 1983. *Methods for Evaluating Stream, Riparian, and Biotic Conditions*. General Technical Report INT-138, U.S. Department of Agriculture, Forest Service, Ogden, UT.

PSU (Penn State University). 2016. *Testing Services Fee Schedule*. Penn State University, College of Agricultural Sciences, Agricultural Analytical Services Lab, University Park, PA. Accessed February 5, 2016. http://www.opsu.edu/dwn/WLAB_Water_Testing_Services.pdf.

Rankin, E.T. 1989. *The Qualitative Habitat Evaluation Index [QHEI]: Rationale, Methods, and Application*. Ohio Environmental Protection Agency, Division of Surface Water, Ecological Assessment Section, Columbus, OH.

Rantz, S.E. et al. 1982. *Measurement and Computation of Streamflow: Volume 2. Computation of Discharge*. Water-Supply Paper 2175. U.S. Geological Survey, Washington, DC. Accessed February 5, 2016. http://pubs.usgs.gov/wsp/wsp2175/pdf/chapter10_vol2.pdf.

Reckhow, K.H. and S.C. Chapra. 1983. *Engineering Approaches forLake Management. Volume 1: Data Analysis and Empirical Modeling*. Butterworth Publishing, Wobrun, MA.

Rice, E.W., R.B. Baird, A.D. Eaton, and L.S. Clesceri, ed. 2012. *Standard Methods for the Examination of Water and Wastewater*. 22nd ed. American Public Health Association, American Water Works Association, Water Environment Federation. American Public Health Association Publications, Washington, DC.

Richards, R.P. 1998. *Estimation of Pollutant Loads in Rivers and Streams: A Guidance Document for NPS Programs*. Prepared for U.S. EPA Region VIII, by Heidelberg University, Water Quality Laboratory, Tiffin, OH. Accessed February 5, 2016. http://141.139.110.110/sites/default/files/jfuller/images/Load_Est1.pdf.

Richards, R.P. and J. Holloway. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research* 23:1939-1948.

Rickly. 2016. Rickly Hydrological Company, Columbus, OH. Accessed February 8, 2016. http://www.rickly.com/.

Roy, A.H., A.D. Rosemond, M.J. Paul, D.S. Leigh, and J.B. Wallace. 2003. Stream macroinvertebrate response to catchment urbanisation (Georgia, U.S.A.). *Freshwater Biology* 48:329–346.

SAS Institute. 2012. *The UNIVARIATE Procedure*. SAS Institute Inc., Cary, NC. Accessed February 5, 2016.

http://support.sas.com/documentation/cdl/en/procstat/63963/HTML/default/viewer.htm#procstat_univariate_sect008.htm.

Selbig, W.R. and R.T. Bannerman. 2011. Development of a depth-integrated sample arm to reduce solids stratification bias in stormwater sampling. *Water Environment Research* 83(4):347-57.

Selbig, W.R., A. Cox, and R.T. Bannerman. 2012. Verification of a depth-integrated sample arm as a means to reduce solids stratification bias in urban stormwater sampling. *Journal of Environmental Monitoring* 14:1138-44.

Spooner, J., R.P. Maas, M.D. Smolen, and C.A. Jamieson. 1987. Increasing the Sensitivity of Nonpoint Source Control Monitoring Programs. In *Symposium On Monitoring, Modeling, and Mediating Water Quality*, American Water Resources Association, Bethesda, Maryland, pp. 243-257.

Spooner, J., S.L. Brichford, D.A. Dickey, R.P. Maas, M.D. Smolen, G.J. Ritter and E.G. Flaig. 1988. Determining the sensitivity of the water quality monitoring program in the Taylor Creek-Nubbin Slough, Florida project. *Lake and Reservoir Management* 4(2):113- 124.

Spooner, J., S.A. Dressing, and D.W. Meals. 2011. *Minimum Detectable Change Analysis*, Tech Notes #7, December 2011. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Stuntebeck, T.D. 1995. *Evaluating Barnyard Best Management Practices in Wisconsin Using Upstream-Downstream Monitoring*. Fact Sheet FS-221-95. U.S. Geological Survey, Madison, WI. Accessed February 5, 2016. http://pubs.usgs.gov/fs/1995/fs221-95/.

Stuntebeck, T.D., Komiskey, M.J., Owens, D.W., and Hall, D.W. 2008. *Methods of Data Collection, Sample Processing, and Data Analysis for Edge-of-Field, Streamgaging, Subsurface-Tile, and Meteorological Stations at Discovery Farms and Pioneer Farm in Wisconsin, 2001–7*. Open-File Report 2008–1015. U.S. Geological Survey, Reston, VA. Accessed April 22, 2016. http://pubs.usgs.gov/of/2008/1015/pdf/ofr2008-1015.pdf.

Teledyne Isco. 2013a. 6712 Full-Size Portable Sampler. Teledyne Isco, Lincoln, NE. Accessed February 5, 2016. http://www.isco.com/products/products3.asp?PL=201101010.

Teledyne Isco. 2013b. Portable Samplers. Teledyne Isco, Lincoln, NE. Accessed February 5, 2016. http://www.isco.com/products/products2.asp?PL=20110&image=Samplers3.jpg.

Turnipseed, D.P. and V.B. Sauer. 2010. *Discharge Measurements at Gaging Stations*. Techniques and Methods 3-A8. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://pubs.usgs.gov/tm/tm3-a8/.

USDA-NASS (U.S. Department of Agriculture-National Agricultural Statistics Service). 2012. The Census of Agriculture. U.S. Department of Agriculture, National Agricultural Statistics Service. Accessed February 8, 2016. http://www.agcensus.usda.gov/Publications/2012/.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2003. *National Water Quality Handbook*. 450–VI–NWQH. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC. Accessed February 5, 2016. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044775.pdf.

USEPA (U.S. Environmental Protection Agency). 1993a. *Evaluation of the experimental Rural Clean Water Program*. EPA-841-R-93-005. U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1993b. *Paired Watershed Study Design*. 841-F-93-009. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

USEPA (U.S. Environmental Protection Agency). 1997a. *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls*. EPA 841-B-96-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-guidance-determining-effectiveness-nonpoint.

USEPA (U.S. Environmental Protection Agency). 1997b. *Techniques for Tracking, Evaluating, and Reporting the Implementation of Nonpoint Source Control Measures – Agriculture*. EPA 841-B-97-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997c. *Techniques for Tracking, Evaluating, and Reporting the Implementation ofNonpoint Source Control Measures – Forestry*. EPA 841-B-97-009. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1998. *Lake and Reservoir Bioassessment and Biocriteria: Technical Guidance Document*. EPA 841-B-98-007. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 5, 2016.

USEPA (U.S. Environmental Protection Agency). 2001a. *EPA Requirements for Quality Assurance Project Plans, EPA QA/R-5*. EPA 240/B-01/003. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001b. *Techniques for Tracking, Evaluating, and Reporting the Implementation of Nonpoint Source Control Measures – Urban*. EPA 841-B-00-007. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2005a. *Microbial Source Tracking Guide Document*. EPA-600/R-05/064. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Accessed February 5, 2016. http://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryID=133523.

USEPA (U.S. Environmental Protection Agency). 2005b. *Standard Operating Procedure for Calibration and Field Measurement Procedures for the YSI Model 6-Series Sondes and Data Logger*. U.S. Environmental Protection Agency, EPA New England - Region 1, The Office of Environmental Measurement and Evaluation, North Chelmsford, MA. Accessed February 5, 2016.

USEPA (U.S. Environmental Protection Agency). 2008a. *Handbook for Developing Watershed Plans to Restore and Protect Our Waters.* EPA 841-B-08-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 10, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/handbook-developing-watershed-plans-restore-and-protect.

USEPA (U.S. Environmental Protection Agency). 2008b. *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*. SW-846. U.S. Environmental Protection Agency, Washington, DC. Accessed February 5, 2016. http://www3.epa.gov/epawaste/hazard/testmethods/sw846/online/.

USEPA (U.S. Environmental Protection Agency). 2010. *Survey of EPA and Other Federal Agency Scientific Data Management Policies andGuidance 2010*. U.S. Environmental Protection Agency, Office of Research and Development (ORD), Office of Science Information Management (OSIM), Washington, DC. Accessed April 24, 2016. http://cendievents.infointl.com/SDM_062910/docs/EPA_Policy_and_Guidance_SDM_Report.pdf.

USEPA (U.S. Environmental Protection Agency). 2011a. *National Lakes Assessment 2012 Field Operations Manual Version 1.0*. EPA 841-B-11-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 5, 2016.

USEPA (U.S. Environmental Protection Agency). 2011b. *Using Microbial Source Tracking to Support TMDL Development and Implementation*. U.S. Environmental Protection Agency Region 10, Watersheds Unit, Seattle, WA. Accessed February 5, 2016. http://www.epa.gov/region10/pdf/tmdl/mst_for_tmdls_guide_04_22_11.pdf.

USEPA (U.S. Environmental Protection Agency). 2011c. *Water Quality Standards: Protecting Human Health and Aquatic Life*. EPA/820/F-11/001. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 5, 2016. https://owpubauthor.epa.gov/scitech/swguidance/standards/upload/WQS_basic_factsheet.pdf.

USEPA (U.S. Environmental Protection Agency). 2012. *Water: Monitoring & Assessment 5.6 Phosphorus*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 5, 2016. https://archive.epa.gov/water/archive/web/html/vms56.html.

USEPA (U.S. Environmental Protection Agency). 2013. *Nonpoint Source Program and Grants Guidelines for States and Territories*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 5, 2016. http://www.epa.gov/sites/production/files/2015-09/documents/319-guidelines-fy14.pdf.

USGS (U.S. Geological Survey). n.d. *About the Hydrologic Instrumentation Facility*. U.S. Geological Survey, Stennis Space Center, MS. Accessed February 5, 2016. http://water.usgs.gov/hif/about/history/.

USGS (U.S. Geological Survey). 2007. Second photographic competition, national surface water conference and hydroacoustics workshop, St. Louis, Missouri, April 2 – 6, 2007. U.S. Geological Survey, Office of Surface Water. Accessed February 5, 2016. https://water.usgs.gov/osw/images/2007_photos/Hydroacoustics.html.

USGS (U.S. Geological Survey). 2008. *Guidelines for Writing Hydrologic Reports*. Fact Sheet FS-0217-95. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://water.usgs.gov/wid/FS_217-95/guide.html.

USGS (U.S. Geological Survey). variously dated. Handbooks for Water-Resources Investigations. Book 9 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://water.usgs.gov/owq/FieldManual/.

Vermont RCWP Coordinating Committee. 1991. *St. Albans Bay Rural Clean Water Program, Final Report, 1991*. University of Vermont, Vermont Water Resources Research Center, Burlington, VT.

Vrana, B., I.J. Allan, R. Greenwood, G.A. Mills, E. Dominiak, K. Svensson, J. Knutsson, and G. Morrison. 2005. Passive sampling techniques for monitoring pollutants in water. *Trends in Analytical Chemistry* 24(10):845-868.

Wagner, R.J., R.W. Boulger, Jr., C.J. Oblinger, and B.A. Smith. 2006. *Guidelines and Standard Procedures for Continuous Water-Quality Monitors—Station Operation, Record Computation, and Data Reporting*. Techniques and Methods 1–D3. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.usgs.gov/tm/2006/tm1D3/.

Wahl, K.L., W.O. Thomas, and R.M. Hirsch. 1995. *Stream-Gaging Program of the U.S. Geological Survey*. Circular 1123. U.S. Geological Survey, Reston, Virginia. Accessed February 5, 2016. http://pubs.usgs.gov/circ/circ1123/collection.html.

Waschbusch, R.J., W.R. Selbig, and R.T. Bannerman. 1999. *Sources of Phosphorus in Stormwater and Street Dirt from Two Urban Residential Basins in Madison, Wisconsin, 1994–95*. Water-Resources Investigations Report 99–4021. U.S. Geological Survey, Middleton, WI. Accessed February 5, 2016. http://wi.water.usgs.gov/pubs/WRIR-99-4021/WRIR-99-4021.pdf.

Whitney, L. 2007. *Upper Grande Ronde Basin Section 319 National Monitoring Program Project Summary Report*. DEQ07-LAB-0058-TR. Oregon Department of Environmental Quality, Laboratory Division, Portland, OR. Accessed February 5, 2016. http://www.deq.state.or.us/lab/techrpts/docs/DEQ07LAB0058TR.pdf.

Wilde, F.D., ed. 2004. Cleaning of Equipment for Water Sampling. Book 9, Chapter A3 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed April 22, 2016. https://water.usgs.gov/owq/FieldManual/chapter3/final508Chap3book.pdf.

Wilde, F.D. 2005. Preparations for Water Sampling. Book 9, Chapter A1 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.water.usgs.gov/twri9A1/.

Wilde, F.D., ed. 2006. Collection of Water Samples (Version 2.0). Book 9, Chapter A4 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed January 28, 2016. http://pubs.water.usgs.gov/twri9A4/.

Wilde, F.D., ed. variously dated. Field Measurements. Book 9, Chapter A6 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed April 22, 2016. http://water.usgs.gov/owq/FieldManual/Chapter6/Ch6_contents.html#Chapter A6.

Wilde, F.D., D.B. Radtke, J. Gibs, and R.T. Iwatsubo. 2009. Processing of Water Samples (Version 2.2). Book 9, Chapter A5 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 5, 2016. http://pubs.water.usgs.gov/twri9A5/.

Wilde, F.D., M.W. Sandstrom, and S.C. Skrobialowski. 2014. Selection of Equipment for Water Sampling (Ver. 3.1). Book 9, Chapter A2 in *Techniques of Water-Resources Investigations*. U.S.

Geological Survey, Reston, VA. Accessed February 5, 2016. http://water.usgs.gov/owq/FieldManual/Chapter2/Chapter2_V3-1.pdf.

Yamahara, K.M., S.P. Walters, and A.B. Boehm. 2009. Growth of Enterococci in unaltered, unseeded beach sands subject to tidal wetting. *Applied and Environmental Microbiology* 75(6):1517-1524.

This page intentionally left blank.

4 Biological Monitoring of Aquatic Communities

By J. B. Stribling, C.J. Millard, J.B. Harcum, and D.W. Meals

4.1 Overview

Biological monitoring uses surveys of resident biota (e.g., fish, benthic macroinvertebrates, periphyton, amphibians) to characterize the structure and function of the assemblage and assess the condition of a waterbody (USEPA 2013). The central purpose of assessing biological condition is to determine the degradation level of an ecosystem (or water body) and the cumulative effects of physical, chemical, hydrologic or biological stressors on aquatic biota. Resident biota reflect the integrated effects of variable magnitudes of these different stressors and stressor types, and thus provide an overall measure of environmental quality. As such, biological assessment is a crucial monitoring tool used by all 50 states and increasingly by tribes (USEPA 2002, 2011).

Biological assessments conducted by state water resource quality programs are often designed to assess regional or state-wide condition as well as conditions in smaller targeted watersheds or projects. During the past 20 years, these water quality agencies have invested resources to develop biological assessment capabilities within their state or tribe. These capabilities include using scientifically defensible and documented field and laboratory methods/protocols, establishing reference sites, and evaluating/developing metrics and indexes most suitable for assessing biological condition, and implementing them in routine monitoring programs or projects. “Index calibration” is a term encompassing data analyses leading to establishing scoring criteria and testing/selection of the suite of metrics making up multimetric indexes (MMI; further discussed below). To the extent practical, watershed projects or NGOs should use MMI that have been regionally calibrated based on broader datasets of known quality. Use of MMI previously established by local, state, or regional agencies requires that the same or similar methods be used for field sampling and laboratory processing for other streams and sites being evaluated relative to BMP or other issues. This approach of using accepted protocols and calibrated metrics and indexes, coupled with sufficient and appropriate quality control (QC) checks, improves defensibility of assessment results and increases confidence in natural resource management decision-making.

We recommend that biological monitoring be coupled with physical/chemical monitoring, which includes the stressor(s) of concern and focus on smaller watersheds (i.e., sub- hydrologic unit code [HUC] 12 watershed level) to document effectiveness of an individual BMP. Biological assessment is a useful tool for evaluating overall ecological condition because it integrates multiple stressors over time; however, it does not directly measure changes in a specific stressor (e.g., decreased sediment loading resulting from riparian buffers or point source discharge). The BMP could still be evaluated as effective in reducing stressor load, even though a positive biological response might not be detected. And, linking detectable changes in broad scale biological condition to a particular small scale BMP is more difficult as the potential increases for unknown and multiple stressors. Case Study 1 from Wisconsin illustrates the use of several types of indicators for monitoring BMP effectiveness (*Evaluating Effectiveness of Best Management Practices for Dairy Operations in the Otter Creek Watershed, Wisconsin*).

CASE STUDY 1: EVALUATING EFFECTIVENESS OF BEST MANAGEMENT PRACTICES FOR DAIRY OPERATIONS IN THE OTTER CREEK WATERSHED, WISCONSIN

- ✓ Dairy farms, barnyard runoff
- ✓ Manure storage, barnyard runoff control, stream bank protection, stream fencing and crossings, stabilization, buffer strips
- ✓ Fish assemblage monitoring
- ✓ Effectiveness evaluation

Located in eastern Wisconsin, the Otter Creek watershed is a 24.6 km^2 sub-basin tributary to the Sheboygan Creek watershed (Figure CS1-1), the latter ultimately feeding into Lake Michigan. Otter Creek has a total stream length of approximately 21 km and is a third-order watershed with a low to medium slope (2.5 - 5.4 m/km) throughout the area with a median wetted width of 4.2 m (Wang et al. 2006). Stream bottom is composed mostly of sand, silt, and clay, with riffle areas of medium gravel. Land use and land cover during the study period was dominated by row crop agriculture (62 percent), forests (14 percent), and to lesser degrees by grasslands and wetlands (10 percent and 6 percent, respectively).

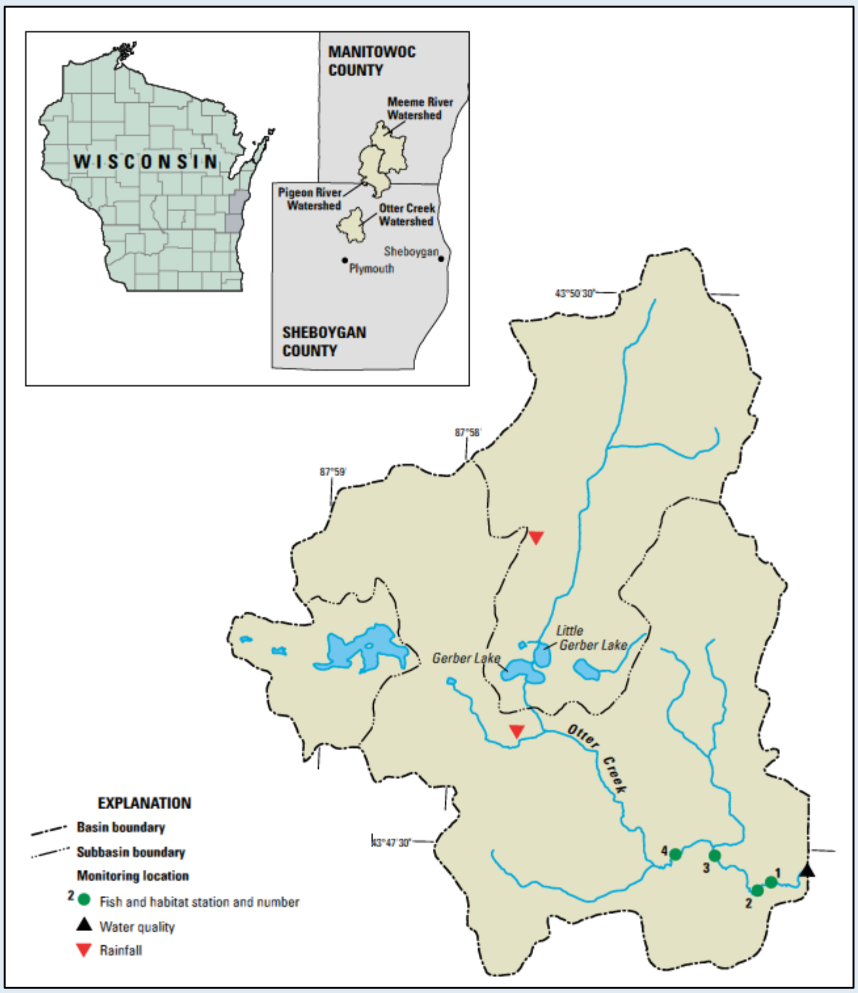


Figure CS1-1. Otter Creek watershed (Corsi et al. 2005)

Corsi et al. (2005) noted that the basin was home to 64 farms averaging approximately 0.5 km^2 in production. In the basin, there were eight barnyards associated with dairy operations, with an average herd size of 45 animals. The Wisconsin Department of Natural Resources reported in 1993 that the primary problems in the basin were direct livestock access to streams, resulting in elimination of bank vegetation, fish habitat degradation, accelerated and extensive bank erosion, and water temperature modification. Other acute problems associated with the livestock were barnyard manure runoff, upland delivery of sediment, and runoff from areas of winter manure-spreading. In addition to degradation of physical habitat, these sources and stressors led to organic and inorganic nutrient over-enrichment, and occasional, if not often, severe depletion of dissolved oxygen (Wang et al. 2006).

The BMPs designed and installed were focused on buffering, reducing, or otherwise eliminating the stressors and included a combination of animal waste management, stream bank protection, and upland management. Waste management practices were developed as facilities that provided improved manure storage, better control of barnyard runoff, and treatment of milkhouse wastewater. Four different types of BMPs were implemented for protection of >1,900 m of stream bank: fencing (2,800 m), stream crossings, grade stabilization, and buffer strips. Upland management included 635 ha (6.4 km^2) of nutrient management and sediment reduction of 250 metric tons/year achieved with changes in crop rotation, reduced tillage, critical area stabilization, grass waterways, and pasture management. Riparian and upland BMPs were implemented during 1993 and 1999.

Monitoring and Sampling Design

The goal of this monitoring project was to evaluate the effectiveness of multiple BMPs on the biological, habitat, and water chemistry characteristics in Otter Creek. Changes, if any, in instream habitat and biological characteristics relative to timing of BMP installation would be interpreted as BMP effectiveness at the watershed scale. Water chemistry evaluated using data from both base- and stormflow sampling would be used to establish changes in stressors impacting the fish assemblage.

Annual fish and habitat evaluations occurred relatively continuously from 1990 to 2002, providing for pre- (1990 to 1992), during (1993 to 1999), and post-installation (2000 to 2002) monitoring at four stations. Stations 1, 3, and 4 were located in areas where streambanks were trampled from cattle with free access while station 2 was located in a wooded riparian area (Figure CS1-1). Streambanks at station 1 and 3 were fenced in 1993 and 1996, respectively, while station 4 was not fenced (Corsi et al. 2005). There were four streams in similar watersheds that also flow to Lake Michigan that were also monitored. Two watersheds, the Meeme and Pigeon Rivers, were monitored as two control watersheds in a paired-watershed monitoring design with a single sampling station each (see section 2.4.2.3). Neshota and Trout Creeks (tributaries to the West River and Duck Creek, respectively) were also monitored with single stations.

Fish sampling and physical habitat evaluations were generally conducted each year at four locations in the lower part of the watershed, during roughly a six-week period spanning August and September. Gradually increasing funding resulted in more complete sampling over time, and thus, the record for some of the sites is better than others. In Otter Creek, 1 station was sampled in 1990, 2 in 1991, and all 4 from 1992-2002. Fish were sampled from a reach length of 35 times the median wetted channel width, which was an actual range from 105 to 234 m. Fish sampling used a single-tow barge electrofisher to cover the sampling reach in a single-pass/no block net approach. All fish captured were identified to species, counted, weighed (total, by species), and released unharmed back to the stream. Data were summarized by reach and sampling event as species lists, as proportions of individuals in the samples omnivores, insectivores, carnivores, simple lithophils, relative stressor tolerance, and as an IBI. The IBI had been previously calibrated for Wisconsin warm-water streams. Habitat features recorded for each sampling event at each location included turbidity, dissolved oxygen, specific conductance, and flow, along with 30 habitat variables, encompassing channel morphology, bottom substrates, cover for fish, bank conditions, riparian vegetation, and land use. These data were summarize by calculating mean and variance for each reach and sampling event, and used to calculate width/depth ratio and total habitat quality index.

Monitoring also included a single-watershed, before/after study of water chemistry. A stream gage installed in 1990 at the base of Otter Creek just upstream of its confluence with the Sheboygan River was equipped to continuously record data on stage and for automatic activation to collect and refrigerate samples with every 0.2-ft. increase and 0.3-ft decrease in stage. Water chemistry constituents measured from these grab samples included TSS, TP, dissolved ammonia nitrogen (NH_3-N), 5-day biochemical oxygen demand (BOD_5), and fecal coliforms. Fixed interval grab samples, analyzed for the same constituents, were also taken throughout the pre- and post-BMP study period. Precipitation was measured at three locations in the watershed, one near the gaging station, and two others further up in the headwaters.

Results

For both fish and habitat data, analysis of covariance (ANCOVA) (see section 7.8) was used to relate the fish assessments and habitat variables and conditions of the Otter Creek sites to the other watersheds over time. The statistical significance of changes was evaluated using the Wilcoxon rank-sum nonparametric test at 95 percent confidence. Changes were measured from different subsets of base- and stormflows, including some investigation of seasonal effects (vegetative vs. nonvegetative).

Post-BMP implementation baseflow samples showed statistically significant lower concentrations of TSS and BOD_5, higher fecal coliform, and no differences in dissolved NH_3-N and TP for the combined seasons (vegetative and nonvegetative), whereas samples from nonvegetative periods exhibited lower BOD_5 and no changes in the other four analytes. TSS concentration was lower during the vegetative season. From several different analyses, BOD_5 was demonstrated to have decreased substantially (by 45 percent median concentrations) in baseflow from the pre- to post-BMP periods. TSS was also found to be lower in post-BMP samples relative to those of pre-BMP, however, only actually evident in the full dataset, and not in the smaller data subsets (i.e., vegetative vs. nonvegetative). Dissolved NH_3-N concentrations decreased from pre- to post-BMP baseflow for the full dataset and nonvegetative season, but there were no differences for the vegetative season. Analyses showed none of these difference to be statistically significant ($p<0.05$), even with a measured 32 percent decrease in dissolved NH_3-N. There was a significant ($p<0.05$) increase in median fecal coliform concentrations of 260 percent for pre- to post-samples, demonstrated in both the full and stratified (vegetative and nonvegetative) datasets.

Analyses using pre- and post-BMP regressions across all monitored storms showed reductions in stormflow constituents, whether combined or stratified by vegetative vs. nonvegetative. For TSS, the pre/post reduction was 58 percent for the combined nonvegetative and vegetative dataset, 41 percent for the vegetative season, and 73% for the nonvegetative season. TP predictions for reductions were 48 percent, 34 percent , and 61 percent , while for dissolved NH_3-N they were 41 percent , 40 percent , and 42 percent , respectively.

Although considerable variation in stream physical habitat was observed in the reference streams, there was clear post-BMP improvement in Otter Creek (Corsi et al. 2005). In particular, there were significant increases ($p<0.05$) in percent cobble and percent gravel, and significant decreases in percent embeddedness, sand, and silt for stations 1, 2, and 3 (Wang et al. 2006). These changes reflected the natural woody buffer (station 2) or exclusion fences (stations 1 and 3). Stream width-to-depth (W/D) ratio and percent bank erosion decreased significantly ($p<0.05$) for stations

1 and 3 where exclusion fences were installed. Other habitat variables that showed significant improvement for one or more of the sample reaches were sediment depth and percent riffles. ANCOVA showed that overall habitat quality improved in Otter Creek only for those reaches that had a natural riparian buffer or had exclusion fences installed. For stations 1, 2, and 3, most of the habitat variables (8 out of 10) improved significantly ($p<0.05$) with BMP implementation, with substrate embeddedness, bank erosion, sediment depth, silt and sand substrates, and W/D ratio decreasing and conductivity, gravel and rubble substrates, and overall habitat scores increasing.

Cumulative fish species were similar for each of the Otter Creek locations, the two control watersheds (Meeme and Pigeon Rivers), and the two additional watersheds (Neshota and Trout Creeks) (33, 29, and 31 species, respectively), and six species dominated in all streams. From pre- to post-BMP installation, fish abundance decreased by 79 percent in Otter Creek and by 65 percent in the control watersheds. When sampling years were considered in combination, the percentages of stressor-tolerant fishes and omnivores *increased* in Otter Creek, while stressor tolerant fishes *decreased* and insectivores increased in control watersheds. In neither Otter Creek nor the control watersheds was there an obvious directional change in percentage of stressor intolerant fishes or IBI scores. ANCOVA showed significance ($p<0.05$) in a decrease of abundance and an increase in percent omnivores; and, for one of the Otter Creek sample locations where riparian pasture was dominant, there were significant decreases in number of fish species and percentage of darter individuals.

Characteristics of physical habitat and water chemistry improved, apparently substantially so, as a result of the several BMPs that were implemented in Otter Creek. However, there was no significant improvement in biological condition as measured by the fish community. This status of the fish assemblage even with habitat and water chemistry suitable for supporting much higher quality led the authors to conclude that there are some broader scale watershed or regional factors preventing more evident, positive changes in the fish community. Wang et al. (2006) speculate that pollution intolerant species might have been largely eliminated in the larger watershed and thus not able to colonize Otter Creek.

Literature

Corsi, S.R., J.F. Walker, L. Wang, J.A. Horwatich, and R.T. Bannerman. 2005. *Effects of Best-Management Practices in Otter Creek in the Sheboygan River Priority Watershed, Wisconsin, 1990-2002*. Scientific Investigations Report 2005–5009. U.S. Geological Survey (in cooperation with the Wisconsin Department of Natural Resources). Accessed February 9, 2016. http://pubs.usgs.gov/sir/2005/5009/.

Wang, L., J. Lyons, and P. Kanehl. 2006. Habitat and fish responses to multiple agricultural best management practices in a warm water stream. *Journal of the American Water Resources Association* 42(4):1047-1062.

Whether indicators being monitored are based on biological, physical, chemical, or hydrologic data, or targeted vs. probability-based site selection, questions required for monitoring designs are similar. Targeted sampling designs are needed to answer questions on BMP effectiveness at a particular location typically using a paired (treatment/control) watershed or upstream/downstream approach. Alternatively, probability-based designs could also be used to evaluate the difference in biological condition from randomly selected sample locations where a particular BMP (treatment) was implemented versus a sample of non-treated (i.e., control) areas. Unlike traditional physical/chemical monitoring programs that take measurements and collect samples frequently throughout a year, most biological monitoring programs take samples once (or, rarely, twice) during specific periods annually because biota do not usually vary dramatically in response to individual transient events. Biological data are converted into indicator values (e.g., such as an MMI, also known as an Index of Biological Integrity [IBI]). As a result of many of these protocol characteristics, variability that could otherwise cause data analysis issues related to seasonality, autocorrelation, and non-normality is controlled (Fore and Yoder 2003). Disaggregation of indexes to individual metrics or even taxa can allow for more detailed interpretation of subtle biological changes relative to BMP effectiveness. However, that level of interpretation requires access to a biologist with appropriate training and experience.

The concept of reference conditions has begun to emerge in the analysis of traditional physical/chemical monitoring data. Reference conditions are those observed in unimpaired or minimally impaired waterbodies in the region of interest and are used as a benchmark against which to measure changes. As previously stated, many states and tribes have established regional reference sites in support of their ongoing assessment programs, and it may be useful and efficient to use them in a watershed monitoring program.

This chapter presents basic information about biological monitoring and its applicability to NPS and watershed projects. Section 4.2 introduces the different types of biological monitoring and common terms used in this chapter with an emphasis on benthic macroinvertebrates and periphyton. An overview of monitoring design and assessment protocols is provided in sections 4.3 and 4.4, respectively.

4.2 Background

Natural biological communities are often diverse, comprising multiple species at various trophic levels (e.g., primary producers, secondary producers, carnivores) and varying degrees of sensitivity to environmental changes. Adverse impacts from NPS pollution or other stressors, such as habitat alteration, can reduce the diversity of the biota, change the relative abundances of different taxa, or alter the trophic structure. Biological surveys of resident biota particularly sensitive to stressors, such as fish, benthic macroinvertebrates, or periphyton, take advantage of this sensitivity as a means to evaluate the collective influence of the stressors on the biota (Cummins 1994).

The central purpose of biological assessment is to characterize the condition of resident biota relative to cumulative effects of stressors as the principal indicator of stream (or water body) condition. Monitoring changes in biological condition can be particularly useful for determining the impacts, depending on the frequency and duration of exposure, of episodic stresses (e.g., spills, dumping, treatment plant malfunctions), toxic nonpoint source pollution (e.g., agricultural pesticides), cumulative pollution (i.e., multiple impacts over time or continuous low level stress), non-toxic mechanisms of impact (e.g., trophic structure changes due to nutrient enrichment), or other impacts that periodic chemical sampling might not detect (USEPA 2011).

4.2.1 Types of Biological Monitoring

Different kinds of biological monitoring are defined by the particular indicators being used and the spatial and temporal scales of questions being addressed. The most common biological indicator groups (or assemblages) used for routine biological monitoring and assessment of freshwater ecosystems in North America are benthic macroinvertebrates, fish, and periphyton (algae). Additionally, programs will often collect data on physical, chemical, and hydrologic features of the systems being evaluated to provide information on environmental factors potentially affecting the biota. The scales of the questions being addressed drive (or should drive) the number and distribution of locations, and the frequency and duration of sampling, that is, the monitoring and sampling design. Further, efforts to control or better understand the natural variability of the biota have led to different kinds of specific field and laboratory methods for taking the samples or performing measurements. Controlling variability of indicator values means that assessment data are of known quality, and leads to improved confidence in and defensibility of management decisions.

The purpose of this document is to provide users the information necessary for applying existing biological indicators (metrics and indexes) to their ecosystems or water bodies of concern. It is not intended to give a comprehensive review of all methods and procedures used for biological assessments. By existing indicators, we mean those MMI or observed/expected (O/E) ratios that have been appropriately calibrated for the region and water body type. The remainder of this section, and those that follow, provide an overview of selected methods.

4.2.1.1 Benthic Macroinvertebrates

Stream environments contain a variety of macro- and microhabitat types including pools, riffles, and runs of various substrate types; snags; and macrophyte beds (Hawkins et al. 1993). Relatively distinct assemblages of benthic macroinvertebrates inhabit various habitats, and it is unlikely that most sampling programs would have the time and resources to sample all habitat types. Decisions on the habitats selected for sampling should be made with consideration of the regional characteristics of the streams. For instance, higher-gradient streams (slope roughly >1:1) often have riffle habitat with hard bottom substrate (cobble or gravel riffles) that serve as excellent habitat for a diversity macroinvertebrates, whereas low-gradient coastal streams lack riffles but have a very productive habitat including woody debris snags (Figure 4-1), leaf packs, undercut banks, and shorezone vegetation. These two different stream types could be sampled with different methods, during different times of the year, or with different biological index periods, but consideration should be given to protocols already being applied in the water body or region. Many routine biological monitoring programs use multi-habitat composite sampling techniques, e. g., the 20-jab method, because the intent is to characterize the biota of the stream reach.

4.2.1.2 Fish

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

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

Figure 4-1. Using a D-frame net to sample woody snag habitat for stream benthic macroinvertebrates

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

For biological assessments of the entire assemblage, the gear and methods used should ensure that a representative sample is collected.

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

4.2.1.2.1 *Length, Weight, and Age Measurements*

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

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

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

4.2.1.2.2 *Fish External Anomalies*

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

4.2.1.3 Periphyton

Periphyton is an assemblage of organisms that adhere to and form a surface coating on stones, plants, and other submerged objects in aquatic habitats. These can take the form of soft algae, algal or filamentous mats, or diatoms. The advantages of using the periphyton assemblage as an indicator include:

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

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

In Montana, Bahls (1993) developed metrics for diatoms that included a diversity index, a pollution index, a similarity index, and a siltation index. Three other metrics—dominant phylum, indicator taxa, and number of genera—were used for soft-bodied algae to support the diatom assessment. Further study and refinement of these metrics has led to the development of diatom biocriteria (Teply and Bahls 2005) and an evaluation of their discrimination efficiency (Teply and Bahls 2007). Periphyton data collections currently comprise routine monitoring of the Statewide Monitoring Network (SWM) and support the determination of designated aquatic life uses.

4.2.2 Linkages to Habitat

The quality of the physical habitat is an important factor in determining the structure of benthic macroinvertebrate, fish, and periphyton assemblages (Southwood 1977). The physical features of a habitat include substrate type, quantity and quality of organic debris (leaf litter, woody materials) in the waterbody, exposure to sunlight, flow regime, and type and extent of aquatic and riparian vegetation. Even though there might not be sharp boundaries between habitat features in a stream, such as riffles and pools, the biota inhabiting each feature are often taxonomically and biologically distinct (Hawkins et al. 1993). Habitat quality is assessed during biological assessment.

Habitat features are generally associated with biological diversity, (Southwood 1977, Raven et al. 1998) and their quality largely determines both the structure and function of benthic macroinvertebrate and fish assemblages. Habitat quality refers to the extent to which a suitable environment for a healthy biota exists. It encompasses five factors: habitat structure, flow regime, energy source, biotic interactions (such as invasive species and disease), and chemical water quality. *Habitat structure* refers to the physical characteristics of stream environments. It comprises channel morphology (width, depth, sinuosity), floodplain shape and size, channel gradient, instream cover (boulders, woody debris), substrate types and diversity, riparian vegetation and canopy cover, and bank stability. *Flow regime* is defined by the patterns of velocity and volume of water moving through a stream over time. *Energy* enters streams as the input of nutrients in runoff or ground water, as organic debris (e.g., leaves) falling into streams, or from photosynthesis by aquatic plants and algae. *Biotic interactions* include issues such as invasive species and disease while *water quality* includes a range of issues from nutrient sources to toxicants.

These factors are interrelated and make stream environments naturally heterogeneous. Habitat structural features that determine the assemblages of macroinvertebrates can differ greatly within small areas—or microhabitats—or in short stretches of a stream. For instance, woody debris in a stream affects the flow in the immediate area, provides a source of energy, and offers cover to aquatic organisms. Curvature (sinuosity) in a stream affects currents and thereby deposition of sediment on the inner and outer banks, in turn influencing the character of the streambed. Rocks and boulders create turbulence, which affects dissolved oxygen levels. Shallow stream reaches where water velocity is relatively high (riffles) provide areas of high dissolved oxygen and a gravel or cobble bottom ; deep, wide portions (pools) are areas of lowered velocity where material can settle out of the water, streambeds are composed of soft sediments, and increased decomposition occurs.

Aspects of habitat structure are separated into primary, secondary, and tertiary groupings corresponding to their influence on small-, medium-, and large-scale aquatic habitat features (Barbour et al. 1999). The status or condition of each aspect of habitat is characterized along a continuum from optimal to poor. An optimal condition would be one that is in a natural state. A less than optimal condition, but one that satisfies most expectations, is suboptimal. Slightly worse is a marginal condition, where degradation is for the most part moderate, but is severe in some instances. Severe degradation is characterized as a poor condition. Habitat assessment field data sheets (see Plafkin et al. 1989) provide narrative descriptions of the condition categories for each parameter. Habitat can be assessed visually, and a number of biological assessment methods incorporate assessments of the surrounding habitat (Ball 1982, Barbour et al. 1999, OEPA 1987, Platts et al. 1983).

A clear distinction between impacts due to watershed (i.e., large-scale habitat), stream habitat, and water quality degradation is often not possible, so it is difficult to determine with certainty the extent to which biological condition will improve with specific improvements in either habitat or water quality. Generally, if the biological community condition varies directly with the habitat quality, then water quality is not the principal factor affecting the biota. The opposite is considered true if the biological condition is degraded relative to the potential of its habitat. Some measures of biological condition in comparison to habitat condition can be used as indicators of organic enrichment or energy source alteration (Barbour and Stribling 1991).

Biological and habitat data collected from numerous sites in good or near-natural condition can be used to determine the type of biological community that should be found in a particular aquatic habitat. This natural condition has been referred to as reflecting biological integrity, defined by Karr and Dudley (1981) as “the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region.” Highly detailed biological assessments are comparisons of biological conditions at a test site to the expected natural community and are thus a measure of the degree to which a site supports (or does not support) its “ideal” or potential biological community (Gibson et al. 1996). Other types of biological assessment involve comparisons of impacted sites to control sites, the latter being sites that are similar to monitored sites but are not affected by the stresses that affect the monitored sites (Skalski and McKenzie 1982). Knowledge of the natural condition is still valuable for accurate data interpretation when control sites are used (Cowie et al. 1991).

4.2.3 Limitations of Biological Assessments

Although biological assessment is useful for detecting and prioritizing severity of aquatic ecosystem degradation, it does not necessarily provide a direct linkage to or measurement of specific stressors. Thus, it usually does not provide definitive information about the cause of observed water body degradation, i.e., the specific pollutants or their sources. Monitoring of chemical water quality and toxicity may also be necessary to design appropriate pollution control programs.

Prior to routine application in monitoring and assessment, it is necessary to calibrate biological indicators for the water body types, geographic and ecological regions, and to structure sampling and analysis programs to appropriate address management objectives. Thus, establishing a biological assessment and monitoring program can require a significant investment of time, staff, and money. However, the majority of these are one-time, up-front investments dedicated to the establishment of reference conditions, standard operating procedures, and a programmatic quality assurance and quality control plan.

Finally, there can often be a lag between the time at which a toxic contaminant or some other stressor is introduced into a water body and a detectable biological response. Consequently, biological monitoring may not always be appropriate for determining system response due to short-term stresses, such as storms. Similarly, there is often a lag time in the improvement of biological condition following habitat restoration or pollution abatement. The extent of this lag time is difficult to predict; but, it should be recognized and anticipated. Other factors also determine the rate at which a biological community recovers, e.g., the availability of nearby populations of species for recolonization following pollution mitigation and the extent or magnitude of ecological damage done during the period of perturbation (Richards and Minshall 1992). Both the possibility of the lack of detectable recovery from perturbation and unpredictable lag times before improvement is noticeable have obvious implications for the applicability of biological monitoring to some NPS pollution monitoring objectives. Table 4-1 summarizes the strengths and limitations of the biomonitoring approach.

Table 4-1. General strengths and limitations of biological monitoring and assessment approaches

Strengths	Limitations
Properly developed methods, metrics, and reference conditions (i. e., calibration) provide a means to assess the ecological condition of a waterbody	Development of regional methods, metrics, and reference conditions takes considerable effort and an organized and well-thought-out design
Biological assessment data can be interpreted based on regional reference conditions where reference sites for the immediate area being monitored are not available	Rigorous bioassessment can be expensive and requires a high level of training and expertise to implement
Bioassessments using two or more organism groups at different trophic levels provide improved confidence in interpretation of assessment results	Basic biological assessment information does not provide information on specific cause-effect relationships
Biological condition is an indicator of cumulative effects from both short- and long-term stressors	There may be a lag time between pollution abatement or BMP installation and community recovery, so monitoring over time is required for trend detection
	Biological assessment does not always distinguish between the effects of different stressors in a system impacted by more than one stressor

4.2.4 Reference Sites and Conditions

Biological condition assessments often compare metrics of observed assemblages to the expectations for those assemblages in the absence of environmental disturbance. Those expectations are based on samples taken from water bodies that are minimally degraded, have low stressor loads, or are considered “best available”. They constitute the reference condition, which is often derived from observations collected from reference sites with minimal levels of disturbance (Hughes 1995, Stoddard et al. 2006, Gibson et al. 1996 [see Figure 4-2]).

A reference condition is a composite characterization of the natural biological condition in multiple ecologically homogeneous reference sites. The overall goal of establishing a reference condition is to describe the natural potential of the biota in the context of natural variation. Minimally disturbed reference sites are those with habitats assumed to fully support a natural biota. The greater the difference is between indicator characteristics of reference and monitored samples, the more disturbed the monitored samples are considered. The disturbance responsible for the difference might be a habitat change, pollution, or some other stress. Another approach is describing expectations relative to a complete gradient of disturbance conditions, and thus, interpreting biological conditions relative to that gradient, the biological condition gradient (BCG) (Davies and Jackson 2006 [see section 4.4.3]). Site classification is integral to the reference condition concept (Gerritsen et al. 2000, Hawkins et al. 2000a). Site classification accounts for natural biological variability prior to evaluating potential effects of human disturbance. The objective of classification is to group water bodies with similar reference biological

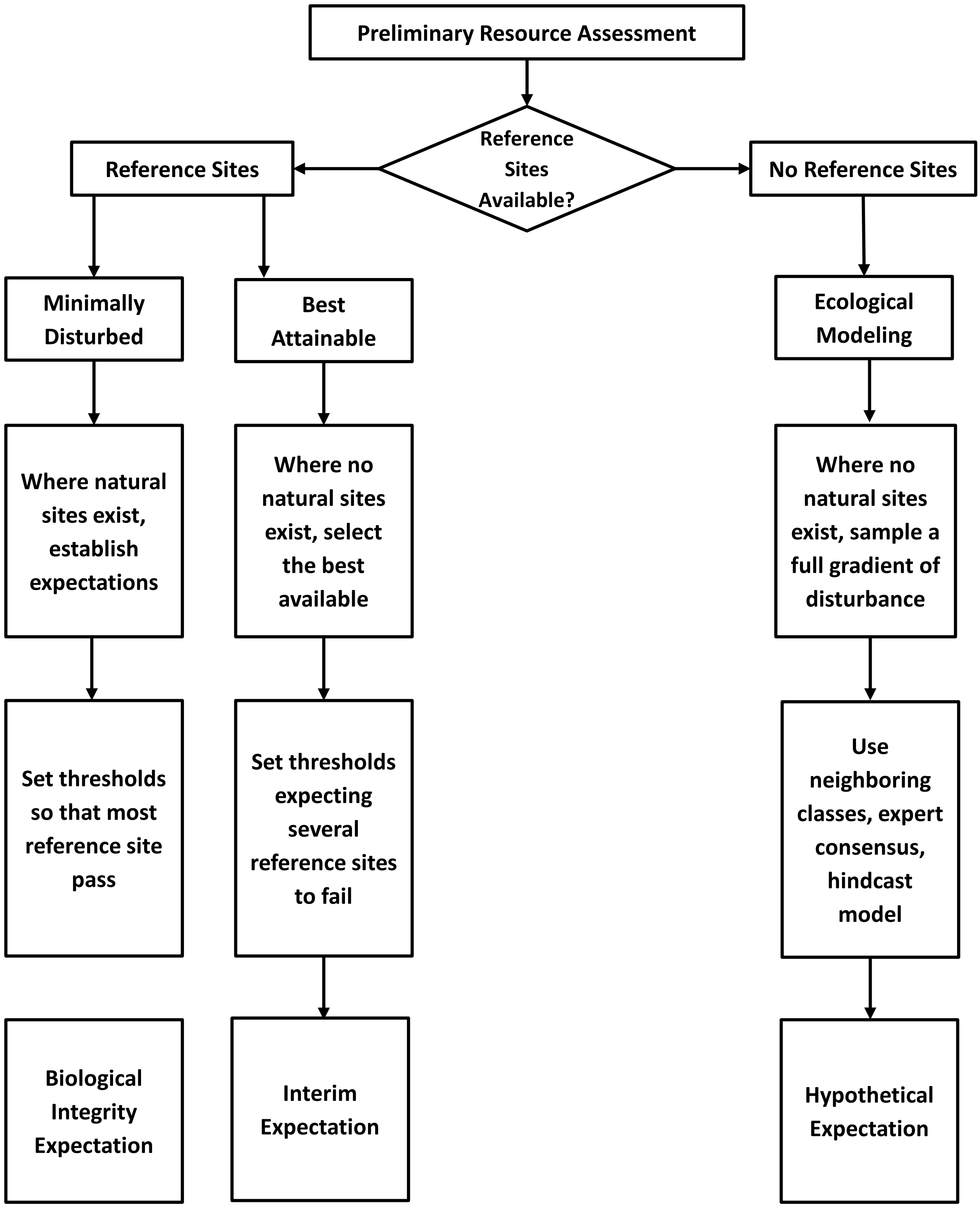


Figure 4-2. Approach to establishing reference conditions (after Gibson et al. 1996)

characteristics together, allowing formulation of precise indicator expectations within each group (Hughes et al. 1986). Site classification can be categorical or continuous. Categorical classification groups distinct site types with similar biological and environmental characteristics. Continuous classification recognizes the gradation of site types. Categorical classification is generally used for MMIs, and continuous classification is often associated with predictive models of observed and expected taxa. Reference sites can be defined identically for either indicator type. Developing reference condition/expectations and site classification are both critical components of index calibration and will have already been accomplished by regional environmental resource management agencies, universities, or other monitoring entities.

In the sense of biological monitoring and assessment, reference sites are traditionally thought of as having been used in the index calibration process, and thus, sampling of them for a particular site specific assessment is not required. In addition to the calibrated metric and index scoring framework used in assessments, individual control sites could add potentially useful information for specific stressor inputs to streams or other water bodies. Control sites could be positioned to be upstream/downstream, or pre- and post-implementation, of input points or zones from the land use/land cover of concern, including BMP or other stressor source control activities. In this situation, it would be necessary to have multiple reaches sampled to represent different assessment areas.

4.3 Biomonitoring Program Design

As described in chapter 2, different monitoring designs might be applicable depending on the objectives of individual monitoring projects with consequent implications for site selection process, number of sites sampled, what features are monitored, and time and frequency of sampling. The sampling design used in NPS biological monitoring might consist of either targeted or probabilistic designs. Targeted monitoring designs are normally chosen for site-specific objectives such as whether biological impairment exists at a given site or whether impairment has been reduced by a watershed project. See section 2.4 for a more detailed discussion of targeted site selection design; however in brief, site locations are selected based on the purpose of the project. Targets might include vulnerable areas with known or suspected perturbations (stressors), planned point source controls, or waterbodies in areas treated with BMPs. NPDES permits, urban stormwater sites, timber harvest areas, rangeland, row crop farming, and construction sites are examples of known stressor sites. Upstream/downstream sampling stations, before-and-after site alterations, or recovery zones (sampling at established distances from sources) are types of sampling locations for known stressor sites. Ecologically sensitive sites that may or may not be affected by stressors and reference sites with minimal disturbance might also be chosen as with targeted monitoring.

While targeted monitoring designs are often used for site-specific purposes, the results from those studies cannot be extended to other sites in the region (Fore and Yoder 2003). Alternatively, probabilistic designs are useful for providing unbiased assessments of conditions across a water body or large geographic area. In a probabilistic sampling program, the entity about which inferences are made is the population or target population and consists of population units. The sample population is the set of population units that are measured. As an example, in a watershed impacted by nonpoint sources, the target population could be the biological condition of all 1st-, 2nd-, and 3rd- order streams. Benthic macroinvertebrates, selected water chemistry, and physical habitat quality are then collected at randomly selected sites drawn from the population of 1st, 2nd, and 3rd order streams. By sampling and statistically evaluating randomly selected population units, inferences can be made about the entire waterbody. The advantages and disadvantages of targeted and probabilistic sampling are summarized in Table 4-2. In some cases, a monitoring program may have a combination of targeted and probabilistic sites.

Table 4-2. Comparison of probability-based and targeted monitoring designs

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

For probability-based designs, simple random sampling is not optimal. It can produce clusters of sampling sites that might not be representative of the larger scale area of interest (e.g., Hurlbert 1984). Therefore, some sort of stratification is preferred for ensuring a dispersed distribution of site locations (Stevens and Olsen 2004). The approach of stratifying the target population and then randomly sampling, referred to as *stratified-random sampling*, is often more efficient than simple random sampling. This is because a target population is recognized to consist of groups that each have internal homogeneity (relative to other groups), and stratifying the target population will tend to minimize within-group variance and maximize among-group variance (Gilbert 1987, Fore and Yoder 2003). Case Study 2 from Maryland provides an example of secondary uses of data and assessment results from stratified-random sampling. Another approach for randomly selecting sites, used by EPA's national aquatic resource surveys, that allows for random sampling while ensuring representation from all relevant site types and locations is the unequal probability Generalized Random Tessellation Stratified (GRTS) spatially-balanced survey design (Stevens and Olsen 2004).

CASE STUDY 2: EFFECTIVENESS OF STORMWATER PONDS IN ENHANCING INSTREAM BIOLOGICAL CONDITIONS IN MARYLAND URBAN WATERSHEDS

- ✓ Urbanized watersheds, elevated flashiness, channel instability, habitat degradation
- ✓ Stormwater retention ponds
- ✓ Multiple watershed monitoring
- ✓ Benthic macroinvertebrates, physical habitat
- ✓ Random site selection, post-stratification
- ✓ Assessing effectiveness

Prince George's County, Maryland covers 1,290 km^2 of the mid-Atlantic Coastal Plain, and is the suburban jurisdiction immediately east of the District of Columbia (Figure CS2-1). It has more than 965 kilometers of stream channels draining to the Patuxent River on the east, Anacostia River in the west and northwest, Potomac River on the southwest, and Mattawoman Creek in the south. The northeast corridor of the U. S. has undergone heavy and nearly continuous urbanization, resulting in an increasing percentage of landscape covered with impervious surfaces. Increased imperviousness reduces the capacity for land to absorb rainfall and causes storm water to be instantly converted to runoff. This can result in accelerated erosion and severe channel instability in watersheds that contain large proportions of impervious cover. Surface runoff can also transport solid trash and contaminants from parking lot asphalt, sidewalks, and rooftops. These stressors can combine to cause substantial harm to aquatic communities. The primary objective for this project was to assess the effectiveness of stormwater detention/retention ponds in protecting and enhancing instream biological condition.

Stream and watershed assessments have been conducted in Prince George's County for nearly two decades and have fully documented spatial patterns of aquatic biological conditions. The initial round of county-wide assessments (1999 to 2003) showed that more than half of the stream length (52.5 percent) was biologically degraded, with the majority of impaired stream reaches located in the western part of the county near major regional transportation thoroughfares. Higher quality streams predominantly rated as "fair" and "good" were found to be in the eastern part of the county along the drainage to the Patuxent River, and in the south, near the border with Charles County

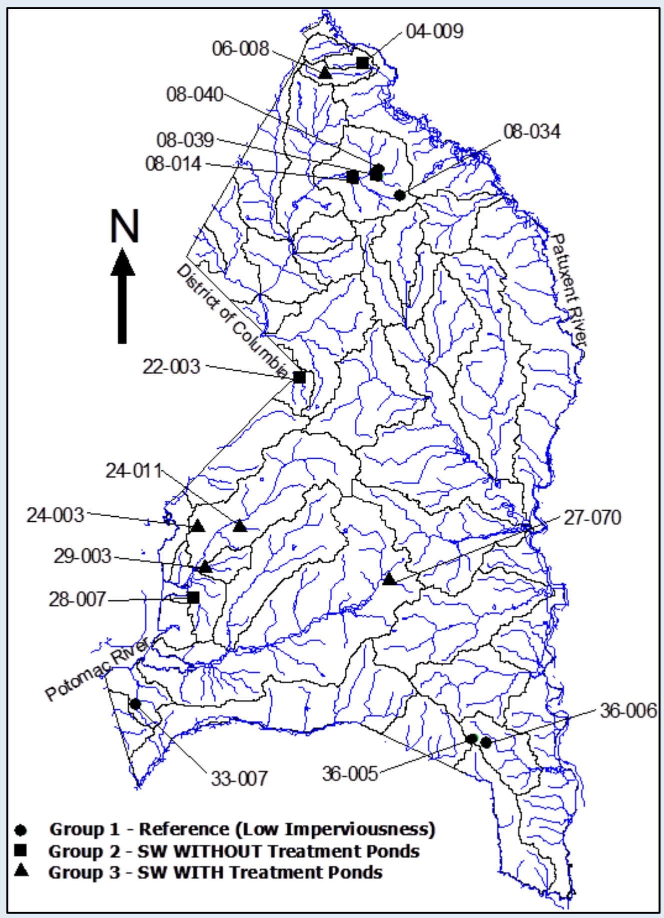


Figure CS2-1. Prince George's County, Maryland. Distribution of sample locations used as part of this analysis.

and Mattawoman Creek. This pattern was expected because greater development intensity typically follows transportation corridors with the attendant increases in impervious surfaces.

One of the principal BMPs used for stormwater management in the 1990s was stormwater detention/retention ponds (SWPs). Many managers designing and installing SWPs had multiple objectives including, but not limited to, collecting and slowing runoff from impervious surfaces, allowing suspended particulates to settle out, collecting trash solids, and providing features that helped foster lower water temperatures and elevated levels of dissolved oxygen. BMP siting decisions were often based on stream problems brought to the attention of county managers by the public, citizens' environmental groups, or stream monitoring data. Thus, it was not always straightforward to identify specific management objectives or goals associated with an individual SWP. Nonetheless, the authors of this study (Stribling et al. 2001) assumed that improving instream biological condition was among the objectives for SWP implementation.

Monitoring and Sampling Design

Routine county-wide monitoring performed by the Prince George's County Department of Environmental Resources (PGDER) to assess stream and watershed biological conditions is based on a long-term, probability-based, rotating basin plan. Stratified (by wadeable stream order) random sampling was used to select sampling sites for this study, with the number of potential sites for each stream order within each of the 41 subwatersheds set proportional to the number of stream km in each order (map scale = 1:100,000). There are 50 to 60 sites sampled per year for biological (benthic macroinvertebrates), selected water chemistry, and physical habitat quality variables. Benthic macroinvertebrates are collected over 100-m channel reaches by making 20 1-m linear sweeps (jabs) with a 500 μ mesh D-frame net distributed among different habitat types (such as snags, leaf packs, vegetated/undercut banks, bottom, riffle/cobble) in proportion to their frequency of occurrence at each site. To minimize the effects of seasonal variability, all sampling occurs during the Maryland Department of Natural Resources' Biological Stream Survey (MBSS)-specified index period, which is March 01 to April 30. Ten percent of the sampling segments are randomly selected for replicate reaches, which provides information necessary for quality control and for calculating field sampling precision. Data on physical habitat quality and water chemistry (pH, conductivity, water temperature, and dissolved oxygen) are collected at each site for their potential in explaining biological condition. Benthic data (number of taxa, number of individuals of each, per sample) are used to calculate the benthic-index of biological integrity (B-IBI) developed by the MBSS. In general, the B-IBI and the physical habitat quality index, as applied by the PGDER program, have 90 percent confidence intervals of ± 0.67 points on a 5-point scale, and ±6.7 points on a 200-point scale, respectively.

The purpose of this study (Stribling et al. 2001) was to determine effectiveness of stormwater detention ponds in protecting and/or enhancing in-stream biological condition. The physical habitat and biological data were segregated into the following three treatment groups and directly compared using percentile distributions of measurement values:

Group 1 Streams with minimal stormwater stressors (*0-5 percent impervious surface*),

Group 2 Streams with substantial stormwater stressors (>12 percent impervious surface) and *without* SWP, and

Group 3 Streams with stormwater stressors (>12 percent impervious surface) and *with* SWP.

Using GIS analysis, upstream drainage areas were delineated for all sites sampled in 2000, and land use/land cover (LULC) determined for each, including calculation of impervious surface. From these data and the existing dataset, all sites were screened to define a set of sites that would reasonably represent each of the groups (Table CS2-1), resulting in five sites per treatment group. Group 1 was represented by sites with drainage areas (DAs) ranging from 31 to 585 ha and imperviousness of 1.0 to 4.7 percent. Group 2 had DAs ranging from 248 – 634 ha and imperviousness of 17 to 34 percent, and Group 3 had DAs from 76 to 568 ha and 11.2 to 34.9 percent imperviousness. The ages (time elapsed since installation) of the SWPs of Group 3 ranged from 7 to 13 years, so all SWPs should have had ample opportunity to "mature" through multiple growing seasons. SWP functionality was not assessed as part of this study.

Table CS2-1. Sampling sites in Prince George's County including drainage area, percent imperviousness in the drainage, and proportions of different land use/land cover types in the drainage area

Site ID	Grp	Site Name	DA (ha)	% Imper	Date of Pond	Percent Land Use/Land Cover								
						AGR	BAR	COM	FOR	HDR	IND	LDR	MDR	OS
08-034	1	Beck Branch	60	1.0	NA	19.5	0	0	80.5	0	0	0	0	0
08-040	1	UT to Upper Beaverdam Creek	148	2.7	NA	27.8	0	1.7	70.5	0	0	0	0	0
33-007	1	UT to Lower Potomac River	284	2.1	NA	12.7	0	0	80.3	0	0	5.2	1.7	0
36-005	1	Black Swamp Creek	585	4.7	NA	18.2	5.9	1.3	64.3	0	0	7.8	2.4	0
36-006	1	UT to Black Swamp Creek	31	2.4	NA	28.3	10.0	0	61.6	0	0	0	0	0
04-009	2	Crows Branch	297	34.1	NA	0	0	7.3	19.9	20.9	0	17.9	29.5	4.4
08-014	2	UT to Upper Beaverdam Creek	476	27.6	NA	2.7	0.7	21.2	57.7	16.9	0	0	0	0.8
08-039	2	UT to Upper Beaverdam Creek	599	17.2	NA	13.1	0.7	21.3	64.7	0	0	0.3	0	0
22-003	2	Watts Branch	248	26.2	NA	0	0	7.5	37.5	2.9	0	9	43.1	0
28-007	2	UT to Broad Creek	634	20.5	NA	6.4	0.1	10.6	47.9	0	0	7.9	26.9	0
06-008	3	Bear Branch	167	20.5	1987	3.9	20.2	3.8	51.3	5.6	15.2	0	0	0
24-003	3	UT to Carey Branch	76	34.9	1987	0	0	0.1	4.6	0	0	13.3	81.9	0
24-011	3	UT to Henson Creek (Broad Creek)	120	31.3	1993	0	0	0	13.6	0	0	13.1	73.3	0
27-070	3	UT to Piscataway Creek	515	11.2	1989	15.7	6.2	0.7	52.2	1.9	0	5.2	18.1	0
29-003	3	Hunters Mill Branch	568	11.7	1987	12.1	0.1	0.5	59.7	0	0.9	2.7	24.1	0

Abbreviations: Grp-treatment group; DA-drainage area; % Imper-percent imperviousness; AGR-agriculture; BAR-bare ground; COM-commercial; FOR-forest; HDR-high density residential; IND-industrial; LDR-low density residential; MDR-medium density residential; OS-open space

Results

As is common in many stream assessments, biological conditions (B-IBI scores and assessments) of the stream groups were plotted against the overall scores for physical habitat quality. Expectations are that biological conditions would be elevated in the presence of good habitat quality, which generally held true with this dataset (Figure CS2-2). Those locations with higher scores and ratings for biological condition tended to be those with better habitat quality and falling in the upper right hand quadrant of the chart; those sites also tended to be in the Group 1 set of sites, that is, with lower stormwater stressors. Those sites with lower biological condition scores and ratings were generally in the central and lower left portions of the chart, and tended to be members of site groups 2 and 3 which were in drainages with substantial impervious surface areas (>11 percent) and the attendant storm water stressors.

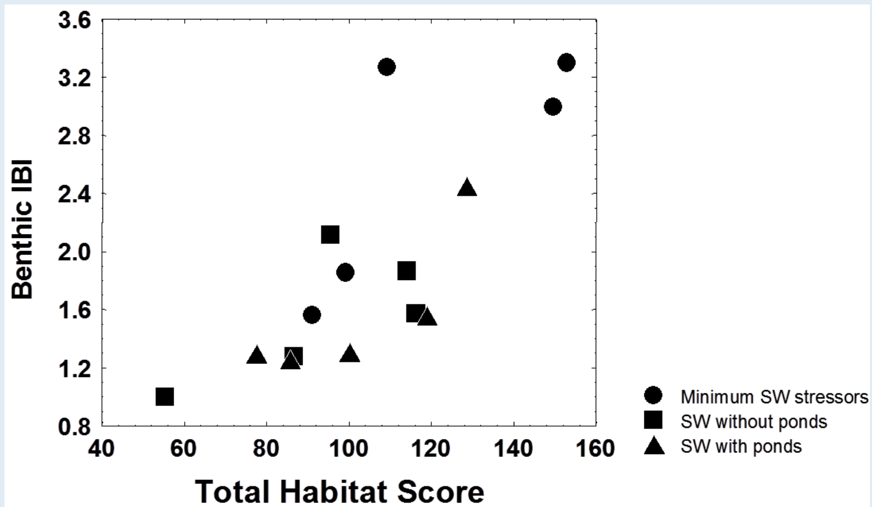


Figure CS2-2. Relationship of biological condition (benthic index of biological integrity [B-IBI]) to physical habitat quality (total habitat score). The 90 percent confidence intervals (CI) are 0.67 points on a 5-point scale for the B-IBI and 6.7 points on a 200-point scale for the habitat quality index.

Both habitat quality and biological condition indicated higher quality for streams with minimal impervious (<5 percent) (Figures CS2-3 and CS2-4). Habitat quality was slightly better in Group 1 streams (with scores ranging from 91 to 150) than in Groups 2 and 3 where scores ranged from 55 to 116 and 78 to 129, respectively (Figure CS2-3). The biological condition of Group 1, however, was more strongly separated from the other two groups (Figure CS2-4). The ranges of B-IBI scores were 1.57 to 3.29 for Group 1, 1.0 to 2.14 for Group 2, and 1.29 to 2.43 for Group 3. There was very little difference between the two groups of streams exposed to stormwater stressors, whether with SWPs (Group 3) or without (Group 2), suggesting that the ponds, in themselves, do not have a strong effect on improving the quality of instream biological conditions. While the SWPs buffered some stressors arising from impervious surface runoff, it is likely that there were other stressors that were unrecognized and not addressed by the SWPs. These other stressors could include upstream or atmospheric chemical contamination, excessive suspended particulates, altered energy input (i.e., leaf litter or woody materials), or habitat alteration. Potential shortcomings of this study include lack of focus on evaluating changes of stressor loads from individual SWPs through, for example, pre- and post-implementation data on stressors for streams where SWPs were implemented; and lack of information regarding the design specifications of the SWPs. The investigators recognized that evaluating the effectiveness of individual BMPs required more intensive load monitoring around the BMP(s) of concern, but, nonetheless, concluded that isolated BMPs were not likely to enhance instream biological condition on their own and that restoration and protection of natural resources requires management on the watershed scale and dealing with multiple and complex stressors and stressor sources.

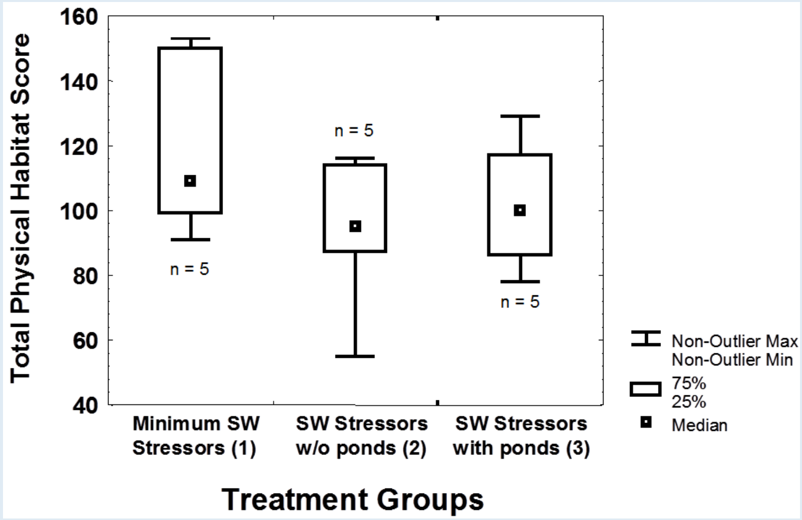


Figure CS2-3. Boxplots of physical habitat quality among each of three stormwater (SW) treatment groups

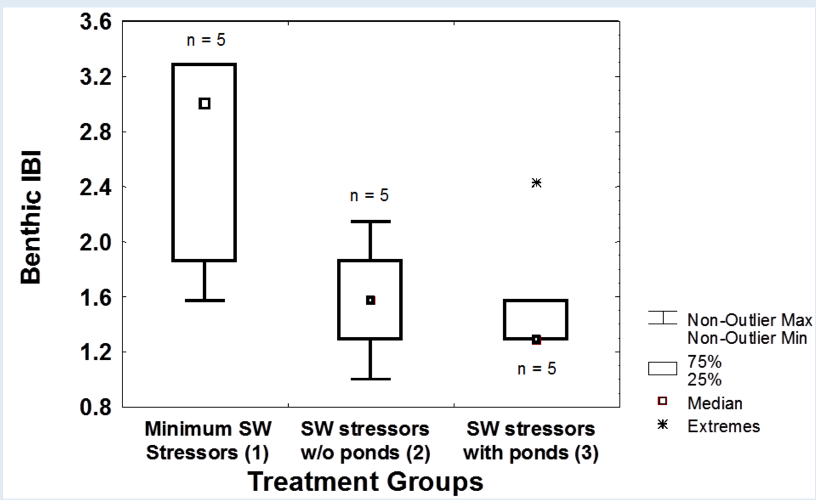


Figure CS2-4. Boxplots of biological condition (benthic index of biological integrity [B-IBI]) among each of three stormwater (SW) treatment group

Literature

Stribling, J. B., E. W. Leppo, J. D. Cummins, J. Galli, S. Meigs, L. Coffman, and M.-S. Cheng. 2001. Relating instream biological condition to BMP activities in streams and watersheds. In *Linking Stormwater BMP Designs and Performance to Receiving Water Impact Mitigation Proceedings of the United Engineering Foundation Conference, August 19-24, 2001, Snowmass Village, Colorado* ed. B. R. Urbonas, pp. 287-304. ISBN 0-7844-0602-2.

Monitoring performed at different spatial scales can provide different types of information on the quality and status of water resources. Conquest et al. (1994) discuss a hierarchical landscape classification system, originally developed by Cupp (1989) for drainage basins in Washington State that provides an organizing framework for integrating data from diverse sources and at different resolution levels. Assessments of waterbodies on a large scale such as an ecoregion, subregion, state, or county provide information on the overall condition of waterbodies in the respective unit. Assessment on a small geographic scale may involve a whole stream, river, or bay or a segment (reach) of the waterbody. A targeted sampling design applies to monitoring waterbodies within a watershed that are exposed to known stressors. Known disturbances, such as point sources, specific NPS inputs, or urban stormwater runoff, can all be targeted for small-scale assessments. At this scale the effectiveness of specific pollution controls, BMP installation/implementation, natural resource management activities, or physical habitat restoration can be monitored. This scale is also where a paired-watershed design can be useful. See Case Study 3 from Pennsylvania for an example of interpreting results from this type of assessment design. Table 4-3 summarizes a waterbody stratification hierarchy for streams and rivers, lakes, reservoirs, estuaries, and wetlands. Further, the sampling site, or the portion of the water body to be sampled, is defined based on technical objectives and programmatic goals of the assessment and/or monitoring activity (Flotemersch et al. 2011). For example, EPA defined a sample reach as 20 times the mean wetted width for its national surveys of lotic waters (streams and rivers) (USEPA 2009); however, many individual states use a fixed 100 m as the sampling reach (Barbour et al. 1999, Carter and Resh 2013).

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

Case Study 3: Effects of Streambank Fencing on Benthic Macroinvertebrates

- ✓ SE Pennsylvania
- ✓ Pasture animal exclusion from stream access
- ✓ Paired-watershed monitoring
- ✓ Nested experimental design

Big Spring Run Basin, a subbasin of Mill Creek Watershed located in Lancaster County, Pennsylvania (Figure CS3-1) is dominated by agricultural land use, much of which is adjacent to aquatic systems. The Mill Creek Basin falls in the Susquehanna River Basin, which ultimately feeds into the Chesapeake Bay. The most common agricultural NPS control measures implemented in the watershed were barnyard runoff control and streambank fencing. This study was designed to provide land managers information on the effectiveness of streambank fencing in controlling NPS pollution. While the project addressed both water quality and biological condition, the emphasis here is on results associated with macroinvertebrate monitoring.

Monitoring and Sampling Design

The objective of this monitoring program was to document the effectiveness of streambank fencing of pasture land on the quality of surface water and near stream ground water. The primary monitoring design was a paired-watershed design, but above/below monitoring was also included to provide multiple opportunities for comparisons to ensure that the effects of fencing could be documented.

The Big Spring Run Basin consists of two similar subbasins ideal for paired-watershed analysis; one was chosen as the treatment basin and the other as the control. The treatment basin was 3.6 km^2 with 4.5 km of stream, of which approximately 70 percent of the streams run through open pasture. The control basin was 4.7 km^2 with 4.3 km of stream and consisting of approximately 70 percent of streams running through open pasture. Elevation and geologic makeup were also nearly identical for both basins, with stream gradients ranging from 0.3 to 0.6 m elevation change for every 30 m of channel. Temperate zone climate was typical for the study basins, with an average precipitation of 104 cm and an average temperature of 11°C. Agriculture accounted for over 80 percent of the land use in each subbasin.

Surface water monitoring stations for the paired-watershed analysis were located at the outlets of the control (C-1) and treatment (T-1) subbasins (Figure CS3-1). Site T-1 was to also be used with T-3 in the treatment subbasin for an above/below study; streambank fencing was to be installed between T-1 and T-3. A site (T-2) was also added at a visually degraded upstream tributary for comparison with C-1 in another paired-watershed analysis. Site T-4 was added to determine the effects of new construction that began two years into the study. Surface water samples were collected every 10 days from April to November (about 25 to 30 samples per site per year) because this was when dairy cows and heifers were pastured. Monthly base-flow samples were collected during the remaining part of the year. Storm event samples were collected at all sites except T-3, with from 35 to 60 percent of the storm events sampled over the entire study period. Monitoring variables included total and dissolved nutrients, suspended sediment concentration, field parameters (low flow only), fecal streptococcus (low flow only), and discharge.

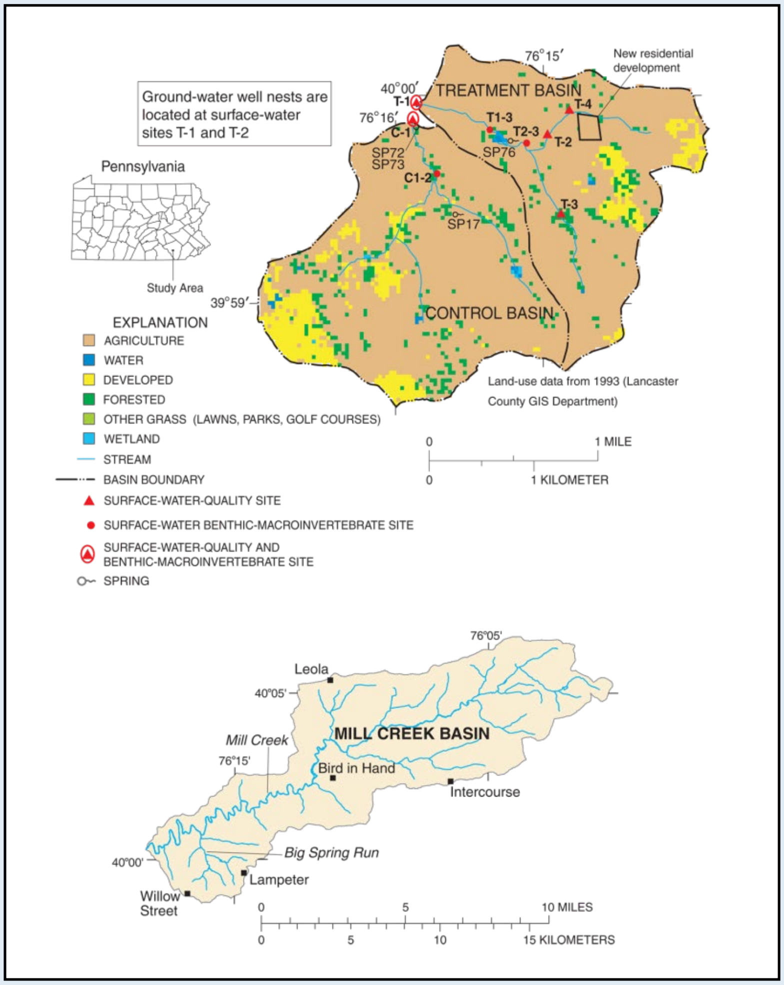


Figure CS3-1. Land-use map of study area and location of surface-water sites, ground-water well nests, and selected springs in the Big Spring Run Basin, Lancaster County, PA

A nested well approach was used to monitor near-stream groundwater parameters, including water quality and level, flow directions, age dating, and chemical quality. Two nested wells were placed in the treatment basin at sites T-1 and T-2. At each location, three shallow wells and one deep well were clustered together. One of the shallow wells at each ground water monitoring location was placed outside of the fenced area as a control. Water level was measured continuously and wells were sampled monthly during periods of little to no recharge. The resulting samples were analyzed for nutrients and fecal streptococcus.

Benthic macroinvertebrate sampling was conducted in May and September of each year at five different locations (Figure CS3-1): one at the outlets of the control and treatment basins (sites T-1 and C-1), two upstream in the treatment basin (sites T1-3 and T2-3), and one upstream in the control basin (site C1-2). Stream pool and riffle habits were sampled using the kick-net method; the USEPA Rapid Bioassessment Protocols (RBP) was used to characterize habitat; and water quality samples and stream measurements were also taken. Most metrics were applied to taxonomic identifications to the family level. The list of metrics includes percent dominant taxon (genus and family), EPT (Ephemeroptera, Plecoptera, and Trichoptera) index, generic EPT/Chironomidae ratio, EPT/total number, percent Chironomidae, shredders/total taxa ratio, scrapers/filterers ratio, Hilsenhoff Biotic Index (HBI) (genus and family), taxa richness (genus and family), and percent Oligochaeta.

All monitoring was conducted both before and after installation of streambank fencing which occurred in the treatment subbasin from May 1997 through July 1997. About 3.2 kilometers of fencing was installed along riparian zones in pastured areas to create a 1.5- to 3.6-m-wide stream buffer strip. Each pastured fenced had an average of two cattle crossings through the stream to allow the animals to migrate between pasture locations and access a water supply. Monitoring was carried out before and after fencing was installed. The pre-treatment period was from 1993 to 1997 and post-treatment monitoring was carried out from mid-July 1997 through June 2001.

Due to their potential effect on the quality and quantity of the water and habitat, basin-wide covariate data were also collected during the study period, such as precipitation, inorganic and organic nutrient applications, and the number of cows present. Precipitation and agricultural data were obtained, respectively, using precipitation gauges (logging at 15-minute intervals), and from farm operators (monthly records).

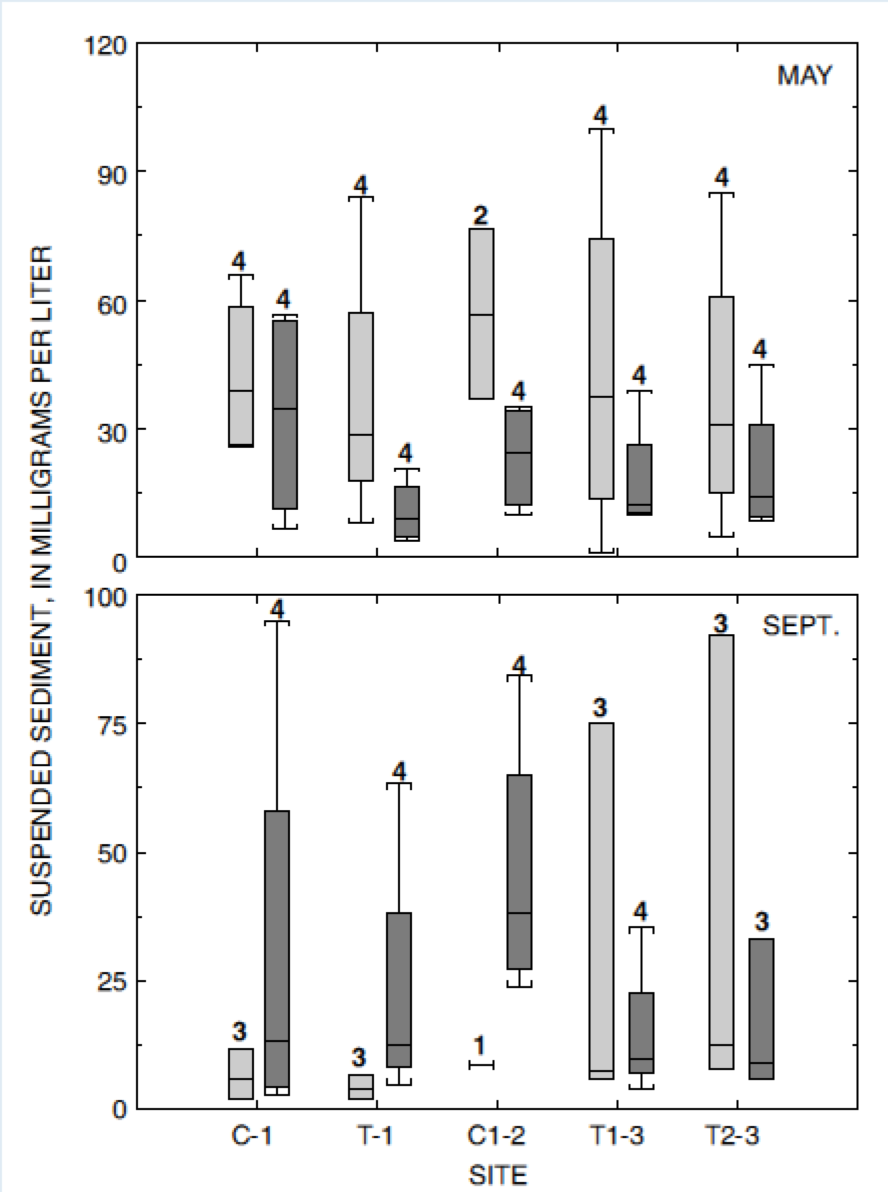


Figure CS3-2. Distribution of concentrations of suspended sediment for May and September sampling events at benthic macroinvertebrate sites for the pre- and post-treatment periods in the Big Spring Run Basin, Lancaster County, PA. Light shaded boxes are pretreatment, dark shaded are post-treatment, C represents Control, and T represents Treatment.

Results

Data collected from the pre-treatment period (1993 to 1997) were compared to those from the treatment and control basins to determine the effectiveness of streambank fencing. Changes in the yields of nutrients and suspended sediment during low flow and storm flow events were quantified using ANCOVA. ANCOVA was also used to quantify changes between pre- and post-treatment concentrations of nutrients, water quality, and fecal streptococcus in collected water samples, as well as nested wells inside and outside of the treatment area. Canonical correspondence analysis (CCA) was used to determine the effects of streambank fencing on instream biological conditions as characterized by benthic macroinvertebrates. A brief overview of major findings from analysis of water chemistry data is presented here, followed by a more detailed summary of results from macroinvertebrate monitoring.

It was concluded that water chemistry results indicated that riparian fencing had fairly consistent effects on suspended sediment but less clear effects on nutrients. Post-treatment period improvements were evident at site T-1 for both nutrients and sediments; however, site T-2 showed reductions only in suspended sediment. The average reduction in suspended sediment yield for the treated sites was about 40 percent. N species at T-1 showed reductions of 18 percent (dissolved NO_3) to 36 percent (dissolved ammonia); yields of TP were reduced by 14 percent. Conversely, site T-2 showed increases in N species of 10 percent (dissolved ammonia) to 43 percent (total ammonia plus organic N), and a 51-percent increase in yield of TP. The different results for nutrients at T-2 and T-1 were attributed to ground water contributions and the failure to implement nutrient management along with the fencing. Shallow ground water flow contributed to stream flow at T-2, but the stream was losing water to the shallow ground water system at T-1. It is believed that an upland agricultural field caused increased dissolved P levels in shallow ground water at T-2, resulting in a transport of P from ground water to the stream that increased stream P levels. In addition, cattle contributed nutrients directly to the stream via excretion at the embedded stream crossing at T-2.

Analysis of the benthic macroinvertebrate samples showed some apparent improvements relative to the control sites in riparian and instream habitat (sites T2-3 and T-1 versus C-1). Some differences in bottom substrate, bank stability, available cover, and scouring and deposition were observed in the downstream and upstream locations within the treatment basin that could potentially be considered slight improvements. Water quality data collected during the benthic macroinvertebrate sampling suggested the overall improvement to instream habitat was due to the decreased load of suspended sediment (Figure CS3-2). The fenced riparian buffer, despite being narrower than what was considered optimal, allowed vegetation to become fully established and bank stability to improve. It was particularly evident at site T2-3 where it became overgrown with vegetation and blocked the stream from view.

The composite benthic index, which combines all metrics, is called the "Macroinvertebrate Aggregated Index" (MAI). For both spring and fall samples, the index showed some improvement for the treatment sites relative to control sites, though, trends were mixed overall. For the treatment basin, sites T1-3 and T2-3 showed no change with spring samples, while the outlet site, T-1, showed a slight 1 unit increase. From the pre-treatment to the post-treatment period, fall index scores changed in the control basin by 1 unit, increasing at C-1 and decreasing at C1-2. Fall scores for the treatment basin also changed over this timeframe, increasing by 2 units for T-1 and T1-3, but decreasing by 1 unit at T2-3.

Disaggregating the index into individual metrics allows evaluation of different components of the benthic macroinvertebrate assemblage. In this dataset, there are different responses by different metrics. No difference was seen for five of the 10 genus-level metrics, which included percent dominant taxa (generic level) (PDTG), EPT taxa, percent EPT taxa, percent shredders, and ratio of scrapers to filterers. Thus, treatment elicited no effect for 50 percent of the metrics. Two of the metrics (EPT/Chironomidae ratio, Hilsenhoff Biotic Index [genus level]; 20 percent) suggested some, or slight, effect of the streambank fencing on treatment sites relative to control; and, distinct effects were seen for the remaining three metrics: percent Chironomidae (Figure CS3-3), taxa richness, and percent Oligochaeta.

Further evaluation of taxa lists for dominance, occurrence, and uniqueness of and by individual taxa can help illuminate differences, in particular, for those taxa that are known to be more pollution-tolerant or sensitive. Spring samples were numerically dominated by worms (Naididae, Tubificidae), scud (Amphipoda: Gammaridae), several different midges (Diptera: Chironomidae: i.e., *Cricotopus, Orthocladius, Dicrotendipes, Micropsectra*), and blackflies (Diptera: Simuliidae: *Simulium*). The fall samples illuminated a shift in the actual taxonomic composition of the site to a greater diversity (i.e., a larger number of taxa) and dominance by largely different taxa, including riffle beetles (Coleoptera: Elmidae: *Dubiraphia, Stenelmis*), net-spinning caddisflies (Trichoptera: Hydropsychidae: *Hydropsyche, Cheumatopsyche*), midges (*Chironomus, Dicrotendipes, Polypedilum, Rheotanytarsus*), and blackflies. The only taxa that retained any kind of dominance for this site across seasons were *Dicrotendipes* and blackflies. The overall differences are driven by seasonality of the system, simply showing a greater diversity during the fall season and not by any changes in stressor load. Although the two outlet sites (C-1, T-1) basically had the same taxa dominating sample data, each had a greater diversity than the upstream sites. Elevated taxonomic/biological diversity can be an indicator of greater diversity and complexity of habitat characteristics; lacking other types of stressor loads, this diversity is likely what is reflected at the outlet sites. The upstream sites in the treatment basin (T2-3, T1-3) consistently showed more diversity than C1-2, but the overall PDTG means for T1-3 and T2-3 increased slightly (<1 percent) and by 13 percent, respectively, from the pre- to post-treatment period. Because differences in assemblage makeup are largely explainable by factors other than what might be introduced by streambank fencing, such as seasonality, and expected physical habitat characteristics, the authors concluded that the treatment did not seem to improve benthic-macroinvertebrate community structure based on PDTG.

Across the full dataset, the dominant family-level taxa in spring samples were Chironomidae, Gammaridae, Naididae, and Tubificidae, all recognized as being semi-tolerant to organic enrichment. The dominant families in fall samples were Gammaridae, Tubificidae, Elmidae, Physidae, Baetidae, Chironomidae, and Simuliidae, all considered as being moderately to very tolerant of organic enrichment. This indicates that the more sensitive taxa were not able to become dominant members of the benthic-macroinvertebrate assemblage after the fences were installed in the treatment basin. Sensitive taxa may not have been present or only a few individuals were present during the post-treatment period because of 1) not enough time for the system to equilibrate to the new conditions, or 2) because these are spring-fed, first- to second-order limestone streams. Limestone streams typically support assemblages including mayflies (Ephemeroptera), midges (Chironomidae), scud (Amphipoda: Gammaridae), and pillbugs/sowbugs (Isopoda: Asellidae), all of which were present in treatment and control basins. Some of the more stressor-sensitive taxa that were present in small numbers included *Promoresia* (Elmidae), *Oxyethira* (Hydroptilidae), *Antocha* (Tipulidae), and two species of Chironomidae (*Pagastia* and *Prodiamesa*).

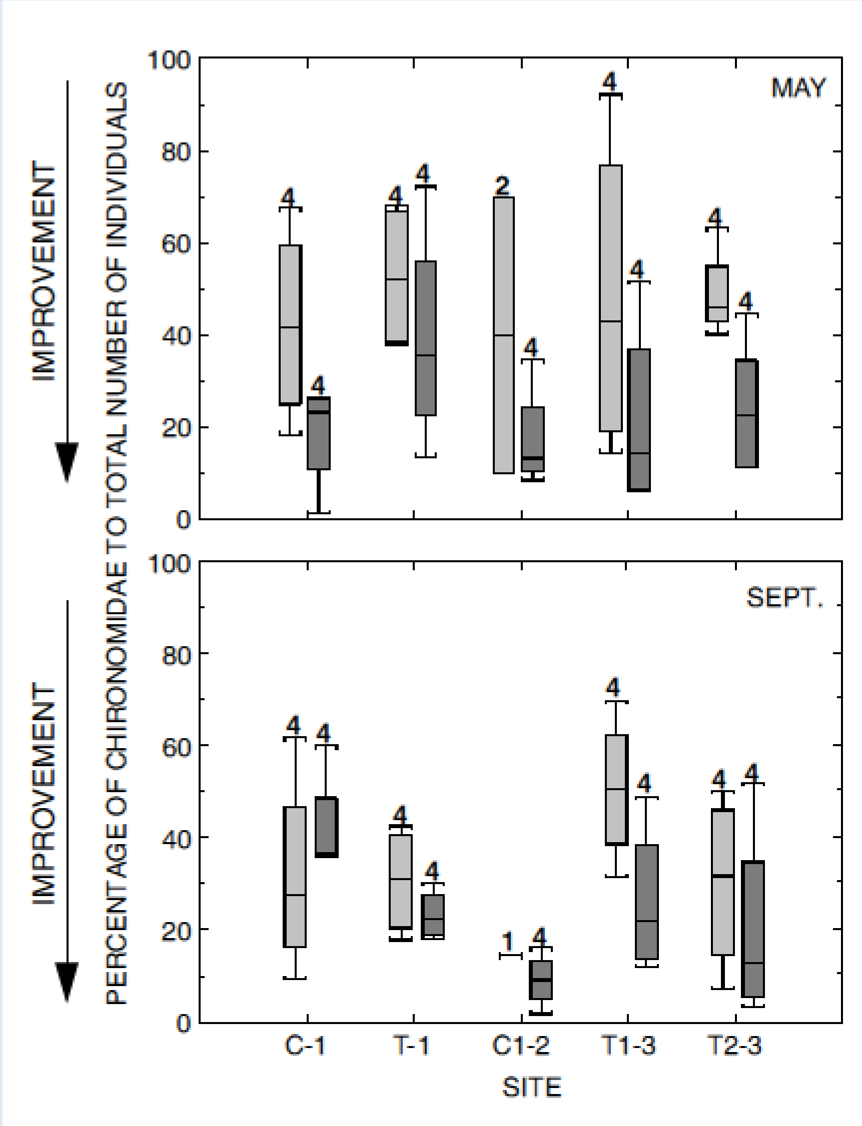


Figure CS3-3. Distribution of the percentage of Chironomidae to total number of individuals for May and September sampling events at benthic-macroinvertebrate sites for the pre- and post-treatment periods in the Big Spring Run Basin, Lancaster County, PA. Light shaded boxes are pretreatment, dark shaded are post-treatment, C represents Control, and T represents Treatment.

Overall, the authors conclude that streambank fencing had a positive influence on the taxonomic diversity of benthic-macroinvertebrates, both at genus and family levels. This positive influence is interpreted as primarily resulting from stabilization of the riparian zone, allowing growth of streamside vegetation to progress, and ultimately allowing better habitat to develop and support more taxa.

Previous studies suggest that for optimal reduction of nutrient loads into nearby aquatic systems, the buffer size should be greater than the 1.5- to 3.6-m buffer used in this study. Because there was such a wide range of what would be considered an adequate buffer size, it was uncertain which nutrient types, if any, could be controlled or reduced with this approach. Study results show that while the fenced streambank buffer was relatively small, it still was substantially effective in improving surface and near-stream shallow ground water quality, and led to some improvement in instream biology. Small-scale stream buffers and exclosure fencing both have limited effectiveness in controlling high nutrient input ultimately transported through subsurface flows; the most pronounced effects of exclosures are in reducing suspended sediment inputs and consequently leading to improved habitat. There may be some effectiveness in controlling excessive nutrient flows, but the benefits are likely minor in comparison to the habitat effects.

Literature

Galeone, D.G., R.A. Brightbill, D.J. Low, and D.L. O'Brien. 2006. *Effects of Streambank Fencing of Pastureland on Benthic Macroinvertebrates and the Quality of Surface Water and Shallow Ground Water in the Big Spring Run Basin of Mill Creek Watershed, Lancaster County, Pennsylvania, 1993-2001*. Scientific Investigations Report 2006-5141. U. S. Geological Survey, Reston, VA.

Galeone, D.G. and E.H. Koerkle. 1996. *Study design and Preliminary Data Analysis for a Streambank Fencing Project in the Mill Creek Basin, Pennsylvania.* Fact Sheet FS-193-96. U. S. Geological Survey, Lemoyne, PA.

Table 4-3. Waterbody stratification hierarchy

Population	Streams	Lakes[b,c]	Reservoirs[b,d]	Estuaries[e]	Wetlands[f]
Operational Sampling Unit (SU)	Channel segment (i.e., a length of river channel into which no tributaries flow)	Self-contained basin	Self-contained basin (hydrologically isolated from other basins)	Self-contained basin	Transect upland or deep water boundaries
Strata (or higher stages) comprising SUs	Ecoregion	Ecoregion	Ecoregion	Biogeographic province	Wetland system type (marine, estuarine, riverine, lacustrine, palustrine)
	Watershed	Size/surface area of lake (km^2)	Size/surface area of reservoir (km^2) (watershed area/basin surface area)	Watershed	Watershed recharge, discharge or both
	Stream/river channel (ordinal/areal)	Lake hydrology (retention time, thermal stratification)	Hydrology (water level fluctuation/ drawdown; retention time stratification)	Watershed area (km^2)	Class (based on vegetative type; substrate and flooding regime; hydroperiod)
	Segment	Characteristic water quality (natural conditions)	Characteristic water quality (natural conditions)	Zones (tidal basin, depth, salinity)	Flooding regime water chemistry soil type
Habitat within SUs	Characteristic water quality (natural conditions) Hydraulic conductivity (homogeneous, heterogeneous, isotropic, anisotropic)				
	Macrohabitat (pool/riffle; Shorezone vegetation; submerged aquatic macrophytes)		Longitudinal zone (riverine; transitional; lacustrine; tail waters (can be more riverine but always associated with dams))		
		Depth zone (eulittoral 6 profundal)	Depth	Substrate/habitat	Subsystem (subtidal, intertidal, tidal, lower perennial, upper perennial, intermittent, littoral, limnetic)
	Microhabitat	Substrate/ microhabitat	Substrate/ microhabitat		Substrate/ microhabitat

[a] Frissell et al. 1986; [b] Gerritsen et al. 1996; [c] Wetzel 1983; [d] Thornton et al. 1990; [e] Day et al. 1989; [f] Cowardin et al. 1979

4.4 Biological Assessment Protocols

Biological indicators are widely recognized as being critical for evaluating ecological conditions, helping identify and prioritize problems, designing controls or other solutions, and in evaluating effectiveness of management efforts. However, prior to being able to make management decisions using these indicators, two things need to occur. First, the indicator must be calibrated; and, second, sampling and analysis must be instituted in a routine and consistent manner to directly address management objectives. To obtain valid assessment results, and to optimize defensibility of decisions based on them, it is imperative that monitoring be founded on data of known quality (Flotemersch et al. 2006b, Stribling 2011). In this section, our discussion focuses on wadeable streams, and we assume that the user has access to indicators that have been calibrated and are applicable to their region and water body (-ies) of concern. Many states have developed MMIs using one or more biological groups, but most typically benthic macroinvertebrates, and sequentially less so, fish and periphyton (algae and diatoms). Carter and Resh (2013) surveyed state agencies about different characteristics of their biological monitoring programs, and, although there are differences in some of the specific techniques, there has also been considerable convergence among methods during the past 10 to 15 years. Monitoring programs that have gone through the index calibration process have worked or are working through technical issues associated with customizing sampling techniques to water body type, prevailing climatic conditions, and programmatic capacity; using field data to characterize environmental/ecological variability; defining mathematical terms of the indicator (metrics and index make-up); and defining thresholds for judging degradation.

Part of customizing field sampling and laboratory analysis methods for a program involves understanding the range of variability of field conditions and the data/assessments that arise from them. One approach for developing such an understanding is to recognize that biological monitoring and assessment protocols are made up of a series of methods (Flotemersch et al. 2006b, Stribling 2011) generally corresponding to different steps of the overall process: field sampling, sample preparation, taxonomic identification, enumeration, data entry, data reduction, and site assessment/interpretation. In this section, we present descriptions of the background, purpose, application, and output of the methods along with relevant procedures for documenting data quality associated with each.

4.4.1 Field Sampling

In the context of the field sampling approach being used for biological assessment, taking or observing organisms from a defined sample location is intended to provide a representation of the biological assemblage supported by that water body, whether benthic macroinvertebrates, fish, or periphyton. These three assemblages are emphasized because they are most commonly used in routine monitoring and assessment programs in the US, and methods for them are relatively well-documented (Barbour et al. 1999, Moulton et al. 2002, Stribling 2011, Carter and Resh 2013).

4.4.1.1 Benthic macroinvertebrates

Benthic macroinvertebrate samples are taken from multiple habitat types, and composited in a single sample container (Figure 4-3). If sampled from transects as per the USEPA national surveys (USEPA 2009), they are collected along 11 transects evenly distributed throughout the reach length, using a D-frame net with 500-μm mesh openings (Klemm et al. 1998, Flotemersch et al. 2006b). An alternative to transects is to estimate the proportion of different habitat types in a defined reach (e.g., 100m), and distribute a fixed level of sampling effort proportional to their frequency of occurrence throughout the reach (Barbour et al. 1999, 2006). Whether using transects or proportional distribution, organic and inorganic sample material (leaf litter, small woody twigs, silt, and sand; also includes all invertebrate specimens) are composited in one or more containers, preserved with 95% denatured ethanol, and delivered to laboratories for processing (Figure 4-4). A composite sample over multiple habitats in a reach is a common protocol feature of many monitoring program throughout the US (Carter and Resh 2013), although some programs choose to keep samples and data from different habitat types segregated.

Figure 4-3. Removing a benthic macroinvertebrate sample from a sieve bucket and placing the sample material in a 1-liter container with approximately 95% ethanol preservative

Figure 4-4. Labelling benthic macroinvertebrate sample containers and recording field data

4.4.1.2 Fish

Fish sampling is designed to provide a sample that is representative of the fish community inhabiting the reach, and which assumed to reasonably represent species richness, guilds, relative abundance, size, and anomalies. The goal is to collect fish community data that will allow the calculation of an IBI and observed/expected (O/E) models. Electrofishing is the preferred method of sampling, involving the operator and (ideally) two netters, and occurs in a downstream direction at all habitats along alternating banks, over a length of 20 times the mean channel width at designated transects (USEPA 2009). Collection of a minimum of 500 fish is the target number of specimens (USEPA 2009), and in the event this is not attained, sampling will continue until 500 individuals are captured or the downstream extent of the site is reached.

4.4.1.3 Periphyton

Periphyton collections are made from shallow areas near each of the sampling locations on the 11 cross-section transects established within the sampling reach and are collected at the same time as the benthic macroinvertebrate samples (USEPA 2009). There is one composite sample of periphyton for each site, from which separate types of laboratory samples can be prepared, if necessary. The different sample type could include a) an ID/enumeration sample to determine taxonomic composition and relative abundances, b) a chlorophyll sample, c) a biomass sample (for ash-free dry mass [AFDM]), or d) an acid/alkaline phosphatase activity [APA] sample). There are potentially other analysis types that could be performed, thus requiring additional sample segregates.

4.4.1.4 Quality control measures

Other than a qualitative judgment that field personnel are adequately trained, have sufficient experience, and have been successfully audited as having completely and accurately applied the correct SOP, the quality of field sampling cannot be determined without sample processing. The consistency of field sampling is a measure of data quality quantified by precision calculations using indicator values – individual metrics, IBI scores, or predictive models - collected from adjacent stream reaches (i. e., two stream channel lengths where the second [B] begins at the endpoint of the first [A]). As a rule-of-thumb, we recommend a site duplication rate of 10 percent, where duplicate locations are randomly selected from the full sample lot, and fieldwork occurs as routine. Terms calculated from the duplicate sample results include median relative percent difference (mRPD), 90 percent confidence intervals/minimum detectable difference (CI90/DD90), and coefficient of variation (CV) (Flotemersch et al. 2006b, Stribling et al. 2008a, Stribling 2011). Depending on programmatic application, natural variability of the landscape the watershed is draining, density and distribution of potential stressor sources, number of field crews, and, of course, budgetary resources, it can be useful to stratify distribution of duplicate reaches. This will allow programmatic measurement quality objectives (MQO) to be established for objective benchmarks for acceptable quality of data. Typical MQO for field sampling precision (Stribling et al. 2008a, Stribling 2011) might be:

- mRPD<15,
- CI90<15 index points on a 100-point scale, and
- CV<10% for a sampling event

Depending on programmatic needs, values exceeding these MQO could highlight samples for more detailed scrutiny to determine causes for the exceedances, and the need for corrective actions.

4.4.2 Sample processing/laboratory analysis

For biological monitoring and assessment programs, sample processing employs procedures for organizing sample contents so that analysis is possible. For benthic macroinvertebrate and periphyton samples, those procedures are laboratory-based; however, for fish, they are performed primarily in the field (USEPA 2004, 2009).

4.4.2.1 Benthic macroinvertebrates

The three aspects of sample processing for benthic macroinvertebrate samples are a) sorting, which serves to separate the organisms from other sample material, specifically, organic detritus inorganic silt, and other materials (Figure 4-5), b) subsampling, which isolates a representative sample fraction from the whole, and c) taxonomic identification, which characterizes the (sub)sample by naming and counting individuals in it.

Figure 4-5. Examining, washing, and removing large components of sample material prior to putting in sample container

4.4.2.1.1 *Sorting and subsampling*

The sorting/subsampling procedure is based on randomly selecting portions of the sample material spread over a gridded Caton screen (Caton 1991, Barbour et al. 1999, Flotemersch et al. 2006b, Stribling 2011), and fully removing (picking) all organisms from the selected fractions. The screen is divided into 30 grid squares, each individual grid square measuring 6 cm x 6 cm, or 36 cm^2 (note that it is *not* 6 cm^2 as indicated in Figure 6-4b of Flotemersch et al. [2006b]). Prior to beginning the sorting/subsampling process, it is important that the sample is mixed thoroughly and distributed evenly across the screen to reduce the effect of organism clumping that may have occurred in the sample container. Depending on the density of organisms in the sample, multiple levels of sorting may be necessary, the purpose of which is to minimize the likelihood that the entire sample to be identified comes from a very small number of grids. Initially, four grids are randomly selected from the 6 x 5 array, removed from the screen, placed in a sorting tray, and coarsely examined. If the density of organisms is high enough that there are many more than the target number in the four selected grids (i. e., greatly exceeding by twofold or more the 100-, 200-, 300-, 500-organisms, or more, depending on the project), that material is re-spread on a second gridded screen and the process repeated (second level sort). This is repeated until it is apparent that the density of specimens will require <u>at least</u> four grids to be sorted to attain the target number (±20%). Once re-spreading is no longer needed, all organisms are removed from the four grids using forceps. If the final rough count is ±20 of the target subsample size, then subsampling is complete; if >20% less than the target subsample size, then additional, single grids of material are moved from the tray, and picked in entirety. This is repeated, one grid at a time, until within 20 percent of the target number. Following picking, the sort residue should be transferred to a separate container labeled with complete sample information, and the words “SORT RESIDUE” clearly visible. Completely record the number of sort levels and grids processed. The sorting and subsampling process should result in at least three containers: a) clean (sub)sample, b) sort residue, and c) unsorted sample remains. Container ‘a’ is provided to the taxonomist for identification and counting, ‘b’ is available for QC sort re-check, and ‘c’ is archived until all QC checks are complete. In the event of certain QC failures, it may be necessary to process portions of the unsorted remains.

Fixed count subsamples - Fixed organism counts vary among monitoring programs (Carter and Resh 2013), with 100, 200, 300 and 500 counts being most often used (Barbour et al. 1999, Cao and Hawkins 2005, Flotemersch et al. 2006a). Flotemersch et al. (2006a) concluded that a 500-organism count was most appropriate for large/nonwadeable river systems, based on examination of the relative increase in richness metric values (< 2%) between sequential 100-organism counts. However, they also suggested that 300-organism count is sufficient for most study needs. Others have recommended higher fixed counts, including a minimum of 600 for wadeable streams (Cao and Hawkins 2005). The subsample count used for the USEPA national surveys is 500 organisms (USEPA 2004); many states use 200 or 300 counts.

4.4.2.1.2 Taxonomic identification

Genus level taxonomy is the principal hierarchical level used by most routine biological monitoring programs for benthic macroinvertebrates (Carter and Resh 2013), although occasionally family level taxonomy is used. For genus level to be attained, most direct observations can be accomplished with dissecting stereomicroscopes with magnification ranges of 7-112x; however, midges (Chironomidae) and worms (Oligochaeta) need to be slide-mounted and viewed through compound microscopes that have magnification ranging 40-1500x, under oil. Slide-mounting specimens in these two groups is usually (though, not always) necessary to attain genus level nomenclature, and sometimes even more coarse level for midges (i.e., less specific). Taxonomic classification is a major potential source of error in any kind of biological monitoring data sets (Stribling et al. 2008b, Bortolus 2008) and the rates of error can be managed by specifying both hierarchical targets and counting rules. Hierarchical targets define the level of effort that should be applied to each specimen but may often not be possible for some specimens due to poor slide mounts, damaged, or their being juvenile (early instars). Further, the requirement for some taxa may be more coarse, such as genus-group, tribe, subfamily, or even family. In any case, the principal responsibility of the taxonomist is to record and report the taxa in the sample and the number of individuals of each taxon. Consistency in the nomenclature used is more important than the actual keys that are used, although, some programmatic SOPs may specify the technical literature. For example, the identification manual “*An Introduction to the Aquatic Insects of North America*” (Merritt et al. 2008) is useful for identifying the majority of aquatic insects in North America to genus level. However, because many taxonomic groups are often (correctly) under perpetual revision and updates, the nomenclatural foundation of many may have changed, thus requiring familiarity of the taxonomist with more current primary taxonomic literature. Merritt et al. (2008) is not applicable to non-insect macroinvertebrate taxa that are often captured in routine sampling, including Oligochaeta, Mollusca, Acari, Crustacea, Platyhelminthes, and others; exhaustive lists of literature for all invertebrate groups are provided by Klemm et al. (1990) and Thorp and Covich (2010). Identification staff may also need information on accepted nomenclature, including validity, authorship, and spelling, all of which could be found in the integrated taxonomic information system (ITIS; http://www.itis.gov/). Although it is a nomenclatural clearinghouse, it should be recognized that it is not completely current for all taxa potentially requiring independent confirmation.

It should be noted that some volunteer monitoring programs, such as the Izaak Walton League (IWL) and the Maryland DNR Stream Waders Program (MDNR), use simpler taxonomic procedures. The IWL uses field identification of a small number of organisms that are of a limited number of kinds, like mayflies, stoneflies, caddisflies, beetles, and mollusks, and just note their presence or absence. The MDNR Stream Waders program, including metrics and index, are based on family level data.

4.4.2.2 Fish (field taxonomic identification)

Identification and processing of fish occur at the completion of each transect (USEPA 2009), where the data recorded include species names, number of individuals of each, length, and DELT anomalies (**D**eformities, **E**roded fins, **L**esions and **T**umors). Taxonomic identification and processing should only be completed on specimens >25 mm total length and by qualified staff. Common names of species should follow those established under the American Fisheries Society's publication, "Common and Scientific Names of Fishes from the United States, Canada and Mexico" (Nelson et al. 2004). Species not positively identified in the field should be separately retained for laboratory identification (up to 20 individuals per species). For programs not using the transect method of sample reach layout, electrofishing will cover all habitat throughout the reach. Further, fish sample vouchers are developed for a minimum of 10% of the sites sampled (USEPA 2012).

4.4.2.3 Periphyton

Two activities making up sample processing for periphyton are further segregated into those for a) soft-bodied algal forms and b) diatoms. Although methods for both are presented by USEPA (2012), diatom procedures are based principally on those of the US Geological Survey National Water Quality Assessment Program (USGS/NAWQA) (Charles et al. 2002). Microscopic diatoms encountered are identified (to lowest possible taxon level), enumerated and recorded. Estimates of the biovolume of dominant species are made using existing parameters, or those found in the literature, and used to determine the biovolume of the sample. Detailed information on the different procedures, especially on the analytical approaches for soft algae using the Sedgewick-Rafter and extended Palmer-Maloney count techniques, can be found in USEPA (2012) and Charles et al. (2002).

4.4.2.4 Quality control measures/data quality documentation

Quality control (QC) for sample processing for these three taxonomic groups is, in some respects, similar, but in others, different. Some of the similarities are that several aspects quality evaluations are based on repeating processes; specifically, duplicating field samples, or repeating of sample processing activities (sorting, identification, and counting). Differences arise out of the fact that there are not always analogous methods for dealing with the different organism types. Specifically, fish are identified in the field, whereas, benthic invertebrates and algae/diatoms are laboratory-identified. Logistical constraints prevent whole-sample re-identification of fish, whereas it is easily done for the other groups. And, subsampling is not done with fish samples, where it is explicitly done for benthic invertebrates, and functionally done for algal/diatom samples.

Sorting QC (benthic macroinvertebrates [only]). – Sorting QC is accomplished through rechecking the sample sort residue from 10% of the samples, randomly selected, and calculating the term 'percent sorting efficiency' (PSE) (Stribling 2011, Flotemersch et al. 2006b). This value reports the number of specimens missed during primary sorting as a proportion of the original number of specimens found. A typical MQO for this is PSE>90%, with the goal of minimizing the number of samples that fail. Individual programs must specify what is acceptable, but generally, the goal should be to have <10% of the samples fail. It is a measure of bias associated with sample sorting.

Taxonomic QC. – As a measure of precision in taxonomic identifications, consistency is quantified by independent re-identification of whole (sub)samples, where those samples are randomly selected (10%, as a rule-of-thumb) from the full sample lot. Sample results from the QC taxonomist are directly compared to those of the primary taxonomist, and differences quantified as 'percent taxonomic disagreement' (PTD) for identifications, as 'percent difference in enumeration' (PDE) for counts, and as 'absolute difference in percent taxonomic completeness' (*abs*[PTC]). Typical MQO for these are PTD<15%, PDE<5%, and *abs*PTC<5% (Stribling 2011).

If MQO thresholds are routinely or broadly exceeded, samples failing should be examined in more detail to determine causes of the problem, and what corrective actions may be necessary.

4.4.3 Data reduction/indicator calculation

Once necessary corrective actions for sample processing and taxonomic identifications have been implemented and effectiveness confirmed, data quality is known and acceptable, sample data are converted into the primary terms to be used for analysis. As stated above, monitoring practitioners usually have access to published MMI for application to sample data, as well as sometimes predictive models and established decision analysis systems. Indicators most often take the form of a multimetric Index of Biological Integrity (IBI; Karr et al. 1986, Hughes et al. 1998, Barbour et al. 1999, Hill et al. 2000, 2003) or a predictive observed/expected (O/E) model based on the River Invertebrate Prediction and Classification System (RIVPACS; Clarke et al. 1996, 2003, Hawkins et al. 2000b, Hawkins 2006). The Illinois Department of Natural Resources used the Macroinvertebrate Biotic Index (MBI) in their analysis of restoration effectiveness in the Waukegan River (see Case Study 4).

4.4.3.1 Multimetric indexes

The purpose of any MMI is to summarize complex biological and environmental information into a form and format that can be used for management decision-making (Karr and Dudley 1981, Karr 1991, Angermeier and Karr 1994), and doing so in a manner that allows uncertainty associated with those decisions to be known and communicated. Index calibration is the empirical process of determining which measures are best suited for that purpose, specifically in terms of their capacity for detecting biological changes in response to environmental variables of concern (pollutants, or stressors). For the purpose of this guidance it is assumed that the calibration procedure (Hughes et al. 1998, Barbour et al. 1999, McCormick et al. 2001) has been completed, and that an MMI is available to monitoring practitioners for application. The reader should be aware that no attempt is made to be comprehensive in discussing either metric diversity or index formulation, or of reviewing supporting technical literature. As such, only selected examples are used below to illustrate different aspects of MMI application.

CASE STUDY 4: BIOLOGICAL AND PHYSICAL MONITORING OF WAUKEGAN RIVER RESTORATION EFFORTS, LAKE COUNTY, ILLINOIS

- ✓ Urbanized watershed
- ✓ Severely degraded stream habitat; channel instability/bank erosion; high velocity runoff
- ✓ Bank stabilization using LUNKERS and riparian re-vegetation
- ✓ Grade control using rock weirs and artificial riffles
- ✓ Benthic macroinvertebrate assemblage monitoring
- ✓ Effectiveness evaluation

The Waukegan River watershed is located on the western shore of Lake Michigan, about 56 km (35 mi) north of Chicago in Lake County (Figure CS4-1). It is approximately 20 km (12.4 mi) long and has a drainage area of 2,994 ha (7,397 ac). The river channel drops from an approximate 222 m (730 ft) headwaters elevation to around 177 m (580 ft) above sea level, before discharging directly into Lake Michigan through Waukegan Harbor on its western shore. The Waukegan River watershed receives a mean annual precipitation of 834 millimeters (mm) (32.8 in) and has a mean annual temperature of 8.8°C (47.8°F). Historical records (circa 1840) indicate substantial marshes in the area, and recent soils studies indicate that wetlands covered approximately 15 percent of the watershed.

The north and south branches of the basin, including the mainstem to Waukegan Harbor, comprise approximately 20 channel km (12.5 mi), excluding Yeoman Creek from the north. The mean channel width ranges from 4.4 to 6.7 m (14.6 – 21.9 ft) and has a mean depth from 0.07 – 0.28 m (0.23 – 0.92 ft) (White et al. 2003). There is more substantial shading from riparian vegetation in the North Branch subwatershed than in the south. The South Branch has a greater discharge than the North Branch, approximately 0.1 cubic meters per second (cms) (3.4 cfs) versus 0.01 cms (0.4 cfs). Dominant substrate types range from sand to large cobble and boulder, with some bedrock. Within the project area, there was one control monitoring site (S2) and three sites where stream restoration was carried out (S1, N1, N2) (Figure CS4-2).

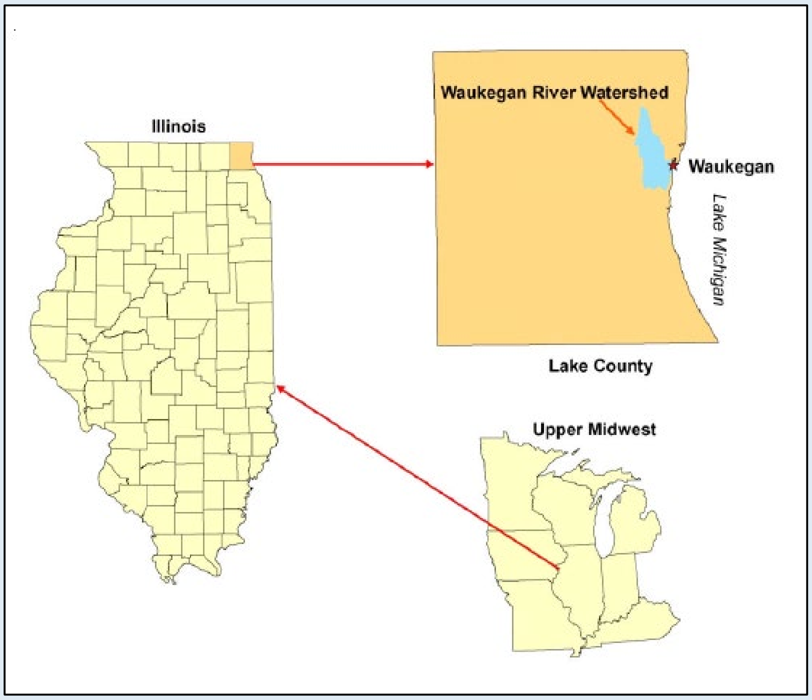


Figure CS4-1. The Waukegan River watershed in northeastern Illinois (White et al. 2010)

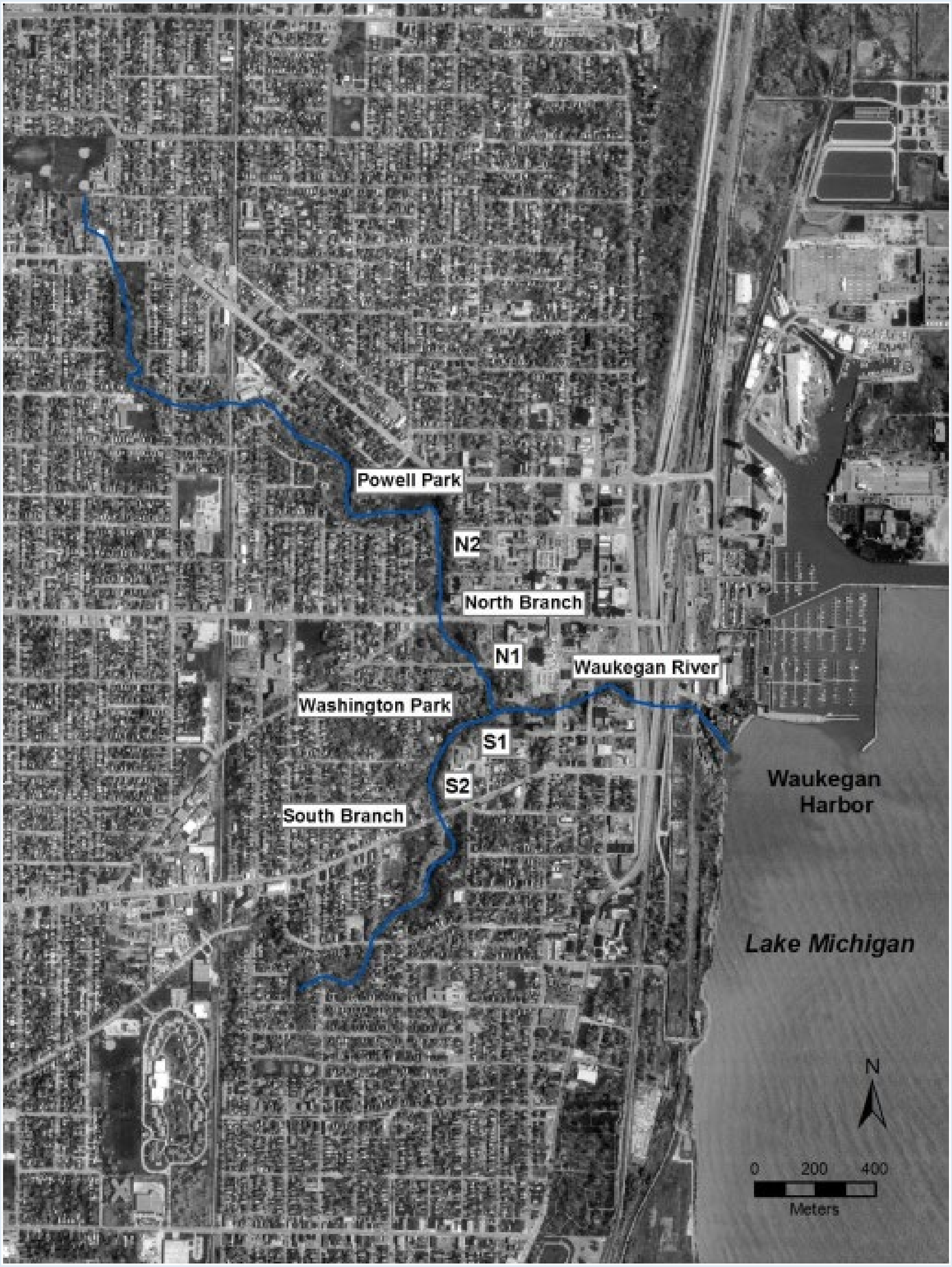


Figure CS4-2. Aerial map taken in 1998 of a portion of the Waukegan River watershed, showing sampling locations and restoration project areas (White et al. 2010)

As of 2005, major land uses in the Waukegan River watershed included approximately 36.7 percent residential; 21.5 percent transportation; 12 percent commercial, retail, government and institutional; and about 20 percent open space, forest, grasslands, and beaches. The remaining land uses are associated with small amounts of disturbed lands (3.6 percent), industrial (2.8 percent), wetlands/water (2.0 percent), and communication/ utilities (1.7 percent) (Lake County SWMC 2008). Approximately 80 percent of the densely urbanized City of Waukegan (2014 population ~89,000) lies within the watershed.

The pace of development and sprawl of Waukegan was substantial throughout the latter part of the 19th century, and until the 1970s and 1980s, when current stormwater regulations began to take effect. Not surprisingly, streams were heavily impacted by urbanization. Minimal management of the stormwater quantity and quality led to flashy stormwater runoff conditions and elevated pollutant loads. Additional water resource issues associated with urban development included combined sewer overflows (CSO), stream channel instability (accelerated vertical and lateral erosion processes), nutrient enrichment, and contamination by metals, pesticides/herbicides, pharmaceuticals and personal care products (PPCPs), and endocrine disruptors (White et al. 2003, 2010). Channel erosion processes were accelerated by flashy stormflows, contributing to degraded physical habitat and decreased capacity of the stream to support the survival and reproduction of stream biota.

Monitoring and Sampling Design

The overall restoration goal in the North Branch and South Branch was to rehabilitate physical habitat and hydrologic conditions to support recovery of benthic macroinvertebrate and fish assemblages (White et al. 2010). Project leaders chose to design and install biotechnical stabilization, a combination of stream bank physical stabilization and riparian re-vegetation, to address the severe channel instability and erosion problems in Washington Park and Powell Park (Figure CS4-2). The objectives of these techniques are complementary. A stream channel experiencing severe lateral and vertical erosion (mass-wasting and down-cutting, respectively), by definition, is losing habitat suitable for stream biota. Damaged or missing riparian vegetation results in diminished root mass to hold soil together, lowered inputs of leaf litter and woody materials (food source and habitat structure), and less shading, which can lead to warmer water and increased photosynthetic activity and algal growth. Combinations of LUNKERS[1], a-jacks, stone, coconut fiber rolls, dogwoods, willows, and grasses were installed at selected locations on the North Branch/Powell Park (1992–93) and on the South Branch/Washington Park (1995).

There were four sample locations, two each on the South Branch and the North Branch (Figure CS4-2). For the South Branch, station S2 was the upstream control reach, and S1 was the downstream treatment reach. On the North Branch, the two sample locations (N1 and N2) were located to coincide with treatment reaches. Wooden LUNKERS were used as the principal rehabilitation feature at N1, while recycled plastic lumber and concrete a-jacks were used for LUNKERS construction at N2. All four locations were sampled annually during spring, summer, and fall over a 13-year period (1994–2006).

This case study focuses solely on responses of benthic macroinvertebrates, although it should be recognized that fish were also evaluated for both branches. Benthic macroinvertebrates, physical habitat, and chemical water quality were sampled, measured, and characterized at each stream location. Three macroinvertebrate samples were taken at each location using a Hess sampler with

[1] 'Little Underwater Neighborhood Keepers Encompassing Rheotaxic Salmonids' (Vetrano 1988)

a 500 micron mesh net. Sample material was preserved in 95 percent ethanol; organisms were sorted to segregate individuals from non-target material, and subsequently identified to genus level.

Sampling data were used to calculate the Macroinvertebrate Biotic Index (MBI). The MBI is based on the Hilsenhoff Biotic Index (HBI; Hilsenhoff 1977, 1982) but uses an 11-point scale, rather than the HBI 5-point scale. Lower MBI scores indicate better or less-degraded water quality. Physical habitat was characterized using the Potential Index of Biological Integrity (PIBI) which incorporates percent substrate particle sizes, magnitude of sediment deposition, pool substrate quality, and substrate stability, a series of hydraulic and morphometric measures, riparian features, and various aspects of instream cover (White et al. 2003). The PIBI was calculated from measurements and field observations made on each of 10 equal-length segments established by the 11-transect method. The purpose of the PIBI is to help illuminate which habitat features, if any, might be limiting survival, growth, and reproduction of stream biota.

Results

Overall, among all four locations, MBI scores ranged from around 5 to just below 10 (good to very poor), with an average around 7.2, or “fair” (see Figure CS4-3). On the South Branch, station S2 exhibited the highest mean score for a single year, 7.5, indicating “fair” stream condition, slightly better than “poor.” MBI scores indicate worsening conditions over time at stations S1 and N1. There was virtually no change over the 13 years for stations S2 and N2, with average MBI scores in the “fair” and “poor” ranges.

All sites were dominated by stressor tolerant taxa, with sample data comprised of 82-89 percent non-biting midges (Insecta: Diptera: Chironomidae), segmented worms (Annelida: Oligochaeta), and aquatic sowbugs (Crustacea: Isopoda: Asellidae). The dominance of these animals in the North and South branches clearly shows stressed or degraded conditions before, during, and after any kind of habitat restoration or other remedial activities. Mean taxa richness (number of distinct taxa) over the sampling period was 10 for site S1, and 8 for the other three locations. Ninety-two percent of the samples fell in the “poor” or “very poor” categories. There were very low numbers of stressor-sensitive EPT taxa (mayflies [Ephemeroptera], stoneflies [Plecoptera], and caddisflies [Trichoptera]) throughout the monitoring period. This is generally indicative of elevated pollutant levels and greater degradation. Some improvement in physical habitat quality was observed at treatment stations S1 and N1, likely due to improvements in bank stability and decreases in overall proportions of percent fines, silt, and mud. The other treatment station, N2, which was bank-armored, remained relatively consistent over the full period of record, as did the non-treatment control, S2.

Improvement in physical habitat quality and overall biological diversity was achieved as a result of these restoration activities, but improvement in biodiversity, primarily relative to the fish assemblage (not discussed in this case study) was only temporary (White et al. 2010). The authors acknowledged that sustainable biological diversity in a damaged watershed will require more complete understanding of landscape and watershed processes, their degree of degradation, and a comprehensive approach to conservation that addresses the system in its entirety. In the case of the Waukegan River watershed, this calls for a systematic approach to correcting other sources of hydrologic and chemical water quality stressors associated with water and sewer management operations, channel and flow alterations, and extensive aquifer drawdown. One result of this project was initiation of a comprehensive watershed plan, selection of a coordinator, development of stakeholder and technical planning committees, and creation of a long-term action plan.

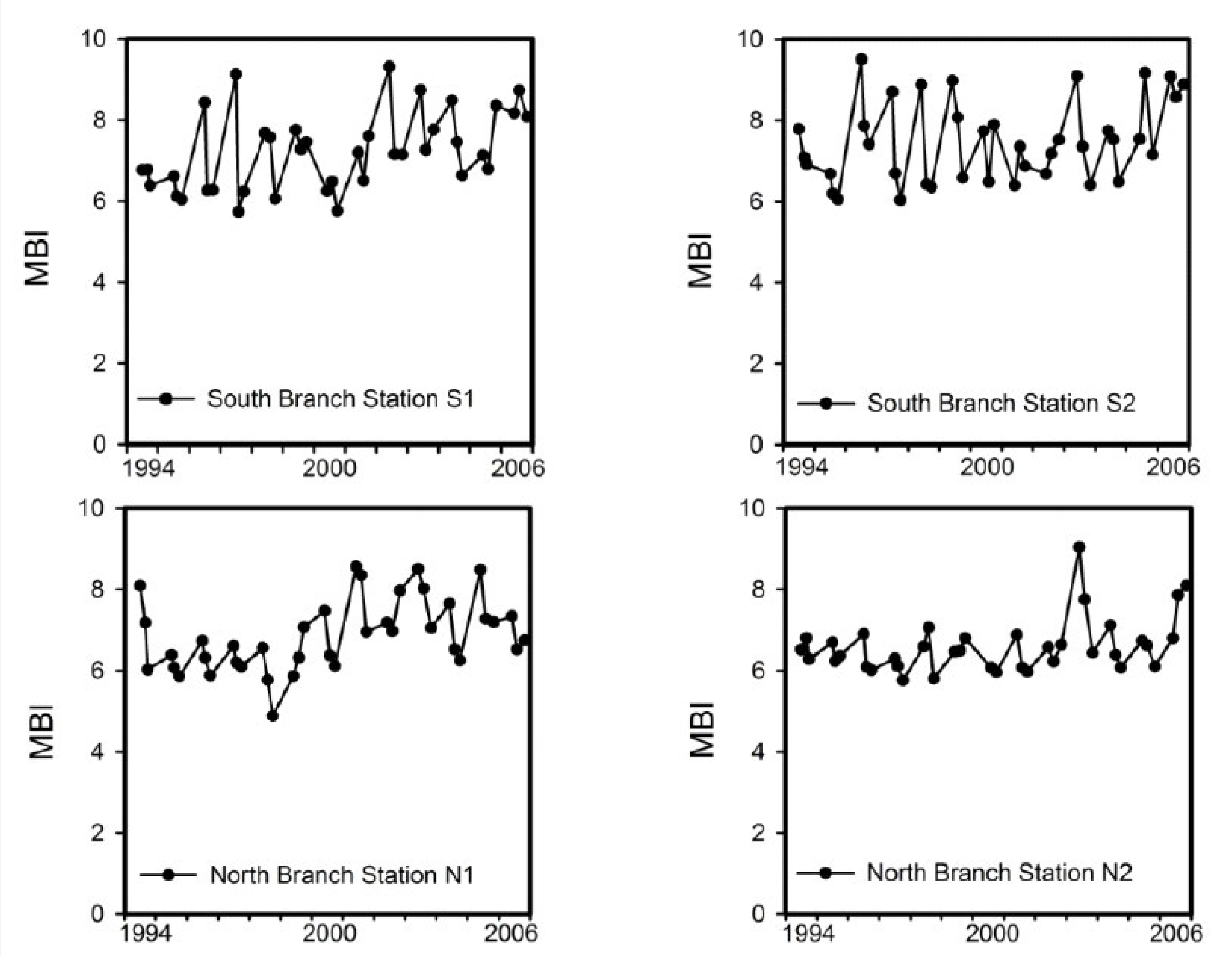


Figure CS4-3. MBI scores from monitoring stations in Waukegan River (White et al. 2010). Assessment classes (narrative ratings) for stream condition based on MBI scores are: very poor, 9.0-11.0; poor, 7.6-8.9; fair, 6.0-7.5; good, 5.0-5.9; very good, <5.0.

Literature

Hilsenhoff, W.L. 1977. *Use of Arthropods to Evaluate Water Quality of Streams.* Technical Bulletin 100. Wisconsin Department of Natural Resources. Madison, WI.

Hilsenhoff, W.L. 1982. *Using a Biotic Index to Evaluate Water Quality in Streams.* Technical Bulletin 132. Wisconsin Department of Natural Resources. Madison, WI.

Lake County SWMC. 2008. *Waukegan River Watershed, Lake County, Illinois. Lake County Stormwater Management Commission, Lake County Department of Information and Technology, GIS & Mapping Division. Map*. Accessed February 9, 2016.

Vetrano, D.H. 1988. *Unit Construction for Trout Habitat Improvement Structures for Wisconsin Coulee Streams.* Administrative Report No. 27. Wisconsin Department of Natural Resources, Madison, WI.

White, W.P., J.D. Beardsley, J.A. Rodsater, and L.T. Duong. 2003. *Biological and Physical Monitoring of Waukegan River Restoration Efforts in Biotechnical Bank Protection and Pool/Riffle Creation.* National Watershed Monitoring Project. Annual Report. *Prepared by:* Watershed Science Section, Illinois State Water Survey, Illinois Department of Natural Resources*. Prepared for:* Illinois Environmental Protection Agency and U.S. Environmental Protection Agency (Region 5).

White, W.P., J. Beardsley, and S. Tomkins. 2010. *Waukegan River, Illinois national nonpoint source monitoring program project. NWQEP NOTES. The NCSU Water Quality Group Newsletter*. No. 133. April 2010. ISSN 1062-9149. North Carolina State University Water Quality Group, Campus Box 7637, North Carolina State University, Raleigh, NC 27695-7637. Accessed February 9, 2016. http://www.ncsu.edu/waterquality/.

4.4.3.1.1 *Metric and index calculations*

Metrics are mathematical terms calculated directly from sample data, with resulting values scored relative to quantitative criteria. The Table 4-5 below presents an example of four different metric sets representing site classes (or bioregions) within a particular US State. Each set of either 5 or 6 metrics forms the basis of an MMI previously calibrated to wadeable streams of the class. Metric values result from direct calculations on raw sample data, taxonomic identifications and counts (list of taxa and number of individuals of each, by sample).

Table 4-4. Metrics and associated scoring formulas for four site classes from an example monitoring and assessment program

Metrics	Scoring formulas
Site class A	
1. Total taxa	100*(*metric value*)/51.5
2. Percent EPT individuals, as sensitive	100*(*metric value*)/39
3. Percent Coleoptera individuals, as sensitive	100*(*metric value*)/10.5
4. Beck's biotic index	100*(*metric value*)/31
5. Percent of taxa, as tolerant	100*(43-[*metric value*])/40
Site class B	
1. Total number of taxa	100*(*metric value*)/51.5
2. Number of EPT taxa	100*(*metric value*)/14
3. Percent individuals *Cricotopus/Orthocladius* + *Chironomus*, of total Chironomidae	100*(45-[*metric value*])/45
4. Percent EPT individuals, as sensitive	100*(*metric value*)/39
5. Number of taxa, as shredders	100*(*metric value*)/7
6. Hilsenhoff Biotic Index	100*(8.5-[*metric value*])/5
Site class C	
1. Total taxa	100*(*metric value*)/51.5)
2. Percent of taxa, as non-insects	100*(46-[*metric value*])/40
3. Percent individuals *Cricotopus/Orthocladius* + *Chironomus*, of total Chironomidae	100*(24-[*metric value*])/24
4. Percent of individuals, as filterers	100*(*metric value*)/70
5. Number of taxa, as sprawlers	100*(*metric value*)/14
6. Hilsenhoff Biotic Index	100*(8.5-[*metric value*])/5
Site class D	
1. Number of Oligochaeta taxa	100*(6-(*metric value*])/6
2. Percent EPT individuals, as sensitive	100*(*metric value*)/15
3. Percent individuals, as Crustacea and Mollusca	100*(*metric value*)/30
4. Percent individuals, as Odonata	100*(16.5-[*metric value*])/16.5
5. Number of taxa, as collectors	100*(20-[*metric value*])/19.5
6 Percent individuals, as swimmers	100*(12-[*metric value*])/12

Many metrics require assigning different characteristics or traits to taxa in the dataset prior to calculation. These characteristics include functional feeding groups (FFGs), habit, and stressor tolerance values. Many states use various literature resources to develop traits databases (e.g., Merritt et al. 2008, Barbour et al. 1999, Vieira et al. 2006, Carter and Resh 2013).

The formulas in the table resulted from the calibration process, and serve to convert the calculated metric value to a normalized, unitless score on a 100-point scale. The multiple metric values are then combined for each sample by simple averaging. Additionally, formulas are developed, in part, so that individual metrics are scored depending on their direction of change in the presence of stressors.

4.4.3.1.2 *Quality control measure*

Metric calculations are typically performed in spreadsheets or relational databases with embedded queries. To ensure that resulting calculations are correct and provide the intended metric values, a subset of them should be recalculated by hand. A reliable approach is to calculate a) one metric across all samples, followed by b) all metrics for one sample. When recalculated values differ from those values in the output matrix, reasons for the disagreement are determined and corrections are made. Reports on performance include the total number of reduced values as a percentage of the total, how many errors were found in the queries, and the corrective actions specifically documented.

4.4.3.2 Predictive models (observed/expected [O/E])

Predictive models are based on the premise that the taxa occurring in a minimally disturbed system can be predicted based on multiple measures of the environmental setting and that if the predicted taxa are not observed in an evaluation site, then disturbance can be suspected. The ratio of the number of observed taxa to that expected to occur in the absence of human-caused stress is an intuitive and ecologically meaningful measure of biological integrity. Low observed-to-expected ratios (O/E << 1.0) imply that test sites are adversely affected by some environmental stressor. The models are commonly called RIVPACS models (River Invertebrate Prediction And Classification System [Wright 1995]) based on observed:expected taxa (Clarke et al. 1996, 2003, Hawkins et al. 2000b). Because they are based on taxa in reference sites, the predictive models are not well suited to assemblages with naturally low diversity (as in oligotrophic fish communities). The loss of reference taxa is difficult to detect when only few taxa are expected.

The number of taxa expected at a site is calculated as the sum of individual probabilities of capture for all taxa found in reference sites in the region of interest. All probabilities greater than a designated threshold are summed to calculate the expected number of taxa (E), and this number is compared to the reference taxa observed (O) at a site. Because these models predict the actual taxonomic composition of a site, they also provide information about the presence or absence of specific taxa. If the sensitivities of taxa to different stressors are known, this information can lead to derived indices and diagnoses of the stressors most likely affecting a site. In addition, taxa can be identified as increasers or decreasers with respect to general environmental stress encountered in the model development data set. A variation of the O/E models measures the Bray-Curtis compositional dissimilarity between an observed and expected assemblage directly, which detects stress-induced shifts in taxonomic composition that leave assemblage richness unchanged (Van Sickle 2008).

The steps that go into building a predictive model include 1) classifying reference sites into biologically similar groups, 2) creating discriminant functions models to estimate group membership of sites from environmental data, 3) establishing taxon-specific probabilities of capture for individual sites,

4) identifying taxa expected and comparing to those observed, 5) estimating model error, and 6) applying the model in test sites. These steps can be automated to allow exploration of model performance for multiple subsets of environmental predictor variables.

4.4.3.3 Quantitative decision analysis systems (biological condition gradient [BCG])

The Biological Condition Gradient (BCG) was described by USEPA (2011) as "a conceptual model that describes how biological attributes of aquatic ecosystems might change along a gradient of increasing anthropogenic stress." The model can serve as a template for organizing field data (biological, chemical, physical, landscape) at an ecoregional, basin, watershed, or stream segment level. The BCG was developed by EPA and other agencies to support tiered aquatic life uses in state water quality standards and criteria. It was developed through a series of workshops, and described fully by Davies and Jackson (2006). The BCG includes a narrative description of ecological condition that can be translated across regions, assemblages, and assessment programs. The descriptions recognize six levels of quality in ecological condition ranging from "1" (most desirable) where natural structural, functional, and taxonomic integrity is preserved to "6" (least desirable) in which there are extreme changes in structure and ecosystem function and wholesale changes in taxonomic composition.

The quantitative decision analysis systems approach explicitly uses the BCG as a scale for biological assessment. It differs from the multimetric and predictive model approaches in that it is not dependent on definition of reference sites (although that can be useful), and development relies on consensus of experts instead of an individual or a few analysts. It is similar to the multimetric approach in its reliance on distinct site classes. It is similar to the predictive modeling approach in its examination of individual taxa (though metrics are also incorporated in the models).

Calibrating a BCG to local conditions begins with the assembly and analysis of biological monitoring data. Following data assembly, a calibration workshop is held in which experts familiar with local biotic assemblages of the region review the data and the general descriptions of each of the BCG levels. The expert panel then uses the data to define the ecological attributes of taxa, and to develop narrative statements of BCG levels based on sample taxa lists. The expert panel is usually convened multiple times to refine decisions, to react to interim results, and to assign BCG levels to new sites. The steps typically taken during a calibration workshop include the following:

1) An overview presentation of the BCG and the process for calibration;
2) A "warm-up" data exercise to further familiarize participants with the process;
3) Assignment of taxa to BCG taxonomic attributes (based on known tolerance and rarity);
4) Description of biota in undisturbed conditions (best professional judgment [BPJ]; regardless of whether such conditions still exist in observed reference sites);
5) Assignment of sites in the data set to BCG levels; and
6) Elicitation of rules used by participants in assigning sites to levels.

Documentation of expert opinion in assigning attributes to taxa and BCG levels to sites is a critical part of the process. Facilitators elicit from participants sets of operational rules for assigning levels to sites. As the panel assigns example sites to BCG levels, the members are polled on the critical information and criteria they used to make their decisions. These form preliminary, narrative rules that explained how panel members make decisions. Rule development requires discussion and documentation of BCG level assignment decisions and the reasoning behind the decisions. During these discussions, records are kept on each participant's decision ("vote") for the site; the critical or most important information for the decision; and any confounding or conflicting information and how this is resolved.

A decision model is then developed that encompasses the taxa attributes and quantitatively replicates the rules used by the expert panel in assigning BCG levels to sites. The decision model is tested with independent data sets as a validation step. A quantitative biological assessment program can then be developed using the rule-based model for consistent decision-making in water quality management.

The decision analysis models can be based on mathematical fuzzy-set theory (citation) to replicate the expert panel decisions. Such models explicitly use linguistic rules or logic statements, e.g., "If taxa richness is high, then condition is good" for quantitative, computerized decisions. The models can usually be calibrated to closely match panel decisions in most cases, where "closely matched" means the model either exactly matched the panel, or selected the panel's minority decision as its level of greatest membership. The decision analysis models can also be cross calibrated to other assessment tools, such as the MMI. Models can be developed as spreadsheet tools to facilitate programmatic application.

4.4.4 Index scoring and site assessment

The site-specific MMI score, as calculated above in section 4.4.3, is compared to degradation thresholds (Table 4-6) to determine whether biological degradation exists relative to minimally degraded reference conditions (Barbour et al. 1999, Stoddard et al. 2006). The range of potential scores in Table 4-6 is 0 (most degraded) to 100 (least degraded). The 90 percent confidence intervals (CI90) are calculated using sample repeats (see section 4.4.2.4). Defining the numeric values of degradation thresholds is an integral phase of index calibration, and is affected by regional and climatic conditions, along with the overall level and consistency of landscape alteration and available data to characterize the broad range of degradation.

Table 4-5. Degradation thresholds to which MMI score are compared for determination of status

Site class	Degradation threshold
A	52.3
B	65.7
C	66.0
D	55.9

The confidence interval (CI), also known as detectable difference (DD) (Stark 1993), is associated with individual MMI scores and represents the magnitude of separation between two values before they can be considered truly different (Stribling et al. 2008a). Reported values falling below the threshold are considered degraded, those above are non-degraded, while site index values falling near a threshold may require additional samples to determine final rating category (Stribling et al. 2008a, Zuellig et al. 2012). Some programs, if not most, also subdivide value ranges above and below the degradation threshold to allow communication of multiple levels of non-degradation and degradation, e.g., very good, good, fair, poor, or very poor.

4.4.5 Reporting assessment results at multiple spatial scales

Depending on the monitoring design, it is possible to use assessment results for several purposes, some of which may have been previously unanticipated. For example, a probability-based design provides assessments that can be aggregated for assessments broader than the individual location from which a sample was taken (Larsen 1997, Urquhart et al. 1998). Simultaneously, each sample from that kind of design provides information useful for interpreting conditions at individual sample locations. Assessments from a targeted design provide information about the sites sampled, and although they cannot be used for broader scale assessments, can assist with confirming the effects of known stressors and stressor sources.

4.4.5.1 Watershed or area-wide

For programs using a stratified random monitoring design, a simple inference model similar to that described by Olsen and Peck (2008) and Van Sickle and Paulsen (2008), can be used to estimate the number of degraded stream miles (D) for a watershed or area-wide region with the formula:

$$D = (N/T)x\ L$$

where:

> N is the number of sites rated by the MMI as degraded,
>
> T is the total number of sites assessed for the sampling unit (subwatershed or watershed group), and
>
> L is the total number of stream miles in the sampling unit.

Total stream channel miles (L) should be estimated with GIS using the National Hydrography Dataset (NHD), or other stream data layer appropriate to the watershed or region of interest. Note that replicate samples taken for QC purposes are not included in these calculations. Results can also be presented as percent degradation (%D) by using the calculation:

$$\%D = (N/T)x100$$

For the Lake Allatoona/Upper Etowah River watershed, site selection and monitoring was stratified by the 53 HUC subwatersheds, and cumulative assessments showed distinctive patterns of degradation (Figure 4-6). More intensive development and imperviousness are closer to transportation corridors.

Trends in %D over time can be evaluated using test such as the Kendall tau test (Helsel and Hirsch 2002). It should be noted that for very small sample sizes (i.e., 3 or 4), all values would need to be consecutively decreasing to reject a one-sided null hypothesis with a p equal to 0.167 and 0.042, respectively.

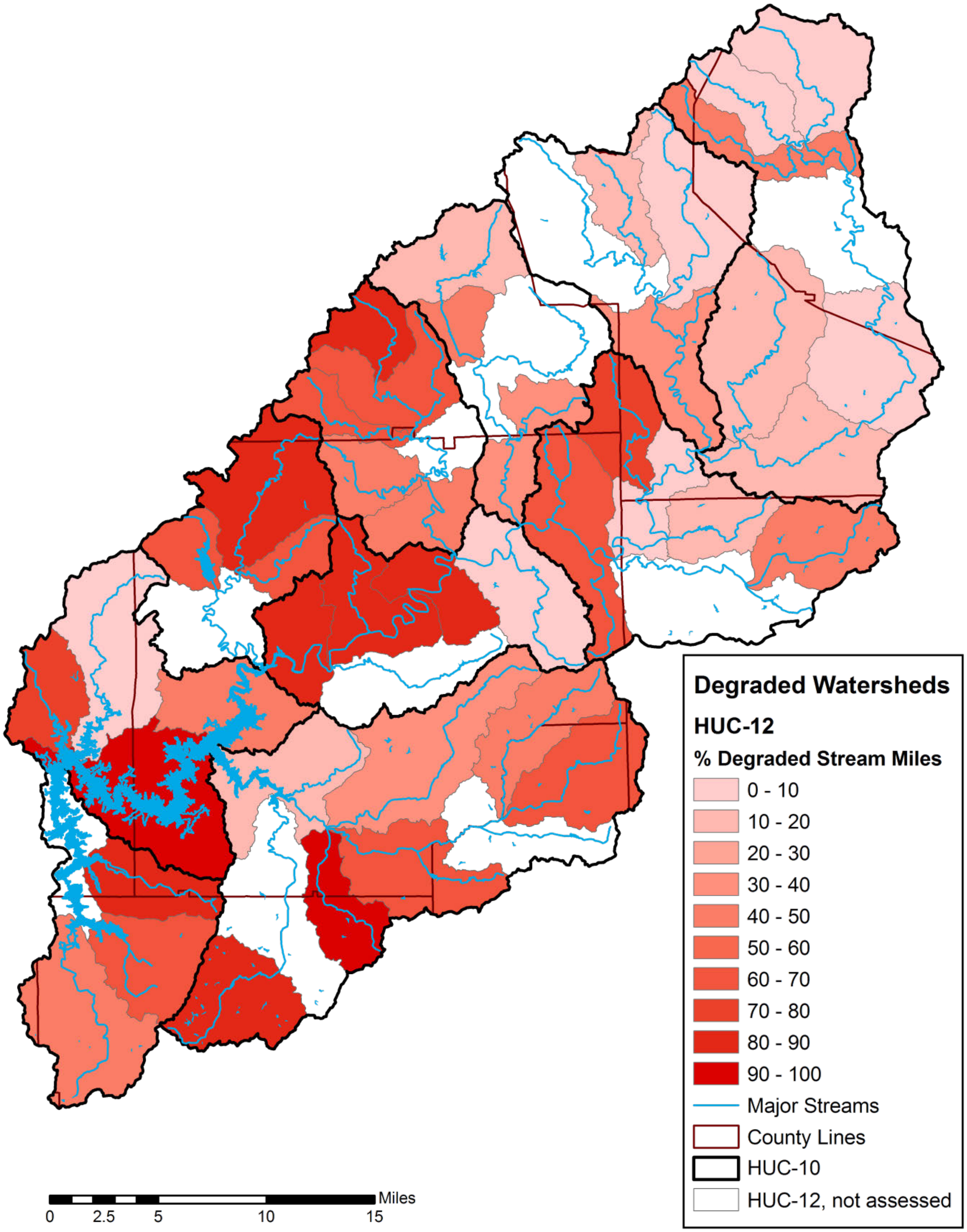


Figure 4-6. Percent degradation of subwatersheds as measured by biological monitoring and assessment, Lake Allatoona/Upper Etowah River watershed (Millard et al. 2011)

4.4.5.2 Stream- or site-specific

The MMI scores and status ratings are approaches useful for summarizing and communicating site specific conditions; this is the application for which they are designed, and for which they are ideally suited. However, a well-organized and functional database allows the index to be disaggregated where individual metrics and even taxa can be evaluated by biologists to help determine those which are most influencing an assessment. Presenting site-specific point assessments from the previous example (Figure 4-7) shows the specific distribution of the most- and least-degraded streams, and more detailed examination can begin to reveal proximity of potential stressor sources (Figure 4-8). At this stage of evaluating watershed-based stream assessments, if necessary, the assessor can turn to the USEPA stressor identification process, also known as "The Causal Analysis/Diagnosis Decision Information System", or CADDIS (http://www.epa.gov/caddis/), to assist in determining the most probable causes of biological degradation. It is using this process, including evaluating the relative dominance of the various taxa that taxon-specific environmental requirements, stressor tolerances, feeding types, and habits, which can lead to more defensible decisions on stressor control actions, such as BMPs or stream/watershed restoration activities. MMI confidence intervals can be computed and used for point comparisons in the same manner as other water quality variables (see section 7.3).

4.4.5.3 Relative to specific sources

Monitoring objectives requiring documentation of instream biological condition relative to a specific and known source of stressors require that sample data be drawn from one or more locations exposed to those stressors. In particular, if the source is an area of specific land use, a type of BMP, or a point source, confidence in the result will likely be enhanced by a thorough and quantitative description of the source. However, it should be recognized that lack of a clear site-specific response to either measured or assumed exposure to stressors does not mean that the biota are non-responsive. Neither does it mean that the BMP is ineffective. The BMP can likely be proven effective at reducing the single or multiple stressors for which it is intended and designed.

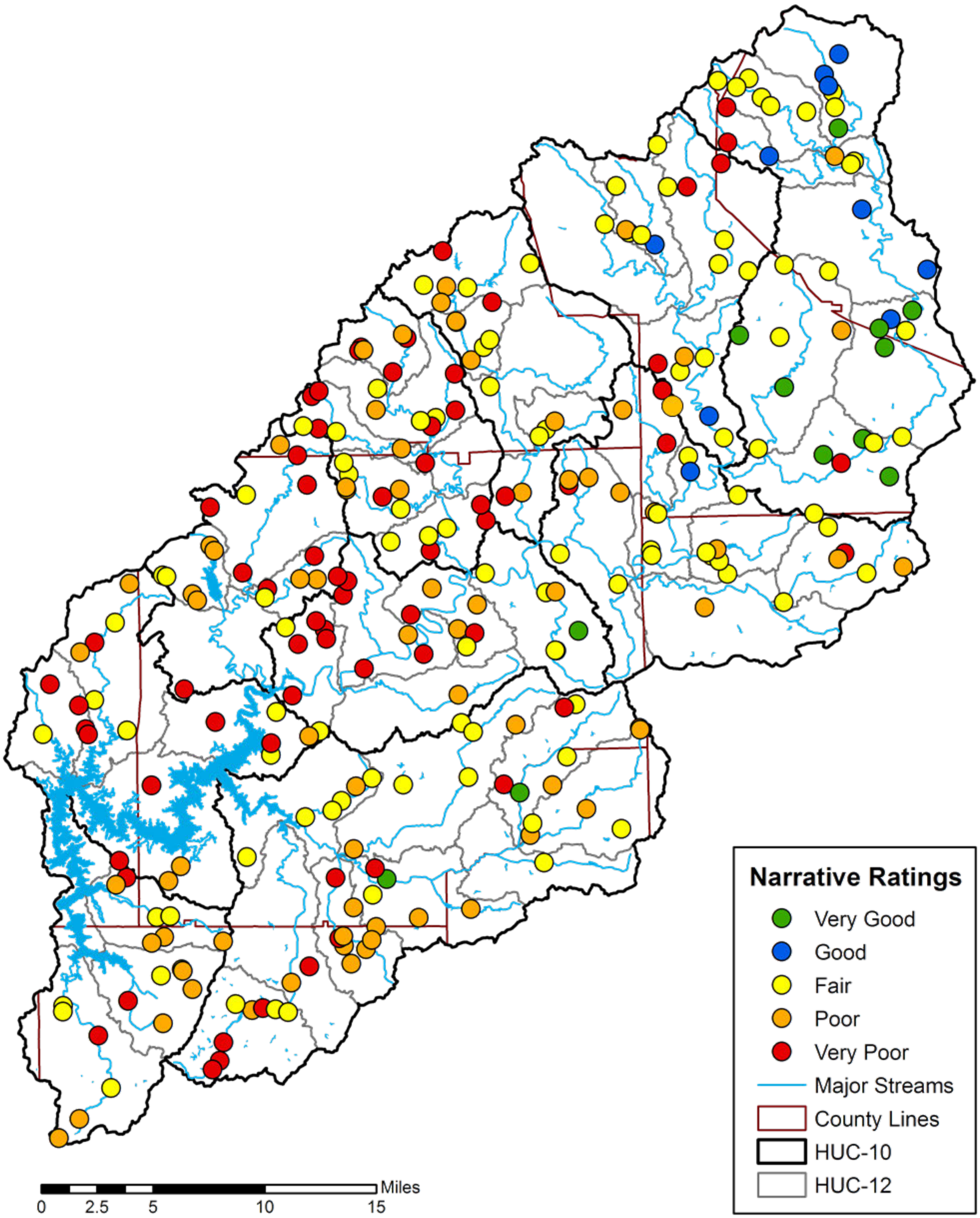


Figure 4-7. Distribution of stream biological assessments in the Lake Allatoona/Upper Etowah River watershed, using a benthic MMI developed by the Georgia Environmental Protection Division (Millard et al. 2011)

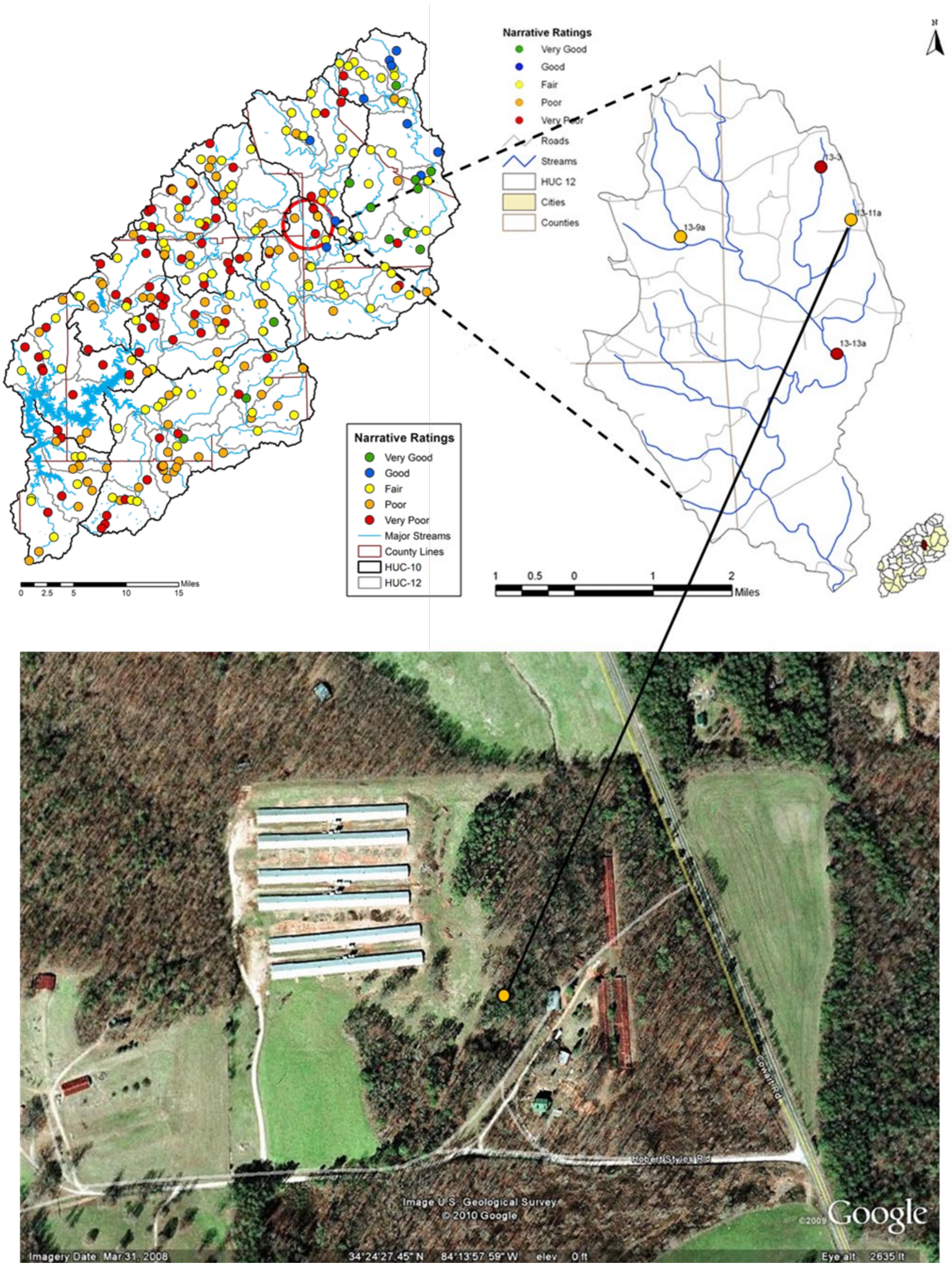


Figure 4-8. More detailed examination of the Yellow Creek subwatershed, Lake Allatoona/Upper Etowah River watershed, Georgia, reveals a sample location, rated biologically as “poor,” is on a stream flowing through a poultry production operation (Millard et al. 2011)

4.5 References

Allen, D.M., S.K. Service, and M.V. Ogburn-Matthews. 1992. Factors influencing the collection efficiency of estuarine fishes. *Transactions of the American Fisheries Society* 121(2): 234–244.

Angermeier, P., and J. Karr. 1994. Biological integrity versus biological diversity as policy directives: protecting biotic resources. *BioScience* 44(10): 690–697.

Bahls, L. 1993. *Periphyton Bioassessment Methods for Montana Streams*. Department of Health and Environmental Science, Water Quality Bureau, Helena, MT.

Ball, J. 1982. *Stream Classification Guidelines for Wisconsin*. Technical Bulletin. Wisconsin Department of Natural Resources, Madison, WI.

Barbour, M.T., and J.B. Stribling. 1991. Use of Habitat Assessment in Evaluating the Biological Integrity of Stream Communities. In *Biological Criteria: Research and Regulation Proceedings of a Symposium*. EPA-440/5-91-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*. 2nd ed. EPA/841-D-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 9, 2016. http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm.

Barbour, M.T., J.B. Stribling, and P.F.M. Verdonschot. 2006. The multihabitat approach of USEPA's rapid bioassessment protocols: benthic macroinvertebrates. *Limnetica* 25(3–4): 229–240.

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* 37(2): 114–118.

Cao, Y., and C.P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. *Journal of Applied Ecology* 42: 954–965.

Carter, J.L., and V.H. Resh. 2013. *Analytical Approaches used in Stream Benthic Macroinvertebrate Biomonitoring Programs of State Agencies in the United States*. Open File Report 2013–1129. U.S. Geological Survey, Reston, VA. Accessed April 22, 2016. http://pubs.usgs.gov/of/2013/1129/pdf/ofr20131129.pdf.

Caton, L. R. 1991. Improved subsampling methods for the EPA rapid bioassessment protocols. *Bulletin of the North American Benthological Society* 8: 317–319.

Charles, D.F., C. Knowles, and R.S. Davis, ed. 2002. *Protocols for the Analysis of Algal Samples Collected as Part of the U.S. Geological Survey National Water-Quality Assessment Program*. Report No. 02-06. Prepared for U.S. Geological Survey, by The Academy of Natural Sciences, Patrick Center for Environmental Research–Phycology Section, Philadelphia, PA.

Clarke, R.T., M.T. Furse, J.F. Wright, and D. Moss. 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics* 23: 311–332.

Clarke, R.T., J.F. Wright, and M.T. Furse. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modeling* 160: 219–233.

Conquest, L.L., S.C. Ralph, and R.J. Naiman. 1994. Implementation of Large-Scale Stream Monitoring Efforts: Sampling Design and Data Analysis Issues. In *Biological Monitoring of Aquatic Systems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Boca Raton, FL.

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe. 1979. *Classification of Wetlands and Deepwater Habitats of the United States*. FWS/OBS-79/31. U.S. Department of the Interior, U.S. Fish and Wildlife Service, Biological Services Program, Washington, DC.

Cowie, G.M., J.L. Cooley, and A. Dutt. 1991. *Use of Modified Benthic Bioassessment Protocols for Evaluation of Water Quality Trends in Georgia*. Technical Completion Report USDI/GS Project 14-08-0001-G1556 (03); ERC-06-91; USGS/G-1556-03. University of Georgia, Institute of Community and Area Development, Athens, GA, and Georgia Institute of Technology, Environmental Resources Center, Atlanta, GA.

Cummins, K.W. 1994. Bioassessment and Analysis of Functional Organization of Running Water Ecosystems. In *Biological Monitoring of Aquatic Ecosystems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Ann Arbor, MI.

Cupp, C.E. 1989. *Stream Corridor Classification for Forested Lands of Washington*. Washington Forest Protection Association, Olympia, WA.

Dauble, D.D., and R.H. Gray. 1980. Comparison of a small seine and a backpack electroshocker to evaluate near shore fish populations in rivers. *Progressive Fish-Culturist* 42: 93–95.

Davies, S.P., and S.K. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16(4): 1251–1266.

Day, J.W., Jr., C.A.S. Hall, W.M. Kemp, and A. Yáñnez-Arancibia. 1989. *Estuarine Ecology*. John Wiley and Sons, Inc., New York, NY.

Dewey, M.R., L.E. Holland-Bartel, and S.T. Zigler. 1989. Comparison of fish catches with buoyant pop nets and seines in vegetated and nonvegetated habitats. *North American Journal of Fisheries Management* 9: 249–253.

Flotemersch, J.E., K.A. Blocksom, J.J. Hutchens, and B.C. Autrey. 2006a. Development of a standardized large river bioassessment protocol (LR-BP) for macroinvertebrate assemblages. *River Research and Applications* 22: 775–790.

Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006b. Concepts and Approaches for the Bioassessment of Non-Wadeable Streams and Rivers. EPA/600/R-06/127. U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Cincinnati, OH. Accessed February 10, 2016.

Flotemersch, J.E., J.B. Stribling, R.M. Hughes, L. Reynolds, M.J. Paul, and C. Wolter. 2011. Reach length for biological assessment of boatable rivers. *River Research and Applications* 27(4): 520–535.

Fore, L., and C.O. Yoder. 2003. *The Design of a Biological Community Trend Monitoring Program for Michigan Wadeable Streams* (DRAFT). MI/DEQ/WD-03/086. Prepared for Michigan Department of Environmental Quality, by Statistical Design and Midwest Biodiversity Institute.

Frissell, C.A., W.J. Liss, C.E. Warren, and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* 10(2): 199–214.

Gerritsen, J., J.S. White, E.W. Leppo, J.B. Stribling, and M.T. Barbour. 1996. *Examination of the Effects of Habitat Quality, Land Use, and Acidification on the Macroinvertebrate Communities of Coastal Plain Streams*. CBWP-MANTA-TR-95-2. Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD.

Gerritsen, J., M.T. Barbour, and K. King. 2000. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. *Journal of the North American Benthological Society* 19(3): 487–496.

Gibson, G.R., M.T. Barbour, J.B. Stribling, J. Gerritsen, and J.R. Karr. 1996. *Biological Criteria: Technical Guidance for Streams and Small Rivers*. EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

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

Gulland, J.A. 1983. *Fish Stock Assessment: A Manual of Basic Methods*. FAO/Wiley Series, Vol. 1. Wiley & Sons, New York, NY.

Hawkins, C.P. 2006. Quantifying biological integrity by taxonomic completeness: evaluation of a potential indicator for use in regional- and global-scale assessments. *Ecological Applications* 16:1277–1294.

Hawkins, C.P., J.L. Kershner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman, and M.K. Young. 1993. A hierarchical approach to classifying stream habitat features. *Fisheries* 18(6): 3–12.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, and R.J. Stevenson. 2000a. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society* 19(3): 541–556.

Hawkins, C.P., R.H. Norris, J.N. Hogue, and J.W. Feminella. 2000b. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10(5): 1456–1477.

Hayes, M.L. 1983. Active Fish Capture Methods. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 123–145. American Fisheries Society, Bethesda, MD.

Helsel, D.R., and R.M. Hirsch. 2002. Statistical Methods in Water Resources. Book 4, Chapter A3 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 10, 2016. http://pubs.usgs.gov/twri/twri4a3/.

Hill, B.H., A.T. Herlihy, P.R. Kaufmann, R.J. Stevenson, F.H. McCormick, and C.B. Johnson. 2000. Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19: 50–67.

Hill, B.H., A.T. Herlihy, P.R. Kaufmann, S.J. Decelles, and M.A. Vander Borgh. 2003. Assessment of streams of the eastern United States using a periphyton index of biotic integrity. *Ecological Indicators* 2: 325–338.

Hubert, W.A. 1983. Passive Capture Techniques. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 95–122. American Fisheries Society, Bethesda, MD.

Hughes, R.M. 1995. Defining Acceptable Biological Status by Comparing with Reference Conditions. In *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*, ed. W.S. Davis and T.P. Simons, pp. 31–47. Lewis Publishers. Boca Raton, FL.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.

Hughes, R.M., P.R. Kaufmann, A.T. Herlihy, T.M. Kincaid, L. Reynolds, and D.P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1618–1631.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2):187-211.

Jerald, A., Jr. 1983. Age Determination. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 301–324. American Fisheries Society, Bethesda, MD.

Karr, J.R. 1991. Biological integrity: a long neglected aspect of water resource management. *Ecological Applications* 1: 66–84.

Karr, J.R., and D.R. Dudley. 1981. Ecological perspective on water quality goals. *Environmental Management* 5: 55–68.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. *Assessing Biological Integrity in Running Waters: A Method and its Rationale*. Special publication 5. Illinois Natural History Survey, Champaign, IL.

Klemm, D.J., P.A. Lewis, F. Fulk, and J.M. Lazorchak. 1990. *Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. EPA-600-4-90-030. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Klemm, D.J., Q.J. Stober, and J.M. Lazorschak. 1992. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. EPA/600/R-92/111. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Klemm, D.J., J.M. Lazorchak, and P.A. Lewis. 1998. Benthic Macroinvertebrates. In *Environmental Monitoring and Assessment Program—Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams*, ed. J.M. Lazorchak, D.J. Klemm, and

D.V. Peck, pp. 147–182. EPA/620/R-94/ 004F. U.S. Environmental Protection Agency, Washington, DC.

Larsen, D.P. 1997. Sample survey design issues for bioassessment of inland aquatic ecosystems. *Human and Ecological Risk Assessment* 3(6): 979–991.

McCormick, F.H., R.M. Hughes, P.R Kaufmann, D.V. Peck, J.L. Stoddard, and A.T. Herlihy. 2001. Development of an index of biotic integrity for the mid-Atlantic highlands region. *Transactions of the American Fisheries Society* 130: 857–877.

Meador, M.R., T.F. Cuffney, and M.E. Gurtz. 1993. *Methods for Sampling Fish Communities as Part of the National Water Assessment Program*. Open File Report 93-104. U.S. Geological Survey, Raleigh, NC.

Merritt, R.W., K.W. Cummins, and M.B. Berg, ed. 2008. *An Introduction to the Aquatic Insects of North America*. 4th ed. Kendall/Hunt Publishing Company, Dubuque, IA. ISBN 978-0-7575-5049-2.

Millard, C.J., N. Jokay, C. Wharton, and J.B. Stribling. 2011. *Ecological Assessment of the Lake Allatoona/Upper Etowah River Watershed - Year 5 of the Long-Term Monitoring Program, 2009–2010*. Prepared for Lake Allatoona/Upper Etowah River Watershed Partnership, Canton, GA, by Tetra Tech, Inc., Owings Mills, MD, and Atlanta, GA.

Moulton, S.R. II, J.G. Kennen, R.M. Goldstein, and J.A. Hambrook. 2002. *Revised Protocols for Sampling Algal, Invertebrate, and Fish Communities as Part of the National Water-Quality Assessment Program*. Open File Report 02-150. U.S. Geological Survey, Reston, VA.

Nelson, J.S., E.J. Crossman, H. Espinosa-Perez, L.T. Findley, C.R. Gilbert, R.N. Lea, and J.D. Williams. 2004. *Common and Scientific Names of Fishes from the United States, Canada, and Mexico*. Special Publication 29. American Fisheries Society, Bethesda, MD.

OEPA (Ohio Environmental Protection Agency). 1987. *Biological Criteria for the Protection of Aquatic Life: Volumes I-III*. Updated 1988, 1989, and 2015. Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, OH. Accessed February 10, 2016. http://www.epa.state.oh.us/dsw/bioassess/BioCriteriaProtAqLife.aspx.

Olsen, A.R., and D.V. Peck. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. *Journal of the North American Benthological Society* 27(4): 822–836.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish*. EPA/440/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Platts, W.S., W.F. Megahan, and G.W. Minshall. 1983. *Methods for Evaluating Stream, Riparian, and Biotic Conditions*. General Technical Report INT-138. U.S. Department of Agriculture, Forest Service, Ogden, UT.

Raven, P.J., N.T.H. Holmes, F.H. Dawson, P.J.A. Fox, M. Everard, I.R. Fozzard, and K.J. Rowen. 1998. *River Habitat Quality: the physical character of rivers and streams in the UK and Isle of Man*. Environment Agency, Bristol, England. ISBN1 873760 42 9.

Richards, C., and G.W. Minshall. 1992. Spatial and temporal trends in stream macroinvertebrate communities: the influence of catchment disturbance. *Hydrobiologia* 241(3): 173–184.

Skalski, J.R., and D.H. McKenzie. 1982. A design for aquatic monitoring programs. *Journal of Environmental Management* 14: 237–251.

Southwood, T.R.E. 1977. Habitat, the templet for ecological strategies? *Journal Animal Ecology* 46: 337–365.

Stark, J.D. 1993. Performance of the macroinvertebrate community index: effects of sampling method, sample replication, water depth, current velocity, and substratum on index values. *New Zealand Journal of Marine and Freshwater Research* 27: 463–478.

Stevens, D.L., and A.R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99(465): 262–278.

Stoddard, J. L., D.P. Larsen, C.P. Hawkins, R.K. Johnson, and R.H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16(4): 1267–1276.

Stribling, J.B. 2011. Partitioning Error Sources for Quality Control and Comparability Analysis in Biological Monitoring and Assessment. Chapter 4 in *Modern Approaches to Quality Control*, ed. A.B. Eldin, pp. 59–84. INTECH Open-Access Publisher. ISBN 978-953-307-971-4. doi: 10.5772/22388. Accessed February 10, 2016. http://www.intechopen.com/books/modern-approaches-to-quality-control/partitioning-error-sources-for-quality-control-and-comparability-analysis-in-biological-monitoring-a.

Stribling, J.B., B.K. Jessup, and D.L. Feldman. 2008a. Precision of benthic macroinvertebrate indicators of stream condition in Montana. *Journal of the North American Benthological Society* 27(1): 58–67.

Stribling, J.B., K.L. Pavlik, S.M. Holdsworth, and E.W. Leppo. 2008b. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society* 27(4): 906–919.

Teply, M., and L. Bahls. 2005. *Diatom Biocriteria for Montana Streams*. Prepared for Montana Department of Environmental Quality, by Larix Systems, Inc., Helena, MT. 2005.

Teply, M., and L. Bahls. 2007. *Statistical Evaluation of Periphyton Samples from Montana Reference Streams*. Prepared for Montana Department of Environmental Quality, by Larix Systemes, Inc., Helena, MT. 2007.

Thornton, K.W., B.L. Kimmer, and F.E. Payne, ed. 1990. *Reservoir Limnology: Ecological Perspectives*. John Wiley and Sons, Inc., New York, NY.

Thorp, J.H., and A.P. Covich, ed. 2010. *Ecology and Classification of North American Freshwater Invertebrates*. Academic Press (Elsevier).

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8(2): 246–257.

USEPA (U.S. Environmental Protection Agency). 2002. *Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers*. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2004. *Wadeable Stream Assessment: Benthic Laboratory Methods*. EPA 841-B-04007. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2009. *National Rivers and Streams Assessment: Field Operations Manual*. EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2011. *A Primer on Using Biological Assessments to Support Water Quality Management*. EPA-810-R-11-01. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2012. *National Rivers and Streams Assessment 2013-2014: Laboratory Operations Manual*. EPA-841-B-12-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2013. *Biological Assessment Program Review: Assessing Level of Technical Rigor to Support Water Quality Management*. EPA-820-R-13-001. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USFWS (U.S. Fish and Wildlife Service). 1991. *Principles and Techniques of Electrofishing*. U.S. Fish and Wildlife Service, Fisheries Academy, Office of Technical Fisheries Training, Kearneysville, WV.

Van Sickle, J. 2008. An index of compositional dissimilarity between observed and expected assemblages. *Journal of the North American Benthological Society* 27(2): 227–235.

Van Sickle, J., and S.G. Paulsen. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. *Journal of the North American Benthological Society* 27(4): 920–930.

Vieira, N.K.M., N.L. Poff, D.M. Carlisle, S.R. Moulton II, M.K. Koski, and B.C. Kondratieff. 2006. A Database of Lotic Invertebrate Traits for North America. Data Series 187. U.S. Geological Survey, Reston, VA. Accessed February 10, 2016. http://pubs.water.usgs.gov/ds187.

Weatherley, A.H. 1972. *Growth and Ecology of Fish Populations.* Academic Press, New York, NY.

Wetzel, R.G. 1983. *Limnology*. Saunders College Publishing, New York, NY.

Wright, J.F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology* 20:181–197.

Zuellig, R.E., D.M. Carlisle, M.R. Meador, and M. Potapova. 2012. Variance partitioning of stream diatom, fish, and invertebrate indicators of biological condition. *Freshwater Science* 31(1): 182–190. doi: 10.1899/11-040.1.

This page intentionally left blank.

5 Photo-Point Monitoring

By S.A. Dressing and D.W. Meals

5.1 Introduction

Good photographs can yield much information, and photography can play important roles in watershed projects because the technology is available to everyone, training is simple, and the cost is relatively low (USEPA 2008, ERS 2010). In the assessment phase, photos can help identify problem areas both within the water resource (e.g., algal blooms, streambank erosion) and within the drainage area (e.g., cattle in streams, discharge pipes). These same photos can also be very helpful in generating interest in the project because they can convey easily understood information to a wide audience. In addition, photos can be used to document implementation of practices including contour strip-cropping, stream buffers, rain gardens, and other practices where physical changes are observable. Finally, photos can be used in project evaluation. For example, photos taken before and after implementation of some types of remedial efforts (e.g., trash removal and prevention) provide an indicator of progress that can be communicated easily to most people.

To be useful, however, photographs must be taken in accordance with a protocol that ensures the photographic database accurately represents watershed conditions and is suitable for meeting stated objectives. This section provides an overview of ground-based photographic, or photo-point, monitoring, including specific elements of an acceptable protocol and example applications.

5.2 Procedure

Photo-point monitoring requires careful planning to ensure that meaningful information is provided to assess condition or trends (Bauer and Burton 1993). Monitoring design begins with a set of clear objectives, and different objectives will generally require different photo points (Hamilton, n.d.).

There are two basic methods of photo-point monitoring – comparison photography and repeat photography – but these methods can be used in combination (i.e., comparison photography repeated over time). Method selection should generally precede other design decisions but choices made in one step of monitoring plan design can affect the options in other steps, so flexibility is necessary. Selection of monitoring areas, identification of the specific features to photograph, camera placement, and the timing and frequency of photography are all typically determined after monitoring objectives and basic method are addressed (after Hamilton n.d.).

Photo-Point Monitoring

- Set objectives
- Select method
- Select monitoring areas
- Establish, mark, and assign identification numbers to photo and camera points
- Identify a witness site
- Record site information and create a site locator field book
- Determine timing and frequency of photographs
- Define data analysis plans
- Establish data management system
- Take and document photos

All photo and camera locations must be marked, monitoring site characteristics must be recorded, and a field book or similar documentation should be created to assist those taking photographs at the sites over time. This is critical if different people will be taking photographs throughout the course of a project. Plans for analysis of the photos and use of any photo-derived information must be determined and documented before photo-point monitoring begins. The data analysis plan will also help determine how best to organize and file photos and metadata.

These various design steps are described in greater detail below.

5.2.1 Setting Objectives

There will most likely be different photo-point monitoring objectives for project assessment, planning, implementation, and evaluation. Objectives for all project phases should be defined as early on in the project as possible, however, to maximize the efficiency of the photo-point monitoring effort. It may be possible, for example, to use photos from problem assessment or planning as pre-implementation photos for tracking implementation.

Realistic objectives begin with an understanding of what is likely to be seen and measured with photographs. Cameras exist that can take photos in both the visible and the non-visible spectrum (e.g., infrared or ultraviolet). For example, aerial photography has been used successfully to identify sediment sources at the watershed scale through correlation of photo density readings from the transparencies of color-infrared photographs with suspended sediment measurements (Rosgen 1973 1976). In addition, Hively et al. (2009a 2009b) combined cost-share program enrollment data with satellite imagery and on-farm sampling to evaluate cover crop N uptake on 136 fields within the Choptank River watershed on Maryland's eastern shore. Thermal infrared (TIR) images acquired from airborne platforms have been used in stream temperature monitoring and analysis programs, detecting and quantifying warm and cool water sources, calibrating stream temperature models, and identifying thermal processes (Faux et al. 2001). TIR imagery has also been used in the mapping of groundwater inflows and the analysis of floodplain hydrology. While such applications are indeed useful, this guidance and the example objectives that follow focus solely on ground-based photography in the visible spectrum.

An array of observable features listed in various guidance documents includes pasture condition, livestock distribution in a meadow, ground cover, tree canopy and health, vegetation density, woody vegetation, native vegetation area, wetland area, native plant richness, large trees, stream profile, streambank stability, streambank cover, fallen woody material and in-stream habitat, farm water flow, gully erosion, hill slope erosion, wind erosion, weed cover and species (Bauer and Burton 1993, ERS 2010, Hall 2001, Shaff et al. 2007). In addition, Hall (2001) provides numerous examples of successes and failures to measure changes in observable features with photo-point monitoring.

Examples of potential objectives for photo-point monitoring at various project stages include the following.

Assessment

- Document trash levels on beaches or in urban settings
- Document stream features
- Document algal blooms in waterbodies
- Identify sources of sediment plumes

- Document livestock activity near waterbodies
- Identify gullies and areas of streambank instability
- Identify areas in greatest need of urban runoff control measures

Planning

- Help locate areas were streambank protection and stream restoration are needed
- Document livestock operation needs to assist in budget development
- Provide evidence of watershed problems and potential solutions for public outreach
- Provide photos to assist the design of urban runoff control measures

Implementation

- Document tree growth in riparian zone over time
- Document implementation of rain gardens
- Document stream restoration activities
- Document and track changes in percent residue at representative agricultural sites across a watershed

Evaluation

- Document changes in streambank cover or stream profile as a result of stream restoration
- Demonstrate the effects of different grazing management systems on pasture condition
- Illustrate how a stream handles high-flow events before and after restoration
- Document changes in beach trash over time

The type and rigor of photo-point monitoring needed to meet these objectives varies. Alternative methods are described below.

5.2.2 Selecting Methods

As defined by Hall (2001), ground-based photo monitoring involves "using photographs taken at a specific site to monitor conditions or change," something that is accomplished by one of two methods: comparison or repeat photography. Comparison photography typically involves the creation of a photo guide from a set of standard photos taken to represent the expected range of an attribute (or condition) of interest (e.g., utilization of grazing plants). Field measurements are taken to establish values for the attribute of interest at levels represented by each of the photos in the guide. Figure 5-1 illustrates the concept whereby the value (percentage of area covered with dots) is determined from field measurement of the attribute of interest (dots/unit area in this conceptual example). The comparison photos in the guide are then used in the field to perform on-site assessment.

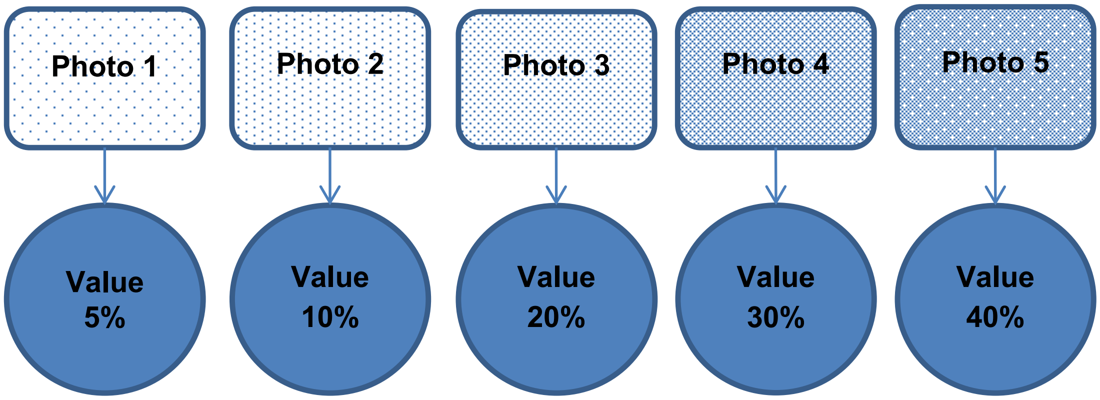


Figure 5-1. Comparison photos

In repeat photography, photos are taken of the subject over time at the same location to document change or monitor activity. Repeat photography has been used to document landscape change, including the advance and retreat of glaciers (Key et al. 2002). This method has also been used extensively to document progress in dam removal (USDA-FS 2007), riparian area protection (Bauer and Burton 1993), and stream restoration projects (Bledsoe and Meyer 2005).

A third type of photography is opportunistic photography. As described by Shaff et al. (2007), opportunistic photos are not taken from a permanently marked location, and they are not part of a repeat photography effort. There is also no photo guide as is used in comparison photography. Examples of subjects that can be addressed with opportunistic photography include a site during construction or an area after a significant natural or human-induced event.

Comparison photography is generally well suited to meeting assessment objectives in cases where photography is an appropriate monitoring approach. Opportunistic photography also usually plays a role in problem assessment. Both methods can be used for qualitative purposes, and comparison photography can be used in quantitative analyses to a limited degree (see “Qualitative” and “Quantitative” below). Opportunistic photography is not designed for quantitative analyses, however. Other information sources (e.g., livestock inventories, street maps, and permitted discharge reports) and monitoring data (e.g., water chemistry, aquatic biology, and habitat) will be needed in combination with photos to meet assessment objectives.

A combination of comparison and opportunistic photography can be helpful in achieving planning objectives, coupled with information from other sources. Opportunistic photos, in particular, can be quite helpful in communicating to the general public and stakeholders the need for restoration or BMPs to achieve watershed objectives. Visual inventories can be helpful in estimating implementation costs but should be used in combination with more traditional approaches to assessing need.

Repeat photography is generally most useful for tracking restoration and implementation of BMPs. Comparison photos can be used to assess such important indicators as the extent that conservation tillage has resulted in increased percent residue. Opportunistic photos can help show how restored stream reaches or urban runoff practices handle high-flow events.

While photo-point monitoring can be very helpful, it should be kept in mind that tracking implementation of rain gardens, for example, does not require photos. Observers could simply record in a database that a rain garden has been implemented at a specific address or global positioning system (GPS) location, but a photograph might add valuable information about the rain garden (e.g., size, location, plant selection and density) that could be explored at a later date if water quality data raise questions about rain garden performance.

Watershed projects cannot rely on photographs as the sole source of information for problem assessment or planning. Project implementation is nearly always tracked by means other than photo-point monitoring, but the addition of photographs can be the best way to document the installation of structural practices (e.g., lagoons, constructed wetlands) or the growth of vegetation associated with stream restoration or grazing management. It is important to keep in mind that photo-point monitoring should always be considered as a cost-effective tool for providing information in conjunction with other monitoring and information gathering efforts. While there are examples where photo-point monitoring is relied on as the primary monitoring method due to budgetary constraints, it is not recommended.

All three photo-point monitoring methods – comparison, repeat, and opportunistic – can support qualitative analyses, and comparison and repeat photography can also be used in quantitative analyses.

5.2.2.1 Qualitative Monitoring

Photographic monitoring methods usually generate qualitative information (e.g., Shaff et al. 2007). Creating a pictorial record of changing conditions, showing major changes in shrub and tree populations, visually representing physical measurements taken at a location, or recording particular events such as floods are typical of the types of photo-point monitoring objectives stated for these projects (ERS 2010). Those who have used photographic monitoring for watershed projects have generally used this method to document implementation of practices, typically the growth of vegetation associated with stream/streambank restoration or grazing management. These qualitative findings have been used most frequently to corroborate findings from more quantitative monitoring methods.

Photos are recommended for long-term monitoring of grassland, shrubland, and savanna ecosystems but simply as a qualitative indicator of large changes in vegetation structure and for visually documenting changes measured with other methods (Herrick et al. 2005 2005a). Photos should not be considered as a substitute for quantitative data; it is very difficult to obtain reliable quantitative data from photos unless conditions are controlled. Bledsoe and Meyer (2005) used photographs to compare changes from year to year, document noteworthy morphologic adjustments, document features of interest at various locations and times during the year, and analyze vegetation establishments as part of monitoring channel stability.

5.2.2.2 Quantitative Monitoring

Quantitative monitoring involves either measurement or counting. When measurement is desired it is important to use meter boards (field rulers mounted vertically) or other size control boards to provide a reference for measurement (Hall 2001 2002). Small frames (1 m^2) have been used for closeup or plot studies, while meter boards and Robel poles are often used for more distant studies. These standard references are captured within the photographs to provide a means of measuring features of interest. Counts of items of interest (e.g., trees of varying heights) can be obtained through visual observation of images. Another alternative for obtaining counts or percentages for quantitative analysis is to count digital image pixels that fall within a specified color range (see Digital Image Analysis below).

Meter boards can also provide a consistent point for camera orientation and a point on which to focus the camera (Hamilton n.d.). Figure 5-2 illustrates the use of a meter board and photo identification card (see section 5.2.13). The following are methods described by Hall (2001) that incorporate varying degrees of quantitative analysis. It should be noted that while these methods all support some level of quantification, documentation of precision and accuracy is generally lacking.

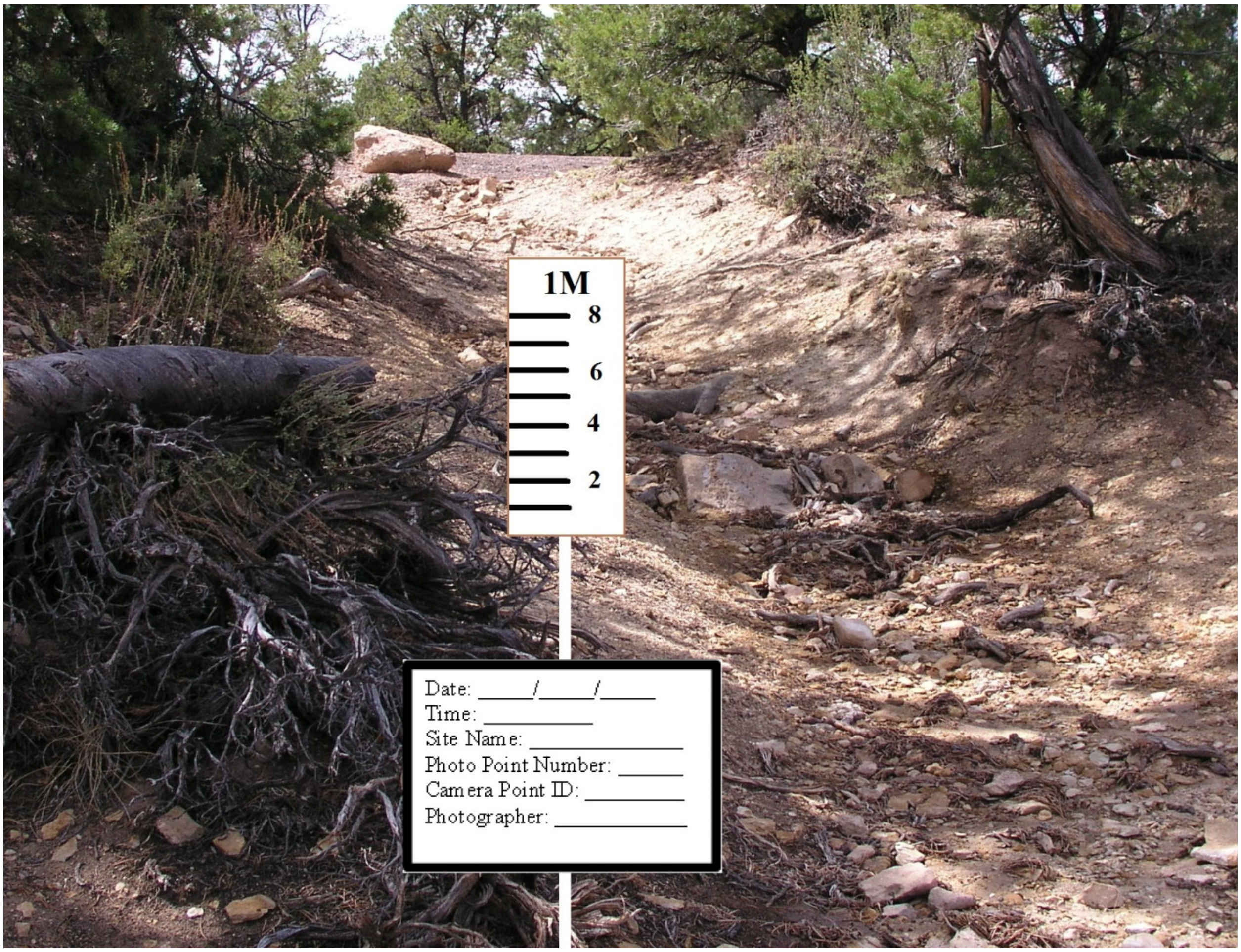


Figure 5-2. Illustration of a photo identification card and a meter board

5.2.2.2.1 Photo Grid Analysis

Photo grid analysis involves placing a standardized grid over a photo and counting the number of intersects between the grid lines and features of interest. When photo grid analysis is planned, it is very important that the distance between the camera and meter board is constant (Hall 2001 2002). It is recommended that the camera height is held constant, but it is only *required* to be constant if the grid is used to track position (in addition to size) of features over time. The size control board should cover at least 25 percent of the photo height, with the optimum range being 35 to 50 percent. The board, however, cannot obstruct the features of interest that will be measured. A level meter control board is preferred because it will match up more easily with a superimposed grid. Vegetation around the front of the meter board should be removed to expose the bottom measurement line to provide maximum precision in grid adjustment.

Hall (2001 2002) notes that both grid precision and observer variability are major factors in determining the ability to measure change. The percentage of photo height taken by the meter board is a very important factor in the precision with which grids are fit. It should be noted that changes in technology (cameras and software) may provide better results than found by Hall. For example, testing showed that a meter board that covers 35 percent of the photo height was 1.3 times more precise than a board that covered 25 percent of the photo height. Testing on observer variability also indicated that, on average, a change >12 percent in intersects for all shrubs (a measurement for grid analysis) was needed to demonstrate change at the 5 percent confidence level. Additional details and examples of photo grid analysis are provided by Hall (2001 2002).

5.2.2.2.2 Transect Photo Sampling

Photo points can also be established along a transect to obtain more quantitative information (Hamilton n.d.). Hall (2001) describes in detail five kinds of photo transects: (1) 1-ft^2 frequency photographed with or without a stereo attachment on the camera, (2) nested frequency using four plot sizes in a 0.5- by 0.5-m frame, (3) 1-m^2 plot frame photographed at an angle, (4) vertical photographs of tree canopy cover, and (5) measurement of herbaceous stubble height using the Robel pole system.

Transect installation is straightforward, requiring skillsets and procedures similar to those for the establishment of photo-point and camera sites (see sections 5.2.4 and 5.2.5). Equipment needs are similar as well. Size control boards are required, and they can serve multiple purposes, including estimation of height of grass and shrubs, orientation (for consistency) and focus (for greatest depth of field) of the camera, and grid analysis (Hall 2001). Key features of the five kinds of photo transects are provided below, but the reader should not select any of these methods until reviewing the detailed discussion of each by Hall.

5.2.2.2.2.1 One-Square-Foot Sampling

This method uses a 1-ft^2 plot placed every 5 ft along a 100-ft transect. The 20 plots are monitored to document changes in species, species density, and frequency as a means to estimate change in vegetation and soil surface conditions. Statistical analysis of data generated by this method is not possible.

5.2.2.2.2.2 Nested Frequency

This method uses a sample frame with four nested plot sizes to document change in species frequency along five 100-ft transects of 20 plots each. Statistical analysis suggests significant change in frequency (the number of times a species occurs in a given number of plots) at the 80-percent level of probability.

5.2.2.2.2.3 Nine-Square-Foot Transects

This plot system uses five 9-ft^2 plots along a 100-ft transect to document changes in species frequency. Photographs are taken of the plot frame at an oblique angle rather than from directly above. Interpretation of change is based not on statistical analysis but on professional judgment and interpretation of the photos.

5.2.2.2.2.4 Tree Canopy

It is recommended that any transect placed in a forest setting should have tree cover sampled because of its effects on the density and composition of ground vegetation. Tree canopies are photographed from ground level by using a camera leveling board or other means to ensure that the camera is pointing directly above. The method requires photographs of tree cover at the 0-, 25-, 50-, 75-, and 100-ft locations on transects used for any of the three methods described above. Because photo grid analysis is used to estimate tree cover, the same focal length must be used for all photos and the long axis of the camera should be perpendicular to the transect.

5.2.2.2.2.5 Robel Pole

A Robel pole is a 4-ft pole with 1-in bands painted in alternating colors (USDA-CES et al. 1999). Vegetation height is measured by photographing the pole from a specific distance and height above the ground. This is accomplished by attaching a 4-m-long line between the 1-m mark on the Robel pole and the top of a 1-m-tall line pole. The Robel pole is placed at the sample location and the line is stretched out. The camera is set on top of the line pole and a photo is taken. By consistently using the 4-m line and 1-m camera height (4-to-1 ratio), the same angle is obtained for all photos.

5.2.2.2.3 Digital Image Analysis

Many of the methods described by Hall (2001) were centered on film-based photography, and they often require a substantial amount of measurement and analysis by hand. Newer methods such as digital image analysis (DIA) use computers to analyze digital images, offering the potential advantages of improved objectivity, accuracy, and precision. In one form of DIA, color images are converted to grayscale (monochrome) images using an algorithm that converts each pixel to white or black based on the color content of the original pixel. The algorithm in this case is designed to select those colors that represent the feature to be counted. For example, Rasmussen et al. (2007) used DIA to determine the proportion of pixels in digital images that were green to estimate crop soil cover in weed harrowing research.

There are significant hurdles to overcome in applying DIA to photo-point monitoring for watershed projects. Factors such as lighting, camera angle, size of the area photographed, and the growth stage of plants should be evaluated to quantify their effects on the accuracy or precision of the method (Rasmussen et al. 2007). It is also important to have a true value to compare against the DIA-based results to assess the accuracy of the method (Richardson et al. 2001).

A significant contribution to DIA made by Rasmussen et al. (2007) was automated determination of the gray-level threshold which defines the difference between vegetation (the subject of interest in their study) and non-vegetation. This is especially important when lighting conditions vary in the field. With this capability, the researchers were able to develop an automated DIA procedure for converting each digital image into a single leaf cover (proportion of pixels that are green) value for analysis. Their research used the MATLAB Image Processing Toolbox (MathWorks 2012) but other options include Mathematica (Wolfram 2012) and a wide range of image processing products developed for a large number of applications.

5.2.3 Selecting Areas to Monitor

The areas selected for photo-point monitoring must be appropriate for the stated objectives and consistent with the data analysis plans (section 5.2.11). Depending on the monitoring objectives, suitable sampling locations may be chosen to represent average or extreme conditions.

For problem assessment where opportunistic photography is used, site selection may be similar to that employed in a synoptic survey for water quality monitoring. Photos may be taken by individuals walking the stream to identify areas of streambank erosion or point source discharges. Photography of sources could involve a windshield-survey approach where photos are taken on a pre-determined route. Each opportunistic photo would need to be properly labeled as described in section 5.2.13.

When tracking project implementation (e.g., BMPs, restoration) or evaluating project success, it is most important to select an area that is most likely to undergo the physical transformations that can and must be tracked in order to support these objectives. Hall (2001) notes that this task may be straightforward (e.g., measuring the impact of stream restoration on the segment restored) or somewhat more complicated (e.g., documenting the impacts of livestock grazing on riparian vegetation). The latter case is more complicated because it requires some knowledge of livestock distribution, areas sensitive to grazing, and grazing patterns. Because it is likely that only a portion of the area of interest can be monitored, it is important to determine up front whether or not the findings can be extrapolated to areas not monitored. This is particularly challenging for photo-point monitoring because statistical analysis of photo-based data is not common. Attribution of sample findings to the broader area of interest would require the sample is representative, there is a measurable variable from the photos, the distribution for that variable is known, and an estimate of the standard deviation is available.

Some may wish to use photo-point monitoring to track BMP-related information in support of a traditional biological or chemical monitoring program. For example, if total suspended sediment concentration or loads are monitored in a predominantly agricultural watershed, it may be useful to track percent residue as an indicator of the extent to which reduced tillage practices have been implemented across the watershed. This could be accomplished in a number of ways including photo-point monitoring of a set of randomly selected field sites. Both comparison (to determine percent residue) and repeat (to track changes in percent residue over time) photography would be used in this application (see section 5.2.2). Again, attribution of sample findings to the broader area would require that the samples are representative, the distribution of percent residue is known, and an estimate of the standard deviation is available.

5.2.4 Identifying Photo Points

Photo points are defined somewhat differently in various guidance manuals, which can lead to confusion when flipping back and forth between manuals. This document adopts the terminology used by Hall (2001), in which the photo point is essentially what you point the camera at when you take the photograph, and the camera point is a permanently marked location for the camera (Figure 5-3). Photo points have also been defined as *permanent or semi-permanent sites set up from where you take a series of photographs over time* (ERS 2010). Despite the different definitions and intermingling of various concepts within these definitions, photo-point monitoring manuals ultimately address the area to be photographed, the location from which the photos are taken, and the camera direction and settings to identify what will be captured in the photos. In simple terms, the photo point is what you point the camera at when you take the photograph.

The area captured in each photo will depend on the monitoring objectives and is controlled by camera settings and the distance between the camera location and the subject. Hall (2001) describes three general types of photos, each of which has an associated scale:

- Landscape – distant scenes with areas generally greater than 10 ha
- General – specific topics monitored on areas 0.25 to 10 ha
- Closeup – specific topics on areas under 0.25 ha

Figure 5-3. Photo illustrating photo points (A and B) and camera points (1 and 2). Photos of A and B are taken from cameras located at 1 and 2.

Landscape photography generally requires a long-term commitment during which repeat photos are taken as infrequently as every 20 years or so (Hall 2001). This timeframe is greater than typically encountered in watershed projects. General and closeup monitoring will be more appropriate for most watershed-scale projects. Hamilton (n.d.) states that general photography can be used to document an entire scene, whereas topic (closeup) photography narrows the target down to specific elements or subjects in the landscape.

Scale is also incorporated within the definitions of photo types found in other guidance documents. For example, one scheme refers to spot, trayback (small truck with short, flat tray in back rather than a typical pickup box), and landscape photographs which generally correspond to Hall's closeup, general, and landscape photos (ERS 2010). Shaff et al. (2007) describes feature, landscape, and opportunistic photos. Landscape photos cover a broader area than feature photos, while opportunistic photos (see section 5.2.2) vary in scale but are generally at the feature or finer scale. The authors also provide guidelines on the type of photography and features to photograph for various restoration activities associated with habitat improvement projects, road projects, water management projects, wetlands, and fish passage improvement.

Be sure to consider the following when establishing photo points:

- The general or specific features that must be photographed to meet the monitoring objectives.
- How representative the photo points are of conditions in the study area.
- Whether the number and type of photo points are sufficient for tracking change.
- Whether changes will be visible at the desired scale.
- Whether the site is accessible and lighting and sight lines are adequate during the entire monitoring period.

5.2.5 Establishing Camera Points

As noted in section 5.2.4, camera points are permanently marked locations for the camera. Hamilton (n.d.) suggests selecting camera points from which multiple photo points can be photographed. The same photo point can also be photographed from multiple camera points, for example, if there is a need to examine the subject matter at different scales or from different angles. If the sizes of objects will be compared in photos taken from multiple camera points, the distance from each camera point to the photo point must be the same. In addition, to avoid shadowing of the photo point, camera points should be located north of photo points when they are close together.

Hall (2001) performed field testing of camera point setups (e.g., distance from photo point and the vertical and horizontal positioning of the camera) to determine the effects of various camera positions and settings on the ability to perform reliable repeat photography. Results of this testing clearly showed the following:

- Distance from the camera to the meter board (or subject) affects both the size and location of objects photographed.
- The vertical and horizontal position of the camera affects the location but not the size of objects photographed.
- Focal length is not a critical issue because images can be enlarged or reduced to a constant area of coverage. Resolution can be lost, however, if images are enlarged or cropped too much, so it is best that the same or similar focal length be used for all photos.

Depending on the study objectives, therefore, camera point setup should provide a constant distance from the camera to the photo point (for size and location considerations), and consistent height and left-right orientation of the camera (for location). It should be noted that in Hall's testing, camera position was shifted both upward and sideways by 40 cm (16 in) from an initial position centered at 1.4 m (55 in) above the ground. Smaller shifts would result in lesser changes in object location.

Figure 5-3 illustrates the location of photo and camera points. Both camera points 1 and 2 would need consistent camera positions if object locations were to be tracked over time. Meter boards can be used to guide camera position when taking photos, with the camera siting always on the top, bottom, or other specific marking on the meter board.

A recommended standard equipment list for establishing photo-point monitoring areas can be found in section 5.3.

5.2.6 Marking and Identifying Photo and Camera Points

Every photo and camera point should be geolocated, photographed, and permanently marked so that those returning to take photos can find the sites with little waste of time. Capturing prominent features such as a ridge line in the photos can help others identify the location and the photo points (Bauer and Burton 1993). Labor is usually the greatest cost associated with monitoring efforts (see chapter 9), and doing whatever it takes to minimize the time needed to find photo-monitoring sites is cost effective. If volunteers perform the monitoring, marking of photo and camera points is essential to efficiently finding the locations so they can spend more time taking and documenting photos and less time searching for sites.

The best material to mark sites depends on the circumstances, but metal fenceposts work well in many cases (Hamilton n.d.). If metal fenceposts are unsuitable due to appearance or other considerations, steel survey stakes driven into the ground may be appropriate provided that metal detecting equipment is available (Hall 2001). If steel stakes are used, they can be covered with plastic pipe for safety, and all stakes can be painted in bright colors to improve visibility (Larsen 2006). Each photo and camera point should be given a unique identification number.

It is very important that the distance between camera points and photo points is measured and documented (Hamilton n.d.). Site location can be facilitated by use of a GPS but marking of photo and camera points will still be necessary in many cases, given that the best resolution for GPS systems is currently about 3-5 meters. Identifiers for opportunistic photos and temporary photo and camera points used for problem assessment and planning should at least include the purpose, address or GPS coordinates, camera direction, date photos were taken, narrative description of what was observed, and photographer name to provide sufficient information to interpret the information obtained and revisit the site if necessary.

5.2.7 Identifying a Witness Site

A witness site is an object that can be easily identified when returning to the monitoring area (Hamilton n.d., Hall 2001). It may be a large rock, a structure, or other feature that is easily identifiable from the road or path to the photo and camera points. It is important to measure and document the distance and direction from the witness site to the camera points, photo points, or both. If possible, it is also helpful to attach a permanent identification tag to the witness site with the distance and direction to the photo and/or camera points inscribed on the tag (Hamilton n.d.). Newer photo-monitoring guidance recommends the use of GPS devices to facilitate finding the photo and camera points (ERS 2010, Shaff et al. 2007). In all cases, however, it is helpful to have photographs of the site and a description of landmarks to help locate and identify important spots within the monitoring area.

5.2.8 Recording Important Site Information

Information about any monitoring site, whether it be chemical, biological, physical, or photographic (permanent or temporary), should be recorded to help future staff understand the reasons for selecting the site and to help in the interpretation of data collected from the site. Maps, aerial photographs, and standardized forms can be used to record date, observer name(s), location, site description, objectives, identification numbers, and locations of the witness site, photo points, and camera points, including distances and directions between points. It is important to indicate whether directions are magnetic or true degrees (Hamilton n.d.), a topic addressed in detail by the U.S. Search and Rescue Task Force (USSARTF n.d.). Standardized forms for all aspects of photo-point monitoring can be found in existing documents (Hall 2001 2002, Shaff et al. 2007).

5.2.9 Determining Timing and Frequency of Photographs

Monitoring frequency should be based primarily on the monitoring objectives, planned data analyses, features to be photographed, and expectations regarding detectable change in those features. Photo-point monitoring for problem assessment and planning can be a one-time activity or may involve multiple photographs taken at various times during the year to characterize seasonal, flow-related, or other significant variability. Efforts to track project implementation or evaluate project success will usually involve multiple years, with the frequency and timing of photos based on an understanding of seasonal and other variability.

Land managers are encouraged to photograph native vegetation at least once per year at the end of the growing season, or twice per year to show seasonal differences (ERS 2010). For restoration projects, the frequency options are generally seasonal, annual, or biennial (Shaff et al. 2007). In addition, photos taken during the high-flow and low-flow seasons should be compared to give some indication of the causes affecting streambank condition. Regardless of the frequency selected, annual changes should be assessed using photos taken at the same time of year.

Although photo-point monitoring for watershed projects is usually qualitative rather than quantitative, the concept of MDC (see section 3.4.2) can still be applied when determining the frequency and duration of photography. In essence, MDC is based on sample variance and the number of independent samples taken over time. Kinney and Clary (1998) used repeat photography to track cattle density (animals/ha) on various vegetation-soil categories in a riparian meadow and used analysis of variance to test for differences in cattle distribution across vegetation-soil categories. Such time-series data could be analyzed to estimate variance (i.e., variability) in the number of cattle in each photograph. This data could then be used in an MDC analysis to estimate how often photographs would need to be taken to detect a significant change in cattle density at a given level of confidence. It is important to note that the authors found autocorrelation in their data due to frequency of photography, something that would have to be addressed in the MDC analysis (see section 3.4.2).

In an assessment of photo grid analysis precision, it was found that variability among different observers was about 12 percent, indicating that a change in mean intersects of that much would be needed to indicate that the change was real at the 5 percent level of confidence (Hall 2001). Monitoring, therefore, would need to continue until a 12 percent change or more was expected.

Absent a rigorous database to support MDC analysis, it is recommended that a qualitative assessment of time needed to see measurable change is performed. Guidelines that can be used to estimate the number of years photo-monitoring should continue to document measurable change include plant growth rates for restoration activities, typical timeframes for construction of urban runoff controls, and historical patterns for adoption of agricultural BMPs.

5.2.10 Creating a Field Book

Hamilton (n.d.) recommends creation of a field book to help others find the monitoring location, witness site, and photo and camera points. Field books should also include copies of the original photo-point photographs, and other important site information recorded as described under section 5.2.8. Advances in GPS, portable computer, and cell phone technology, however, may reduce the need for a physical field book, but a printed version should be created as a backup.

5.2.11 Defining Data Analysis Plans

It is essential to establish plans for analysis before taking the photos. As described in section 5.2.1 and 5.2.2, photo-point monitoring objectives can range from highly qualitative to quantitative, and data analysis plans need to be worked out in advance to ensure that information collected through photo-point monitoring will be sufficient to achieve these objectives.

Although statistical analysis of photo-based data for watershed projects is uncommon, examples exist that could be applied to watershed projects. For example, quantitative analysis of differences in grazing patterns in various areas of a riparian meadow was performed by Kinney and Clary (1998) using analysis of variance. Photos were analyzed to count the number of cattle within each of five vegetation-soil categories that were delineated within the study area and superimposed on individual photographs. Through this method, researchers created a database with counts that were converted to a density measure that was associated with both year and class variables (e.g., vegetation-soil category, pasture number).

In another example where statistical analysis was applied to photo-derived data, digital image analysis was compared against subjective analysis (SA) and line-intersect analysis (LIA) in determining the percentage of turf cover on study plots (Richardson et al. 2001). For DIA, the percentage of green pixels in images of turfgrass taken from a digital camera mounted on a monopod was calculated to determine the turf coverage percentage in each of the images. The DIA approach was shown to be very accurate through calibration with turf plugs of known cover, and DIA also performed far better than either SA or LIA in determining the percent cover of study plots. The variance for DIA was only 0.65, while the variances for LIA and SA were 13.18 and 99.12, respectively.

As described in section 5.2.2, both the photo grid analysis and nested frequency methods support statistical analysis (Hall 2001). For example, demonstration of regression analysis of grid intersects from annual photography over a 20-year period appeared to be useful.

If these or other monitoring approaches that support statistical analysis are planned, it is essential that the statistics to be performed are identified, the data needs to support the statistical analyses are documented, and plans are developed at the beginning of a project to obtain the needed information from photo-point monitoring. Because statistical analysis of photo-derived data is uncommon for watershed projects, it is essential that a statistician is involved in the design of the monitoring effort.

5.2.12 Establishing a Data Management System

Data management systems are described in detail in section 3.9. The basic requirements and safeguards associated with a data management system for water quality data also apply to photo-point monitoring data sets. These include an organized and readily accessible filing system, quality assurance and quality control procedures, working interfaces between data files and data analysis software, and backup systems.

It is recommended that backup archives are kept at a location separate from the original data (Hamilton n.d.).

As with water quality monitoring data records, information on monitoring objectives, designs, and locations must also be recorded and associated with the photos taken at each site. All information recorded on forms should be included in the database and linked to photos as appropriate.

If necessary, hard copies of photos can be stored in manila folders in filing cabinets or above-floor boxes and should be labeled clearly with locational information, date, time, and camera and photo point identifiers (Bechtel 2005, Larsen 2006, Shaff et al. 2007). Digital images and files will need to be stored in a computer database housed on a computer or computer network, and it is recommended that file names provide the same information contained in the labels on the paper photos (Bechtel 2005, Shaff et al. 2007). Software such as GPS Photo Link can be used to process the GPS information onto the images (Larsen 2006). Digital information should be backed up on CDs or other "permanent" storage devices, and networks should be backed up nightly (Bechtel 2005). Photo-point monitoring will usually be performed far less frequently than storm-event monitoring, for example, but the file sizes associated with photographs may create data storage challenges that should be considered early on in the project.

Whether photos are used for qualitative or quantitative analyses, it is important that standard procedures are established and followed. For example, photos used in a river continuity assessment in New Hampshire were taken in accordance with a standard operating procedure that was incorporated within a quality assurance project plan (Bechtel 2005). The QAPP identified equipment needs and the roles and duties of team members, provided general instructions, and gave details on all important aspects of selecting sites and taking the photos. In addition, volunteers were trained in photo documentation, and standardized forms were provided to ensure consistency.

5.2.13 Taking and Documenting Photographs

Whether photo points are temporary or permanent, opportunistic or part of a trend assessment, certain guidelines should be followed to ensure that the photos support the monitoring objectives. It should be clear from the following recommendations, some of which are slightly at odds with each other, that photography is part art, part science (Bechtel 2005, ERS 2010, Shaff et al. 2007):

- Closeup photos should be taken from the north facing south to minimize shadows.
- Both medium and longer distance photos should be taken with the sun behind the photographer.
- Recommendations on the best times for taking photos vary, with some choosing early in the morning, late in the afternoon, or on slightly overcast days to reduce shadows and glare, and others wanting clear days between 9 a.m. and 3 p.m.. Photos taken before 9 a.m. and after 3 p.m. can result in increased shadowing and a different color cast that could conceal some features.
- Some recommend camera settings that give the greatest depth of field, while others simply recommend using the camera's auto settings.
- Report the true compass bearing (corrected for declination) if possible.

Additional guidelines apply when the monitoring plan involves repeat photography. For example, consistency is essential for trend assessment, and the following information taken from a variety of sources should be recorded with each photograph to ensure such consistency (Bechtel 2005, ERS 2010, Hall 2001, Hamilton n.d., Larsen 2006, Shaff et al. 2007):

- When shooting repeat photography it is helpful to compare the view through the camera with a copy of the original photo to create comparable photos. Camera settings should be the same as those documented when the original photo was taken.
- Document the type of camera and lens used, digital resolution, tripod and camera height, lens focal length or degree of zoom, light conditions, compass direction of the photo, and the distance from the camera to the one-meter board or center of the photo area.
- Document whether the camera is held horizontally or vertically.
- Record the date, location, compass bearing, and management history since the last photo was taken (e.g., description of observable progress in achieving restoration or BMP goals).
- Describe the scene or subject and record that information.
- Hold the camera at eye level, positioning it so the one-meter board is centered in the middle of the photo. Try to include some skyline in the photo to help establish the scale of the area. Photo identification cards should be placed within the camera's field of view for each photograph to embed relevant information into the picture. Figure 5-2 illustrates one approach to positioning of the 1-m board and photo-identification card. The recommended content for each card is illustrated in Figure 5-4. Some of this information (e.g., date and time) can be embedded using digital camera options, and these options are likely to improve over time.
- Blue paper should be used for photo identification cards. Alternative approaches may include laminated cards or small chalk boards.
- Framing of the photo should ensure that the photo identification card does not obscure features of interest.
- The angle from which the photo is taken should be consistent. When taking photos at a height of about 3 m from a trayback, tripod, or step ladder, a downward angle of 15 degrees is recommended to illustrate ground condition and features, (e.g., the amount of feed available in a pasture).

Date: _____/_____/_____ Time: __________
Site Name: ___________________________
Photo Point Number: ______________
Camera Point ID: _____________
Photographer: _______________________

Figure 5-4. Photo identification card

Logistical considerations for repeat photography include the following:

- Photo-documentation teams should consist of two people for both safety and logistical concerns (Bechtel 2005, Herrick 2005).

- Once at the site, it is estimated that it will take about 3 min per photo from a single camera point (Herrick 2005).
- Landowner permission may be needed for some monitoring locations, and it is advisable to check on the legality of taking photos of private property in your jurisdiction before monitoring begins. There may also be gates for which keys or combinations are needed to gain access to the photo points. It is important that landowners be notified before photos are taken and that keys or combinations for gates are in hand.

A recommended standard equipment list for photo-taking events can be found in section 5.3. Larsen (2006) recommends using GPS Photo Link, a software program that "links" digital photos to the GPS coordinates. This software program is now marketed as GeoJot+ Core (GeoSpatial Experts 2016). A geo-location feature is available on some current digital camera models. There are a wide range of GPS receivers now available, with most enabling the user to take precise position coordinate readings and record details about each position in an attribute table that can be downloaded to a computer (ERS 2010). In addition, GIS software usually supports display of digital images, and there are numerous options for property mapping software that can be found on the Internet (ERS 2010).

5.3 Equipment Needs

Methods described by Hall (2001 2002) are still largely relevant today but equipment has changed considerably in the past decade. Most cameras in use today are digital, with resolutions far exceeding the 2 megapixel cameras described by Hall. Storage cards are larger and faster as well, and batteries last far longer than they did just five years ago. The many improvements in camera technology have increased the capabilities of photo-point monitoring by increasing the amount and quality of information contained in each photo, increasing the number of photos that can be taken and stored under a single battery charge, improving the options for time-lapse and programmed photography, and greatly enhancing the capabilities for photo interpretation and analysis with computer software.

Because camera technology will continue to improve, it is recommended that an initial step in designing a photo-point monitoring effort should be to survey currently available cameras and associated hardware and software to assess the possibilities for photographic data collection and analysis, the potential for unattended time-lapse photography (e.g., how long will batteries last at various resolutions and frequency of taking photos), the ability to retrieve photos from a remote location through a computer link or to rapidly upload images directly from the camera to a remote website, and the cost of various options. Coordination with others (e.g., USDA) may be an excellent way to obtain access to integrated technology for photo-point monitoring. For example, software such as GPS Photo Link[1] has been used by NRCS to link photos to GPS coordinates and create data files that include the photos, coordinates, and other descriptive information (GeoSpatial Experts 2004). Technology should not drive study objectives but it is common sense to assess the extent to which available technology can be used to meet or augment study objectives. With labor the major cost in many monitoring efforts, there may be attractive options for using more technology and less labor to keep costs down.

The following items should also be considered in standard equipment lists for site establishment and subsequent photo-taking visits (Bechtel 2005, Hamilton n.d., Herrick et al. 2005a, Larsen 2006):

[1] Now marketed as GeoJot+ Core (Geospatial Experts 2016).

Site Establishment

- Camera (and extra batteries)
- GPS unit or map of monitoring areas
- Clipboard, data forms (site description/location, camera location and photo points), and pencils OR field computer with data entry software (extra battery for field computer if used)
- Compass
- Level (for permanently mounted meter boards)
- Hammer or post driver
- Keys and gate combinations (if needed)
- Measuring tape
- Rebar (3 ft) or other states for marking transect ends (if used)
- Shovel
- Whiteboard (and marker), chalkboard (and chalk), or photo-point ID cards
- Fenceposts
- Stakes or posts made of wood, fiberglass, plastic, rebar, or steel (point markers)
- Meter board
- Spray paint
- PVC pole (1.5 m or 5 ft long) or tripod for mounting camera at fixed height

Each Photo-Taking Visit

- Camera (and extra batteries)
- Compass
- Level
- Timepiece
- GPS unit or map of monitoring areas
- Site locator field book or field computer with copies of original photos and site information (extra battery for field computer if used)
- Clipboard, data forms (site description/location, camera location and photo points), and pencils or field computer with data entry software (e.g., GPS-photo ID software)
- Whiteboard, chalkboard, or photo-point ID cards
- Thick marking pen
- PVC pole (1.5 m or 5 ft long) or tripod for mounting camera at fixed height
- Keys and gate combinations (if needed)
- Measuring tape
- Metal detector (if needed for stake location)

- Ruler (optional – for scale on close-ups)
- Spray paint

5.4 Applications of Photo-Point Monitoring

5.4.1 Comparison Photos

Comparison photography has been used in a number of applications associated with grazing. In one example cited by Hall (2001), the height and weight of grasses and forbs were measured, and a height-weight curve was developed and used to estimate percent utilization based on height measurements (Kinney and Clary 1994). The utilization level of an individual plant was determined by matching its residual stubble to a photo in the guide and then assigning the percent utilization value for that photo to the plant. Average utilization in an area was estimated from a number of individual plants (e.g., 50 to 100). It should be noted that the quality of estimates developed with this method depends substantially on the level of detail in the photo guide. It may be necessary to develop seasonal or species-specific guides depending on the level of accuracy and precision needed for the study. The authors concluded that about 25 random plant height measurements should give mean plant height estimates within 5 percent of the mean at 95 percent confidence.

Comparison photos have also been used to provide a quick approximation of percent residue under various conservation tillage practices (Eck and Brown 2004, Hickman and Schoenberger 1989, Shelton et al. 1995). Percent cover can usually be estimated within 10 to 20 percent of the actual cover when using the photo-comparison method. When using this method to estimate percent residue it is important to find a representative area of the field, look straight down at the residue if it is flat or at an angle if it is standing residue, and compare the observed residue cover with photos of known cover. Interpolation between photos may be necessary, and it is recommended that the results of three or more observations from different representative locations on the field be averaged for a better estimate.

The Queensland BioCondition Assessment Framework specifies a quantitative approach to photo-point monitoring to assess terrestrial biodiversity, incorporating a 100-m vegetation transect and spot (close-up) and landscape photos taken in accordance with a detailed protocol (Eyre et al. 2015). Despite the attention to detail regarding the taking of photographs, no analysis of the photographs is described, and photos are only recommended, not required. The related method for establishing reference sites for biocondition assessment states only that spot photos can be useful to capture the variability in ground cover within sample locations (Eyre et al. 2011).

5.4.2 Repeat Photography

Repeat photography has been used for a range of purposes in a large number of NPS projects including wetland restoration, streambank restoration, and fencing (OEPA n.d., Oregon DEQ 2002, Shaff et al. 2007). The Jordan Cove, CT, Section 319 National Nonpoint Source Monitoring Program (NNPSMP) project took weekly photos as homes were constructed and documented all development changes in the suburban lot. Weekly observation of construction activities allowed documentation of water quantity effects such as storage of water in cellar excavations and rainfall ponding on pavement (Clausen 2011). The Morro Bay Section 319 NNPSMP project in California documented implementation of BMPs with photo-point monitoring (CCRWQCB 2012b). In the Maino Ranch study area of the Morro Bay project, photo-point monitoring failed to document changes in stream channels as a result of fencing and other practices designed to control cattle movement through pastures (CCRWQCB and CPSU 2003). This result agreed, however, with the findings from the monitoring of stream channel stability and stream profiles from fall 1993 through spring 2001.

Photo-monitoring of pre- and post-construction conditions is used to document the success of all erosion control projects on rural roads in Santa Cruz County, California (CCRWQCB 2012a). A report on Section 319 projects funded in NM from 1998 to 2008 showed that 11 of 127 projects used photo-point monitoring for project evaluation, and many others used photos to assist in problem documentation (NMED 2009). Of the 11, nine photographed vegetation to track progress associated with range/grazing management and/or riparian restoration, one tracked road reclamation, and the other used photo-point monitoring to document improvement from trail reconstruction.

Photo-point monitoring at Chinamans Beach, Australia, was used to gain understanding of the movement and accumulation of wrack (piles of seaweed) on the beach (MMC n.d.). Photos collected two times per week over a 12-week period helped determine the need for and best approach to beach raking. Supplemental information on tides, weather, and activities in the area was used to help interpret the photos but all observations were qualitative.

Photo-documentation was a major component of assessment monitoring for the South Fork Palouse River riparian area restoration project (PCEI 2005). Permanent photo monitoring stations were established along the restoration site to document both vegetation establishment success and streambank stability. Using the methods of Hall (2001), bank stability was evaluated with photos taken twice per year (in March following high-flows and in July under base-flow conditions) at three photo points located along the restored site. Permanent meter stakes installed at the top of the bank at each location served as visual reference points for photo monitoring and as references to measure erosion. Vegetation establishment success (changes in growth and production) was also tracked through photo monitoring, with photos taken during the first week of August and then yearly for 10 years following restoration.

The NRCS has published guidance on photo-point monitoring as a qualitative method for documenting short-term and long-term effects of a prescribed grazing plan (Larsen 2006). In support of this guidance, the Nebraska NRCS developed a field office guide to demonstrate the use of GPS Photo Link[2], a software program that “links” digital photos to the GPS coordinates (GeoSpatial Experts 2004).

Kinney and Clary (1998) used time-lapse photography to demonstrate differences in time spent by cattle on several pastures within a riparian meadow. Cattle location was classified by five broad plant community-soil groups. Photographs were taken at 20-min intervals during daylight hours, a frequency at which auto-correlation was observed. Information obtained from the photos was reduced to number of cattle per unit area, and analysis of variance was performed on number of animals per ha per plant-soil site per photograph, with pasture and year used as explanatory variables that would account for differences in animal stocking densities. The authors were able to show statistically significant differences in cattle densities among site categories overall and for three different animal positions (standing head down, standing head up, and lying down).

Photo-documentation is very popular among volunteer monitoring groups. For example, the SOLVE Green Team in Oregon uses photo point monitoring to track progress at watershed restoration sites (SOLVE 2011). The Missouri Stream Team uses photo-point monitoring to supplement water quality and other stream monitoring activities (MST n.d.).

[2] Now marketed as GeoJot+ Core (GeoSpatial Experts 2016).

5.5 Advantages, Limitations, and Opportunities

Photo-point monitoring can potentially be used for a variety of purposes, including problem assessment and planning, tracking BMP implementation, providing supporting information for traditional water quality monitoring, discovering unexpected events, serving as surrogates for water quality parameters, and serving as direct measures of water quality conditions (Figure 5-5).

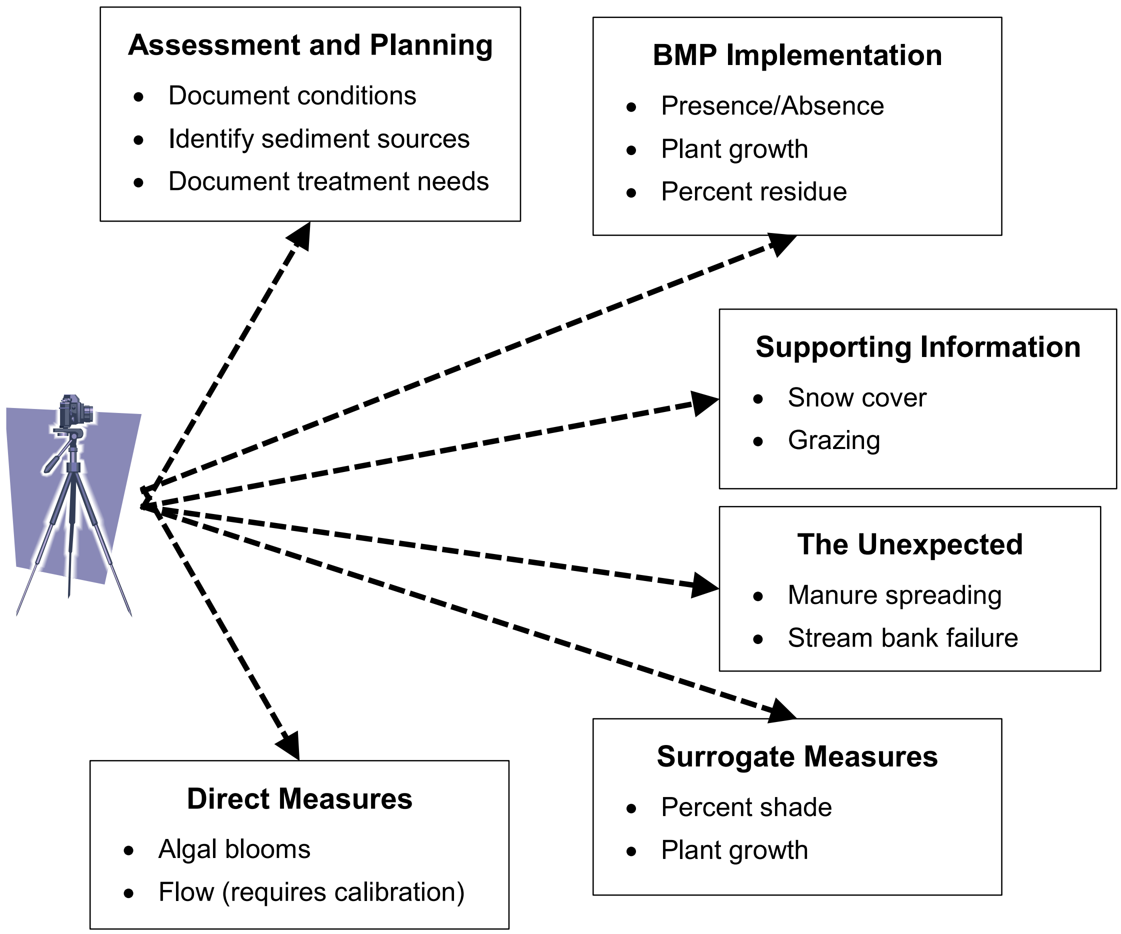


Figure 5-5. Various potential applications of photo-point monitoring

5.5.1 Advantages

Every monitoring option has advantages and limitations, and Hamilton (n.d.) identified the following strengths of photo-point monitoring:

- Uses readily available equipment.
- Is an effective communication tool for public education.
- Is a method of providing landscape context for a study area.
- Is a standardized evaluation procedure for comparing multiple locations.

- Is a method to document rates of change.

In addition to these observations, photo-point monitoring is less expensive than most other watershed project monitoring options.

5.5.2 Limitations

Some weaknesses of photo-point monitoring were also identified by Hamilton (n.d.):

- Only limited quantitative data can be obtained.
- Bias in photo point placement may occur.
- It may be difficult to use in dense vegetation.
- Photo points can be lost or obscured over time.

An additional limitation of photo-point monitoring for watershed projects is that, in most cases, it cannot be used to evaluate progress in achieving water quality objectives. Further, statistical approaches to using photo-derived data remain to be developed for use by those who apply photo-point monitoring techniques.

5.5.3 Opportunities

Recognizing the inherent advantages and limitations of photo-point monitoring, there are many opportunities to use this tool for watershed projects. Several of these opportunities have been realized, while others are suggested only for consideration, with full understanding that any method must be tested and evaluated before being adopted.

Photo-point monitoring can be very helpful in assessing watershed problems. For example, it was used in a volunteer-led river continuity assessment of the Ashuelot River water in New Hampshire (Bechtel 2005). Photos were taken at each dam site (at the downstream end) and at both the upstream and downstream ends of stream crossings. The QA officer used the photographs to ensure that information recorded regarding bridge and culvert type made sense. Photos were also used as part of the permanent inventory record.

Photo-point monitoring for western grazing lands has been found to be an easy and inexpensive way to provide an excellent visual representation of conditions at a given point in time. These photographs were considered only as supplementary data, however, not sufficient alone to evaluate objectives (Bauer and Burton 1993). Photographs could be used to indicate a trend in woody vegetation, streambank stability, and streambank cover, but the authors noted that vegetation “expression” as seen in photographs was not the same as vegetation “succession” needed for stream ecosystem health.

At the farm-scale, researchers at the University of Wisconsin-Platteville have applied photo-point monitoring to farm-scale research. Photos have been used for a variety of applications as seen in the sidebar (Busch and Mentz, 2012).

As an example of new applications of photo-point monitoring, it is feasible that photo-point monitoring could be used to track flow provided that a stage-discharge relationship is first established. While this may at first seem to offer no advantage over visual observation of a staff gage, tracking stage with photographs could offer the advantages of 24-hour surveillance and safety during high-flow events.

Cameras would need to be positioned in secure locations, however, and remote transmission of photos may be required.

The greatest opportunity for photo-point monitoring at the watershed scale, however, may be an improvement in the quantification of variables of interest and statistical analysis of photo-derived data. All monitoring is limited by sample size and representativeness but interpretation of water chemistry monitoring data, for example, is supported by a long history of statistical analysis. Photo-point monitoring for watershed projects has almost no history of statistical analysis. Numeric data are needed for statistical analysis. The primary challenge for those who want to pursue low-cost photo-point monitoring options for project evaluation is to develop more quantitative data and put that data through statistical analyses to create a record of achievement and potential.

Photographic Data Collected at UW-Platteville Pioneer Farm

Researchers at the University of Wisconsin-Platteville have applied photo-point monitoring to farm-scale research. Photographs are used to identify areas of concern, record field conditions within research project areas, monitor the locations of grazing cattle, record unusual or atypical events, and support QA/QC efforts in the surface-water runoff monitoring program. Photographs can be especially useful to convey information to off-site researchers.

Time-lapse photos are taken on a 24-hr interval at surface-water gauging stations to create a record of field conditions within monitored areas. These photographs are useful in determining soil cover, plant canopy, snow cover, and crop growth throughout the year- especially at times when runoff events occur. Moreover, photographs of surface water runoff sample bottles are taken after collection and prior to lab analysis (Figure 1). While bottle photos provide only qualitative information, such as relative sample color, this information, along with time-lapse photos can help confirm results when laboratory test results are in question. Photos of the bottle tops are used as part of the chain of custody record and project QA/QC, providing an accurate record of samples shipped for analysis.

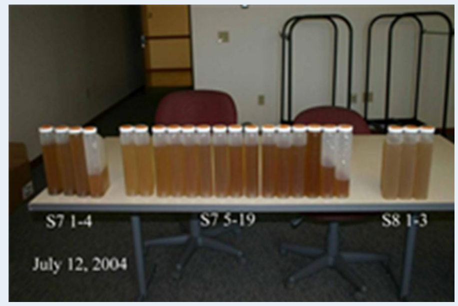


Figure 1. Sample bottles

Daily time-lapse photos have also been used both to identify paddocks where cattle are grazing in riparian corridors, and to record pasture vegetation height and density. In studies where the location of grazing cattle needs to be recorded daily, landscape photographs can identify the paddocks in which cattle are grazing on a daily basis (Figure 2). Plot photos of pasture vegetation have been used to create a visual record of pasture condition and grass height for runoff studies as well (Figure 3).

Figure 2. Grazing cattle

Figure 3. Pasture vegetation

Photographs are often taken to record extreme events and unusual field observations. For example, photographs have been taken of high-flow events where water depth was greater than the flume height and runoff water flowed over the wing walls holding the flume (Figure 4). Information from these photos can be used to confirm recorded maximum stage readings, and estimate discharge by providing information that can be used to calculate cross-section flow area that occurs above the flume.

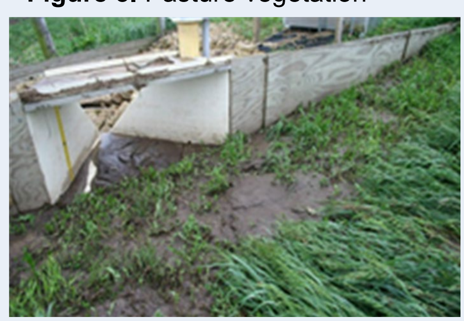


Figure 4. Flume

5.6 References

Bauer, S.B. and T.A. Burton. 1993. Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management on Western Rangeland Streams. EPA 910/R-93-017. U.S. Environmental Protection Agency, Water Division, Seattle, WA. Accessed February 10, 2016.

Bechtel, D.A. 2005. *River Continuity Assessment of the Ashuelot River Watershed Quality Assurance Project Plan*. The Nature Conservancy, New Hampshire Chapter, Concord, NH. Accessed February 10, 2016. http://des.nh.gov/organization/divisions/water/wmb/was/qapp/documents/ashuelot.pdf.

Bledsoe, B.P., and J.E. Meyer. 2005. *Monitoring of the Little Snake River and Tributaries Year 5 – Final Report.* Colorado State University, Department of Civil Engineering, Ft. Collins, CO. Accessed February 10, 2016. http://www.wildlandhydrology.com/assets/Monitoring_of_the_Little_Snake_River_Year_5_Final_Report.pdf.

Busch, D. and R. Mentz. 2012. Photographic Data Collected at UW-Platteville Pioneer Farm. University of Wisconsin-Platteville, Platteville, WI.

CCRWQCB (Central Coast Regional Water Quality Control Board). 2012a. *Water Quality Success Stories: Rural Roads Erosion Control Assistance in Santa Cruz County, California.* Central Coast Regional Water Quality Control Board, San Luis Obispo, CA. Accessed February 10, 2016. http://www.waterboards.ca.gov/centralcoast/about_us/docs/success/rural_roads.pdf.

CCRWQCB (Central Coast Regional Water Quality Control Board). 2012b. *Water Quality Success Stories: The Morro Bay National Monitoring Program - a Ten-Year Rangeland Management Practices Study.* Central Coast Regional Water Quality Control Board, San Luis Obispo, CA. Accessed February 10, 2016. http://www.waterboards.ca.gov/centralcoast/about_us/morrobay.shtml.

CCRWQCB (Central Coast Regional Water Quality Control Board) and CPSU (California Polytechnic State University). 2003. *Morro Bay National Monitoring Program: Nonpoint Source Pollution and Treatment Measure Evaluation for the Morro Bay Watershed 1992-2002 Final Report*. Prepared for U.S. Environmental Protection Agency, by Central Coast Regional Water Quality Control Board and California Polytechnic State University – San Luis Obispo, CA. Accessed February 10, 2016. http://www.elkhornsloughctp.org/uploads/files/1149032089MorroBayEstuaryProgram.pdf.

Clausen, John C., University of Connecticut, Department of Natural Resources and the Environment. 2011. Telephone conversation with Steve Dressing, Tetra Tech, Inc., regarding Jordan Cove Section 319 National Monitoring Program project. Accessed February 10, 2016. See http://jordancove.uconn.edu/index.html for additional details.

Eck, K.J. and D.E. Brown. 2004. *Estimating Corn and Soybean Residue Cover.* AY-269-W. Purdue University Cooperative Extension Service, West Lafayette, IN. Accessed February 10, 2016. http://www.extension.purdue.edu/extmedia/AY/AY-269-W.pdf.

ERS (Environment and Resource Sciences). 2010. *Land Manager's Monitoring Guide – Photopoint Monitoring.* State of Queensland Department of Environment and Resource Management, Environment and Resource Sciences, Brisbane. Accessed February 10, 2016. http://www.granitenet.com.au/assets/files/Landcare/rabbits_photopoint_indicator_aug2010.pdf.

Eyre, T.J., A.L. Kelly, and V.J. Neldner. 2011. *Method for the Establishment and Survey of Reference Sites for Biocondition Version 2.0*. State of Queensland Department of Environment and Resource Management, Biodiversity and Ecological Sciences Unit, Brisbane. Accessed February 10, 2016. https://www.qld.gov.au/environment/assets/documents/plants-animals/biodiversity/reference-sites-biocondition.pdf.

Eyre, T.J., A.L. Kelly, V.J. Neldner, B.A. Wilson, D.J. Ferguson, M.J. Laidlaw, and A.J. Franks. 2015. *Biocondition: a Condition Assessment Framework for Terrestrial Biodiversity in Queensland. Assessment Manual. Version 2.2*. State of Queensland Department of Environment and Resource Management, Biodiversity and Ecosystem Sciences, Brisbane. Accessed February 10, 2016. https://www.qld.gov.au/environment/assets/documents/plants-animals/biodiversity/biocondition-assessment-manual.pdf.

Faux, R.N., P. Maus, H. Lachowski, C.E. Torgersen, and M.S. Boyd. 2001. *New Approaches for Monitoring Stream Temperature: Airborne Thermal Infrared Remote Sensing.* U.S. Department of Agriculture, Forest Service, Remote Application Center, Integrating & Monitoring Steering Committee, San Dimas, CA. Accessed February 10, 2016. http://www.fs.fed.us/eng/techdev/IM/rsac_reports/TIR.pdf.

GeoSpatial Experts. 2004. *Introduction to GPS-Photo Link.* U.S. Department of Agriculture, Natural Resources Conservation Service, NE. Accessed February 10, 2016. GPS-Photo_Link_step_by_step.ppt

GeoSpatial Experts. 2016. *Geotagging and Photo Mapping Software*. Geospatial Experts, Inc., Thornton, CO. Accessed February 10, 2016. http://www.geospatialexperts.com/geotagging.php.

Hall, Frederick C. 2001. *Ground-Based Photographic Monitoring.* General Technical Report PNW-GTR-503. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR. Accessed February 11, 2016. http://www.fs.fed.us/pnw/publications/pnw_gtr503/.

Hall, Frederick C. 2002. *Photo Point Monitoring Handbook.* General Technical Report PNW-GTR-526. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR. Accessed February 10, 2016. http://www.fs.fed.us/pnw/pubs/gtr526/.

Hamilton, R.M. n.d. *Photo Point Monitoring, aWeed Manager's Guide to Remote Sensing and GIS — Mapping & Monitoring*. U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center, Salt Lake City, UT. Accessed February 10, 2016. http://www.fs.fed.us/eng/rsac/invasivespecies/documents/Photopoint_monitoring.pdf.

Herrick, J.E., J.W. Van Zee, K.M. Havstad, L.M. Burkett, and W.G. Whitford. 2005. *Monitoring Manual for Grassland, Shrubland and Savanna Ecosystems - Volume I: Quick Start*. U.S. Department of Agriculture, Agricultural Research Service Jornada Experimental Range, Las Cruces, NM. Accessed February 10, 2016. http://www.ntc.blm.gov/krc/viewresource.php?courseID=281&programAreaId=148.

Herrick, J.E., J.W. Van Zee, K.M. Havstad, L.M. Burkett, and W.G. Whitford. 2005a. *Monitoring Manual for Grassland, Shrubland and Savanna Ecosystems - Volume II: Design, Supplementary*

Methods And Interpretation, U.S. Department of Agriculture, Agricultural Research Service Jornada Experimental Range, Las Cruces, NM. Accessed February 10, 2016. http://www.ntc.blm.gov/krc/viewresource.php?courseID=281&programAreaId=148.

Hickman, J.S. and D.L. Schoenberger. 1989. *Wheat Residue.* Bulletin L-781. Kansas State University, Cooperative Extension Service, Manhattan, KS. Accessed February 10, 2016. http://www.ksre.ksu.edu/bookstore/pubs/L781.pdf.

Hively, W.D., M. Lang, G.W. McCarty, J. Keppler, A. Sadeghi, and L.L. McConnell. 2009a. Using satellite remote sensing to estimate winter cover crop nutrient uptake efficiency. *Journal of Soil and Water Conservation* 64(5):303-313.

Hively, W.D., G.W. McCarty, and J. Keppler. 2009b. Federal-state partnership yields success in remote sensing analysis of conservation practice effectiveness: results from the Choptank River Conservation Effects Assessment Project. *Journal of Soil and Water Conservation* 64(5):154A.

Key, C. H., D. B. Fagre, and R. K. Menicke. 2002. Glacier Retreat in Glacier National Park, Montana. In *Satellite Image Atlas of Glaciers of the World, Glaciers of North America - Glaciers of the Western United States.* U.S. Geological Survey Professional Paper 1386-J, R.M. Krimmel, pp. J365-J381. Accessed February 10, 2016. http://nrmsc.usgs.gov/files/norock/products/GCC/SatelliteAtlas_Key_02.pdf.

Kinney, J.W. and W.P. Clary. 1994. A *Photographic Utilization Guide for Key Riparian Graminoids.* General Technical Report INT-GTR-308. U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, UT. Accessed February 10, 2016. http://www.fs.fed.us/rm/pubs_int/int_gtr308.html.

Kinney, J.W. and W.P. Clary. 1998. *Time-Lapse Photography to Monitor Riparian Meadow Use.* USDA Forest Service Research Note RMRS-RN-5. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Boise, ID. Accessed February 10, 2016. http://www.fs.fed.us/rm/pubs/rmrs_rn005.pdf.

Larsen, D. 2006. *Minimum Standards for Nebraska NRCS Photo-Point Monitoring*. Range and Pasture Technical Note No.16. U.S. Department of Agriculture, Natural Resources Conservation Service. Accessed February 10, 2016. http://efotg.sc.egov.usda.gov/references/public/NE/NE_TECH_NOTE_16.pdf.

MathWorks. 2012. *MATLAB Image Processing Toolbox.* MathWorks, Natick, MA. Accessed February 10, 2016. http://www.mathworks.com/products/image/.

MMC (Mosman Municipal Council). n.d. *Chinamans Beach Photo Point Monitoring Program Spring/Summer 2004 & Autumn/Winter 2005.* Mosman Municipal Council, Mosman, Australia.

MST (Missouri Stream Team). n.d. *Stream Team Photo Point Monitoring.* Missouri Stream Team, Jefferson City, MO. Accessed February 10, 2016. http://www.mostreamteam.org/Documents/STActivities/ppmonitoring.pdf.

NMED (New Mexico Environment Department). 2009. *Watershed Protection Section Clean Water Act §319(H) Projects 1998 – 2008*. New Mexico Environment Department, Surface Water Quality Bureau.

OEPA (Ohio Environmental Protection Agency). n.d. *Ohio Section 319 Success Story: Wetland Restoration in the Sandusky Watershed.* Ohio Environmental Protection Agency, Columbus, OH. Accessed February 11, 2016. http://www.epa.ohio.gov/portals/35/nps/319DOCS/Wetlands%20Restoration%20in%20Northern%20Sandusky%20Watershed.pdf.

Oregon DEQ (Department of Environmental Quality). 2002. *Oregon 2002 Nonpoint Source Pollution Program Annual Report.* Oregon Department of Environmental Quality, Water Quality Division, Portland, OR. Accessed February 11, 2016. http://www.deq.state.or.us/wq/nonpoint/docs/annualrpts/rpt02.pdf.

PCEI (Palouse-Clearwater Environmental Institute). 2005. *South Fork Palouse River - Lower Watershed Restoration Project Final Report.* Palouse-Clearwater Environmental Institute, Moscow, ID.

Rasmussen J., M. Nørremark, and B.M. Bibby. 2007. Assessment of leaf cover and crop soil cover in weed harrowing research using digital images. *Weed Research* 47, 299–310.

Richardson, M.D., D.E. Karcher, and L.C. Purcell. 2001. Quantifying turfgrass cover using digital image analysis. *Crop Science* 41:1884-1888.

Rosgen, David L., 1973. The Use of Color Infrared Photography for the Determination of Sediment Production. In *Fluvial Process and Sedimentation: Proceedings of Hydrology Symposium*, Canadian National Research Council, Edmonton, Alberta, May 8-9, 1973, pp. 381-402.

Rosgen, David L., 1976. The Use of Color Infrared Photography for the Determination of Suspended Sediment Concentrations and Source Areas. In *Proceedings of the 3rd Inter-Agency Sediment Conference, Water Resources Council*, Denver, Colorado, March 22-25, 1976, Chapter 7, pp. 30-42. Accessed February 11, 2016. http://www.wildlandhydrology.com/assets/The_Use_of_Color_Infrared_Photography-FISC.pdf.

Shaff, C., J. Reiher, and J. Campbell. 2007. *OWEB Guide to Photo Point Monitoring.* Oregon Watershed Enhancement Board, Salem, OR. Accessed February 11, 2016. http://www.oregon.gov/OWEB/docs/pubs/PhotoPoint_Monitoring_Doc_July2007.pdf?ga=t.

Shelton, D.P., P.J. Jasa, J.A. Smith, and R. Kanable. 1995. G95-1132 *Estimating Percent Residue Cover.* Paper 784. Historical Materials from University of Nebraska-Lincoln Extension. Accessed February 11, 2016. http://digitalcommons.unl.edu/extensionhist/784/.

SOLVE. 2011. *Adventures in Monitoring.* SOLVE Green Team, Oregon. Accessed February 11, 2016. http://solvgreenteam.wordpress.com/tag/photo-point-monitoring/.

USDA-CES (U.S. Department of Agriculture, Cooperative Extension Service), USDA-NRCS (U.S. Department of Agriculture, Natural Resources Conservation Service), and USDOI-BLM (U.S. Department of Interior-Bureau of Land Management). 1999. *Utilization Studies and Residual Measurements*. Interagency Technical Reference 1734-3. U.S. Department of Agriculture, Cooperative Extensive Service and Natural Resources Conservation Service and U.S. Department of the Interior, Bureau of Land Management, Denver, CO. Accessed February 11, 2016. http://www.blm.gov/nstc/library/pdf/utilstudies.pdf.

USDA-FS (U.S. Department of Agriculture, Forest Service). 2007. *Removal of Marmot Dam.* U.S. Department of Agriculture, Forest Service, Mt. Hood National Forest, Sandy, OR. Accessed February 11, 2016. http://www.fs.usda.gov/detail/mthood/home/?cid=STELPRDB5200643 and http://or.water.usgs.gov/projs_dir/marmot/index.html.

USEPA (U.S. Environmental Protection Agency). 2008. *Handbook for Developing Watershed Plans to Restore and Protect Our Waters.* EPA 841-B-08-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 10, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/handbook-developing-watershed-plans-restore-and-protect.

USSARTF (United States Search and Rescue Task Force). n.d. *Compass Basics,* United States Search and Rescue Task Force, Elkins Park, PA. Accessed February 11, 2016. http://www.ussartf.org/compass_basics.htm.

Wolfram. 2012. *Mathematica.* Wolfram Research, Champaign, IL. Accessed February 11, 2016. http://www.wolfram.com/mathematica/?source=nav.

This page intentionally left blank.

6 Monitoring Challenges and Opportunities

By D.W. Meals and S.A. Dressing

Monitoring is the foundation of water quality management and provides essential information about the resource. Carefully done, monitoring can answer important questions and contribute to a successful NPS watershed project. However, monitoring can also be challenging and offer numerous pitfalls.

Sections 6.1 and 6.2 of this chapter highlight some of the problems that can hinder watershed monitoring efforts from the planning stage through execution. Opportunities to enhance and expand the impact and utility of monitoring data are discussed in sections 6.3 and 6.4.

6.1 Monitoring Pitfalls

Too many watershed monitoring projects have reported little or no improvement in water quality after extensive implementation of BMPs in the watershed. Reasons for this outcome are numerous and varied and may include:

- Mistakes in understanding of pollution sources
- Improper selection of BMPs
- Poor experimental design
- Uncooperative weather
- Lag time between treatment and response

There are numerous ways that a monitoring effort can fail to achieve its objectives. Reid (2001) examined 30 U.S. monitoring programs and classified reasons for failure into design flaws and procedural problems. Design flaws are errors or shortcomings inherent in the monitoring plan that prevent monitoring from obtaining appropriate data, answering fundamental questions, or otherwise achieving its goals. Serious design flaws can doom a monitoring project from the start and no amount of hard work or added resources can salvage it. Procedural problems are problems in execution of a program that can cause even the best design to fail. Unlike design problems, procedural problems can be overcome by applying additional resources, more personnel, better training, or good management.

A list of the top reasons for monitoring failure drawn from Reid (2001) and experience with numerous NPS monitoring projects includes both design and procedural problems.

6.1.1 Design Flaws

- **Inadequate problem identification/analysis.** In some cases, the source of NPS problems is unclear. For example, *E. coli* bacteria can come from livestock, domestic pets, septic systems, or wildlife. Without accurate identification of the pollutant source (*E. coli* in this case), monitoring is unlikely to be able to document a response to treatment effectively.
- **Fundamental misunderstanding of the system.** Effective monitoring of pollutant load or delivery requires an understanding of how the pollutant moves through the watershed. Monitoring in the

wrong place or on the wrong pathway will doom a program to failure. If nitrate-N moves mainly through ground water, for example, monitoring of surface runoff or streamflow is unlikely to yield good results. Similarly, if most suspended sediment at a watershed outlet comes from stream channels and banks, edge-of-field monitoring will not be effective.

- **Inability of the monitoring plan to measure what is needed.** If a sampling station is mis-located – upstream of a critical tributary inflow, for example – samples taken cannot record the pollutant load delivered in that inflow.
- **Insufficient study duration.** Significant lag time between land treatment and water quality response is common (see section 6.2, below). No matter how well-executed, a three-year monitoring program cannot document a response to BMPs if the response takes ten years to become evident because of legacy pollutants or slow watershed processes.
- **Statistically weak design**. As discussed in section 2.4, monitoring design must be carefully selected to achieve program objectives, be they load measurement, change in pollutant concentration, or response to land treatment, notably in the context of weather-driven variability characteristic of NPS pollution. A statistically weak design – such as a single watershed before and after or side-by-side watersheds – cannot control for weather variations and is unlikely to be able to attribute observed changes in water quality to a specific cause.

6.1.2 Procedural Problems

- **Lack of training or enthusiasm of field staff.** If a field technician is unable or unwilling to collect essential data because of lack of knowledge or initiative, critical data may be lost. In extreme cases, individuals can compromise a data record by cutting corners as illustrated in Figure 6-1. A simple time plot of recently obtained laboratory results revealed a pattern that indicated a sampling irregularity, thus triggering an investigation into the cause before further damage could be done.
- **Failure to collect collateral information.** Often, collateral information is required to properly interpret monitoring data. Information on stream stage, for example, may be essential to understand if a water sample was collected on the rising or falling limb of the hydrograph. Failure to record stage at the time of sample collection will greatly reduce the meaning of the sample result.
- **Bad or misunderstood technology.** Modern field or laboratory instruments make it easy to collect a great deal of monitoring data. However, if a field instrument is deployed for long periods without maintenance or calibration, or if a laboratory instrument is not calibrated and tested regularly, the resulting bad data will seriously impair a monitoring program.
- **Failure to evaluate data regularly.** As noted in section 3.10.2 and illustrated in Figure 6-2, it is essential to examine monitoring data frequently to catch problems early. Two dramatic changes in the apparent pattern of TKN concentration were caused by laboratory actions. Replacement of a defective probe in a lab instrument changed the range and sensitivity of the analytical results (point labeled #1). Later a change in lab method significantly raised the detection limit (point labeled #2). These two phenomena required rejection of almost a year of TKN data, but if the problems had not been noted in a data review, serious bias would have been introduced into the monitoring results for a seven-year monitoring effort (Meals 2001).
- **Protocol changes.** Whether in field or laboratory settings, consistent operating procedures are essential to generating consistent monitoring data. Although long-term monitoring programs should strive for consistency in methods and procedures, sometimes it is necessary to replace or upgrade

instruments or change analytical methods. Without careful documentation and extensive comparative analysis, changes in monitoring or analytical procedures can introduce spurious changes in resulting data, changes that do not reflect conditions in the water resource.

- **Personnel change.** Complex monitoring activities – such as those involving GIS or sophisticated laboratory instruments – require a high level of expertise and/or training. Frequent personnel changes can result in loss of such expertise, with a consequent loss of data or of data accuracy, especially if transitions are not managed properly.
- **Lack of institutional integration.** Most watershed monitoring projects involve multiple participants, with responsibility for different activities sometimes spread across several institutions. If the different departments or agencies do not share information or talk to each other regularly, critical information may be overlooked and the monitoring program may suffer.

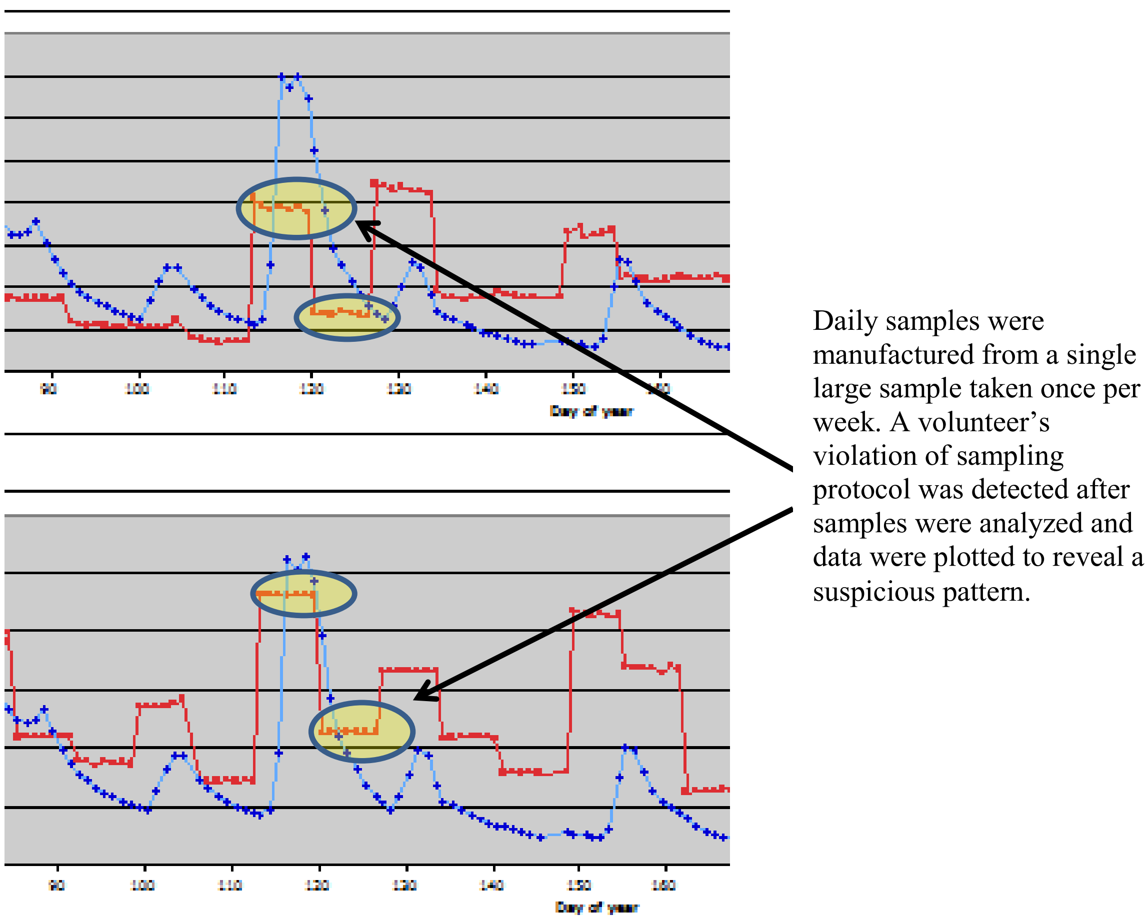


Figure 6-1. Detection of violation of sampling protocol (R.P. Richards, Heidelberg University, Tiffin, OH)

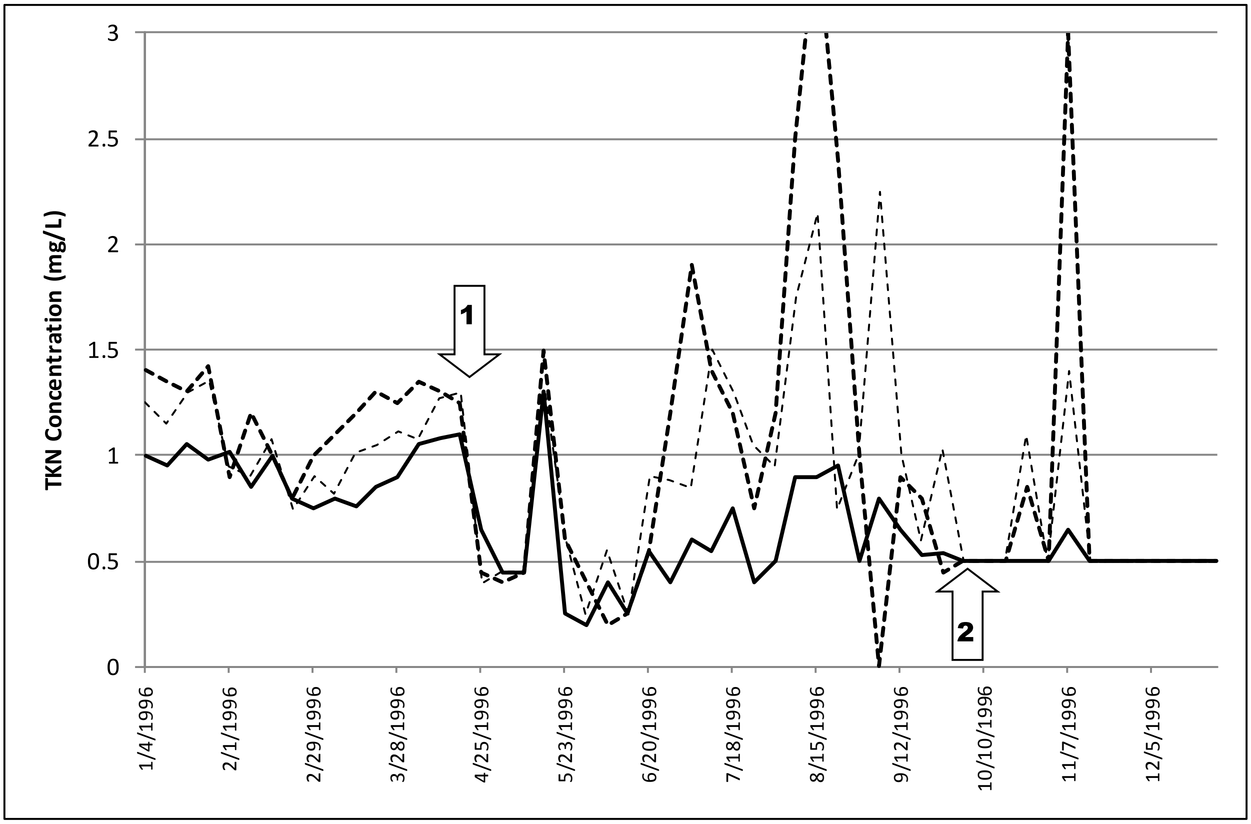


Figure 6-2. Effects of changing (1) a defective probe and (2) a laboratory method detection limit (Meals 2001)

Because design flaws may doom a monitoring project from the start, it is essential to follow the steps in designing a monitoring program discussed in chapters 2 and 3. Procedural problems can be addressed with additional resources, training, and good management during the course of a monitoring program, but such corrections require constant vigilance to identify the problems before they cause too much damage.

6.2 Lag Time Issues in Watershed Projects

One important reason NPS watershed projects may fail to meet expectations for water quality improvement is lag time. Lag time can be thought of as the time elapsed between installation or adoption of management measures at the level projected to reduce NPS pollution and the first measurable improvement in water quality in the target waterbody. Even in cases where a program of management measures is well-designed and fully implemented, water quality monitoring efforts (even those designed to be “long-term”) may not show definitive results if the monitoring period and sampling frequency are not sufficient to address the lag between treatment and response. Lag time issues have been explored in detail in a recent review (Meals et al. 2010).

Project management, watershed processes, and components of the monitoring program itself influence the lag between treatment and response (Figure 6-3). Any or all of these may come into play in a watershed project.

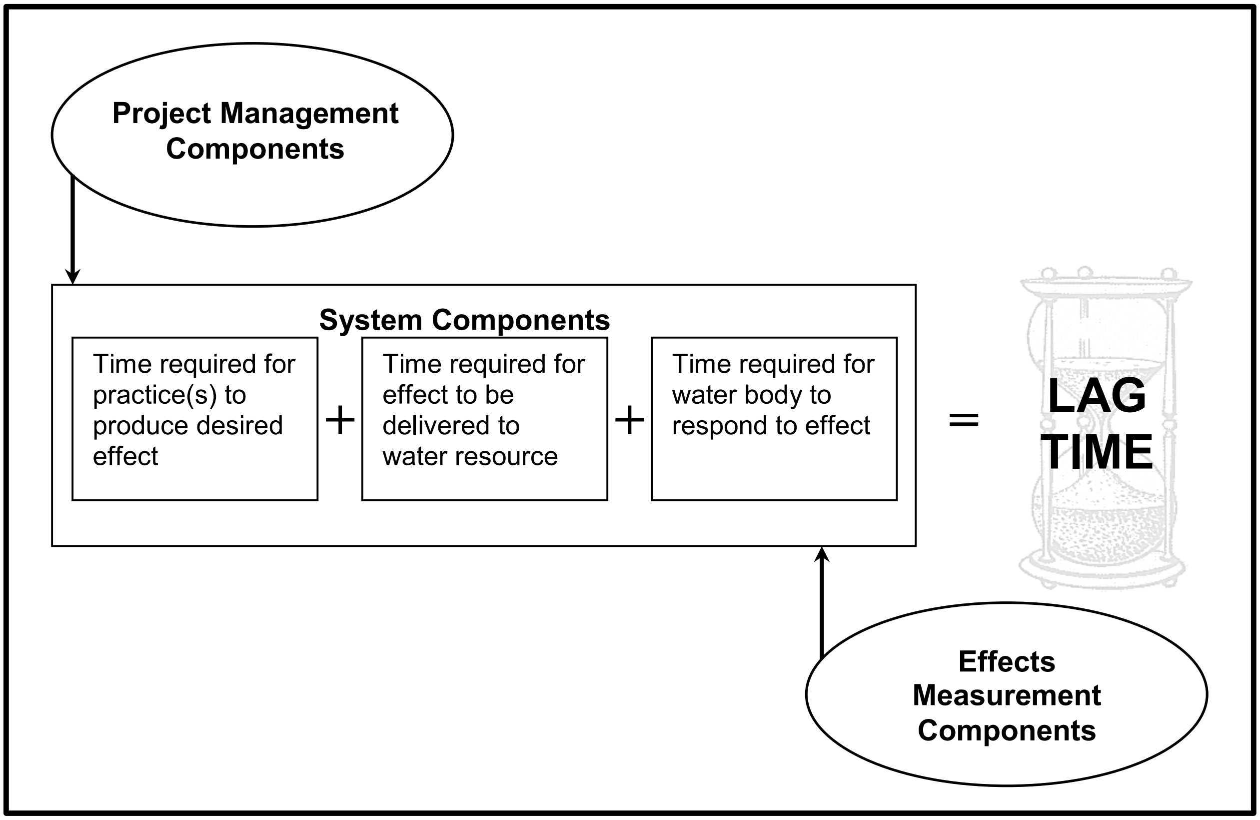


Figure 6-3. Lag time conceptual model

6.2.1 *Project Management Components*

The time required for planning and implementation of a NPS watershed project often causes the public to perceive a delay between the decision to act and results of that decision. A project may be funded and announced today, but it will be some time before that project will be fully planned and implementation begins. It might even take years, considering the essential time required to identify NPS pollution sources and critical areas, design management measures, engage landowner participation, and integrate new practices into cropping and land management cycles. Although such planning delays are not part of the physical process of lag time, stakeholders will often perceive them as part of the wait for results.

6.2.1.1 Time Required for an Installed or Adopted Practice to Produce an Effect

BMPs are installed in watersheds to provide a wide range of effects to protect or restore the physical, chemical, and biological condition of waterbodies, including:

- Change hydrology
- Reduce dissolved pollutant concentration or load
- Reduce particulate/adsorbed pollutant concentration or load
- Improve vegetative habitat
- Improve physical habitat

The time required for a BMP to be fully installed and become operational influences how quickly it can produce an effect. Some NPS control measures may become functional quickly. Installation of livestock exclusion fencing along several Vermont streams over a three-month period resulted in significant nutrient concentration and load reductions and reductions of fecal bacteria counts in the first post-treatment year as the fences immediately prevented manure deposition in the stream (Meals 2001). However, other NPS management measures, especially vegetative practices where plant communities need time to become established, may take years to become fully effective. For example, in a Pennsylvania study of a newly-constructed riparian forest buffer, the influence of tree growth on nitrate–N removal from groundwater did not become apparent until about ten years after tree planting (Newbold et al. 2009).

Lag time between BMP implementation and reduction of pollutant losses at the edge-of-field scale varies by the pollutant type and the behavior of the pollution source. Erosion controls such as cover crops, contour farming, and conservation tillage tend to have a fairly rapid effect on soil loss from a crop field as these practices quickly mitigate the forces contributing to detachment and transport of soil particles (Nearing et al. 1990). However, the response time of runoff P to nutrient management is likely to be much slower. It may take years to "mine" excess P out of the soil through crop removal to the point where dissolved P in runoff is effectively reduced (Zhang et al. 2004, Sharpley et al. 2007).

6.2.1.2 Time Required for the Effect to be Delivered to the Water Resource

Practice effects initially occur at or near the practice location, yet managers and stakeholders usually want and expect the impact of these effects to appear promptly in the water resource of interest in the watershed. The time required to deliver an effect to a water resource depends on a number of factors, including:

- The route for delivering the effect
  - Directly in (e.g., streambed restoration) or immediately adjacent to (e.g., shade) the water resource
  - Overland flow (particulate pollutants)
  - Overland and subsurface flow (dissolved pollutants)
  - Infiltration groundwater and groundwater flow (e.g., nitrate)
- The path distance
- The path travel rate
  - Fast (e.g., ditches and artificial drainage outlets to surface waters)
  - Moderate (e.g., overland and subsurface flow in porous soils)
  - Slow (e.g., infiltration in absence of macropores and groundwater flow)
  - Very slow (e.g., transport in a regional aquifer)
- Hydrologic patterns during the study period
  - Wet periods generally increase volume and rate of transport
  - Dry periods generally decrease volume and rate of transport

Once in a stream, dissolved pollutants like N and P can move rapidly downstream with flowing water to reach a receiving body relatively quickly. However, sediment and attached pollutants (e.g., P and some synthetic chemicals) can take years to move downstream as particles are repeatedly deposited, resuspended, and redeposited within the drainage network by episodic high flow events. This process can delay sediment and P transport (when attached P constitutes a large fraction of the P load) from headwaters to outlet by years or even decades. Substantial lag time could occur between reductions of sediment and P delivery into the headwaters and measurement of those reductions at the watershed outlet.

Pollutants delivered predominantly in groundwater such as nitrate-N generally move at the rate of groundwater flow, typically much more slowly than the rate of surface water flow. For example, about 40% of all N reaching the Chesapeake Bay travels through groundwater before reaching the bay. Phillips and Lindsey (2003) estimated that N loads associated with groundwater in the Chesapeake Bay Watershed would have a median lag time of ten years for water quality improvements to become evident. Groundwater nitrate concentrations in upland areas of Iowa were still influenced by the legacy of past agricultural management conducted more than 25 years earlier (Tomer and Burkart 2003).

6.2.1.3 Time Required for the Waterbody to Respond to the Effect

The speed with which a waterbody responds to the effect(s) produced by and delivered from management measures in the watershed introduces another increment of lag time. For example, hydraulic residence time (or the inverse, flushing rate) is an important determinant of how quickly a waterbody may respond to changes in nutrient loading. Residence times in selected North American waterbodies range from 0.6 year for Chesapeake Bay to 3.3 years for Lake Champlain to 191 years for Lake Superior to more than 650 years for Lake Tahoe. Simply on the basis of dilution, it will likely take considerably longer for water column nutrient concentrations to respond to a decrease in nutrient loading in Lake Superior than in Chesapeake Bay.

Apparent lag time in water quality response may also depend on the indicator evaluated or the impairment involved, especially if the focus is on biological water quality. A relatively short lag time might be expected between reductions of *E. coli* bacteria inputs and reduction in bacteria levels in the receiving waters because the bacteria generally do not persist as long in the aquatic environment as do heavy metals or synthetic organic chemicals. Such response has been demonstrated in estuarine systems where bacterial contamination of shellfish beds has been reduced or eliminated through improved waste management on the land in less than a year (BBNEP 2008). Improved sewage treatment in Washington, D.C. led to sharp reductions in point source P and N loading to the Potomac River Estuary in the early 1970s (Jaworski 1990). The tidal freshwater region of the estuary responded significantly over the next 5 years with decreased algal biomass, higher water column dissolved oxygen levels, and increased water clarity.

In contrast, lake response to changes in incoming P load is often delayed by recycling of P stored in aquatic sediments. When P loads to Shagawa Lake (MN) were reduced by 80% through tertiary wastewater treatment, residence time models predicted new equilibrium P concentrations within 1.5 years, but high in-lake P levels continued to be maintained by recycling of P from lake sediments (Larsen et al. 1979). Even more than 20 years after the reduction of the external loading, sediment feedback of P continued to influence the trophic state of the lake (Seo and Canale 1999). Similarly, St. Albans Bay (VT) in Lake Champlain failed to respond rapidly to reductions in P load from its watershed. From 1980 through 1991, a combination of wastewater treatment upgrades and intensive implementation of dairy waste management BMPs through the Rural Clean Water Program brought about a reduction of P loads to this eutrophic bay. However, water quality in the bay did not improve significantly. This pattern was attributed to internal loading from sediments highly enriched in P from decades of point and NPS inputs (Meals 1992). Although researchers at that time believed that the sediment P would begin to decline over time as the internal supply was depleted, subsequent monitoring has shown that P levels have not declined over the years as expected (LCBP 2008). Recent research has confirmed that a substantial reservoir of P continues to exist in the sediments that can be transferred into the water under certain chemical conditions and nourish algae blooms for many years to come (Druschel et al. 2005). In effect, this internal loading has become a significant source of P, one that cannot be addressed by management measures on the land.

Macroinvertebrate or fish response to improved water quality and habitat conditions in stream systems requires time for the organisms to migrate into the system and occupy newly improved habitat. Significant lag times have been observed in the response of benthic invertebrates and fish to management measures implemented on land, including in the Middle Fork Holston River project (Virginia), where IBI scores and *Ephemeroptera-Plecoptera-Trichoptera* (EPT) scores did not improve, even though the project accomplished substantial reduction in the sediment, N, and P loadings (VADCR 1997). The lack of increase in the biological indicator scores indicates a system lag time between the actual BMP implementation and positive changes in the biological community structure. This lag could depend in part on the amount of ecological connectivity with neighboring healthier aquatic systems that could provide sources of appropriate organisms to repopulate the restored habitats. In several Vermont streams, the benthic invertebrate community improved within 3 years in response to reductions of sediment, nutrient, and organic matter inputs from the land (Meals 2001). However, despite observed improvement in stream physical habitat and water temperature, no improvements in the fish community were documented. The project attributed this at least partially to a lag time in community response exceeding the monitoring period.

6.2.2 Effects Measurement Components of Lag Time

Watershed project managers are routinely pressed for results by a wide range of stakeholders. The fundamental temporal components of lag time control how long it will take for a response to occur, but the effectiveness of measuring the response may cause a further delay in recognizing it. The design of the monitoring program is a major determinant of our ability to discern a response against the background of the variability of natural systems.

In the context of lag time, sampling frequency with respect to background variability is a key determinant of how long it will take to document change. In a given system, taking n samples per year provides a certain statistical power to detect a trend. If the number of samples per year is reduced, statistical power is reduced (the magnitude by which is influenced by the degree of autocorrelation), and it may take longer to document a significant trend or to state with confidence that a concentration has dropped below a water quality standard. Simply stated, taking fewer samples a year is likely to introduce an additional "statistical" lag time before a change can be effectively documented.

6.2.2.1 The Magnitude of Lag Time

The magnitude of lag time is difficult to predict in specific cases and generalizations are difficult to make. A few examples, summarized in Table 6-1, illustrate some possible time frames for several categories of lag times.

Table 6-1. Examples of lag times reported in response to environmental impact or treatment

Parameter(s)	Scale	Impact/Treatment	Response lag	Reference
Sediment	Large watershed	Extreme storm events	8-25 yr	Marutani et al. 1999
Sediment	Large watershed	Cropland erosion control	19 yr	Newson 2007
Chloride	Large aquifer	Road salt	> 50 yr	Bester et al. 2006
NO_3-N	Small watershed	N fertilizer rates	> 30 yr	Tomer and Burkart 2003
NO_3-N	River basin	N fertilizer rates	> 50 yr	Bratton et al. 2004
NO_3-N	Large watershed	Nutrient management	≥ 5 yr	STAC 2005
NO_3-N	Small watershed	Nutrient management	15-39 yr	Galeone 2005
NO_3-N	Small watershed	Prairie restoration	10 yr	Schilling and Spooner 2006
NO_3-N	Small watershed	Riparian forest buffer	10 yr	Newbold et al. 2009
Soil test P	Field	P fertilizer rates	8-14 yr	McCollum 1991
Soil test P	Field	P fertilizer rates	10-14 yr	Giroux and Royer 2007
Soil and runoff P	Plot/field	Poultry litter management	> 5 yr	Sharpley et al. 2007
P	Lake	WWTP upgrade	> 20 yr	Larsen et al. 1979
P	Lake	WWTP upgrade/agricultural BMPs	> 20 yr	LCBP 2008
P, N, *E. coli*	Small watershed	Livestock exclusion	≤ 1 yr	Meals 2001
Fecal bacteria	Estuary	Waste management	< 1 yr	BBNEP 2008
Fecal bacteria	Estuary	Waste management	1 yr	Spooner et al. 2011
Macroinvertebrates	Small watershed	Livestock exclusion	3 yr	Meals 2001
Macroinvertebrates	Small watershed	Mine waste treatment	10 yr	Chadwick et al. 1986
Fish	First order stream	Habitat restoration	2 yr	Whitney and Hafele 2006
Fish	Small watershed	Acid mine drainage treatment	3-9 yr	Cravotta et al. 2009

6.2.3 How to Deal with Lag Time

In most situations, some lag time between implementation of BMPs and water quality response is inevitable. Although the exact duration of the lag can rarely be predicted, in many cases the lag time will exceed the length of typical monitoring periods, making it problematic to document a water quality response. Several possible approaches are proposed to deal with this challenge.

6.2.3.1 Recognize Lag Time and Adjust Expectations

It usually takes time for a waterbody to become impaired and it will take time to accomplish the clean-up. Failure to meet quick-fix expectations may cause frustration, pessimism, and a reluctance to pursue further action. It is up to scientists, investigators, and project managers to do a better job explaining to all stakeholders in realistic terms that current water quality impairments usually result from historically poor land management and that immediate solutions should not be expected.

6.2.3.1.1 Characterize the Watershed

Before designing a NPS management program and an associated monitoring program, investigate important watershed characteristics likely to influence lag time. Determining the time of travel for groundwater movement is an obvious example. Watershed characterization is an important step in the project planning process (USEPA 2008) and such characterization should especially address important aspects of the hydrologic and geologic setting, as well as documentation of NPS pollution sources and the nature of the water quality impairment, all of which can influence observed lag time in system response.

6.2.3.1.2 Consider Lag Time Issues in Selection, Siting, and Monitoring of Best Management Practices

First and foremost, proper BMP selection must be based on solving the problem and ensuring that landowners have the capability and willingness to implement and maintain the BMPs. Lag time can be an important factor in the final design of BMP systems by ensuring that when down-gradient BMPs are installed, they are ready to handle the anticipated runoff or pollutant load from up-gradient sources. In addition, when projects include targeted BMP monitoring to document interim water quality improvements, recognition of lag time may require an adjustment of the approach to targeting the management program. When designing a program for projects that include BMP-specific monitoring, potential BMPs should be evaluated to determine which practices might provide the most rapid improvement in water quality, given watershed characteristics. For example, practices such as barnyard runoff management that affect direct delivery of nutrients into surface runoff and streamflow may yield more rapid reductions in nutrient loading to the receiving water than practices that reduce nutrient leaching to groundwater, when groundwater time of travel is measured in years. Fencing livestock out of streams may give an immediate water quality improvement, compared to waiting for riparian forest buffers to grow. Such considerations, combined with application of other criteria such as cost effectiveness, can help determine priorities for BMP implementation in a watershed project.

Lag time should also be considered in locating management practices within a watershed. Managers should consider the need to demonstrate results to the public, which may be easier at small scales, along with the need to achieve water quality targets and consequently wider benefits at the large watershed scale. Where sediment and sediment-bound pollutants from cropland erosion are primary concerns, implementing practices that target the largest sediment sources closest to the receiving water may provide a more rapid water quality benefit than erosion controls in the upper reaches of the watershed. Where groundwater transport is a key determinant of response, application of a groundwater travel time model before application of management changes could help managers understand when to anticipate a water quality response and communicate this issue to the public. At best, the model will support targeting the application of an initial round of management measures to land areas where the effects are expected to be transmitted to receiving waters quickly. An example of this can be found in Walnut Creek, Iowa (Schilling and Wolter 2007).

It is important to point out that factoring lag time into BMP selection and targeting is not to say that long-term management improvements like riparian forest buffer restoration should be discounted or that upland sediment sources should be ignored. Rather, it is suggested that planners and managers may want to consider implementing BMPs and treating sources likely to exhibit short lag times first to increase the probability of demonstrating some water quality improvement as quickly as possible. "Quick-fix" practices with minimum lag time must be complemented by other needed practices to ultimately yield permanent reductions in pollutant loads.

6.2.3.1.3 Monitor Small Watersheds Close to Sources

In cases where documentation of the effects of a management program on water quality is a critical goal, lag time can sometimes be minimized by focusing monitoring on small watersheds, close to pollution sources. Lag times introduced by transport phenomena (e.g., groundwater travel, sediment flux through stream networks) will likely be shorter in small watersheds than in larger basins. In the extreme, this principle implies monitoring at the edge of field or above/below a limited treated area, but small watersheds (e.g., < 1500 ha) can also yield good results. In the NNPSMP, projects monitoring BMP programs in small watersheds (e.g., the Morro Bay Watershed Project in California, the Jordan Cove Project in Connecticut, the Pequea/Mill Creek Watershed Project in Pennsylvania, and the Lake Champlain Basin Watersheds Project in Vermont) were more successful in documenting improvements in water quality in response to change than were projects that took place in large watersheds (e.g., the Lightwood Knot Creek Project in Alabama and the Sny Magill Watershed Project in Iowa) in the 7- to 10-year time frame of the NNPSMP (Spooner et al. 2011).

Monitoring programs can be designed to get a better handle on lag time issues. Monitoring indicators at all points along the pathway from source to response or conducting periodic synoptic surveys over the course of a project will identify changes as they occur and document progress toward the end response. Supplementing a stream monitoring program with special studies can help project managers understand watershed processes, predict potential lag times, and help explain delays in water quality improvement to stakeholders. In the Walnut Creek (IA) watershed, no changes in stream suspended sediment loads were documented, despite extensive conversion of row crop land to prairie and reductions in field erosion predicted by RUSLE (Revised Universal Soil Loss Equation). This was explained largely by a 22-mile stream survey showing that streambank erosion contributed more than 50% of Walnut Creek sediment export (Spooner et al. 2011).

6.2.3.1.4 Select Indicators Carefully

Some water quality variables can be expected to change more quickly than others in response to management changes. As documented in the Jordan Cove (CT) NNPSMP Project (1996–2005), peak storm flows from a developing watershed can be reduced quickly through application of stormwater infiltration practices (Clausen 2007). NNPSMP projects in California, North Carolina, Pennsylvania, and Vermont demonstrated rapid reductions in nutrients and bacteria by reducing direct deposition of livestock waste in surface waters through fencing livestock out of streams (Spooner et al. 2011).

Improvements in stream biota, however, often come beyond the time frame of many watershed-scale monitoring efforts, but a number of NNPSMP projects have documented success with biological monitoring. As noted in section 6.2.1, Meals (2001) found that the benthic invertebrate community in Vermont streams improved within 3 years in response to livestock exclusion practices, but improvements in the fish community were not documented. Whitney and Hafele (2006) noted improvements in the fish community within two years of a habitat restoration effort, and Cravotta et al. (2009) documented the gradual return of fish to streams within a few years after treatment to neutralize acid mine drainage.

Despite these successes, many other watershed-scale projects have failed to document improvements by monitoring macroinvertebrates and fish. This may simply argue for a more sustained monitoring effort to document a biological response to land treatment. Failing that, however, selection of indicators that have relatively short lag times where possible will make it easier (and quicker) to demonstrate success. Simple numbers of macroinvertebrates, for example, may respond before more complex community indices show change. See chapter 4 for additional details and illustrative case studies on biological monitoring approaches.

6.2.3.1.5 Design Monitoring Programs to Detect Change Effectively

Monitor at locations and at a frequency sufficient to detect change with reasonable sensitivity. Assess background variability before the project begins and conduct a minimum detectable change analysis as described in section 3.4.2 to determine a sampling frequency sufficient to document the anticipated magnitude of change with statistical confidence (Spooner et al. 1987, Richards and Grabow 2003). Although lag time will still be a factor in actual system response, a paired-watershed design (Clausen and Spooner 1993, King et al. 2008), where data from an untreated watershed are used to control for weather and other sources of variability, is one of the most effective ways to document water quality changes in response to improvements in land management. If a monitoring program is intended to detect trends, evaluate statistical power to determine the best sampling frequency for the project. See Meals et al. (2011) and section 7.8.2.4 for additional information on trend analysis.

Target monitoring to the effects expected from the BMPs implemented, in the sequence that those effects are anticipated. For example, when the ultimate goal is habitat/biota restoration in an urban stream, if BMPs are implemented first that will alter peak stormflows, design the monitoring program to track changes in hydrology. After the needed hydrologic restoration is achieved, monitoring can be redirected to track expected changes in channel morphology. Once changes in channel morphology are documented, monitoring can then focus on assessment of habitat and biological community response. Response of stream hydrology is likely to be quicker than restoration of stream biota and would therefore be a valuable—and more prompt—indicator of progress.

6.3 Integrating Monitoring and Modeling

Monitoring and modeling are the primary tools for assessment of NPS watershed projects. By providing essential data about the resource, water quality monitoring has long been the foundation of water quality management. Monitoring can, however, be expensive and technically challenging and requires careful design and execution to achieve objectives. Modeling, on the other hand, is indispensable in evaluating alternative scenarios and in forecasting water quality over time. Modeling is also technically demanding, and application of a model in the absence of observed data can contribute to legitimate skepticism and uncertainty about model results that can compromise the utility of modeling for watershed management. To meet the demands of future watershed programs, it is essential that we integrate the strengths of both tools.

6.3.1 The Roles of Monitoring and Modeling

Both monitoring and modeling have distinctive roles to play in watershed projects. In many cases these roles are complementary, but in some cases one tool is used as a substitute for the other for various reasons including budgetary constraints.

6.3.1.1 Monitoring

Monitoring plays many key roles in watershed projects:

- Identify and document water quality problems and impairments

- Assess compliance with water quality standards and other regulations
- Establish baseline conditions
- Provide credibility to project planning
- Provide data to support modeling
- Document water quality change
- Assess program or project effectiveness
- Provide information for adaptive management
- Inform stakeholders
- Contribute to behavior change by documenting actual watershed conditions

Monitoring can provide fundamental knowledge about the generation, fate, and transport of NPS pollutants. Monitoring data provide hard evidence of water quality impairment and represent the best evidence of water quality restoration. When successful, monitoring can effectively document water quality response to land treatment, e.g., reductions in nutrient and sediment loads resulting from livestock exclusion in Vermont (Meals 2004) and reductions in nitrate loading to streams from prairie restoration in Iowa (Schilling and Spooner 2006).

Water quality monitoring also presents important challenges in watershed projects. Over the past decades, many projects have failed to show water quality response through monitoring. Such failure can be attributed to shortcomings in both design (e.g., failure to measure what is needed, inadequate sampling frequency) and execution (e.g., failure to evaluate data regularly, inadequate staff training, poor institutional integration) (Reid 2001). As noted throughout this guidance, monitoring must be conducted under appropriate objectives with a statistical design that can meet those objectives. Monitoring must be conducted at a frequency adequate to meet objectives (e.g., to document change) and for an adequate duration (e.g., to overcome lag time). Water quality monitoring must be executed effectively, with careful attention to procedural issues like collection of collateral information, regular data evaluation, and institutional coordination.

6.3.1.2 Modeling

Modeling also plays a number of critical roles in watershed projects:

- Provide initial estimates of flow and pollutant loads
- Link sources to impacts and evaluate relative magnitudes of sources
- Identify critical areas for management
- Predict pollutant reductions and waterbody response to management actions
- Support informed choices among alternative actions
- Analyze cost-effectiveness of alternatives
- Address issues of lag time in system response to treatment
- Guide monitoring design
- Help build knowledge of natural processes and response to treatment

- Provide opportunities for collaborative learning and stakeholder involvement

Modeling can forecast future response to alternatives too numerous or time-consuming to monitor effectively. Modeling provides the means to assemble, express, and test the current state of knowledge and point the way for future investigations. Model applications for watershed evaluation range from the simple to the very complex. An Oklahoma project used SIMPLE (*Spatially Integrated Models for Phosphorus Loading and Erosion*) to identify high-risk P sources in the Peacheater Creek watershed to design a land treatment plan (Storm et al. 1996). A recent Vermont project used SWAT (*Soil and Water Assessment Tool*) to identify critical source areas for NPS P in a large agricultural watershed (Winchell et al. 2011). National CEAP Cropland Studies in the Upper Mississippi River Basin (USDA-NRCS 2012), the Chesapeake Bay region (USDA-NRCS 2011a), and the Great Lakes system (USDA-NRCS 2011b) used SWAT and other models to quantify the effects of conservation practices currently present on the landscape in the regions and to project potential benefits that could be gained by implementation of additional conservation treatment in under-treated agricultural acres.

Modeling also presents significant challenges in watershed projects. Some data are always required – for model parameterization, calibration, and validation – and inadequate supporting data can significantly degrade model performance. Technical and financial resources are required for modeling that may be difficult to assemble and sustain. Modeling may be impaired by inappropriate or outdated information (e.g., soil surveys, use of Curve Numbers), or by lack of fundamental understanding of how agroecosystems or urban stormwater processes function. The credibility of model application may be threatened by lack of appropriate algorithms for simulating conservation or urban stormwater management practices and by failure to adequately analyze uncertainties associated with model results. Model results nearly always require analysis and interpretation to be useful; failure to provide such support can lead to justifiable skepticism about model results. The Chesapeake Bay model, for example, has been criticized for overstating environmental achievements in contradiction to monitoring data (GAO 2005, Powledge 2005). Disputes or misunderstandings over pollutant loads simulated by the SPARROW model in the Mississippi River Basin have generated economic and political conflict over source identification and choices of alternatives for remediation (Robertson et al. 2009).

6.3.2 Using Monitoring and Modeling Together

Clearly, monitoring and modeling are not mutually exclusive and can be better integrated in watershed protection and restoration projects. Each tool has its own strengths and weaknesses and neither can by itself provide all the information needed for water quality decision-making or program accountability. Integration of monitoring and modeling should address these elements:

Use the strengths of both tools.

- Monitoring is the best tool for project evaluation, but modeling simulations and extrapolations can play an important role in projecting whether project success is likely.
- Modeling can provide guidance on where and how the on-the-ground monitoring is best conducted.
- Modeling is better than monitoring for comparing numerous scenarios and extrapolating effects into the future.
- Data collected through monitoring are essential for calibration and validation of models, and for establishing credibility for modeling-derived information.

- The validity of model application and the type of questions that are addressed must be corroborated by watershed stakeholders.
- Models are underutilized for collaborative learning purposes. Their use within collaborative frameworks must be promoted to incorporate feedback from stakeholders while demonstrating how decisions at the field-scale affect the environment.

Begin with project objectives and design the monitoring-modeling program to do what can be done well to meet those objectives.

- Begin with a clear set of objectives. Determine if the objectives need to be quantitative (e.g., reduce N load by 40%), if they need to incorporate time frames and scales for which accountability is needed (e.g., reduced N load at a tributary mouth or at each HUC-12), and if there is a need to attribute changes to activities on the land (e.g., in response to implementing specific management measures at a specified level).
- Establish a clear set of evaluation objectives. Define the specific questions to be answered with monitoring (measure N load reductions with a minimum detectable change of 20%) and with modeling (measure and project N load reductions within ±15% of actual loads). Incorporate the needed time frames and scales within the objectives, and ensure that monitoring and modeling objectives are complementary. For example, the monitoring objective might be to measure N load reductions with a minimum detectable change of 20% in select smaller watersheds within 10 years and assess with an MDC of 30% long-term N load trends at mouths of larger watersheds and the state line. The evaluation objective for modeling might be to estimate and project N load reductions within 15% of actual loads in select smaller watersheds within 10 years and estimate and project within 15% of actual long-term N load trends at mouths of larger watersheds and the state line. Address uncertainty at the outset and include uncertainty in all monitoring and modeling reporting.
- Select a model based on project needs – models selected solely by cost or convenience before setting objectives are unlikely to be satisfactory.
- Create a monitoring program that will collect the number and frequency of samples that are required to provide useful information – monitoring designs based solely on budget may yield data that cannot serve project objectives.

Select the appropriate designs.

- Establish the monitoring design(s). Address overall experimental design (e.g., long-term trend, upstream-downstream) and specify the elements of monitoring scale, sample type, station locations, sampling frequency, collection and analysis methods, land use/land treatment monitoring, and data management (see chapters 2 and 3).
- Select the modeling approach. Determine which model(s) to use, input data requirements and availability, model testing locations and procedures, and procedures for output analysis. Make certain that adequate technical skill and support are available for the selected approach.

Pay attention to source data.

- Availability of data at consistent scales and of known quality is essential to an integrated monitoring-modeling effort.
- Spatially- and temporally-explicit land treatment and agricultural management data are necessary for both water quality monitoring and watershed modeling.

- Identify common needs of monitoring and modeling. Share precipitation, land use, land treatment, and other data. Use monitored flow and water quality data to calibrate and validate the model(s).

Evaluate the suitability of both monitoring data/programs and proposed model(s) for the project in the project planning stage, before a project is funded and underway.

- Evaluate existing and planned monitoring data for quality, consistency, and suitability for project purposes.
- Evaluate candidate watershed models for applicability to watershed characteristics, technical competence, and resources necessary to apply and support modeling in the project.
- Verify that important watershed characteristics (e.g., claypan soils) and conservation and stormwater management practice functions can be adequately represented in the selected model.

Integrate data analysis and reporting.

- Combine systems for discharge calculations, loads calculated from monitoring data, and land use/land treatment data.
- Link monitoring data to a GIS framework used for modeling.
- Provide for compatibility between monitoring data and model(s) to permit efficient use of monitoring data for model calibration and validation.
- Facilitate analysis of small-scale monitoring and modeling to develop input parameters for large-scale model application(s).

Include a documentation plan for both monitoring and modeling.

- Use a formal Quality Assurance Project Plan (QAPP) to guide and document all aspects of the monitoring and modeling efforts.
- Lay out the purpose of model application and the justification for the selection of a particular model.
- Document the model name and version and the source of the model.
- Identify and document model assumptions.
- Document data requirements and sources of data sets to be used.
- Provide estimates of the uncertainty associated with modeling and monitoring results, particularly when they are used to quantify the environmental benefits of practices.

Develop a communication strategy. Control expectations from the beginning by addressing monitoring and modeling uncertainty explicitly. Avoid overly optimistic projections.

Be aware of potential differences in precision and accuracy of modeling results vs. monitoring data. Monitoring data may be used to identify trends or changes in water quality (see sections 7.7 and 7.8); such trends are identified in the context of statistical confidence, based largely on the characteristics of the monitoring program (see MDC, section 3.4.2). Model predictions, however, may show changes in water quality without the benefit of statistical trend analysis and thus suggest very small trends that cannot be verified by monitoring data. Monitoring data may, for example, support a MDC of 20% for phosphorus concentration, while a model may predict a 7% reduction. This situation is not necessarily contradictory, but calls for a bit of realistic caution in application and interpretation of model results.

Finally, in practical terms, project water quality monitoring and watershed modeling activities must be closely coordinated so that information from each effort can be collected, shared, and combined at appropriate times to meet project goals. Preliminary model runs to identify critical subwatersheds, for example, can also be used to help select monitoring station locations. Similarly, water quality data that are analyzed in a timely fashion as described in section 3.10.2 are more likely to be available at the right time for model calibration and validation.

6.4 Supporting BMP and Other Databases

6.4.1 General Considerations

Monitoring is often performed to develop a better understanding of BMP effectiveness, characterize reference conditions over broad geographic areas, determine effluent characteristics, or address other purposes not directly related to problem assessment or watershed project evaluation. In some cases this monitoring can be done in conjunction with problem assessment or project evaluation to maximize the return on resources expended, but this monitoring is often done separately.

The basic steps presented in chapters 2 and 3 should also be applied to development of monitoring plans in support of BMP and other databases. Some of the specifics may not apply, however, such as watershed characterization or monitoring of meteorological variables in cases where urban stormwater BMPs are assessed in a laboratory setting. Pollutant transport mechanisms and pollutant source activities may be of little interest in monitoring designed to establish reference conditions. Still, the focus on objectives must be the driving force behind all monitoring design.

For new databases, decisions need to be made regarding the types and quality of data that will be included. Development of a QAPP (see chapter 8) is an important first step in defining data needs and data quality expectations for the database.

When monitoring to support existing databases, it is essential that data requirements are reviewed and understood before the monitoring plan is developed to ensure that suitable data will be collected. For example, those managing the International Stormwater BMP Database have developed guidance with recommended BMP monitoring protocols that are directly related to requirements of the database, and have established a recommended protocol for evaluating BMP performance (Geosyntec and WWE 2009). This database is described in section 6.4.2.

Databases may have specific requirements for monitoring designs (e.g., above/below), sampling type (grab or composite), sampling frequencies, specific variables (e.g., EPA Method 365.4 for total P), and other monitoring details, as well as requirements for reporting information on the study conditions and features. For example, it may be required that designs for BMPs are reported in accordance with industry standards, or that a specific level of detail be reported for soils or crops. All of these requirements need to be reviewed and understood before monitoring begins.

Data format, approaches to data analysis, and data transmittal requirements may also be specified. Questions and issues associated with these requirements need to be addressed up front to prevent problems later.

The single most important step to take when monitoring in support of database development is for those performing the monitoring to communicate with those managing the databases to ensure that monitoring, data analysis, reporting, and data management requirements are understood and that the proposed monitoring plan is suitable before monitoring begins.

6.4.2 International Urban Stormwater BMP Database

The International Stormwater BMP Database (www.bmpdatabase.org/) is a database of over 530 BMP studies, performance analysis results, tools for use in BMP performance studies, monitoring guidance, and other study-related publications. The overall purpose of the project is to provide scientifically sound information to improve the design, selection, and performance of BMPs. Data obtained from BMP studies are expected to help create a better understanding of factors influencing BMP performance.

The database is focused on field studies of post-construction, permanent BMPs (International Stormwater BMP Database 2013). Data entry requirements are specified in a user's guide (WWW and Geosyntec 2010). Options for BMPs include structural BMPs, non-structural BMPs, low-impact development sites, and composite BMPs. Monitoring results may include precipitation, flow, water quality, and settling velocity.

Guidance is provided on approaches to determining BMP performance using concentrations, loads, and volume reductions (Geosyntec and WWW 2009). Comparison of the average value of the Event Mean Concentrations (EMC) or storm loads for the outlet as compared to the inlet is emphasized. Examining the cumulative distribution of each of the outlet and inlet storm EMCs allows for more detailed examination of the efficiency at different inlet loadings. This approach, the Effluent Probability Method (Strecker et al. 2003, Erickson et al. 2010), is described in more detail in section 7.7.2.

The database structure and contents may be downloaded from the project website and used solely for the following purposes (International Stormwater BMP Database 2013):

- Research and analysis related to BMP performance and costs, characterization of urban runoff, characterization of receiving water impacts, and characterization of the ability of BMPs to meet water quality goals or criteria.
- Use of database structure and/or data entry spreadsheets to track performance data for regional, state, watershed or local purposes or for subsequent upload to the International Stormwater BMP Database.

6.5 References

BBNEP (Buzzards Bay National Estuary Program). 2008. *Turning the Tide on Shellfish Bed Closures in Buzzards Bay During the 1990s*. Buzzards Bay National Estuary Program, East Wareham, MA. Accessed February 12, 2016. http://www.buzzardsbay.org/shellclssuccess.htm.

Bester, M.L., E.O. Frind, J.W. Molson, and D.L. Rudolph. 2006. Numerical investigation of road salt impact on an urban wellfield. *Ground Water* 44:165–175.

Bratton, J.F., J.K. Böhlke, F.M. Manheim, and D.E. Krantz. 2004. Submarine ground water in Delmarva Peninsula coastal bays: ages and nutrients. *Ground Water* 42:1021–1034.

Chadwick, J.W., S.P. Canton, and R.L. Dent. 1986. Recovery of benthic invertebrate communities in Silver Bow Creek, Montana following improved metal mine wastewater treatment. *Water, Air, & Soil Pollution* 28:427–438.

Clausen, J.C. 2007. *Jordan Cove Watershed Project Final Report*. University of Connecticut, College of Agriculture and Natural Resources, Department of Natural Resources Management and Engineering. Accessed January 8, 2016. http://jordancove.uconn.edu/jordan_cove/publications/final_report.pdf.

Clausen, J.C. and J. Spooner. 1993. *Paired Watershed Study Design*. 841-F-93-009. Prepared for S. Dressing, U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 12, 2016.

Cravotta, C.A., III., R.A. Brightbill, and M.J. Langland. 2009. Abandoned mine drainage in the Swatara Creek basin, southern anthracite coalfield, Pennsylvania, USA: 1. streamwater-quality trends coinciding with the return of fish. *Mine Water and the Environment* 29(3):176-199.

Druschel, G., A. Hartmann, R. Lomonaco, and K. Oldrid. 2005. *Determination of Sediment Phosphorus Concentrations in St. Albans Bay, Lake Champlain: Assessment of Internal Loading and Seasonal Variations of Phosphorus Sediment-Water Column Cycling*. Final Report prepared for Vermont Agency of Natural Resources, by University of Vermont, Department of Geology, Burlington.

Erickson, A.J., P.T. Weiss, J.S. Gulliver, and R.M. Hozalski. 2010. Analysis of Long-Term Performance. In *Stormwater Treatment: Assessment and Maintenance*. ed. J.S. Gulliver, A.J. Erickson, and P.T. Weiss. University of Minnesota, St. Anthony Falls Laboratory. Minneapolis, MN. Accessed February 12, 2016. http://stormwaterbook.safl.umn.edu/pollutant-removal/analysis-long-term-performance.

Galeone, D.G. 2005. Pequea and Mill Creek watersheds section 319 NMP project: effects of streambank fencing on surface-water quality. *NWQEP Notes* 118:1-6, 9. North Carolina State University Cooperative Extension, Raleigh. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/notes118.pdf.

GAO (Government Accountability Office). 2005. *Chesapeake Bay Program: Improved Strategies are Needed to Better Assess, Report, andManage Restoration Progress*. GAO-06-96. US Government Accountability Office, Washington, D.C.

Geosyntec and WWE (Wright Water Engineers, Inc.). 2009. *Urban Stormwater BMP Performance Monitoring*. Prepared for U.S. Environmental Protection Agency, Water Environment Research Foundation, Federal Highway Administration, and Environmental and Water Resources Institute of the American Society of Civil Engineers, by Geosyntec Consultants and Wright Water Engineers, Inc., Washington, DC. Accessed February 12, 2016. http://www.bmpdatabase.org/Docs/2009%20Stormwater%20BMP%20Monitoring%20Manual.pdf.

Giroux, M., and R. Royer. 2007. Long term effects of phosphate applications on yields, evolution of P soil test, saturation, and solubility in two very rich soils. *Agrosolutions* 18:17–24. Inst. de recherche et de développement en agroenvironnement (IRDA), Québec, Canada. (In French, with English abstract.)

International Stormwater BMP Database. 2013. Version 03 24 2013. Prepared for Water Environment Research Foundation (WERF), American Society of Civil Engineers (ASCE)/Environmental and Water Resources Institute (EWRI), American Public Works Association (APWA), Federal Highway Administration (FHWA), and U.S. Environmental Protection Agency (EPA), by Wright Water Engineers, Inc. and Geosyntec Consultants, Washington, DC. Accessed February 17, 2016. http://www.bmpdatabase.org/.

Jaworski, N. 1990. Retrospective of the water quality issues of the upper Potomac estuary. *Aquatic Sciences* 3:11–40.

King, K.W., P.C. Smiley, B. Baker, and N.R. Fausey. 2008. Validation of paired watersheds for assessing conservation practices in Upper Big Walnut Creek Watershed, Ohio. *Journal of Soil and Water Conservation*. 63(6):380-395.

Larsen, D.P., J. Van Sickle, K.W. Malueg, and P.D. Smith. 1979. The effect of wastewater phosphorus removal on Shagawa Lake, Minnesota: phosphorus supplies, lake phosphorus, and chlorophyll *a*. *Water Resources* 13:1259–1272.

LCBP (Lake Champlain Basin Program). 2008. *State of the Lake and Ecosystem Indicators Report-2008.* Lake Champlain Basin Program, Grand Isle, VT. Accessed February 17, 2016. http://www.lcbp.org/media-center/publications-library/state-of-the-lake/.

Marutani, T., M. Kasai, L.M. Reid, and N.A. Trustrum. 1999. Influence of storm-related sediment storage on the sediment delivery from tributary catchments in the Upper Waipaoa River, New Zealand. *Earth Surface Processes and Landforms* 24:881–896.

McCollum, R.E. 1991. Buildup and decline of soil phosphorus: 30-year trends on a Typic Umprabuult. *Agronomy Journal* 83:77–85.

Meals, D.W. 1992. Water Quality Trends in the St. Albans Bay, Vermont Watershed Following RCWP Land Treatment. In *The National Rural Clean Water Program Symposium*, Orlando, FL, September 1992. ORD EPA/625/R-92/006. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. pp. 47–58.

Meals, D.W. 2001. *Lake Champlain Basin Agricultural Watersheds Section 319 National Monitoring Program Project, Final Project Report: May, 1994-September, 2000*. Vermont Department of Environmental Conservation, Waterbury, VT.

Meals, D.W. 2004. Water Quality Improvements Following Riparian Restoration In Two Vermont Agricultural Watersheds. In *Lake Champlain: Partnership and Research in the New Millennium*, ed. T. Manley et al., pp. 81-96. Kluwer Academic/Plenum Publishers.

Meals, D.W., S.A. Dressing, and T.E. Davenport. 2010. Lag time in water quality response to best management practices. *Journal of Environmental Quality* 39:85–96.

Meals, D.W., J. Spooner, S.A. Dressing, and J.B. Harcum. 2011. *Statistical Analysis for Monotonic Trends*, Tech Notes #6, September 2011. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 16, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Nearing, M.A., L.J. Lane, E.E. Alberts, and J.M. Laflen. 1990. Prediction technology for soil erosion by water: status and research needs. *Soil Science Society of America Journal* 54:1702-1711.

Newbold, J.D., S. Herbert, B.W. Sweeney, and P. Kiry. 2009. Water quality functions of a 15-year-old riparian forest buffer system. *NWQEP Notes* 130:1-9. North Carolina State University Cooperative Extension, Raleigh. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/notes130.pdf.

Newson, J. 2007. Measurement and modeling of sediment transport in a northern Idaho stream. Masters thesis, University of Idaho, College of Graduate Studies, Biological and Agricultural Engineering, Moscow, ID.

Phillips, S.W. and B.D. Lindsey. 2003. *The Influence of Ground Water on Nitrogen Delivery to the Chesapeake Bay*. Fact Sheet FS-091–03. U.S. Geological Survey, MD-DE-DC Water Science Center, Baltimore, MD. Accessed March 15, 2016. http://md.water.usgs.gov/publications/fs-091-03/index.html.

Powledge, F. 2005. Chesapeake Bay restoration: a model of what? *BioScience* 55(12):1032-1038.

Reid, L.M. 2001. The epidemiology of monitoring. *Journal of American Water Resources Association* 37(4):815-820.

Richards, R.P., and G.L. Grabow. 2003. Detecting reductions in sediment loads associated with Ohio's conservation reserve enhancement program. *Journal of American Water Resources Association* 39:1261–1268.

Robertson, D.M., G.E. Schwarz, D.A. Saad, and R.B. Alexander.2009. Incorporating uncertainty into the ranking of SPARROW model nutrient yields from Mississippi/Atchafalaya River Basin watersheds. *Journal of American Water Resources Association* 45(2):534-549.)

Schilling, K.E., and J. Spooner. 2006. Effects of watershed-scale land use change on stream nitrate concentrations. *Journal of Environmental Quality* 35:2132–2145.

Schilling, K.E., and C.F. Wolter. 2007. A GIS-based groundwater travel time model to evaluate stream nitrate concentration reductions from land use change. *Environmental Geology* 53:433–443.

Seo, D. and R. Canale. 1999. Analysis of sediment characteristics and total phosphorus models for Shagawa Lake. *Journal of Environmental Engineering* 125(4), 346–350.

Sharpley, A.N., S. Herron, and T. Daniel. 2007. Overcoming the challenges of phosphorus-based management in poultry farming. *Journal of Soil and Water Conservation* 62:375–389.

Spooner, J., C.J. Jamieson, R.P. Maas, and M.D. Smolen. 1987. Determining statistically significant changes in water quality pollutant concentrations. *Lake Reservoir Management* 3:195–201.

Spooner, J., L.A. Szpir, D.E. Line, D. Meals, G.L Grabow, D.L. Osmond and C. Smith. 2011. *2011 Summary Report: Section 319 National Monitoring Program Projects, National Nonpoint Source Watershed Project Studies*. North Carolina State University, Biological and Agricultural Engineering Department, NCSU Water Quality Group, Raleigh, NC. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/319monitoring/toc.html.

STAC (Scientific and Technical Advisory Committee). 2005. *Understanding "Lag Times" Affecting the Improvement of Water Quality in the Chesapeake Bay*. STAC Publ. 05–001. Chesapeake Bay Program, Edgewater, MD. Accessed March 15, 2016.

Storm, D.E., G.J. Sabbagh, M.S. Gregory, M.D. Smolen, D. Toetz, D.R. Gade, C.T. Haan, and T. Kornecki. 1996. *Basin-Wide Pollution Inventory for the Illinois River Comprehensive Basin Management Program-Final Report*. Prepared for U.S. Environmental Protection Agency and Oklahoma Conservation Commission, by Oklahoma State University, Departments of Biosystems and Agricultural Engineering, Agronomy, and Zoology, Stillwater, OK. Accessed March 15, 2016.

Strecker, E.W., M.M. Quigley, and B. Urbanas. 2003. A Reassessment of the Expanded EPA/ASCE National BMP Database. In *Proceedings of the World Water and Environmental Congress 2003,* June 23-26, 2003, Philadelphia, PA. ed. P. Bizier and P. DeBarry, ISBN 0-7844-0685-5, American Society of Civil Engineers, Reston VA. pp. 555-574. Accessed March 15, 2016.

Tomer, M.D., and M.R. Burkart. 2003. Long-term effects of nitrogen fertilizer use on ground water nitrate in two small watersheds. *Journal of Environmental Quality* 32:2158–2171.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2011a. *Assessment of the Effects of Conservation Practices on Cultivated Cropland in the Chesapeake Bay Region*. U.S. Department of Agriculture, Natural Resources Conservation Service. Accessed February 16, 2016. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?&cid=stelprdb1041684.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2011b. *Assessment of the Effects of Conservation Practices on Cultivated Cropland in the Great Lakes Region*. U.S. Department of Agriculture, Natural Resources Conservation Service. Accessed February 16, 2016. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?&cid=stelprdb1045403.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2012. *Assessment of the Effects of Conservation Practices on Cultivated Cropland in the Upper Mississippi River Basin*. U.S. Department of Agriculture, Natural Resources Conservation Service. Accessed February 16, 2016. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?&cid=nrcs143_014161.

USEPA (U.S. Environmental Protection Agency). 2008. *Handbook for Developing Watershed Plans to Restore and Protect Our Waters.* EPA 841-B-08-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 15, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/handbook-developing-watershed-plans-restore-and-protect.

VADCR (Virginia Department of Conservation and Recreation). 1997. Alternative Watering Systems for Livestock-the Middle Fork Holston River Builds on Success. In *Section 319 Success Stories: Volume II – Highlights of State and Tribal Nonpoint Source Programs*. EPA 841-R-97-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 15, 2016.

Whitney, L., and R. Hafele. 2006. Stream restoration and fish in Oregon's Upper Grand Ronde river system. *NWQEP Notes* 123:1-6, 9-12. North Carolina State University Cooperative Extension, Raleigh. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/notes123.pdf.

Winchell, M., D. Meals, S. Folle, J. Moore, D. Braun, C. DeLeo, K. Budreski, and R. Schiff. 2011. *Identification of Critical Source Areas of Phosphorus with the Vermont Sector of the Missisquoi Bay Basin*. Final Report to Lake Champlain Basin Program. Accessed March 15, 2016. http://www.lcbp.org/techreportPDF/63B_Missisquoi_CSA.pdf.

WWE (Wright Water Engineers, Inc.) and Geosyntec. 2010. *International Stormwater Best Management Practices (BMP) Database User's Guide for BMP Data Entry Spreadsheets - Release Version 3.2.* Prepared for Water Environment Research Foundation, Federal Highway Administration, Environmental and Water Resources Institute of the American Society of Civil Engineers, U.S. Environmental Protection Agency, and American Public Works Association, by Wright Water Engineers, Inc. and Geosyntec Consultants, Denver, CO. Accessed February 12, 2016. http://www.bmpdatabase.org/Docs/2010%20BMP%20Database%20User's%20Guide.pdf.

Zhang, T.Q., A.F. MacKenzie, B.C. Liang, and C.F. Drury. 2004. Soil test phosphorus and phosphorus fractions with long-term phosphorus addition and depletion. *Soil Science Society of America Journal* 68:519–528.

This page intentionally left blank.

7 Data Analysis

By J. Spooner, J.B. Harcum, D.W. Meals, S.A. Dressing, and R.P. Richards

7.1 Introduction

This chapter of the guidance examines options for planning and analyzing data collected in nonpoint source watershed studies. The emphasis of this chapter is on projects at the watershed or subwatershed level, although evaluation of individual BMPs is also addressed. These analysis approaches complement the watershed project design considerations discussed in section 2.4 of this guidance.

Specifically, this chapter discusses the following topics:

- Exploratory data analysis
- Data transformations that might be necessary to prepare data for valid statistical analysis
- Methods to deal with extreme values, censored data, and missing data
- Data analysis methods for water quality problem assessment
- Data analysis methods for project planning
- Data analysis methods for assessing BMP or watershed project effectiveness
- Techniques for load estimation

The reader may wish to refer to chapter 4 (Data Analysis) of the 1997 guidance (USEPA 1997b) which was written largely to provide a primer on statistical methods for analysis of data generated by nonpoint source watershed projects. The 1997 guidance addresses various topics on statistical analysis in considerable detail, including estimation and hypothesis testing, characteristics of environmental data, and basic descriptive statistics. In addition, the 1997 guidance compares parametric and nonparametric tests, recommends appropriate methods for routine analyses, and provides numerous examples of the application of various statistical tests. Additional resources for data analysis approaches are also available in various *Tech Notes* and other publications (see References).

7.2 Overview of Statistical Methods

A wide range of parametric and nonparametric methods exists for analyzing environmental data. In some cases, graphical methods will be suitable to meet analysis objectives; more rigorous statistical analysis approaches may be best otherwise. This section provides a brief overview and summary of key features of these various methods. Readers should consult the 1997 guidance (USEPA 1997b) and additional sources (e.g., statistics textbooks and software packages) for greater detail.

Recommended statistical methods are summarized in Table 7-1 through Table 7-6 based on watershed project phase or need because experience indicates that this type of grouping will be practical for many involved in such efforts. Methods in these tables are recommended, but the tables do not include all possible alternative approaches. Additional discussion and illustrative examples follow in sections 7.3 through 7.8. Because of its importance to many watershed projects, especially those addressing TMDLs, pollutant load estimation is addressed separately in section 7.9. While most of the methods described in this chapter are more commonly applied to water chemistry, flow, and precipitation data, many can also be applied to biological data as well. Recommended approaches for analyzing biological data are described in detail in chapter 4, and some examples are also provided in this chapter.

7.2.1 Exploratory Data Analysis and Data Transformations

It is often necessary to work with a mix of information and data during the initial stages of watershed projects. A major first task involves gathering and organizing available information and data, followed by an initial examination of the data to help identify water quality problems, pollutants, sources, and pathways. Exploratory data analysis techniques are well suited to this project phase, and should also be applied as a first step to all data subsequently collected by the project. Exploratory data analysis is also a critical first step in beginning to analyze water quality data from watershed projects that are underway, before undertaking more complex analysis.

Exploratory data analysis provides basic information about the data record, including the data distribution and an assessment of missing and extreme values. The presence of autocorrelation and seasonal cycles should also be evaluated. EDA can also be useful to examine clusters in the data or relationships between variables and/or sample locations.

Table 7-1 summarizes exploratory data analysis methods by analytical objective. The type of method (parametric, nonparametric, graphical), basic data requirements (e.g., distribution, independence), and major cautions and concerns are also included in the table.

Table 7-1. Exploratory data analysis methods (see discussion, section 7.3)

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Describe behavior of variable(s)	Univariate statistics (e.g., range, mean, median, interquartile range, variance)	P, N	Minimal	Mean is sensitive to extreme values; median may be preferred measure of central tendency.
Evaluate distribution and assumptions of independence and constant variance	Plots (histogram, probability, lag-n autocorrelation, cumulative distribution functions); skewness, kurtosis; Durbin-Watson statistic to detect presence of autoregressive lag 1 pattern; Shapiro-Wilk test; Kolmogorov-Smirnov test	P, N, G	Minimal to moderate	Data transformations to satisfy likely statistical testing assumptions should be examined. Autocorrelation functions (ACF) which examine auto correlation at each lag require equal time-space data and appropriate software.
Identify extreme values and anomalies	Plots (e.g., time series, boxplots) Compute frequency or proportion of observations exceeding threshold value; cumulative frequency or duration plots	G, P, N	Minimal	Outliers should not be deleted if error cannot be confirmed.

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Observe seasonal or other cycles	Plots (time series, seasonal boxplots)	G	Minimal	More intensive techniques are generally required to confirm and quantify trends.
	Examination of autocorrelation pattern			Use software that can generate autocorrelation function (ACF) graphs (see section 7.3.6).
Find clusters or groupings	Cluster analysis, principal components analysis, canonical correspondence analysis, discriminant function analysis	P, N, G		Factors determining groupings may be difficult to discern or interpret.
Preliminary comparison of two or more locations or time periods	Boxplots	G	Minimal	Visual comparisons should be confirmed by numerical tests.
Examine relationships between variables	Correlation, regression	P	Data must be normally distributed to apply parametric analysis	Graphical analysis should be used to confirm and understand numerical correlation coefficient. Correlation does not guarantee causation.
	Spearman's rho or rank correlation coefficient	N	Can be used when both independent and dependent variables are ordinal or when one variable is ordinal and the other is continuous	
	Bivariate scatterplots LOWESS smoother	G	Minimal	Visual comparisons should be confirmed by numerical tests.

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

Table 7-2 summarizes methods that can be applied to adjust (e.g., transform) data based on the requirements of methods (e.g., normal distribution required for parametric analyses) to be used in the next phase of data analysis. This table also identifies methods that can be used to address problems caused by unexpected events, including washed out monitoring equipment, floods, droughts, ice, failed BMP implementation plans, and equipment and laboratory errors.

Table 7-2. Methods for adjusting data for subsequent analysis (see discussion, section 7.3)

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Obtain a normal distribution (for parametric approaches)	$\log_{10}$ and $\log_e$ (ln) are most commonly used transformations in water resources	P	Original data values must be positive and non-zero.	Other transformations (e.g., Box-Cox) may be required to achieve normal distribution. Very small numbers and legitimate zero values may require a different transformation (e.g., $\log_{10}$(value + n). Transformations will not correct issues of independence. Back-transformations may be difficult to interpret.
	arc-sine square root transformation	P	Used for proportions	
	If distribution assumptions cannot be met, adopt methods resistant to errors in results caused by deviations from the assumption of normality	N	Minimal	Nonparametric procedures may still have other assumptions that must be met for usage. If distributional assumptions can be met, then parametric tools tend to be more powerful.
Accommodate extreme values	Use methods resistant to errors in results caused by extreme values such as: nonparametric trend tests or frequency analyses	N, G	Moderate	If the data are missing due to right censoring (too high to measure), techniques discussed in section 7.4 should be considered.
	Data stratification (e.g., by seasons, base flow, storm, and floods)		Moderate	
	Use covariate/ explanatory variable such as flow to help 'explain' the influence of extreme values			
	Utilize log transformed data to minimize skewness caused by the extreme values	P	Minimal	
Manage missing data	Data aggregation to create uniform time intervals by averaging or using the median value	P	Minimal	Missing values are ignored in most nonparametric and parametric tests; however, some tests require equal spacing of observations. Data aggregation to accommodate missing data or changes in data frequency must be done with care.
	Estimate missing values based upon regression relationship from other sites or events		Regression relationship with data from similar basin (e.g., flow). Sometimes it may also be appropriate to use the flow/concentration relationship at the same station to estimate missing concentration data	Only use when the data meet the assumptions for regression analysis and the sample size is large enough that the regression relationship is reliable.

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Adjust for autocorrelation	Aggregation of data to less frequent observations	N	Minimal	Aggregation must be consistent (e.g., monthly mean of n daily observations), not mix of different sample frequencies.
	Use of parametric time series analysis techniques available in many statistical software tests	P	Generally equally time-space data observations	Software may correct for both autocorrelation and seasonality.
	Adjust the standard error for the trend (difference or slope) to accommodate for the reduced effective degrees of freedom		Need to calculate the autocorrelation coefficient at lag 1 for this adjustment (see section 7.3.6)	
Adjust for seasonality or other cycles	Use non-parametric trend tests that adjust for seasonality		Generally the month of year is needed for the input data set	
	Add explanatory variables that 'explain' the season affect		e.g., add data columns representing seasonal components for seasonal cycle (e.g., sin/cos terms) or monthly indicator variables	
	Use time series models that incorporate a lag term(s) to incorporate for seasonal cycles into statistical models			

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

7.2.2 Dealing with Censored Data

Censored values are usually associated with limitations of measurement or sample analysis, and are commonly reported as results below or above measurement capacity of the available analytical equipment. Table 7-3 summarizes techniques to use when dealing with censored data.

Table 7-3. Methods to deal with censored data (see discussion, section 7.4)

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Accommodate censored data (i.e., values less than detection or reporting limits)	Use parametric (e.g., maximum likelihood estimation (MLE) and robust regression on order statistics (ROS)) or nonparametric procedures designed to accommodate censored data.	P, N, G	Knowledge about analytical detection limits, practical quantitation limits, and data reporting conventions is required to interpret the meaning of censored data.	Although common, substitution of half the detection limit is not recommended as more robust tools are readily available.

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

7.2.3 Data Analysis for Water Quality Problem Assessment

Problem assessment is generally considered the first phase of a watershed project. Data analysis at this stage typically involves using historical data to assess whether water quality standards are being met or whether designated beneficial uses of waters are threatened, and the causes (e.g., pollutants) and sources of identified problems. More refined problem assessment will include determination of pollutant pathways and critical areas needing restoration or BMPs. Methods to support these types of analyses are summarized in Table 7-4.

Table 7-4. Data analysis methods for problem assessment (see discussion, section 7.5)

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Summarize existing conditions	Univariate statistics (e.g., mean, median, range, variance, interquartile range) for different sampling locations, time series analysis for long-term trends and seasonality, and regression analysis comparing pollutant concentrations or loads to hydraulic variables	P,N	Minimal to moderate	To compare locations within or across watersheds, data from different locations must be consistent and comparable (e.g., synoptic survey, multiple sampling stations).
	Boxplots and/or time series plots for different sampling locations	G		
Assess compliance with water quality standards	Identification of extreme values with boxplots or time series plots; calculation of means (arithmetic or geometric) over specific time period(s)	P	Minimal to moderate	Criteria for determining impairment vary (e.g., single observation exceedance vs. geometric mean over n observations); both monitoring program and data analysis must be tailored to regulatory requirements.
	Frequency or probability plots, duration curves	G		
Identify major pollutant sources	Correlation or regression analysis or Kendall's Tau for monotonic association of water quality constituent(s) vs. subwatershed characteristic(s) (e.g., total P concentration vs. manured acres)	P, N, G	Concurrent data from monitored subwatersheds: subwatershed land use and/or management data	Correlation does not guarantee causation; consider transport and other pollutant delivery mechanisms.
	Compare boxplots or bivariate scatterplots from monitored subwatersheds with distinctive land use and/or management; ANCOVA analysis	G, P		
Define critical areas	t-Test, ANOVA, Kruskall-Wallis, cluster analysis to identify significant differences in pollutant concentration/load among multiple sampling points	P, N	Concurrent data from monitored subwatersheds:parametric or nonparametric analysis can be used depending on data distribution	Conditions determining pollutant generation (e.g., storm event, season, management schedule) must be considered in drawing conclusions about critical areas. Modeling may be useful.

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

7.2.4 Project Planning Data Analysis

Project planning involves both land treatment and monitoring design. Decisions regarding project duration, BMP and restoration needs and scheduling, and implementation tracking and monitoring should all be supported by information and appropriate analysis. The quality of information available will vary from project to project. In many cases, the analysis and decisions will have to rely on historical data (perhaps collected for other purposes) or on data from other sites in the region. The methods summarized in Table 7-5 are recommended to assist with various aspects of project planning.

Table 7-5. Data analysis methods for project planning (see discussion, section 7.6)

Analytical Objective	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
Determine pollutant reductions needed to meet water quality objectives	Mass balance/TMDL Receiving waterbody relationships Load-duration curves Reference watershed	P, G		
Estimate BMP treatment needs	Compare estimated pollutant reduction efficiencies of planned BMPs with reductions needed	P	Appropriate local or published values on BMP pollutant reduction efficiencies	Published efficiencies do not generally account for interactions in multiple-BMP systems or pollutant transport or delivery issues beyond edge of field/BMP site. Modeling may be a better approach.
Estimate minimum detectable change (MDC)	MDC calculation (Spooner et al. 2011a)	P	Mean and variance of water quality variable(s) of interest; parameters of planned monitoring program (e.g., sampling frequency)	If MDC is larger than anticipated response to treatment, may need to re-evaluate extent of planned land treatment and/or duration of water quality monitoring. If data are unavailable from subject watershed, data from elsewhere must be used.
Locate monitoring stations	Identify major pollutant sources, critical areas as in Table 7.5 if data are available	P	Concurrent data from subwatersheds (e.g., from a synoptic survey)	Conditions determining pollutant generation (e.g., storm event, season, management schedule) must be considered.
	Target land areas of particular land use/management and/or expected treatment implementation	G	Land use and management data, estimates of treatment adoption	Station location depends on many other factors, including project objectives, monitoring design, and site requirements.

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

7.2.5 BMP and Project Effectiveness Data Analysis

Table 7-6 includes recommended methods for assessing the effectiveness of BMPs and watershed projects. In general, the analytical objective of both kinds of efforts is to document change in pollutant concentrations or loads or both in response to BMP implementation. These methods are linked to monitoring designs that are described in section 2.4. Methods for assessing BMP and project effectiveness using biological data are presented in chapter 4.

Table 7-6. Data analysis methods for assessing BMP or watershed project effectiveness (see discussion, sections 7.7 and 7.8)

Analytical Objective	Monitoring Design Used	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
BMP efficiency	Plot	ANOVA	P	Data must meet assumptions for parametric statistics to apply; otherwise use nonparametric test	Plot data may not easily extrapolate to field or watershed scale.
		Kruskal-Wallis	N		
	Input/output	Paired t-Test, Wilcoxon, or Mann-Whitney tests of input vs. output EMCs (Event Mean Concentrations) or loads	P, N	Data must meet assumptions for parametric statistics to apply; otherwise use nonparametric test	Representing change in load or concentrations as a percent reduction may not be representative for low input concentrations or loads.
		Effluent probability	N		
Watershed project effectiveness	Paired watershed	ANCOVA, paired t-Test, Wilcoxon Rank Sum, Mann-Whitney	P, N	Data from control and treatment watersheds must exhibit significant linear relationship. Conditions (e.g., precipitation, discharge) must be in similar range during calibration and treatment periods.	Quality of relationship between control and treatment watersheds determines level of change that can be detected. Addition of covariates to paired regression model may improve ability to document response to treatment.
	Above/below-Before/after	t-Test of input vs. output EMCs or loads, ANCOVA, Wilcoxon Rank Sum, Mann-Whitney	P, N	Data must meet assumptions for parametric statistics to apply; otherwise use nonparametric test	Change in pollutant concentration or load measured at the below station may be difficult to detect if concentrations or loads at the above station are high.
	Single Watershed Monotonic Trend	Linear regression on time Multiple linear regression on time and covariates Linear regression on time, covariates, and periodic functions	P	Numerous techniques are available, depending on objectives, available data on covariates, seasonality	Trend analysis is most effective with data sampled consistently at fixed locations and fixed time intervals for period sufficient to overlap seasonal or management cycles that do not represent real trends. Covariates such as stream flow, season, etc. are essential to assist with isolating trends due to BMPs.
		Mann-Kendall Mann-Kendall on residuals from regression on covariates Seasonal Kendall	N	Numerous techniques are available, depending on objectives, available data on covariates, seasonality	

Analytical Objective	Monitoring Design Used	Recommended Method	Method Type*	Data Requirements	Major Cautions and Concerns
	Single Watershed Step Trend	t-Test before and after step, Wilcoxon Rank Sum, Mann-Whitney	P, N	Data must meet assumptions for parametric statistics to apply; otherwise use nonparametric test	Selection of step change point in time must be made *a priori* and related to watershed activities, e.g., onset of treatment. Covariates such as stream flow, season, etc. are essential to assist with isolating trends due to BMPs.
	Multiple watersheds	t-Test or Wilcoxon Rank-Sum test ANOVA or Kruskal-Wallis test Regression analysis	P, N	Data must meet assumptions for parametric statistics to apply; otherwise use nonparametric test	Watersheds need to fall into 2 groups (e.g., treated and untreated) for t-Test or Wilcoxon Rank-Sum test. For more than two groups use ANOVA or Kruskal-Wallis.
		Boxplots of results from watershed groupings (e.g., treated/untreated)	G	Minimal	Visual comparisons should be confirmed by numerical tests.
	Linking land treatment to water quality changes	Correlation, regression of pollutant concentration or load on land treatment metric(s)	P, N	Requires quantitative monitoring data on land treatment. Use of explanatory variables (e.g., precipitation, animal populations) may strengthen analysis.	Water quality and land treatment data must be collected on comparable spatial and temporal scales. Monitored pollutants must match pollutants addressed by implemented BMPs.

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

7.2.6 Practice Datasets

This chapter presents a wide range of parametric and nonparametric methods, including several illustrative examples. Because practice is the best way to learn how to apply these methods, example datasets and eight problems are provided to allow readers to test their skills. Using their own statistics software, readers are encouraged to apply the tests indicated in Table 7-7 to the example datasets listed in the fourth column. The objective and statistical tests are listed in the second and third columns of the table. The specific problems and the answers are given in the files identified in the last column.

Table 7-7. Practice datasets

Problem Number	Objective	Test	Dataset in Sampledata.xlsx	Problem and Answer File
1	Test for conformance to normal distribution	Graphical, skewness, kurtosis, Shapiro-Wilk, Kolmogorov	1	normality.pdf
2	Characterize data	Descriptive statistics	1	description.pdf
3	Compare two groups	t-Test	1	2groups.pdf
		Wilcoxon/Kruskal-Wallace		
4	Compare input/output for a BMP	Paired t-Test	2	pairedtests.pdf
		Wilcoxon Rank Sum Test		
5	Compare three groups	ANOVA	1	3groups.pdf
		Kruskal-Wallace		
6	Examine relationships between variables/stations	Correlation	1	correlationregress.pdf
		Simple linear regression		
7	Assess change due to treatment in paired-watershed design	ANCOVA	1	pairedancova.pdf
8	Calculate MDC for a single station	Minimum detectable change	3	mdc.pdf

All files are available at: https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed

7.3 Exploratory Data Analysis (EDA) and Data Adjustment

After a monitoring program is up and running, it is never too soon to begin to evaluate the data. Basic data evaluation should not wait until the end of the project or when a report is due; regular examination of the data should be part of ongoing project activities. A carefully designed monitoring program will have the right kind of data, collected at appropriate times and locations to achieve the objectives, and a plan for analyzing the data.

Describing and summarizing the data in a way that conveys their important characteristics is one purpose of EDA. When deciding how to analyze any data set, it is essential to consider the characteristics of the data themselves. Evaluation of characteristics like non-normal distribution and autocorrelation will help determine the appropriate statistical analysis. Some common characteristics of water quantity and quality data (Helsel and Hirsch 2002) include:

- A lower bound of zero – no negative values are possible.
- Presence of outliers, extreme low or high values that occur infrequently, but usually somewhere in the data set (outliers on the high side are common).
- Skewed distribution, due to outliers or influential data.
- Non-normal distribution.
- Censored data – concentration data reported below some detection limit or above a certain value.
- Strong seasonal patterns.
- Autocorrelation – consecutive observations strongly correlated with each other.

- Dependence on other uncontrolled or unmeasured variables – values strongly co-vary with such variables as streamflow, precipitation, or sediment grain size.

As such, the overall goal of data exploration is to uncover the underlying structure of a data set and set the stage for more detailed analysis, including hypothesis testing. Specific objectives for data exploration might include:

- To find potential problems with data quality such as data entry error, lab or collection errors
- To find extreme values and potential anomalies
- To describe the behavior of one or more variables
- To test distribution and assumptions of independence and constant variance
- To see cycles and trends
- To find clusters or groupings
- To make preliminary comparisons of two or more locations or time periods
- To examine relationships between variables

At the start, check the data for conformity with original plans and QA/QC procedures. Use the approved project Quality Assurance Project Plan (QAPP) as a guide; see section 8.3 for details on preparing a QAPP. A key part of EDA is to verify the data entered in the data sets are valid and not anomalies due to data entry, lab, or collection errors.

Understanding how the data behave with respect to such features as distribution(s), cycles, clusters, seasonality, and autocorrelation assists with selecting the appropriate statistical tests to evaluate achievement of project goals. Data analysis to address project goals will involve more thorough statistical analysis that will be guided by understanding of the data set through EDA.

A secondary reason for doing exploratory data analysis is to start to make sense of the data actually collected. The purpose of EDA is to get a feel for the data, develop ideas about what it can tell, and how to draw some preliminary conclusions. EDA is similar to detective work – sifting through all the facts, looking for clues, and putting the pieces together to find suggestions of meaning in the data.

This process of data exploration differs from traditional hypothesis testing. Testing of hypotheses always requires some initial assumption or prediction about the data, such as "The BMP will reduce phosphorus loads." Although formulating and testing hypotheses is the foundation of good data analysis, the first pass through of the data should not be too narrowly focused on testing a single idea. Hypothesis testing is discussed in section 7.6.1, which focuses on data analysis for project planning. EDA is an approach to data analysis that postpones the usual assumptions about what kind of model the data follow in favor of the more direct approach of allowing the data themselves to reveal their underlying structure. EDA uses a variety of techniques, both numerical and graphical, to open-mindedly search for new, perhaps unexpected, insights into the data. Approaches to EDA for aquatic system biological data have been described by EPA as part of the Causal Analysis/Diagnosis - Decision Information System (CADDIS) (USEPA 2010).

Data exploration is a necessary first step in analyzing monitoring data. Unless initial exploration reveals indications of patterns and relationships, there is unlikely to be something for further analysis to confirm.

J. W. Tukey (1977), the founder of exploratory data analysis, said, "EDA can never be the whole story, but nothing else can serve as the …first step."

For more information, refer to *Tech Notes 1: Monitoring Data Exploring Your Data, The First Step* (Meals and Dressing 2005).

7.3.1 Steps in Data Exploration

Data exploration is a process of probing more deeply into the dataset, while being careful to stay organized and avoid errors. Here are some typical steps in the process of EDA (modified from Jambu 1991), although not all of them may apply to every situation.

1. **Data management.** In the process of working with the data, files will be created. These files should be updated, checked, and validated at regular intervals. The importance of data screening and validation cannot be overemphasized. This should always be done before embarking on specific analyses, plotting, or other procedures. Be as sure as possible that the data are free from entry errors, typos, and other mistakes before proceeding.
2. **One-dimensional analysis.** The first step in really exploring the data is often to simply describe or summarize the information one variable at a time, independent of other variables. This can be done using basic statistics on range, central tendency, and variability, or with simple graphs like histograms, pie charts, or time plots. This kind of information is always useful to put data in context, even though more intensive statistical analysis will be pursued later.
3. **Two-dimensional analysis.** Relationships between two variables are often of great interest, especially if there is a meaningful connection suspected (such as between suspended sediment and phosphorus) or cause and effect process (such as between rainfall and streamflow). Relationships between two sampling locations (such as treatment and control watersheds) or between two time periods (like spring snowmelt and summer) are often of interest. Graphical techniques like scatter plots and numerical techniques like correlation are often used for this purpose.

Because graphs summarize data in ways that describe essential information more quickly and completely than do tables of numbers, graphics are important diagnostic tools for exploring the data. There is no single statistical tool that is as powerful as a well-chosen graph (Chambers et al. 1983). Enormous amounts of quantitative information can be conveyed by graphs and the human eye-brain system is capable of quickly summarizing information, simultaneously appreciating overall patterns and minute details. Graphs will also be essential in ultimately conveying project results to others. With computers and software available today, there are no real constraints to graphing data as part of EDA. Graphical display options are described in section 4.3 of the 1997 guidance (USEPA 1997b).

There are more advanced steps in data exploration including analysis of multiple variables and cluster analysis (section 7.3.8). Also, see chapter 4 of the 1997 guidance (USEPA 1997b) for background on some of these methods.

The project goals and the type of monitoring should guide exploration. If monitoring occurs at a single point while upstream BMPs are implemented gradually, trends may be of the greatest interest. If sampling for phosphorus above and below a land treatment area, a comparison of phosphorus concentrations at the two stations might be necessary. For an erosion problem, a relationship between streamflow and suspended solids concentrations before and after land treatment might be of interest.

The following sections present some specific techniques for exploring data.

7.3.2 Describe Key Variable Characteristics

In most cases, the data should be examined to summarize key characteristics and to determine if the data satisfy statistical assumptions required for parametric statistical analyses. Data that do not meet parametric statistical assumptions should be transformed or nonparametric tests should be used. Key characteristics that are meaningful include central tendency, variability, and distribution.

7.3.2.1 Central Tendency

- The **mean** is computed as the sum of all values divided by the number of values. The mean is probably the most common data summary technique in use; however, an extreme value (either high or low) has much greater influence on the mean than does a more ‘typical’ value. Because of this sensitivity to extremes, the mean may not be the best summary of the central tendency of the data.
- The **median,** or 50th percentile, is the central value of the distribution when the data are ranked in numerical order. The median is the data value for which half of the observations are higher and half are lower. Because it is determined by the order of observations, the median is only slightly affected by the magnitude of a single extreme value. When a summary value is desired that is not strongly influenced by a few extremes, the median is preferable to the mean.

Both the mean and median should be calculated for comparison.

7.3.2.2 Variability

- The sample **variance**, and its square root the **standard deviation**, are the most common measures of the spread (dispersion) of a set of data. These statistics are computed using the squares of the difference between each data point and the mean, so that outliers influence their magnitudes dramatically. In data sets with major outliers, the variance and standard deviation may suggest a much greater spread than exists for the majority of the data. This is a good reason to supplement numerical statistics with graphical analysis.
- The **coefficient of variation (CV)**, defined as the standard deviation divided by the mean, is a relative measure of the variability (spread) of the data. The CV is sometimes expressed as a percent, with larger values indicating higher variability around the mean. Comparing the CV of two data groups can suggest their relative variability.
- The **interquartile range** (IQR) is defined as the 75th percentile minus the 25th percentile. Because it measures the range of the central 50 percent of the data, it is not influenced at all by the 25 percent of the data on either end and is relatively insensitive to outliers.

7.3.2.3 Skewness

Water resources data are usually skewed, meaning that the data values are not symmetric around the mean or median, as extreme values extend out farther in one direction. Streamflow data, for example, are typically right-skewed because of occasional high-flow events (Figure 7-1). When data are skewed, the mean is not equal to the median, but is pulled toward the long tail of the distribution by the effects of the extreme values. The standard deviation is also inflated by the extreme values. Because highly skewed data restrict the ability to use hypothesis tests that assume the data have a normal distribution, it is useful to evaluate the skewness of the data. The **coefficient of skewness** (g) is a common measure of skewness; a right-skewed distribution has a positive g and a left-skewed distribution has a negative g. There are multiple measures of skewness with varying possible ranges. Interpretation of skewness values calculated by Excel, for example, is aided by estimating the standard error of skewness with the following simplified[1] equation for large (<5 percent difference from true value for n≥30) samples (Elliott 2012):

$$Standard\ Error = \sqrt{6/n}$$

where n is the sample size. For n=24, the standard error of skewness is 0.5 using the simplified equation. A skewness value of more than twice this amount (i.e., less than -1 or greater than 1 in this case) indicates a skewed distribution, but a value between -1 and 1 is not proof that the data are normally distributed. Other tests such as goodness-of-fit tests (below) must also be performed to determine if the distribution is normal.

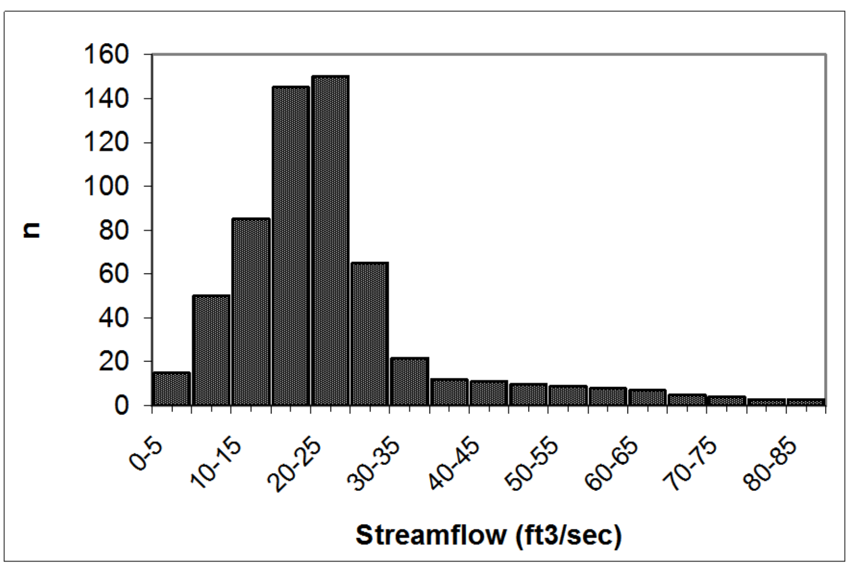


Figure 7-1. Right-skewed distribution

[1] The true standard error of skewness is calculated as: $\sqrt{6n(n-1)/(n-2)(n+1)(n+3)}$

7.3.2.4 Data Distribution

Many common statistical techniques for hypothesis-testing (parametric tests) require, among other characteristics, that the data be normally distributed. It is common practice to apply tests such as the Shapiro-Wilk test or the Kolmogorov-Smirnov (KS) test to evaluate the normality of the data; both of these tests are commonly available in statistical software. The probability plot correlation coefficient (PPCC) can also be used to test for normality. PPCC is essentially a correlation coefficient between the data values and their normal score (i.e., data on probability paper) and the interpretation of the PPCC is similar to that for the correlation coefficient r. This procedure is outlined by Helsel and Hirsch (2002) in section 4.4 and in Appendix Table B.3 which gives critical values for accepting/rejecting the normal assumption.

Histograms are familiar graphs, where bars are drawn whose height represents the number or fraction of observations falling into one of several categories or intervals (see Figure 7-1). Histograms are useful for depicting the shape or symmetry of a data set, especially whether the data appear to be skewed. However, histogram appearance depends strongly on the number of categories selected for the plot. For this reason, histograms are most useful to show data that have natural categories or groupings, such as fish numbers by species, but are more problematic for data measured on a continuous scale such as streamflow or phosphorus concentration.

Quantile plots (also called cumulative frequency plots) show the percentiles of the data distribution. Many statistics packages calculate and plot frequency distributions; instructions for manually constructing a quantile plot can be found in Helsel and Hirsch (2002) and other statistics textbooks. Quantile plots show many important data characteristics, such as the median or the percent of observations less than or greater than some critical threshold or frequency. With experience, an analyst can discern information about the spread and skewness of the data. Figure 7-2 shows a quantile plot of *E. coli* bacteria in a stream; the frequency of violation of the Vermont water quality standard can be easily seen (the standard was exceeded ~65 percent of the time). Flow and load duration curves (see section 7.9.3) are useful tools for visualizing the distribution of streamflows or pollutant loads across a full range of conditions.

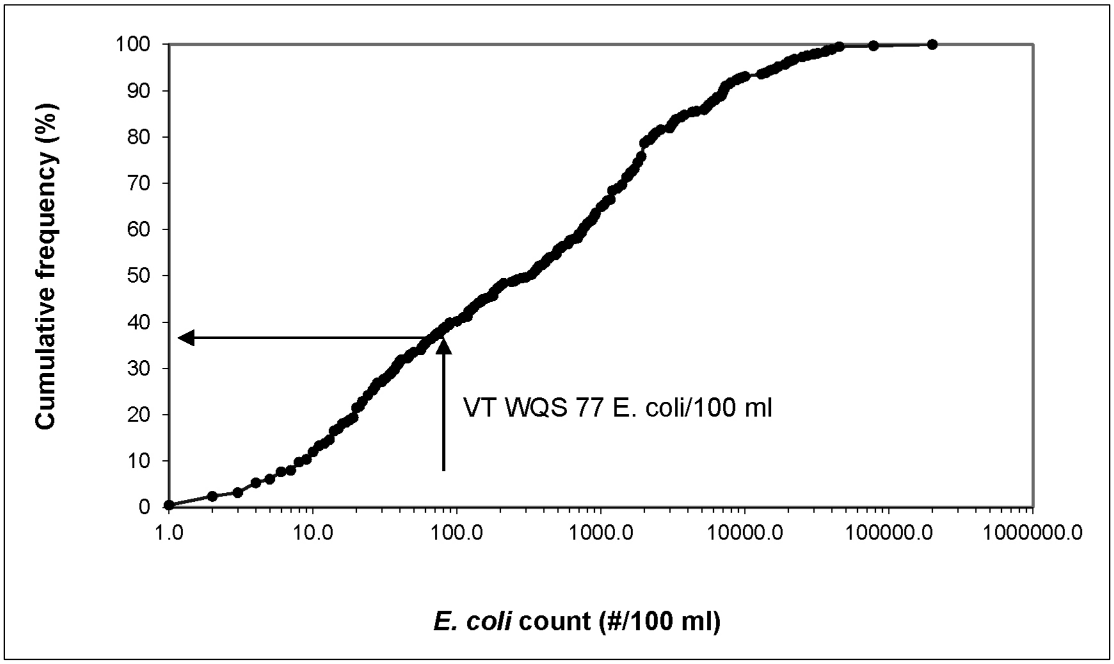


Figure 7-2. Quantile plot or cumulative frequency plot of *E. coli* data, Berry Brook, 1996 (Meals 2001)

A boxplot presents a schematic of essential data characteristics in a simple and direct way: central tendency (median), spread (interquartile range), skewness (relative size of the box halves), and the presence of outliers are all indicated in a simple picture. There are many variations and styles of boxplots, but the standard boxplot (Figure 7-3) consists of a rectangle spanning the 25th and 75th percentiles, split by a line representing the median. Whiskers extend vertically to encompass the range of most of the data (e.g., the 5th and 95th percentiles), and outliers beyond this range are shown by dots or other symbols. The definition of whiskers and outliers may differ among graphing programs; standard definitions can be found in statistics textbooks (e.g., Cleveland 1993; Helsel and Hirsch 2002). When boxplots are presented, the definitions of the rectangle, whiskers, and outlier symbols should be clearly specified.

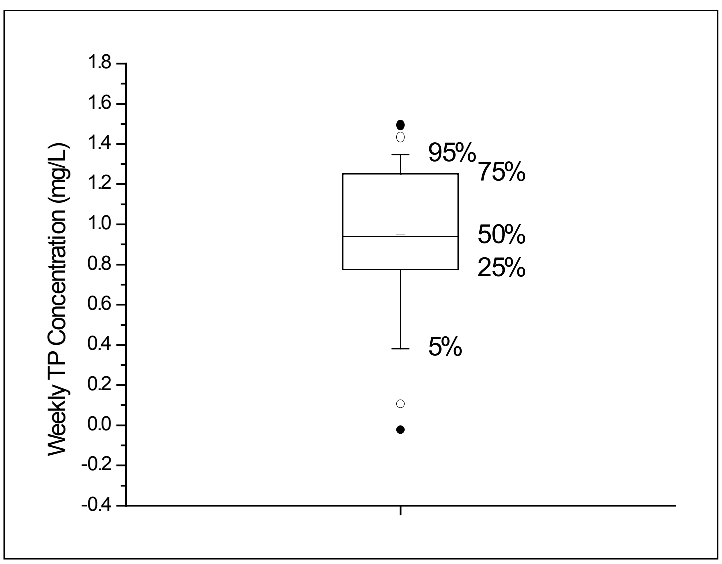


Figure 7-3. Boxplot of weekly TP concentration, Samsonville Brook, 1995 (Meals 2001)

7.3.2.5 Transformations to Handle Non-normal Data with Parametric Statistical Tests

Evaluations conducted thus far may suggest that the data do not conform to a normal distribution. In cases where it is desirable or convenient to use statistical tools that require normally distributed data sets or have a constant variance, transformation may reduce skewness and result in a data set that is more normally distributed. Transformation is simply defined as applying the same mathematical operation to all records in the dataset. Helsel and Hirsch (2002) provide a summary of common transformations. Statistical software packages will often come with Box-Cox transformation tools that allow the analyst to identify the best transformation for achieving normality, although logarithmic (e.g., $\log_{10}$ or $\log_e$) transformation is certainly the most common strategy (Box and Cox 1964). Regardless of which transformation is used, the data analyst should verify that the transformation results in a dataset that satisfies applicable assumptions.

Subsequent analysis of log-transformed data must be done with care, as quantities such as mean and variance calculated on the transformed scale are often biased when transformed back to the original scale. The geometric mean (the mean of the log-transformed data back-transformed to the arithmetic scale), for example, differs from the mean of the untransformed distribution. Furthermore, results of statistical analysis may be more difficult to understand or interpret when expressed on the transformed scale. Typically, when analysis is performed on the log transformed data, the final statistical results are converted to express the results as a percentage change (see Spooner et al. 2011a for additional details on this approach).

Do not assume that a transformation will solve all the problems with the data distribution. Always test the characteristics of the transformed data set again. Violations of the assumption of a normal distribution can lead to incorrect conclusions about the data when parametric tests are used in subsequent hypothesis testing. With that said, some parametric trend tests are robust to some deviation from normality. From a practical standpoint it is best to be consistent. For example, if a log transformation is merited for TP concentrations at most locations in a particular data set, then log transforming all TP for all site locations is a practical course of action.

If transformed data cannot satisfy the assumptions of parametric statistical analysis, consider nonparametric techniques for data analysis. With regard to hypothesis testing, there are a host of nonparametric tests that are robust against non-normality. These tests are often based on the ranks of the data and the influence of a few extreme values is reduced. However, keep in mind that while the normality assumption is relaxed, nonparametric tests have other assumptions (constant variance and independence of data observations) that must be met for their results to be valid. If distributional assumptions can be met, then parametric tools tend to be more powerful. Many nonparametric procedures are described in section 4.11.3 in the 1997 guidance and recommended in Table 7-1 through Table 7-6.

7.3.3 Examination for Extreme, Outlier, Missing, or Anomalous Values

7.3.3.1 Extremes and Outliers

Extreme values are frequently encountered in NPS monitoring efforts and include the exceptionally high and low flow values associated with floods and droughts, respectively. Suspended sediment concentrations may be exceptionally high during spring runoff when cropland fields are bare or when streambank slumping occurs. Very low pesticide levels may be observed with increasing time elapsed since application on cropland. In some cases, the extremes may be more important for water quality than are typical conditions. For example, the extreme values in some lake variables (e.g., Secchi disc readings, turbidity, and pH), the duration of the extreme values, and the season may be the dominant influence on the extent to which lakes support designated beneficial uses. In streams, it is often the extreme low dissolved oxygen condition that determines the character of the biological community the stream can support. Extreme concentrations in toxic contaminants such as pesticides may also be more important than the mean values with respect to acute toxicity to aquatic biota. Nevertheless, extreme concentrations can have an inordinate effect on some statistical analyses, and the analyst must consider these issues when selecting data analysis tools.

On the other hand, outliers can result from measurement or recording errors and this should be the first thing checked (e.g., check lab and field logs). *If no error can be found, an outlier should never be rejected just because it appears unusual or extreme. All samples considered valid after exploratory analysis contain information that should be considered when analyzing monitoring data.* Different subsets of the same dataset may reveal varying aspects of the condition of the water resource. For example, extreme conditions may be most important when considering violations of water quality standards or load allocations from a watershed. Annual or monthly loads may not completely illuminate the severity of a problem, whereas high loads during extreme flow conditions may account for most of the pollutant load. It is commonly observed that the majority of annual pollutant export occurs during a small proportion of the time. Identifying these extremes and understanding the conditions under which they occur may be a key to understanding and interpreting watershed monitoring results.

One approach for identifying and summarizing extreme values is to describe the situation by computing the frequency or proportion of observations exceeding some threshold value (e.g., a water quality criterion). Cumulative frequency or duration plots are also useful to visualize the influence of extreme values on a dataset. In addition, determine whether most or all of the extreme values can be attributed to certain conditions in the watershed (e.g., spring runoff, cropland tillage). In these cases, it might be more useful to stratify the dataset by season or management condition. In this way, monitoring results can be analyzed by season, and values that were "extreme" in the dataset as a whole may be more easily interpreted in their respective season(s).

Histograms can be useful to illustrate exceedances of standards, targets, and goals by setting categories or classes that are outside the standard or target. Quartile plots and boxplots are also useful tools to evaluate the presence of extreme values.

Boxplots can be a useful visual tool for highlighting extreme values in environmental data. They show both the spread and the range of the data. Important values visualized by boxplots include the mean (or the median), and standard error limits (or 25th and 75th percentiles). Values falling outside these 'limits' depict values that are from the tails of the data distribution.

Plotting the data in sequence with date as the horizontal axis are time series plots. Figure 7-4 shows a time series plot of weekly phosphorus concentration data from three stream stations. It is clear that around the middle of the year, something occurred that led to dramatic spikes in P concentration at Station 2, a phenomenon demanding further investigation. Field investigation revealed concentrated overland flow from a new CAFO upstream.

To analyze data sets with extreme values, consider using non-parametric trend tests. If documenting the number or occurrence of extreme values is an objective (e.g., for evaluation of violations of water quality standards or pesticide spikes), frequency analyses are useful. Stratifying the data by seasons or flow conditions (e.g., base flow, storm flows, and flooding) may be helpful in evaluating conditions and trends within each flow regime. Using flow as an explanatory variable/covariate in trend analysis may be helpful to explain the influence/importance of the extreme values. Using the log transformation often minimizes the skewness caused by the extreme values and enables the use of parametric trend techniques. If the data are missing due to right censoring (too high to measure), techniques discussed in section 7.4 should be considered.

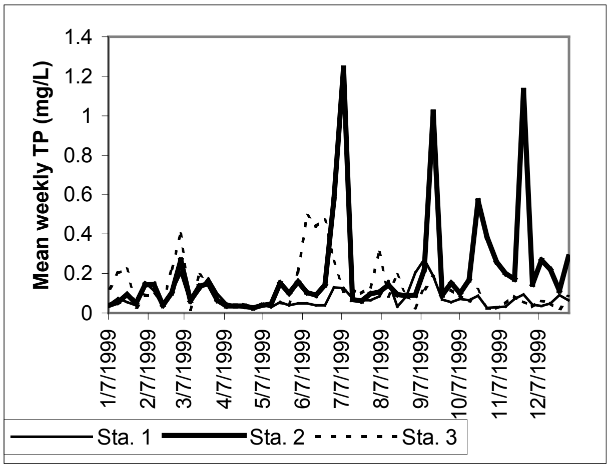


Figure 7-4. Time plot of weekly TP concentration, Godin Brook, 1999 (Meals 2001)

7.3.3.2 Anomalous Values

Plotting the data can also reveal data errors or anomalies. Figure 7-5 shows a time series plot of total Kjeldahl nitrogen (TKN) data collected from three Vermont streams. Something happened around May, 1996 that caused a major shift in TKN concentrations in all three streams. In addition, it is clear that after October, no values less than 0.5 mg/L were recorded. In this case, this shift was not the result of some occurrence in the watersheds, but an artifact of a faulty laboratory instrument, followed by the establishment of a lower detection limit of 0.50 mg/L. Discovery of this fault, while it invalidated a considerable amount of prior data, led to correction of the problem in the lab and saved the project major headaches down the road.

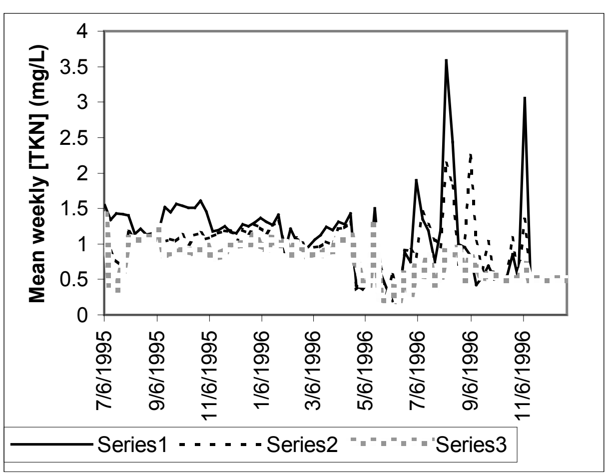


Figure 7-5. Time plot of TKN data from three stream stations, 1995-1996 (Meals 2001)

7.3.3.3 Missing Data

The reality of any watershed project monitoring program is that samples will be missed, equipment will fail or be overwhelmed, droughts and floods will occur, and sample analysis limitations will be exposed, resulting in missing and extreme (both high and low) values. However, if data are missing because of extreme conditions (e.g., streamflow was too high to obtain a measurement or water was so low a sample could not be drawn), then missing data may also represent extreme conditions.

The presence of a few missing values in a data series is not generally a major cause for concern, although some parametric tests (e.g., trend analyses that include autocorrelation errors using time series) require equal spacing of observations[2]. One way to cope with extensive missing data is to aggregate data to a longer, uniform time interval by averaging or using the median value of a group of data points. Daily observations, for example, could be aggregated to weekly means or medians. Such an operation would have an additional potential benefit of reducing autocorrelation (see section 7.3.6). A downside to this approach, however, is a reduced significance level due to fewer degrees of freedom. Do not aggregate data when there is a systematic change in sampling. For example, if the early data were collected as monthly observations and the more recent data were collected as quarterly data, it is not correct to aggregate the monthly data to quarterly averages and then perform analyses. This is because the averaging calculation changes the variability of that portion of the record in comparison to the remainder of the record, resulting in a violation of "identically distributed" assumption of most (including nonparametric) hypothesis tests. In these cases, the analyst will need to subsample from the more intensely monitored data set to best mimic the sampling from the less sampled portion of the data.

For loading analyses that require flow data, it is expected that the missing flow data due to equipment failure could be estimated by evaluating regression relationships with flow from nearby basins. On the other hand, flows that exceed the weir capacity or reach a stage so high that the technician cannot access the site are exceptional events. Certainly one approach to addressing this data gap is to apply the previously mentioned regression relationship with a nearby station. Another approach might be to treat these observations as "greater than the maximum flow" and apply methods appropriate for censored data described in section 7.4.

7.3.4 Examination for Frequencies

For categorical data such as watershed area in different land uses or number of aquatic macroinvertebrates in certain taxonomic groups, data can be effectively summarized as frequencies in histograms or pie charts. Figure 7-6 shows a pie chart of the percent composition of orders of macroinvertebrates in a Vermont stream, clearly indicating that dipterans dominate the community.

---

[2] Some statistical software such PROC AUTOREG in SAS yield valid trend results with autocorrelated data with missing data points, as long as the input record contains equal spaced time intervals (e.g., weekly).

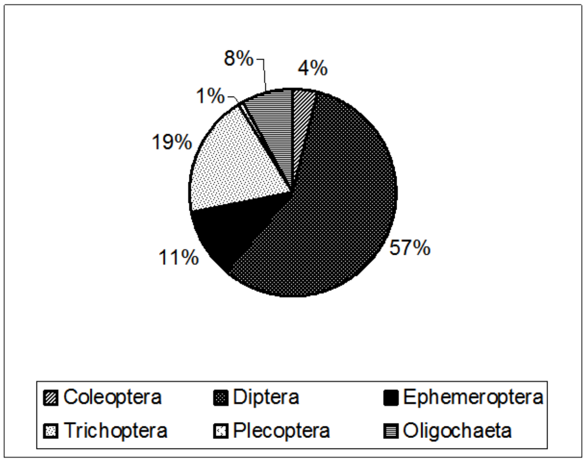


Figure 7-6. Percent composition of the orders of macroinvertebrates, Godin Brook, 2000 (Meals 2001)

7.3.5 Examination for Seasonality or Other Cycles

Monitoring data often consist of a series of observations in time, e.g., weekly samples over a year. One of the first, and the most useful, things to do with any time series data is to plot it. Plotting time series data can provide insight into seasonal patterns, trends, changes, and unexpected events more quickly and easily than tables of numbers.

Figure 7-7 shows a time series plot of *E. coli* counts in a Vermont stream. The extreme range of the counts (five orders of magnitude) and the pronounced seasonal cycle are readily apparent, with the lowest counts occurring during the winter. It is easy to see the times of year when the stream violates the water quality standard for bacteria.

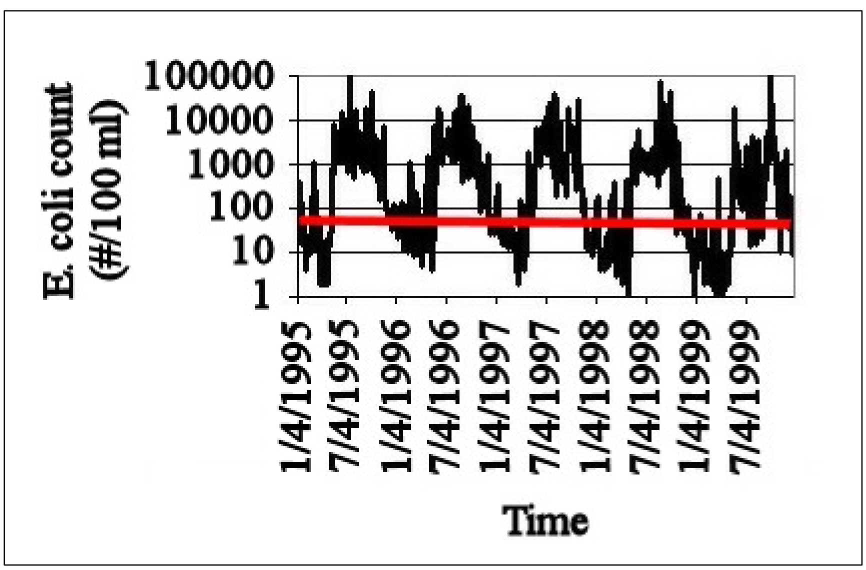


Figure 7-7. Time series plot of weekly *E. coli* counts, Godin Brook, 1995-1999 (Meals 2001). Red line indicates Vermont WQS of 77 *E. coli*/100 ml.

7.3.6 Autocorrelation

Because many hypothesis-testing statistical techniques require that residuals from the statistical tests be independent, it is useful to check the data set for autocorrelation during EDA. Typically, if the data points exhibit autocorrelation, so will the residuals from a statistical test which does not correct for autocorrelation.

Time series data collected through monitoring of water resources often exhibit autocorrelation (also called serial correlation or dependent observations) where the value of an observation is closely related to a previous observation (usually the one immediately before it). Autocorrelation in water quality observations is usually positive in that high values are followed by high values and low values are followed by low values. For example, streamflow data often show autocorrelation, as numerous high wet-weather flows tend to occur in sequence, while low values follow low values during dry periods.

Terms Used in this Section

Lag: the difference in time steps by which one observation comes after another. The lag value is the number of time steps.

Autocorrelation: the correlation between lagged values in a time series (data collected over equal intervals of time, can also be spatial distances)

Correlation Coefficients, ρ_j: a set of correlations for each lag. The autocorrelation coefficient for lag 1 is the correlation between each data in a time series and its previous (lag 1) observation. The autocorrelation coefficient, ρ_j, for lag j is the correlation between each datumin a time series and the observation that lags by j time steps.

Autoregressive: situation where past values (or nearby values for spatial analyses) have an effect on current values. For example, when most of the correlation between the lag variables is between each current value and the immediately preceding value, it is a first-order autoregressive process denoted as AR(1). AR(2) is second order, where previous two values effect the current value, etc. Autoregressive, order 1, AR(1) is common for weekly and monthly water quality samples.

Moving Average: an averaging of a fixed number of consecutive observations, with or without weights. Moving average models are denoted MA(1), MA(2), …MA(q) to indicate the order or maximum lag for consecutive observations that are averaged.

ARIMA (autoregressive integrated moving average) models: time series models that include both autoregressive terms and/or moving average terms

Autocorrelation Function (ACF): the set of correlations (e.g., autocorrelation coefficients) between each value in a series of values (e.g., x_t) and the lagged values within the same series (e.g., x_{t-1}, x_{t-2}, etc.). Alternatively stated, this is the pattern of correlation coefficients vs. lag value. This is generally depicted as a graph of each lag and its autocorrelation coefficient with a standard error bar to help determine the statistical significance of each of the correlation coefficients for each lag. The pattern/shape of the ACF, along with the PACF, is used to assist in determining if the data follow an AR, MA, or ARIMA pattern, and by what order (lag). For example, a seasonal AR(1) series has a large ρ_1, with subsequent ρ_j's trailing off, and a strong seasonal lag correlation.

Partial Autocorrelation Function (PACF): the correlation between two variables, taking into account the relationships of other variables to these two variables. The PACF for an AR(1) series drops to 0 after lag 1).

Autocorrelation usually results in a reduction of the effective sample size (degrees of freedom). It therefore affects statistical trend analyses and their interpretations. As the magnitude of autocorrelation increases, the effective sample size decreases, and the true standard error is therefore greater than if autocorrelation is incorrectly ignored. Adjustment for autocorrelation is needed so that the power of detecting a difference or trend is not incorrectly inflated. For data sets with high autocorrelation, a larger sample size (e.g., longer monitoring duration) than would be necessary in the absence of autocorrelation may be required to correctly detect significant changes or trends.

Autocorrelation is often significant in very frequent data collection, such as that done with recording sensors (e.g., temperature, turbidity). Daily, weekly, and monthly samples also exhibit autocorrelation, but usually to a lesser extent. The time interval between independent samples differs with the water resource and variable. The magnitude of autocorrelation in surface water quality concentrations is usually quite large for samples collected more frequently than monthly (Loftis and Ward 1980a and 1980b, Lettenmaier 1976, Lettenmaier 1978, Whitfield and Woods 1984). Loftis and Ward (1980a and 1980b) verified that some surface water quality samples collected less frequently than once a month may be considered independent if the seasonal variation is removed, although Whitfield (1983) found significant autocorrelation between stream discharge samples taken as much as 60 days apart. Compared to surface water data series, ground water data series tend to retain significant autocorrelation, even with longer sample intervals. Similarly, a ground water data series tends to have greater autocorrelation when compared to surface water data series taken at the same time intervals. This may be due to slower water movement and mixing in ground water as compared to surface waters.

There are numerical techniques to test for autocorrelation, but a simple graphical method can suggest whether data have significant autocorrelation: the lag plot. A lag plot is a graph where each data point is plotted against its predecessor in the time series, i.e., the value for day two and the value for day one are plotted as an x, y pair, then day three, day two, and so on. Different time lags can be examined. A “lag-1“ plot uses each data value paired with its immediate predecessor (t2, t1), a “lag-2” plot uses each data value paired with the value observed two steps previously (t3, t1), and so on. Random (independent) data should not exhibit any identifiable structure or pattern in the lag plot. Non-random structure in the lag plot indicates that the underlying data are not random and that autocorrelation may exist. Figure 7-8 shows a lag-1 plot of weekly streamflow data, suggesting that autocorrelation needs to be addressed.

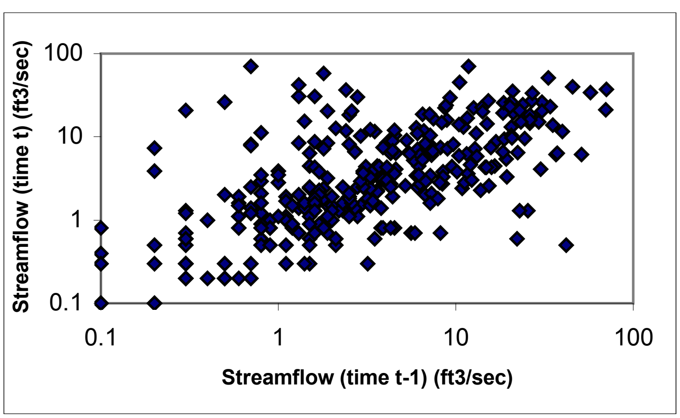


Figure 7-8. Lag-one plot of streamflow observations, Samsonville Brook, 1994 (Meals 2001)

Autocorrelation can be expressed numerically by calculating the correlation between observations separated by j lag time periods. The autocorrelation corresponding to the jth lag is the correlation between the observation at a given time and the observation taken j observation periods earlier. It is denoted by ρ_j or $\rho(j)$. For example, for the first lag (j=1), ρ_1 represents the autocorrelation of data points one time period removed. The time period is a function of the sample frequency and corresponds to the length of time between samples (e.g., daily, weekly, monthly). The range of values for ρ_j is -1 to +1, where +1 represents a perfect positive autocorrelation and -1 is a perfect negative correlation. The sample estimate of autocorrelation is given by r_j (in practice, ρ_j is often used to depict the sample autocorrelation coefficients).

Time series generally exhibit patterns indicated by the pattern of autocorrelation coefficients at various lags. These patterns reveal key characteristics about the data that should be incorporated into subsequent trend analyses. For weekly and less frequent water quality sample collection, the autoregressive, lag 1 or AR(1) data structure is usually appropriate. In this case, most of the autocorrelation can be explained by the correlation between each observation and its previous observation. Moving Average (MA) data structures occur when an observation is only related to the observations up to the lag value (q) and not observations before[3]. Rarely is a MA structure alone useful with water quality samples. However, for some daily or more frequent sampling, a combination of AR and MA data structures become appropriate, known as ARIMA (AutoRegressive Integrated Moving-Average) models.

One common test for autocorrelation is the Durbin-Watson (DW) test. The DW test is appropriately used when the data exhibit first order (lag 1) autoregressive (AR(1)) behavior. AR(1) is common with water quality data collected weekly, biweekly, or monthly. Daily or samples collected more frequently usually exhibit ARIMA autocorrelation structures. Even so, the DW test can be useful to indicate the presence of autocorrelation with such samples as well. The DW test may also be used to test for independence (i.e., the absence of autocorrelation) in the residuals from regression models.

Many statistical software packages offer tools for examining autocorrelation. For example, the Autocorrelation Function (ACF) is the set of all the lag j autocorrelations and is usually depicted as a plot of each lag autocorrelation versus the lag number (Figure 7-9 from Minitab (2016) and Figure 7-10 from JMP (SAS Institute 2016b)) for the same data set. Visual inspection of the ACF is useful to detect the presence of autocorrelation and define the structure of the autocorrelation. Typically, the lag autocorrelation confidence limits (approximately two-standard deviation errors) are also shown on the ACF graphs. This helps analysts determine if the autocorrelation coefficient at lag j is significant. Seasonal patterns show up as cycles in the ACF. As a point of comparison, Figure 7-11 shows a time series plot of independent data (i.e., zero correlation) together with its ACF graph.

Another useful graph is the Partial Autocorrelation Function (PACF) which is included as the last chart in Figure 7-9 and in the last column of Figure 7-10. The PACF is the partial amount of R-square (i.e., correlation) gained due to the additional lag term added to the right hand side of the model (Box and Jenkins 1976). Patterns of the PACF that show dramatic decrease to non-significant values after a lag j, indicate an autoregressive series of order (lag) j. For a qth order moving average model, MA(q), the theoretical ACF function drops off to 0 after lag q with an exponentially decaying PACF value between lag 0 and lag q.

[3] j and q both refer to the number of lags, j for AR and q for MA.

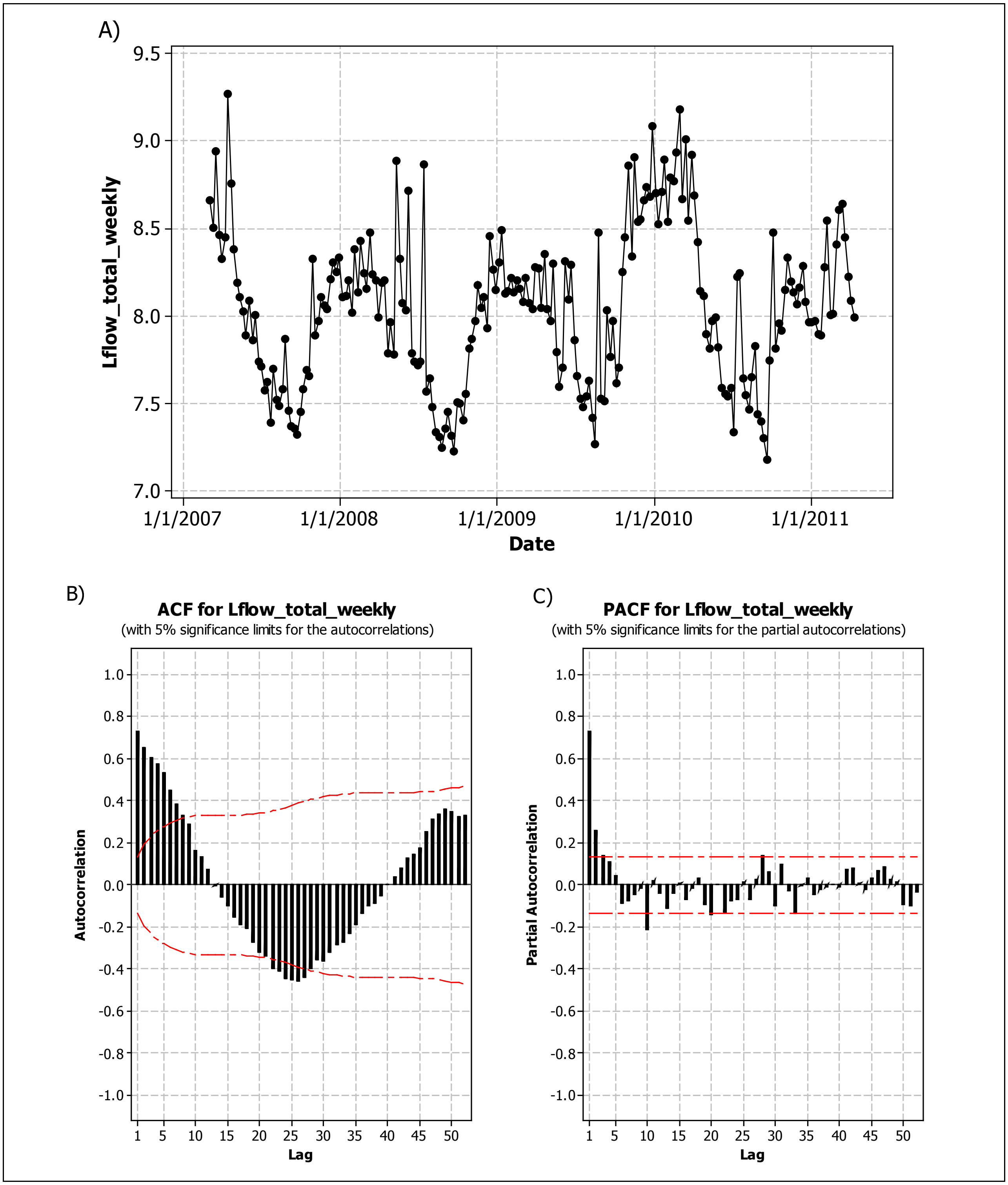
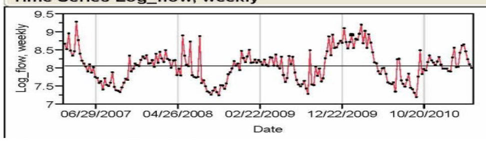


Figure 7-9. A) Time series plot, B) autocorrelation function (ACF) graph, and C) partial autocorrelation function (PACF) graph of Log(10) weekly flow from the Corsica River National Nonpoint Source Monitoring Program Project generated by Minitab. The steps are: Stat > Time Series > Autocorrelation (or Partial Autocorrelation). Identify the time series variable and enter number of lags. Select options for storing ACF, PACF, t statistics, and Ljung-Box Q statistics as desired. Press ok.

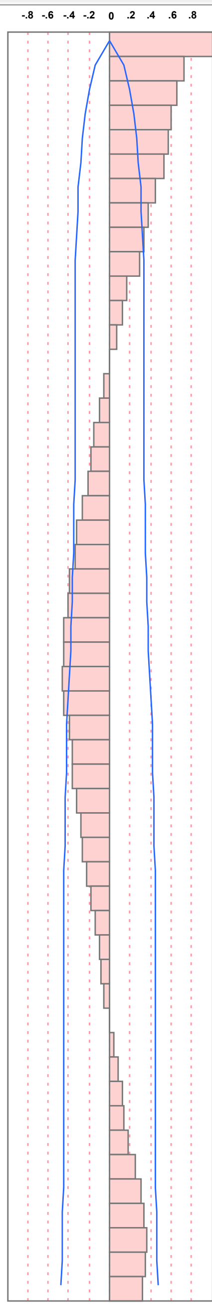
Where(:Site == "OMS")

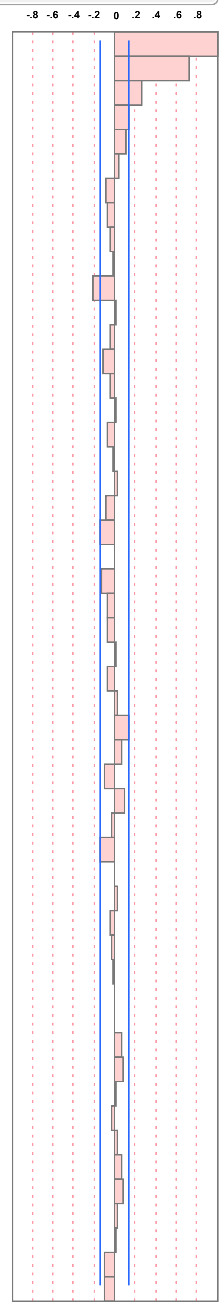
Time Series Log_flow, weekly



Mean	8.0490084
Std	0.441348
N	215
Zero Mean ADF	-0.438002
Single Mean ADF	-5.785516
Trend ADF	-5.827099

Time Series Basic Diagnostics





Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.7317	116.717	<.0001*	1	0.7317
2	0.6566	211.151	<.0001*	2	0.2609
3	0.6064	292.066	<.0001*	3	0.1394
4	0.5755	365.302	<.0001*	4	0.1079
5	0.5361	429.163	<.0001*	5	0.0462
6	0.4528	474.932	<.0001*	6	-0.0874
7	0.3834	507.901	<.0001*	7	-0.0753
8	0.3287	532.259	<.0001*	8	-0.0450
9	0.2876	550.997	<.0001*	9	-0.0135
10	0.1617	556.950	<.0001*	10	-0.2111
11	0.1322	560.945	<.0001*	11	0.0197
12	0.0745	562.219	<.0001*	12	-0.0382
13	-0.0052	562.226	<.0001*	13	-0.1126
14	-0.0587	563.026	<.0001*	14	-0.0410
15	-0.1002	565.368	<.0001*	15	0.0103
16	-0.1540	570.931	<.0001*	16	-0.0673
17	-0.1864	579.118	<.0001*	17	-0.0149
18	-0.2055	589.122	<.0001*	18	0.0317
19	-0.2705	606.542	<.0001*	19	-0.0903
20	-0.3216	631.287	<.0001*	20	-0.1377
21	-0.3377	658.711	<.0001*	21	0.0030
22	-0.3961	696.634	<.0001*	22	-0.1259
23	-0.4121	737.907	<.0001*	23	-0.0735
24	-0.4455	786.378	<.0001*	24	-0.0672
25	-0.4486	835.793	<.0001*	25	0.0136
26	-0.4584	887.653	<.0001*	26	-0.0673
27	-0.4414	935.997	<.0001*	27	0.0275
28	-0.3959	975.103	<.0001*	28	0.1414
29	-0.3568	1007.04	<.0001*	29	0.0630
30	-0.3607	1039.85	<.0001*	30	-0.0988
31	-0.3183	1065.53	<.0001*	31	0.0995
32	-0.2836	1086.04	<.0001*	32	-0.0267
33	-0.2722	1105.03	<.0001*	33	-0.1348
34	-0.2271	1118.32	<.0001*	34	-0.0004
35	-0.1887	1127.55	<.0001*	35	0.0333
36	-0.1332	1132.18	<.0001*	36	-0.0431
37	-0.0974	1134.67	<.0001*	37	-0.0216
38	-0.0882	1136.71	<.0001*	38	-0.0093
39	-0.0509	1137.40	<.0001*	39	0.0010
40	0.0053	1137.41	<.0001*	40	-0.0070
41	0.0393	1137.82	<.0001*	41	0.0768
42	0.0823	1139.65	<.0001*	42	0.0802
43	0.1281	1144.10	<.0001*	43	0.0091
44	0.1443	1149.78	<.0001*	44	-0.0234
45	0.1756	1158.25	<.0001*	45	0.0327
46	0.2529	1175.90	<.0001*	46	0.0716
47	0.3126	1203.04	<.0001*	47	0.0847
48	0.3348	1234.36	<.0001*	48	0.0263
49	0.3594	1270.67	<.0001*	49	0.0170
50	0.3476	1304.83	<.0001*	50	-0.0954
51	0.3273	1335.30	<.0001*	51	-0.0988

Figure 7-10. Autocorrelation Function (ACF) graph of weekly flow from the Corsica River National Nonpoint Source Monitoring Program Project generated by JMP. The steps are: Click “Analyze” tab, select “Modeling” followed by “Time Series.” Select Y time series (LFLOW) and X time series (Date).

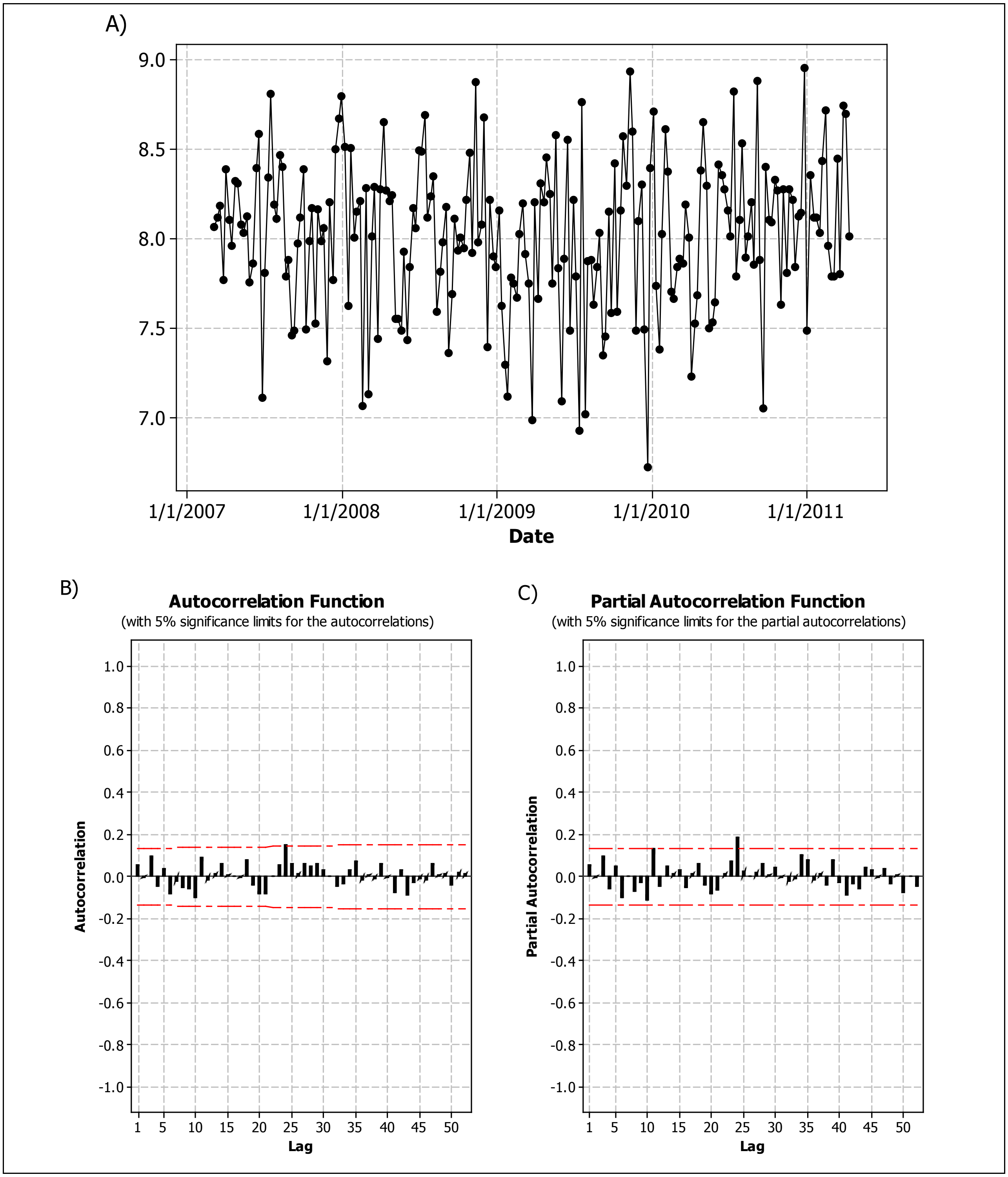


Figure 7-11. A) Time series plot, B) autocorrelation function (ACF) graph, and C) partial autocorrelation function (PACF) graph of data with zero autocorrelation (i.e., independent data with respect to time)

An autoregressive error pattern of order 1, AR(1) means that an observation is correlated with the previous observation. And, because each previous observation is related to the observation prior to it, each observation is related to all past values, but the highest correlation is with the most recent observation. A theoretical[4] AR(1) time series structure is identified by an ACF pattern that trails off bounded by an exponential decay after the first lag and the PACF dropping to 0 after lag 1 (or lag j for higher order AR series).

The patterns for both ACF and PACF in Figure 7-9A and Figure 7-9B are typical of a water quality data set with AR(1) and a strong seasonal pattern (some might argue for an AR(2) in this case which speaks to the fact that interpretation of the patterns is required, with analysts often relying on the preponderance of evidence across monitoring sites). The lag autocorrelations for weekly flow data from the Corsica River (MD) NNPSMP project in these figures do show some significant autocorrelation coefficients. The ρ_j falling outside of the red/blue lines are significant at the 95 percent confidence level. Significant autocorrelation for lag 1, as well as a strong seasonal autocorrelation pattern is evident.

Readers should consult statistics textbooks and software packages for greater detail on this and other methods to test for autocorrelation.

7.3.6.1 Methods to Handle Autocorrelation

Autocorrelation in analysis of time series data can sometimes be reduced by aggregating data over different time periods, such as weekly means rather than daily values. Use of weekly means preserves much of the original information of a daily data series, but separates data points far enough in time so that autocorrelation is reduced. When aggregating data, it is important to use a consistent procedure, e.g., using the weekly mean of 7 daily values for each week in the year, rather than mixing weekly means for some weeks with single grab samples for other weeks. Aggregation has disadvantages including: reducing the degrees of freedom and potential power of a statistical test and dampening out the potentially important high or low data.

Several statistical packages can incorporate a time series error term in the statistical model to address autocorrelation. For example, PROC AUTOREG in SAS (SAS Institute 2016d) can be used for linear regression when the error terms are autoregressive. Similar tools are available in Minitab's time series tools (i.e., Stat > Time Series) or R's statistics package.

Alternatively, if the data exhibit AR(1), which is typical for water quality data collected weekly, biweekly, or monthly, an adjustment can be made to the standard error of the trend (step or slope) terms. The correction factor was derived by Matalas and Langbein (1962) and simplified with a large sample size approximation by Fuller (1976):[5]

$$std.dev._{corrected} = std.dev._{uncorrected}\sqrt{\frac{1+\rho}{1-\rho}}$$

---

[4] Patterns from water quality sampling data will resemble theoretical patterns but will usually deviate in some way, requiring that the analyst develop a feel for interpreting such graphics.

[5] The exact formula is given by $std.dev._{corrected} = std.dev._{uncorrected}\sqrt{\frac{1+\rho}{1-\rho} - \frac{2}{n}\frac{\rho(1-\rho^n)}{(1-\rho)^2}}$ where n is the sample size.

Where ρ = autocorrelation coefficient at lag 1
Std. dev = the standard deviation of the trend term (e.g., standard error of the difference between mean values between two time periods or standard error of the slope of a linear regression).

See Spooner et al. (2011a) for additional details on this approach.

7.3.6.2 Methods to Handle Autocorrelation Caused by Seasonality

When the data exhibit seasonal cycles, incorporation of explanatory variables can be added to parametric methods to allow for adjustment of seasons. Four common approaches are used. One is to add 1 or 2 cycles by using sine and cosine terms to a linear regression model, for example, as described in *Tech Notes 6: Statistical Analysis for Monotonic Trends* (Meals et al. 2011). This approach assumes that the sine or cosine terms realistically simulate annual or semiannual seasonal cycles.

A second approach is to incorporate seasonality into the time series model. An ARIMA time series model could be used that incorporates a time series model with seasonal lag value ("differencing value" or "d"[6] in an ARIMA model, ARIMA(p,d,q)) corresponding to the length of the seasonal cycle. For example, an annual cycle will appear as a strong positive autocorrelation at lag 12 when the data series consists of monthly values or at lag 4 for quarterly values. As noted above, readers should consult statistics textbooks and software packages for greater detail on ARIMA models.

A third approach is to simply add monthly (or other seasonal) indicators to each observation in the dataset and incorporate these indicator variables in a regression model. The number of indicator variables needed is S-1[7]. For example S-1 would be 11 when the cycle is annual, but where the same months behave similarly over the years. Each indicator variable (X_1 through X_{11}) is assigned a value of 0 or 1, as indicated below:

X_1 = "1" for "January" but "0" otherwise
X_2 = "1" for "February" but "0" otherwise
…
X_{11} = "1" for November" but "0" otherwise
Note: December values would all be depicted by "0" values for X_1-X_{11}

After the indicator variables are added to the dataset, regress Y_t on the indicator variables and other independent variables (e.g., time).

A fourth approach to address seasonality is to use non-parametric tests that can handle monthly seasonality. The Seasonal Wilcoxon Rank Sum Test or Seasonal Mann-Whitney Rank Sum Test compares two or more groupings (e.g., seasonal t-test or analysis of variance). The Seasonal Kendall Test incorporates seasonal components when testing monotonic trends. Both parametric and non-parametric trend tests are featured in section 7.8.2.4. There is also a variant of the Kendall tau test (seasonal Kendall tau test with serial correlation correction (Hirsch and Slack 1984)) that can handle seasonality while also adjusting for autocorrelation.

---

[6] *Differencing* is a term used in time series analyses, where *d* is the order of differencing which creates a new time series, W_t, whose values at time t is the difference between x(t) and x(t+d). W_t then becomes the series used in the time series analysis.
[7] Where S would represent the number of time periods (e.g., months, seasons).

7.3.7 Examination of Two or More Locations or Time Periods

Comparison of two or more variables with EDA can mean comparing different data sets, such as stream nitrogen concentrations above and below a feedlot or phosphorus concentrations from a control and a treatment watershed, or comparing data from the same site over two different time periods, such as phosphorus loads from calibration vs. treatment periods.

The characteristics that make boxplots useful for summarizing and inspecting a single data set make them even more useful for comparing multiple data groups representing multiple sites or time periods. The essential characteristics of numerous groups of data can be shown in a compact form. Boxplots of multiple data groups can help answer several important questions, such as:

- Is a factor (location, period) significant?
- Does the median appear to differ between groups?
- Does apparent variability differ between groups?
- Are there outliers? Where?

Boxplots are helpful in determining whether central values, spread, symmetry and outliers differ among groups. If the main boxes of two groups, for example, do not substantially overlap on the vertical scale, there may be a reason to suspect that the two groups differ significantly (note that such difference should be tested using quantitative statistical techniques). Interpretation of boxplots can help formulate hypotheses about differences between groups. Figure 7-12 shows a boxplot of total suspended solids concentrations in three Vermont streams. The plot suggests that TSS concentrations may tend to be slightly lower at Station 3 compared to the other two stations; however, because the boxes overlap, it is unlikely that any comparison of medians would result in statistically significant differences.

Inferences about differences between locations or time periods resulting from graphical evaluation of the data must be confirmed by more rigorous hypothesis testing analyses (see sections 7.7 and 7.8).

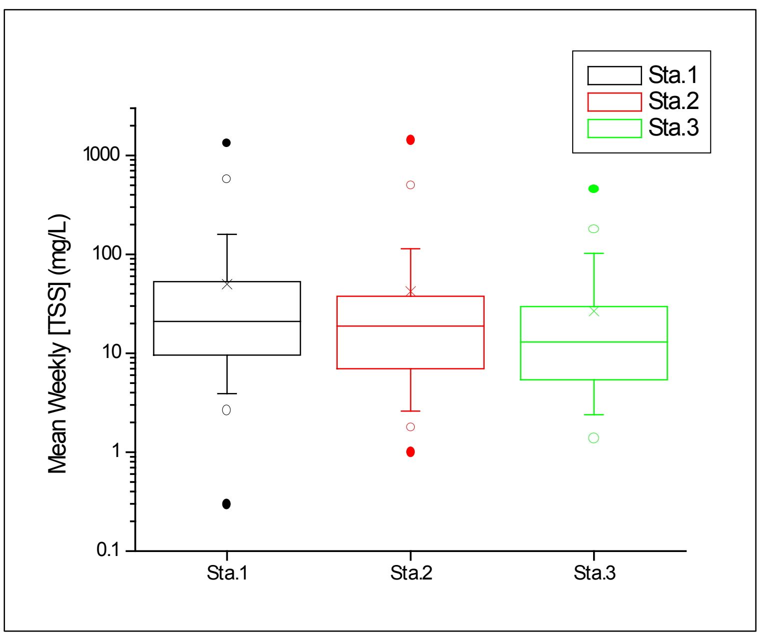


Figure 7-12. Boxplots of TSS concentration for three stream stations, 1998 (Meals 2001)

7.3.8 Examine Relationships between Variables

Looking at how variables relate to each other is a way to begin to consider causality, i.e., is the behavior of one variable the result of action by another. Such ideas can suggest sets of variables to evaluate together. For example, if variable B (e.g., suspended sediment concentration) goes down as variable A (e.g., acres of reduced tillage) goes up, has the BMP program improved water quality? Examination of correlations between different variables observed simultaneously (e.g., SSC and total P or turbidity and SSC) can suggest relationships that might change with BMP programs or indicate where one variable could serve as a surrogate for another. Graphical analysis (e.g, scatterplots of variable A vs. variable B) can suggest meaningful correlations that would need to be confirmed with more rigorous statistical tests.

The two-dimensional scatterplot is one of the most familiar graphical methods for data exploration. It consists of a scatter of points representing the value of one variable plotted against the value of another variable from the same point in time. Scatterplots illustrate the relationship between two variables. They can help reveal if there appears to be any association at all between two variables, whether the relationship is linear, whether different groups of data lie in separate regions of the scatterplot, and whether variability is constant over the full range of data.

Figure 7-13 shows a scatterplot of phosphorus export in a control and a treatment watershed in Vermont. Note that the data are plotted on a log scale to obtain a linear relationship. There is a strong positive association between P in the two streams. This simple scatterplot indicates that it is probably worth proceeding with more rigorous statistical analysis to evaluate calibration between the two watersheds in a paired-watershed design. As with this example, it is common that the relationship between variables is exponential. In such cases, the log transformation allows the relationship to be expressed linearly and evaluated using linear regression.

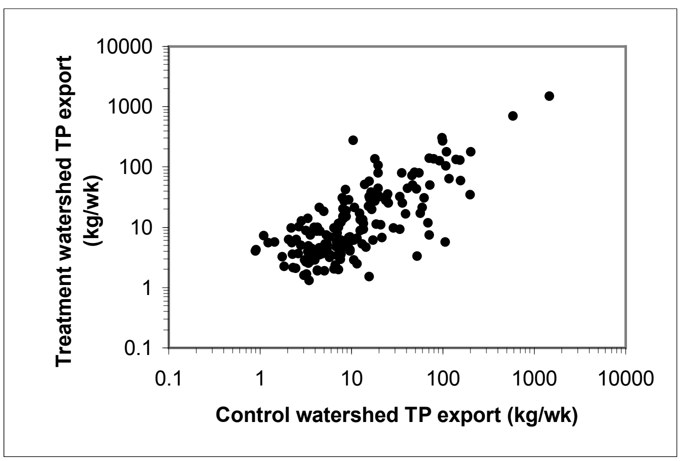


Figure 7-13. Scatterplot of weekly TP export from control and treatment watersheds, calibration period (Meals 2001)

Figure 7-14 shows another scatterplot examining the relationship between streamflow and *E. coli* counts in another Vermont stream. In a nonpoint source situation, a positive association between streamflow and bacteria counts may be expected, as runoff during high flow events might wash bacteria from the land to the stream. In this case, however, it does not require application of advanced statistics to conclude from Figure 7-14 that there is no such association (in fact the correlation coefficient r is close to zero). However, recall that EDA involves an open-minded exploration of many possibilities. In Figure 7-15, the data points have been distinguished by season. The open circles represent data collected in the summer period and there still appears to be no association between streamflow and *E. coli* counts. The solid circles, representing winter data, now appear to show some positive correlation (r = 0.45) between streamflow and bacteria counts, with high bacteria counts associated with high flows. This picture suggests that something different is happening in winter compared to summer with respect to streamflow and *E. coli* in this watershed, a subject for further investigation.

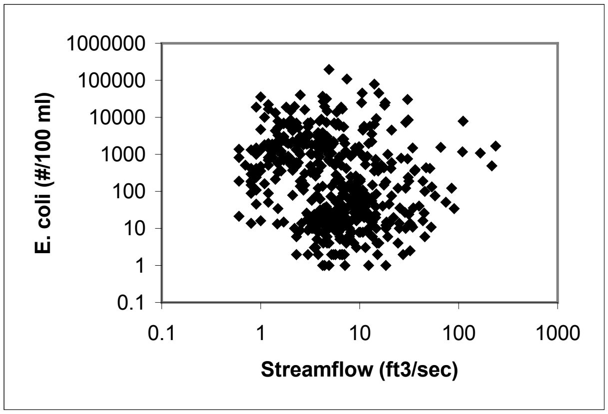


Figure 7-14. Scatterplot of *E. coli* vs. streamflow, Godin Brook, 1995-1998, all data combined (Meals 2001)

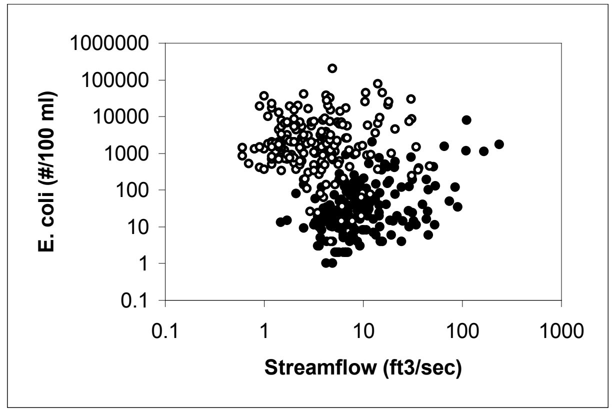


Figure 7-15. Scatterplot of *E. coli* vs. streamflow, Godin Brook, 1995-1998, where solid circles = winter, open circles = summer (Meals 2001)

In looking for correlations in scatterplots, choose the variables carefully. A common mistake is the comparison of variables that are already related by measurement or calculation. An example of such spurious correlation is the comparison of streamflow with load. Because load is calculated as concentration multiplied by flow, a scatterplot of flow vs. load has a built-in correlation that means very little, even though it looks good in a scatterplot. Also remember that correlation does not guarantee causation – just because two variables are correlated does not mean that the variation in one is caused by variation in the other.

There are many numerical techniques available to examine and test the relationship between two or more variables. In EDA, the simplest technique is correlation, which measures the strength of an association between two variables. The most common measure of correlation is Pearson's *r*, also called the *linear correlation coefficient*. If the data lie exactly on a straight line with positive slope, *r* will equal 1; if the data are perfectly random, *r* will equal 0. For Pearson's *r*, both variables should be normally distributed and continuous (Statistics Solutions 2016). The test also assumes a straight-line relationship between the variables and constant variance (homoscedasticity). Pearson's *r* is sensitive to outliers.

Other measures of correlation that are less sensitive to outliers include the nonparametric Kendall's *tau* and Spearman's *rho* (Spearman's rank correlation coefficient). Spearman's *rho* makes no assumptions about the distribution of the data and is an appropriate test when the variables are at least ordinal and the variables are monotonically related (Statistics Solutions 2016). With ordinal variables, the ordering of values is known but the differences between them are not quantified (e.g., Excellent, Good, Fair, Poor).

Measures of correlation are easily calculated by most statistical software packages and are described in chapter 4 of the 1997 guidance (USEPA 1997b). It must be cautioned that whenever a numerical correlation is calculated, the data should also be plotted in a scatterplot and examined visually as described above. Many different patterns can result in the same correlation coefficient. Never compute a correlation coefficient and assume that the data follow a simple linear pattern.

There are several methods of simultaneously evaluating variables that are likely related to each other. Cluster analyses group variables and/or observations into similar categories usually based on an agglomerative hierarchical algorithm which is the most common clustering pattern used in water quality analyses. In this clustering procedure, each observation begins as an individual "cluster." The similarities or distances between these clusters are measured using one of several options, including Euclidian distance and correlation coefficients. The closest two clusters are then merged into a new cluster. Distances are calculated again using the updated set of clusters, and the process repeated until only one cluster remains. The result of this analysis is a sequence of groupings that can be represented in a cluster tree or dendrogram. The analyst can then perform a visual analysis to infer potential groupings and relationships among variables. It is important to note that cluster analysis does not consider multicollinearity between the variables. Cluster analysis conducted as part of EDA might be used to explore and define site or time groupings that would be useful to explore in later analysis.

Other multivariate techniques that can be applied in subsequent analysis include principal components analysis, canonical correlation, and discriminant analysis (SAS Institute 1985). These methods are discussed further in section 7.5.2.5.

7.3.9 Next Steps

Data exploration results (knowledge of how data are distributed, their characteristics, and their relationships) will help illustrate any needs to adjust the data to enable the appropriate subsequent statistical tests. In addition, hypotheses can be refined to facilitate more advanced statistical techniques. section 7.4 describes methods for accounting for censored data. Sections 7.5 through 7.9 present various advanced procedures for analyzing data for a range of purposes. Section 7.10 presents a list of tools and other resources for data analysis.

7.4 Dealing with Censored Data

7.4.1 Types of Censoring

Monitoring programs such as those analyzing for pesticides, metals, or other constituents often present at very low concentrations may report lab results where concentration is below the detection limit of the analysis. Bacteriological tests may report very high results as "too numerous to count" (TNTC). Such data – typically reported as "<" or ">" (left- and right-censored, respectively) some value – are referred to as "censored" data.

Censored values are usually associated with limitations of measurement or sample analysis, and are commonly reported as results below or above measurement capacity of the available analytical equipment. Results that are indistinguishable from a blank sample are normally reported as less than the detection limit (DL). The true values of these *left-censored* observations are considered to lie between zero and the DL. Depending on the laboratory, some results greater than the DL may be identified as less than the quantitation limit (QL) or reported as a single value and given a data qualifier to indicate the value is less than the QL. Typically, results reported as less than the QL indicate that the analyte was detected (i.e., greater than the detection limit), but at a low enough concentration where the precision was deemed too low to reliably report a single value. These *interval-censored* observations are considered to lie between the DL and QL.

Left- and interval-censored observations are less commonly encountered when working with sediment and nutrients because they are usually present at levels above their QLs. However; left censoring is common when toxics and pesticides are being analyzed.

An example of *right-censoring* includes microbiological analyses with misestimated dilution resulting in TNTC (too numerous to count) and exceedance of flow gage limits during floods. Right-censoring may also be encountered when lakes and estuaries are monitored for light penetration via Secchi depth and the result is reported as *visible on bottom*, i.e., the Secchi disk is observable on the bottom.

Helsel (2012) provides a seminal discussion of varying reporting limits and concerns with some data censoring practices. This guidance recommends that detection limits and quantitation limits be stored with the measurements and each result be clearly qualified to indicate its relation to the DL or QL as appropriate.

7.4.2 Methods for Handling Censored Data

There is no single ideal method for managing censored data in statistical analyses. When comparing various methods, this guidance recommends that analysts use methods that minimize bias and error. Extensive research in water resources as well as other fields of science such as survival analysis (e.g., how long does a cancer patient live after treatment) has considered numerous techniques. One deficiency over the last 20 years has been the lack of readily available tools for widespread use, making many of these tools out of reach for general use. Efforts continue to improve upon the availability of these tools. The most notable is a compilation of methods and recommendations developed by Helsel (2012) with additional information provided at *Practical Stats*. Much of the remaining discussion in this section is derived from Helsel's book (Helsel 2012) and the reader is encouraged to review his book for a more in-depth discussion.

7.4.2.1 Past Methods

With improved tool access, past methods for accommodating censored observations can be avoided. The most notable past method is simple substitution. This involves the replacement of censored observations with zero, ½DL, or DL. Although simple substitution is commonly used (and even recommended) in some state and federal government reports as well as some refereed journal articles, there is no real theoretical justification for this procedure. Substitution may perform poorly compared to other more statistically robust procedures, especially where censored data represent a high proportion of the entire dataset. More egregiously, some reports have simply deleted observations less than the detection limit. Some past researchers have recommended simply reporting the actual measured concentrations even if the concentrations are below the DL (Gilliom et al. 1984). This approach has not gained traction as laboratories are reluctant to implement such a practice, although Porter et al. (1988) suggested that an estimate of the observation error could be reported to better qualify the measurement. While simple substitution might be convenient for initial exploratory analyses using spreadsheet tools, more robust procedures are available and are recommended.

7.4.2.2 Using Probability Distribution Theory to Estimate the Summary Statistics

In environmental sciences, two common methods considered for estimating summary statistics from censored data sets include maximum likelihood estimation (MLE) and robust regression on order statistics (ROS). Both methods ultimately rely on a distributional assumption and both methods allow for multiple detection limits and estimation of confidence intervals. The reader is referred to Helsel (2012) for a more detailed discussion.

MLE uses the uncensored observations, the proportion of censored observations, and a distributional assumption to compute estimates of summary statistics. A lognormal distribution is commonly assumed with water quality data; however, commercial software will usually allow a variety of assumptions to be considered.

The robust ROS procedure (Helsel and Cohn 1988) relies on fitting a regression line to a normal probability plot of the uncensored observations and is applicable for multiple censoring levels. If the uncensored data do not fit a normal distribution, the analyst can transform the uncensored data with lognormal or other appropriate transformation. The process of selecting the best transformation is similar to that if all data were uncensored and diagnostics are typically available in current statistical software. The regression is then used to impute values for the censored data. The imputed and uncensored data are then, if necessary, transformed back to their original data scale, allowing summary statistics to be estimated using standard techniques. Confidence intervals for the mean and standard error estimates can be computed using bootstrapping (e.g., Helsel 2012). In summary for the mean, a random sample (with replacement) is selected from the site data. These data are passed through the robust ROS procedure described above, and a resulting mean is computed. The process of selecting a random sample, implementing the robust ROS procedure and computing a resulting mean is repeated, say, 1,000 times. Confidence limits are then empirically selected from this set of 1,000 means (e.g., the 5th and 95th percentile of these 1,000 means would be the 90 percent confidence interval on the mean).

The MLE tool can be applied to less-thans and TNTC in the same data set. Helsel (2012) provides recommendations for which method to use based on the number of observations and degree of censoring. Notably, no method works well when the degree of censoring exceeds 80 percent. In the situations where the censoring level exceeds 80 percent, Helsel (2012) recommends reporting information on the percent of observations above a meaningful threshold and no further summary statistics. For all summary statistics with censored data, this guidance recommends reporting the maximum detection limit, number of observations, and number of censored observations with all summary statistics.

7.4.2.3 Hypothesis Testing with Censored Data

There are a variety of nonparametric hypothesis tests that can be directly used with raw data sets that have censored observations and generally rely on the rank (or order) of the data. These tests include the Mann-Whitney test (two random samples), Wilcoxon (paired samples), and Kruskal-Wallis (several random samples), and Kendall and Seasonal Kendall tau (monotonic trends). In these tests, censored observations are treated as tied values, no different from cases where ties might occur between uncensored observations. Consider the ordered data set of <1, <1, 1.5, 4, 8, 9, 10, and 10. The two censored observations (of <1) are less than all the other observations, but are treated as tied to each other. The handling of the two "*<1*'s" is no different than the two *10*'s which are both greater than all the other values, but tied with each other. One deficiency of these tests is that they are limited to a single detection limit (e.g., the tests do not have a method to compare "<1" and "<2"). To apply the above nonparametric tests with data sets that have multiple detection limits, the analyst will need to re-censor the data to the highest detection limit. Note: do not use the previously described ROS procedure to impute values for censored data and then apply one of the nonparametric tests described in this paragraph (or parametric tests), as erroneous results might be computed because the rank of the imputed values were calculated based upon the order of data set entry, which is not related to any true ranking of the actual water quality values.

An alternative approach is to apply MLE regression tools that are designed for multiply censored dependent variables. Similar to simple regression or multiple regression, relationships between singly- or multiply-censored dependent variables can be established with independent variables. Indicator variables can be used to set up groupings to expand the MLE regression tool for comparing two or more groups or seasonal/explanatory adjustments as well.

7.5 Data Analysis for Problem Assessment

7.5.1 Problem Assessment – Important Considerations

One of the most critical steps in controlling NPS pollution is to correctly identify and document the existence of a water quality problem. The water quality problem may be defined either as a threat to or impairment of the designated use of a water resource. Impairments are generally defined and identified as violations of water quality standards (WQS). Water quality standards define the goals for a waterbody by designating its uses, setting criteria to protect those uses, and establishing provisions such as antidegradation policies to protect waterbodies from pollutants. A WQS consists of four basic elements:

1. **A designated use of the water body.** States and Tribes specify appropriate water uses to be achieved and protected, taking into consideration the use and value of the waterbody for public water supply, for protection of fish, shellfish, and wildlife, and for recreational, agricultural, industrial, and navigational purposes. In designating uses for a water body, States and Tribes consider the suitability of a water body for the uses based on the physical, chemical, and biological characteristics of the water body, its geographical setting and scenic qualities, and economic considerations.
2. **Water quality criteria**. Water quality criteria are science-based numeric pollutant concentrations or narrative requirements that, if met, will protect the designated use(s) of the water body. Criteria may be based on physical, chemical, or biological characteristics. Numeric criteria may, for example, establish limits for concentrations of toxic pollutants to protect human health or aquatic life. Narrative criteria stating that a water body must be "free from" toxic contaminants can serve as a basis for limiting the toxicity of waste discharges to aquatic life.
3. **An antidegradation policy.** Water quality standards include an antidegradation policy that maintains and protects existing uses and water quality conditions necessary to support such uses, maintains and protects high quality waters where existing conditions are better than necessary to protect designated uses, and maintains and protects water quality in outstanding national resource waters. Except for certain temporary changes, water quality cannot be lowered in such waters.
4. **General policies.** States and Tribes may adopt policies and provisions regarding implementation of water quality standards, such as mixing zones, variances, and low-flow policies. Such policies are subject to EPA review and approval.

Water quality monitoring to support problem assessment is usually focused on documenting violations of WQS in time (e.g., frequency of exceedance) and space (e.g., geographic extent of exceedance). Water quality data for such purposes may be collected by an ongoing monitoring program (e.g., a state ambient monitoring program) or by a reconnaissance study designed to provide a preliminary, low-cost overview of water quality conditions in the area of interest (see section 2.4.2.1). The EPA ATTAINS database is the repository for information from state integrated reporting (IR) on water quality conditions under sections 305(b), 303(d), and 314 of the Clean Water Act, and the Reach Address Database contains state IR geospatial data. ATTAINS includes state-reported information on support of designated uses in assessed waters, identified causes and sources of impairment, identified impaired waters, and TMDL status.

A detailed discussion of monitoring designs has been presented in chapter 2 of the 1997 guidance (USEPA 1997b). Some designs appropriate for problem assessment have been discussed in section 2.4 of this guidance. In general, monitoring designs appropriate for collecting data to support NPS problem assessment include:

- **Synoptic surveys** designed to determine the magnitude and geographic extent of WQS violations, often used to identify pollutant source areas within a watershed;
- **Above/below monitoring**, wherein a potential pollutant source area is bracketed between upstream and downstream sampling points to assess the impact of the source area on pollutant levels; and
- **Trend monitoring** designed to collect long-term time-series data at one or more watershed sampling points that are useful in determining the frequency and magnitude of exceedance of WQS.

Both above/below (if pre- and post BMP data is collected) and trend monitoring designs can also be applied to other monitoring objectives such as project effectiveness evaluation using permanent monitoring stations equipped with automatic sampling equipment and continuous flow measurement devices.

Grab samples with instantaneous flow measurements for a few sampling events may be sufficient for initial problem assessment and source identification, but monitoring data for problem assessment should include both baseflow and stormwater monitoring necessary to fully characterize the system. Storm sampling is useful for documenting the delivery of pollutants by runoff and overland flow, critical considerations for waters impacted by NPS. Combined with hydrologic data, basic climatic information can be used to evaluate the seasons or times of the year when pollutant levels are highest or lowest and when high flow events, drought, or other factors affect water quality. Note that concentration data alone without concurrent flow or stage data are often of limited utility.

Biological monitoring is used widely in water quality assessments and EPA provides information and links to resources addressing various aspects of the application of aquatic life criteria in water quality assessments. Chapter 4 of this guidance is devoted to biological monitoring. The discussion below, however, emphasizes the use and application of statistical analysis to chemical and physical monitoring data for which there is a greater body of literature. See chapter 7 of *Handbook for Developing Watershed Plans to Restore and Protect Our Waters* (USEPA 2008) for a broad discussion of approaches to assessing water quality problems and identifying causes and sources of those problems using a wide range of information sources.

7.5.2 Data Analysis Approaches

7.5.2.1 Summarize Existing Conditions

In a single stream or subwatershed, one monitoring location may be sufficient for problem assessment. More often, sampling at two or more locations is necessary to evaluate existing conditions of the watershed. Concurrently, sampling at two or more locations can aid in identification of subwatersheds that merit further evaluation for pollution reductions or water resource protections.

When data from different locations in a watershed or different sampling time periods are consistent and comparable (e.g., from a synoptic survey or from multiple watershed stations in the same monitoring regime), a first step is to summarize existing conditions using univariate statistics – mean, median, range, variance, interquartile range – for different sampling locations. If differences over time or flow conditions are evident, it may be useful to group the data into separate baseflow and wet-weather strata or by season. If enough samples have been collected (i.e., at least three), existing water quality can be compared across multiple sites. Visual comparisons between sites can be depicted graphically using boxplots. Figure 7-16 shows a set of boxplots for one year of weekly conductivity data from three small watershed trend stations in Vermont (Meals 2001). Conductivity at site WS1 appears to be substantially lower than that observed at the other two stations; conductivity at WS2 tended to be somewhat higher than that observed at WS3, with more frequent high extreme values. Mean or median values can be compared between two sites using the unpaired Student's t-Test or a nonparametric equivalent such as the Wilcoxon Rank Sum Test (also known as the Mann-Whitney Rank Sum Test). More than two sites can be compared using Analysis of Variance or the Kruskal-Wallis k Sample Test. Adjustments for seasons or hydrologic explanatory variables should be considered by employing appropriate statistical tests such as Analysis of Covariance or the Seasonal Wilcoxon Rank Sum Test (also known as the Mann Whitney Rank Sum Test). If the data between two sites are paired, differences can be tested using the paired Student's t-Test or the Wilcoxon Signed Rank Sum Test. Paired tests are generally more powerful and should be used when enabled by collecting samples at the same time period at two sites.

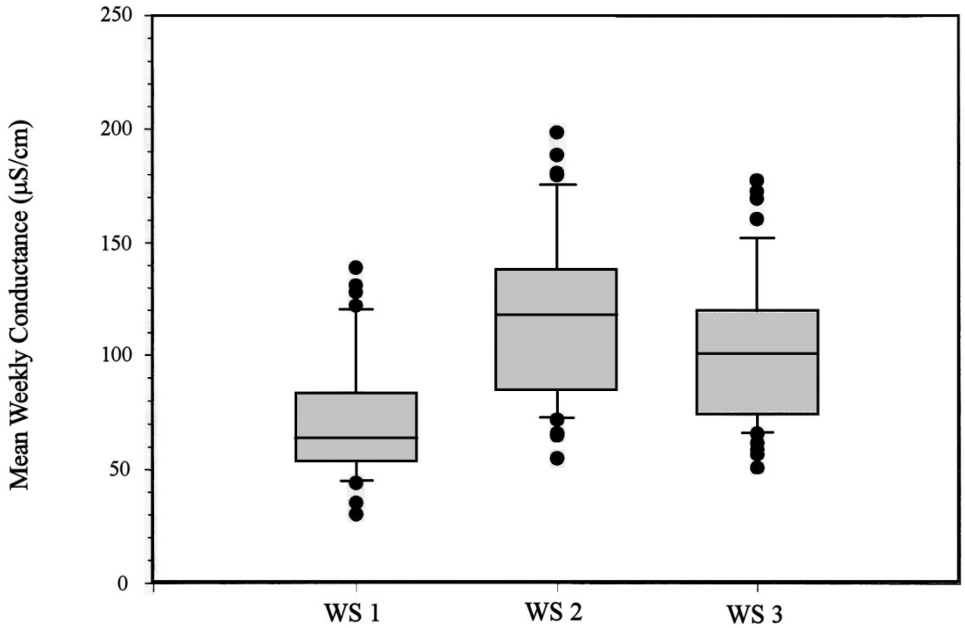


Figure 7-16. Boxplots of conductivity at three Vermont monitoring stations, October 1999 – September 2000 (Meals 2001)

Time series plots can visually reveal relationships over time and between locations. Figure 7-7 from section 7.3, for example, shows very clearly the seasonal cycle in *E. coli* counts in a Vermont stream, and Figure 7-4 reveals a different behavior at Station 2 compared to other stations regarding P concentrations. Time series statistical analyses can reveal autocorrelation and seasonality (see section 7.3.6).

Regression analysis between variables of primary interest (e.g., pollutant concentration/loads) and explanatory variables such as stream discharge can assist in documenting hydraulic relationships at a single monitoring location or between subwatersheds. Establishing relationships among variables can be very helpful in project planning as well. Scientists involved in the Upper Grande Ronde (OR) NNMP project, for example, explored relationships between fish and environmental factors via multivariate analysis and found that management and restoration activities that focus on reducing the maximum annual stream temperature would be the most effective in creating stream conditions that support salmonids (Drake 1999).

7.5.2.2 Assess Compliance with Water Quality Standards

Water quality data can be evaluated for violation of water quality standards (WQS). Note that specific requirements for documenting impairment in a regulatory sense may vary by circumstance. For some states and for some pollutants, a single observation exceeding a WQS may be sufficient to designate impairment. In other cases, determination of impairment must be based on violation of a WQS over a defined period of time or number of observations. A WQS for bacteria to support shellfishing may, for example, be based on a geometric mean of a number of different samples collected over a 30-day period, rather than on a single sample. Sanitary surveys in North Carolina, for example, include a shoreline survey to identify potential pollutant sources, a hydrographic and meteorological survey, and a bacteriological survey (NCDENR 2016). Both the monitoring program and data analysis must be tailored to the regulatory requirements that apply to the watershed under study.

A data series should be plotted and the pattern evaluated for exceedance of WQS; plots of a time series at a single station or boxplot of multiple stations can be examined. Figure 7-17 shows how a time series plot can illustrate both the frequency and magnitude of violations of WQS. The dashed line represents the water quality criterion for chronic exposure; all of the observations exceed that level. The red line marks the acute criterion and shows that several observations exceeded that concentration. Moreover, most of the excursions above the acute criterion occurred around April, suggesting a seasonal aspect to the impairment. This kind of pattern may support inferences about pollutant source activity.

One way to evaluate the frequency or probability of violating WQS is to use probability plots or duration curves. Figure 7-18 shows a cumulative frequency plot of three years of *E. coli* data from a Vermont agricultural watershed (Meals 2001). In this case, it can be seen that compliance with the Vermont WQS of 77 cfu/100 ml *E. coli* occurred about 36 percent of the time and the stream was therefore considered impaired for *E. coli* about 64 percent of the time. If the USEPA criterion of 235 cfu/100 ml were applied, the stream would be in compliance with that criterion about 48 percent of the time and impaired about 52 percent of the time.

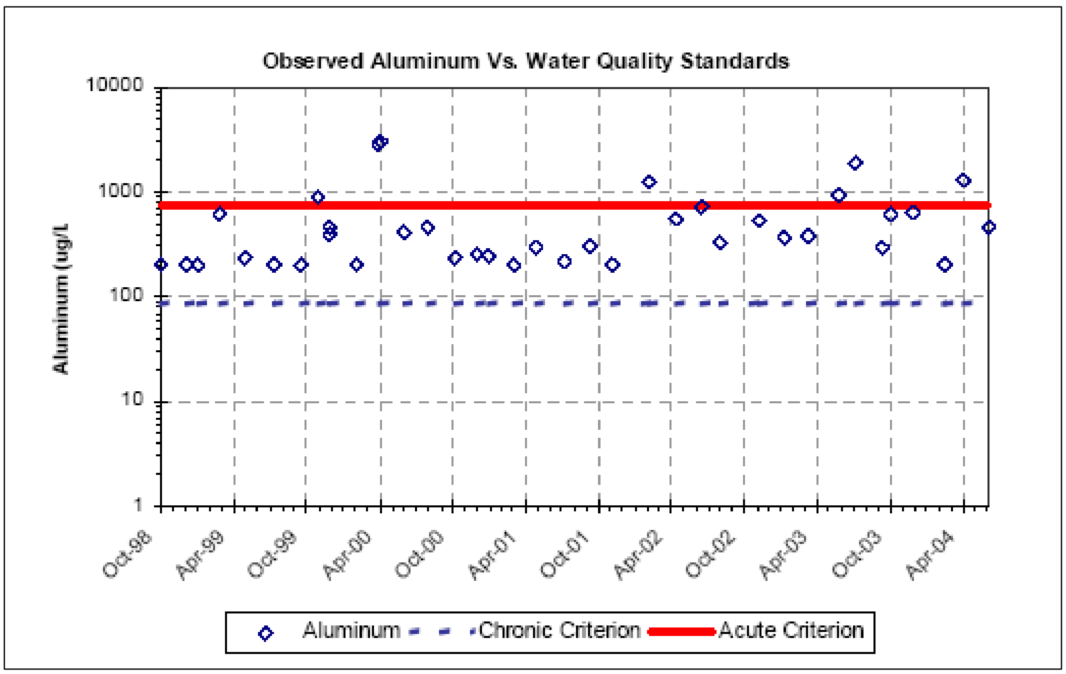


Figure 7-17. Example time series plot of observed aluminum concentrations compared to water quality criteria

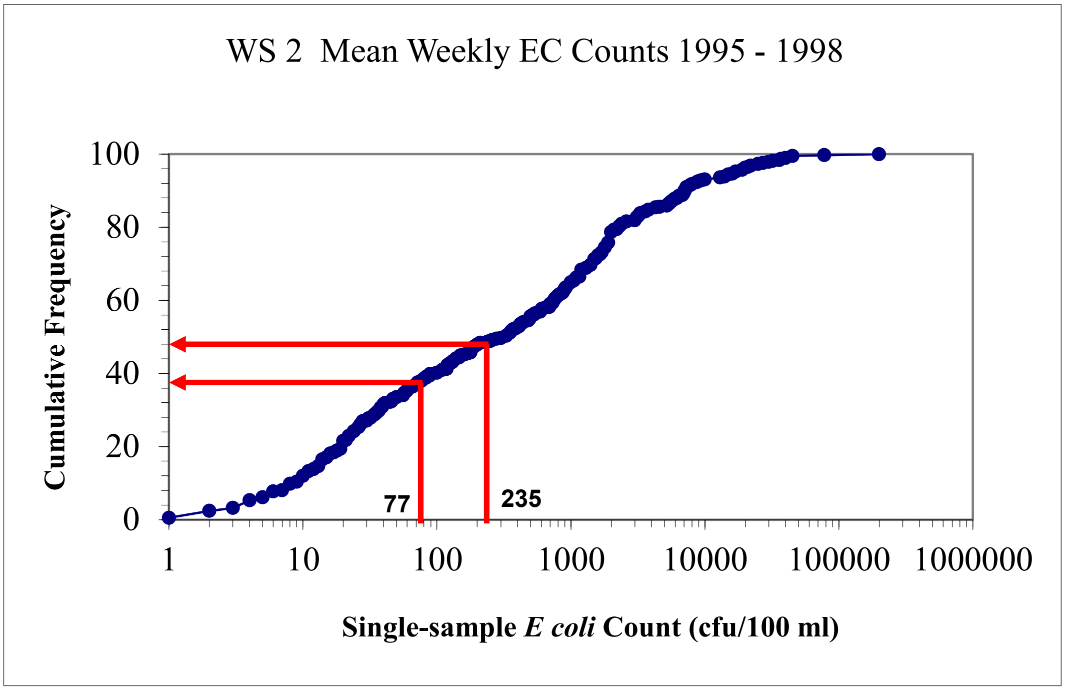


Figure 7-18. Cumulative frequency plot of three years of *E. coli* data from a Vermont stream (adapted from Meals 2001). Red lines represent frequency of observations at or below the VT WQS of 77 cfu/100 ml and the frequency of observations at or below the EPA criterion of 235 cfu/100 ml.

7.5.2.3 Identify Major Pollutant Sources

Cost-effective treatment of watersheds to address the pollutants and other causes of water quality problems requires knowledge of the sources contributing to the problems. Commonly used approaches to identifying and characterizing sources use both water quality and land-based information at varying levels of detail and quality (USEPA 2008). This section describes methods for analyzing water quality and associated monitoring data to characterize and aid in the prioritization of pollutant sources as part of the watershed planning process. See section 4.4.5 for an example of using biological monitoring in the Lake Allatoona/Upper Etowah River (GA) watershed.

Data from a synoptic survey or from regular monitoring of several subwatersheds combined with data on land use, management, or other land-based characteristics can inform understanding of major pollutant sources in a watershed. Correlation or regression analysis can be applied to explore relationships between pollutant concentrations and subwatershed characteristics, e.g., total P (TP) concentrations vs. manured cropland or suspended sediment concentration vs. cropland in cover crops. Annual mean or median values for pollutant concentrations could be compared to annual data on land use/management activities because concentrations will vary widely between individual events against land characteristics that are relatively constant within a single year or crop season. However, this simplification will not reveal seasonal and hydrologic variability in water quality or responses to short term land use changes such as animal numbers or fertilization. Where suitable knowledge of land use or land management is available, it may be more useful to provide water quality summary data for different periods that reflect distinctly different land use/management conditions (e.g., after spring manure applications vs. remainder of the year) during the monitoring period.

Boxplots or bivariate scatterplots can be compared between monitoring sites that reflect distinctive land use or management, thereby suggesting important pollutant source activities. If sufficient data from different subwatersheds or sampling stations exist, analysis of variance (ANOVA), or the nonparametric Kruskal-Wallis k Sample test can be used to test for significant differences in pollutant concentrations between sites and then compare these findings to differences in land use between the drainage areas sampled (graphical or tabular summaries). Analysis of covariance (ANCOVA) should be considered in cases where data are sufficient to test for differences among sites or seasons with adjustment for covariates such as precipitation or flow. See sections 4.6 and 4.8 of the 1997 guidance (USEPA 1997b) for a discussion of ANOVA and ANCOVA.

If flow data are available with concentration data, load estimates can be calculated to compare the magnitudes of pollutant sources (see section 7.9 for load estimation methods). The spatial and temporal resolution possible for load estimates will be determined by the number and location of sampling sites and the time frame and frequency of sampling events, respectively. Source-specific or subwatershed loads will generally be more helpful than loads at the watershed outlet, and in many cases seasonal loads or a classification of event vs. baseflow loads will be very helpful in the watershed project planning phase (see section 7.6).

It should be noted that correlation does not guarantee causation. Specifics of pollutant source activity and transport/delivery mechanisms must be considered to focus in on causation. Time of travel studies for various points in the watershed, for example, can be helpful in better characterizing the relationship between various sources or subwatersheds and downstream water quality. USGS describes methods for measuring time of travel (Kilpatrick and Wilson 1989).

7.5.2.4 Define Critical Areas

Data collected in the problem assessment phase can be used to help define critical source areas for pollutants, knowledge that is key to understanding the watershed, prioritizing land treatment, and evaluating project effectiveness. With concurrent data from monitored subwatersheds or tributaries (e.g., from a synoptic survey), statistical tests such as the Student's t Test or ANOVA can be used to identify significant differences in pollutant concentration or load among multiple sampling points. Such data can be displayed graphically in a map to show watershed regions that may be major contributors of pollutants. Figure 7-19, for example, shows a map of NO_2+NO_3-N concentrations from an April, 2003 synoptic survey in the Corsica River (MD) watershed (Primrose 2003). Nitrate/nitrite concentrations were found to be excessive in four subwatersheds, high in sixteen, and moderately elevated in seventeen others. Benchmarks for determining excessive/high/moderate or similar categories can be based on numeric water quality criteria or reference watershed values. If flow data were also available, it would be possible to estimate loads and compare subwatersheds on the basis of absolute (e.g., kg TP) or areal (e.g., kg TP/ha) loads. Figure 4-3 of section 4.4.5 illustrates how biological monitoring data from the Lake Allatoona/Upper Etowah River (GA) watershed were used for site-specific assessments of biological condition.

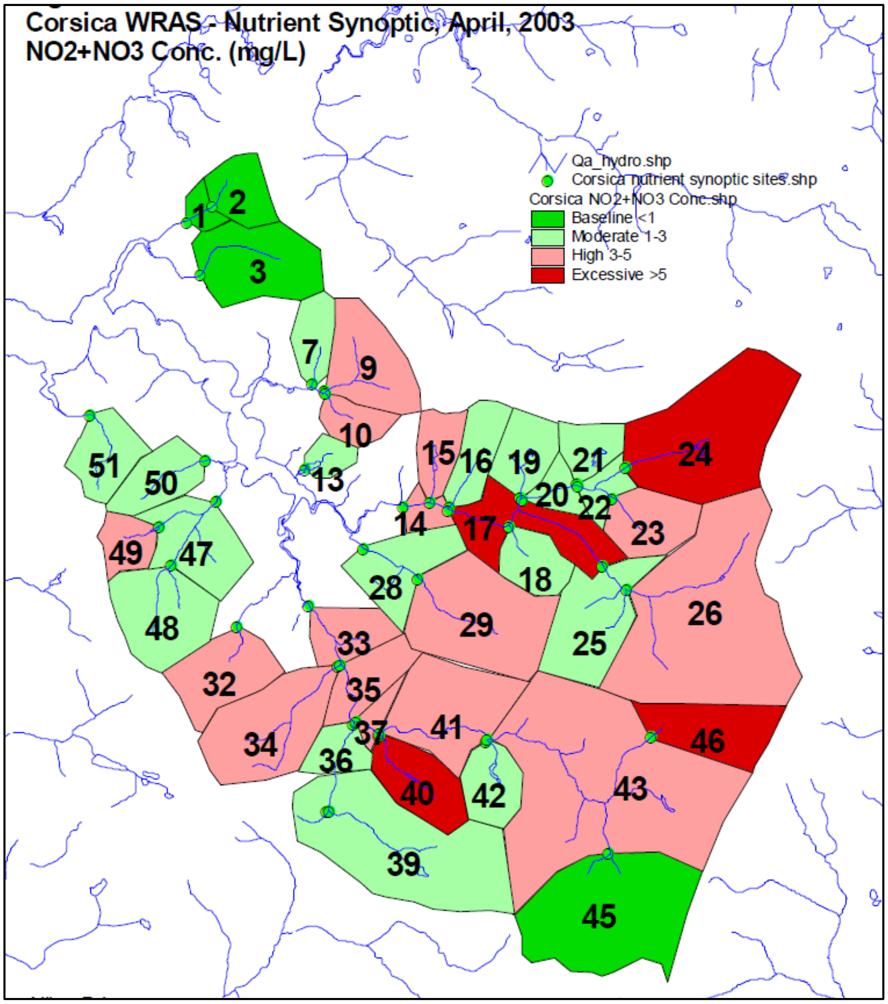


Figure 7-19. Map of synoptic sampling results from 41 stations in the Corsica River Watershed (Maryland) for NO_2+NO_3-N concentration (Primrose 2003). Pink and red shaded subwatersheds represent drainage areas contributing high (3-5 mg/L) and excessive (>5 mg/L) NO_2+NO_3-N concentrations, respectively.

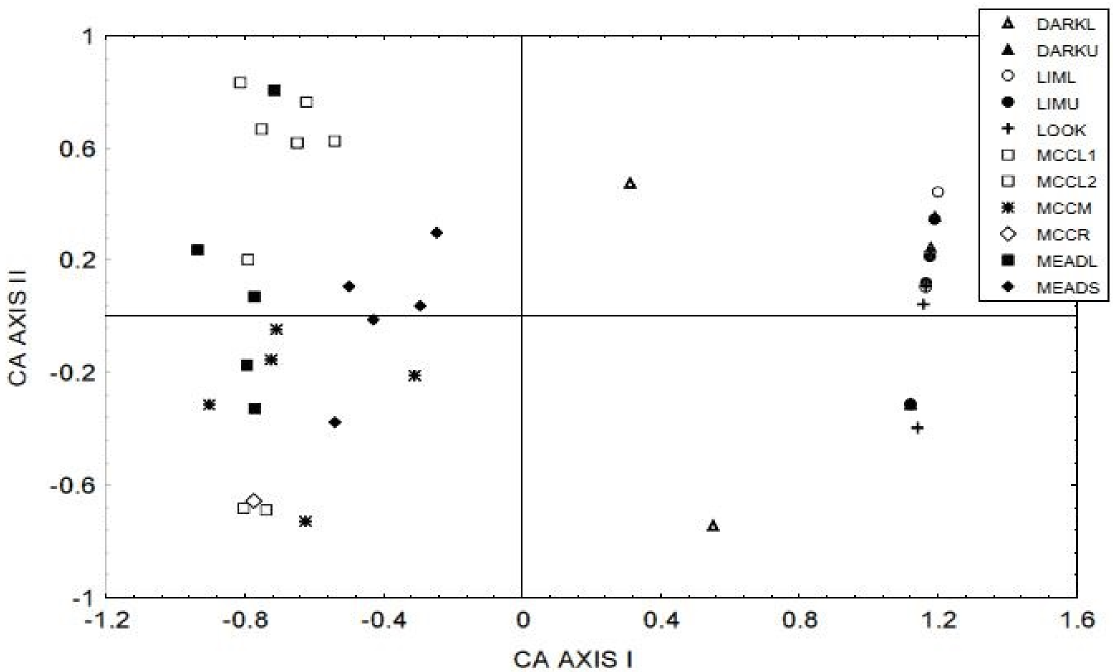
Assessment of critical areas using a small set of water quality data has some limitations. Conditions determining pollutant generation (e.g., storm event, season, management schedules) must be considered in drawing conclusions about critical areas. Data collected during the active crop growth season may show a very different situation from data collected in winter, although for source identification purposes, it may be preferable to sample during the most critical times of year. The data mapped in Figure 7-19, for example, were collected in April, during or immediately following the spring planting and fertilizer application season when N losses from recently applied fertilizers might be expected to be high. Secondly, the spatial resolution of source area identification is limited by the resolution of the sampling network. Detailed site evaluation and/or modeling may be required to identify critical source areas on a finer scale.

Another problem with using only a small set of water quality samples to determine critical areas is that some sources are by default removed from consideration. For example, the role of streambanks and stream channels in delivering sediment and sediment-bound pollutants such as P is often only partially understood at the beginning of watershed projects. The Sycamore Creek (MI) NNMP project, for example, focused on no-till and continuous cover to reduce sediment loads, but later concluded that the stream channel stabilization implemented in one subwatershed must have been at least as important as no-till in reducing suspended solids loads (Suppnick 1999). Solutions to sedimentation problems in Lake Pittsfield (IL) progressed from an initial emphasis on no-till, terraces, and waterways (1979-1985), to numerous water and sediment control basins and a single large sedimentation basin (1992-1996), and then to stream restoration using stone weirs and streambank vegetation (1998) when it was learned that massive bank erosion was increasing sediment yield (Roseboom et al. 1999). See section 4.4.5 for a detailed example of using biological monitoring in the Lake Allatoona/Upper Etowah River (GA) watershed.

7.5.2.5 Additional Approaches

In most cases, projects in the planning phase have limited information with which to perform statistical analyses, particularly advanced procedures. Where such data exist, however, multivariate statistical procedures such as factor analysis, principal component analysis, canonical correlation analysis, and cluster and discriminant analysis can be used to define (and perhaps subsequently adjust for) complex relationships among variables such as precipitation, flow, season, land use, or agricultural activities that influence NPS problems. Spatial and temporal patterns can be revealed with these techniques. Scatterplots of ordination scores can be a useful method to summarize multivariate datasets and visualize spatial and temporal patterns.

Ordination techniques can also be powerful during the EDA phase when looking for patterns and structure in the data. The upper Grande Ronde basin project, for example, used correlation and canonical correspondence analysis to determine which environmental variables are largely responsible for differences in fish assemblages between reference and impaired sites (Drake 1999). Figure 7-20 shows a correspondence analysis plot showing intermediate/impaired sites and reference sites ordinating on the left and right side of the origin (Drake 1999). Scatterplots such as Figure 7-20 can be a useful way to summarize multivariate datasets and visualize these spatial and temporal patterns. With such variables identified, the next step was applying principal component analysis to determine if these variables could be used to track stream improvements over time. These statistical procedures are discussed briefly below. The reader is referred to statistics textbooks and other resources for additional information. Further, it is recommended that these procedures are performed by or in consultation with a trained statistician.



DARKL; Dark Canyon Creek – lower site
DARKU; Dark Canyon Creek– Upper site
LIML; Limber Jim Creek– Lower site
LIMU; Limber Jim Creek– Upper site
LOOK; Lookout Creek
MCCL1 & 2; McCoy Creek – Lower site 1 & 2
MCCM; McCoy Creek – Middle site
MCCR; McCoy Creek – Restored reach
MEADL; Meadow Creek – Lower site
MEADS; Meadow Creek at Starkey

Figure 7-20. Correspondence analysis biplot of Grande Ronde fish data (Drake, 1999)

Principal component analysis (PCA) is a multivariate technique for examining linear relationships among several quantitative variables, particularly when the variables are correlated to each other. PCA can be used to determine the relative importance of each independent variable and determine the relationship among several variables. Given a data set with p numeric variables, p principal components or factors can be computed. Each principal component (or factor) is a synthesized variable that is a linear combination of the original variables (SAS Institute 1985). The first principal component explains the most variance in the original data, while the second principal component is uncorrelated with (i.e., orthogonal to or statistically independent from) the first principal component and explains the next greatest proportion of the remaining variance. This process is continued until there are p statistically independent principal components that explain as much of the variance as possible. The results of PCA can often be enhanced through factor analysis, which is a procedure that can be used to identify a small number of factors that explain the relationships among the original variables. One important aspect of factor analysis is the ability to transform the factors (i.e., reconfigure the linear combinations of original variables) from PCA so that they make more sense scientifically. The SAS procedures PROC PRINCOMP and PROC FACTOR can be used for these analyses (SAS Institute 2010).

Principal component analyses and factor analysis can be used in regression analysis to reduce the number of variables or degree of freedoms (d.f.) by using a subset of the principal components (factors) that explain the majority of the variance of the data set instead of using all of the original variables. This essentially reduces the degrees of freedom used, but incorporates most of the information from each of the explanatory variables, hence increasing the validity and power of the regression analysis. Using PCA to incorporate many explanatory variables into a regression model is superior to other techniques that arbitrarily drop explanatory (X) variables; those may incorrectly drop the more important variables due to multicollinearity between the X's. In principle, PCA and factor analysis could be beneficial to projects in a number of other ways, including helping investigators focus problem assessments on the most important indicators and stressors, aiding in the selection of water quality and land use/treatment variables to be used in the monitoring program, and guiding BMPs toward the most important pollutant sources.

Canonical correlation analysis (CCA[8]) is a technique for analyzing the relationship between two sets of multiple variables (e.g., a set of nutrient variables and a set of biomass-related variables). This multivariate approach examines said relationship "by finding a small number of linear combinations from each set of variables that have the highest possible between-set correlations" (SAS Institute 1985). These linear combinations of variables from each set are synthetic variables called 'canonical variables' and the coefficients of the linear combinations (which are similar to Pearson r) are referred to as the 'canonical weights' (SAS Institute 1985). The first canonical correlation is the correlation between the canonical variables from each set that maximizes the correlation value in accounting for as much as possible of the variance in the variable sets. The second canonical correlation is between a second set of canonical variables, is uncorrelated with the first canonical variables, and produces the second highest correlation coefficient. Additional correlations are established until all variance is explained or the maximum number of canonical correlations has been used (i.e., the number of variables in the smaller set). As such, the canonical variables are similar to principal components in summarizing total variation (SAS Institute 1985).

In simple terms, CCA can be used in problem assessment to look for relationships between sets of grouped variables to help better understand existing water quality problems or the relationships between land use/management variables (e.g., imperviousness, acreage receiving manure) and pollution variables (e.g., discharge, pollutant concentrations) to help guide decisions on BMP selection and placement. There are several output statistics (e.g., significance, correlations, coefficients) in CCA, and the reader is referred to statistical textbooks and other sources for additional details. It should be noted, however, that while many correlations may be output from a specific analysis, only the strongest correlations should be considered for interpretation.

Discriminant analysis is used to assess relationships between a categorical (grouping) variable (e.g., presence or absence of a fish species) and multiple quantitative (predictor) variables (e.g., pH, temperature, D.O.). The category options (e.g., present or absent) are assigned a priori—normally verification of the a priori grouping is performed during discriminant function analysis. Discriminant analysis can be used to verify the observational groupings defined by each cluster (see section 7.3.8) or other defined grouping based on the values of the quantitative variables. This type of analysis is referred to as `classificatory discriminant analysis' and is probably the most common application of discriminant analysis in water quality research. The SAS procedures DISCRIM (parametric) and NEIGHBOR (nonparametric) can be used to perform classificatory discriminant analyses (SAS Institute 1985).

Discriminant analyses can also be used to define a subset of quantitative variables that best describes the differences among the groups; see, for example, the SAS procedure STEPDISC (SAS Institute 1985). Canonical discriminant analysis is equivalent to canonical analysis described above except that a set of quantitative variables is related to a set of classification variables (SAS Institute 1985). Principal

[8] Canonical correspondence analysis is also often abbreviated as CCA.

component analysis is used as an intermediate step in the calculation of the canonical variables. The SAS procedure CANDISC can be used to perform canonical discriminant analyses (SAS Institute 1985).

Cluster and discriminant analyses can be used to understand and adjust for relationships among water variables. For example, spatial heterogeneity and homogeneity can be revealed. This may be necessary to study the transport of a pollutant in a system or to remove the spatial component in order to detect changes over time.

In many cases, watershed projects use simulation models to help with problem assessment and planning. Water quality models that include land use/land treatment and are calibrated using water quality data from the watershed or similar watershed(s) can also assist with identification of critical pollutant sources. The reader is referred to USEPA's watershed project planning guide (USEPA 2008) and TMDL modeling website for additional information on water quality models.

7.6 Data Analysis for Project Planning

Existing data or data collected specifically in support of a developing watershed project may play important roles in project planning, including determination of land treatment needs and design of a water quality monitoring program. These and other aspects of watershed planning are addressed in detail in *Handbook for Developing Watershed Plans to Restore and Protect Our Waters* (USEPA 2008).

7.6.1 Estimation and Hypothesis Testing

Project planning – including setting clear project goals – should result in the articulation of hypotheses that can be tested using appropriate statistical tests. The hypothesis must be stated in quantitative terms that can be adequately addressed by statistical analyses and must be directly related to the stated water quality monitoring goals.

The *null hypothesis* (H_o) is a specific hypothesis about a population that is being tested by analyzing the collected sample data. In water quality studies, the null hypothesis is generally a statement of no change, no trend over time or space, or no relationship(s). In contrast, the *alternative hypothesis* (H_a or H_1) is generally the opposite of the null, e.g., a statistically significant change, a trend over time or space, a relationship between 2 or more variables.

The general approach to hypothesis testing is to:

1. State the null and alternative hypotheses. For example:
   - H_o – There is no statistically significant trend over 10 years in TP at the subwatershed stream outlet
   - H_a – There is a statistically significant trend over 10 years in TP at the subwatershed stream outlet
2. Determine a parameter (e.g., mean, median, slope/trend over time) that would provide a point estimate to test if the sample data follow a distribution that would be expected if the null hypothesis was true, or more importantly, to test if there is evidence that the data come from an alternative population.
3. Design a sampling plan that would collect data to test if there is statistical evidence to reject the null hypothesis and accept the alternative hypothesis.

4. Analyze the sample data to calculate the sample point estimate and its confidence interval based upon the collected data variability.
5. Compare the confidence interval to the point estimate under the null hypothesis to determine if there is statistical evidence to reject the null and accept the alternative hypothesis (e.g., statistical evidence that a trend has occurred over time).

It should be noted that if the null hypothesis is not rejected, it is inappropriate to state that the null hypothesis is accepted. Instead, failure to reject the null or failure to detect significant differences or trends is the proper way to state such results. Failure to reject the null could be due to high sample variability, low sample size, or no real differences or trends. The chance of documenting a true difference or trend with statistical significance is improved by increasing sample frequency and longevity, and by using a monitoring design that will isolate the change/trend, while accounting for some of the high variability in data values observed in natural water quality systems. Effective monitoring designs are described in chapters 2-4.

There are two types of errors in hypothesis testing:

1. Type I: The null hypothesis (H_o) is rejected when H_0 is really true.
2. Type II: The null hypothesis (H_o) is not rejected when H_0 is really false.

The probability of making a Type I error is equal to the significance level (α). The probability of a Type II error is β. The power of a test (1- β) is the probability of correctly rejecting Ho when Ho is false. While the significance level is often taken for granted to be 0.05, a different value might be more appropriate for some NPS studies.

7.6.2 Determine Pollutant Reductions Needed

To set goals for a watershed project, it is important to estimate the pollutant reduction required to meet water quality objectives, usually to meet WQS. There are several approaches to developing such estimates:

- **Mass balance/TMDL**. In a TMDL setting, a load reduction goal is established based on a mass balance approach. Monitoring data are used to estimate the pollutant load a waterbody can receive while complying with WQS. The pollutant load reduction goal for a watershed project becomes the difference between the current load and the TMDL which is defined by:

$$TMDL = WLA + LA + MOS$$

Where WLA is the Waste Load Allocation (the allowable point source load);

LA is the Load Allocation (the allowable nonpoint source load); and

MOS is the Margin of Safety to account for uncertainty in the other estimates.

Note that the LA term (NPS load) is often estimated by difference and is not subdivided by source type. The pollutant load reduction goal for a watershed project focused on agricultural sources, for example, will not necessarily address the full difference between current load and LA because there may be other significant nonpoint sources in the watershed such as urban and residential nonpoint sources. TMDLs are frequently based on modeling analysis, but also use available water quality data to the extent possible.

Detailed information on TMDL analysis is available through USEPA (2013). See Case Study 5 for an illustration of how water quality data can be used in the development of a watershed-scale mass balance. The accuracy of this approach, however, depends on the quality and representativeness of the data used in the analysis. In Case Study 5, for example, because internal P loading is being computed based on estimates of the other terms, underestimation of external P loading will lead to an equal overestimate of internal P loading, thus confounding interpretation of the effects of alum application. For this and other reasons, the adaptive management approach is a cornerstone of TMDL implementation. As additional data are collected, mass balances should be revisited.

- **Receiving waterbody relationships.** Numerous tools exist to evaluate the impacts of pollutant loads on waterbodies that may be helpful in estimating pollutant load reduction goals. In lakes, for example, there are many analytical procedures and modeling tools to relate phosphorus load to lake eutrophication, including the “Vollenweider models” (Vollenweider 1976, Vollenweider and Kerekes 1982) and BATHTUB (Walker 1999). Such tools may be used to “back-calculate” permissible phosphorus loads to lakes. Other receiving water models may be used for similar purposes in other types of waterbodies, e.g., QUAL2K, CONCEPTS, and WASP. All of these models can employ available monitoring data to both establish model parameter values and to conduct calibration and validation. Additional information on models useful in this kind of analysis can be found in the USEPA TMDL Modeling Toolbox. Many of these models need to be calibrated with water quality collected from the study watershed or similar watershed(s).
- **Load duration curves.** A flow or load duration curve is a cumulative frequency plot of mean daily flows or daily loads at a monitoring station (e.g., a watershed trend station or tributary outlet) over a period of record, with values plotted from their highest value to lowest without regard to chronological order (see section 7.9.3). For each flow or load value, the curve displays the corresponding percent of time (0 to 100) that the value was met or exceeded over the specified period – the flow or load duration interval. Extremely high values are rarely exceeded and have low flow duration interval values; very low values are often exceeded and have high flow duration interval values. An estimate of the pollutant reductions needed is obtained by comparing a load duration curve developed from monitored loading data against a similar curve with loads estimated as the product of monitored flows and the pollutant concentration established in a WQS. Detailed information on the application of load duration curves to pollutant load reduction estimates can be found in *An Approach for Using Load Duration Curves in the Development of TMDLs* (USEPA 2007).

CASE STUDY 5: MASS-BALANCE APPROACH USED FOR ESTIMATING PHOSPHORUS LOADS

Western Ohio

- ✓ Treated a large, shallow lake with aluminum sulfate to reduce internal phosphorus loads
- ✓ Used the mass-balance approach to estimate internal phosphorus loads pre- and post-treatment

Grand Lake St. Marys (GLSM) is located in the Grand Lake St. Marys watershed in western Ohio (Figure CS5-1). GLSM is a large (5,000 ha), man-made, shallow (mean depth: 1.6 m) lake originally constructed as a "feeder reservoir" for the Miami-Erie Canal (Hoorman et al. 2008; ODNR 2013; Tetra Tech, Inc. 2013). Over 90 percent of the watershed is in cropland with associated livestock operations. Cyanobacteria blooms in GLSM result both from external and internal phosphorus loading (Tetra Tech, Inc. 2013).

The lake was treated with aluminum sulfate (alum) in June 2011 (23.6 mg Al/L, 49.6 g/m2) and in April 2012 (21.5 mg Al/L, 45.2 g/m2) to reduce internal phosphorus loads. The combined treatments totaled approximately 70 percent of the recommended treatment for the lake (recommended treatment was 86 mg Al/L, 120 g/m2). Monitoring data from 2012 were compared against monitoring data collected between 2010 and 2011 to analyze the results of the treatments (Tetra Tech, Inc. 2013). While the assessment also included analysis of algal biomass and aluminum in the water column and sediments, this summary focuses on total phosphorus (TP).

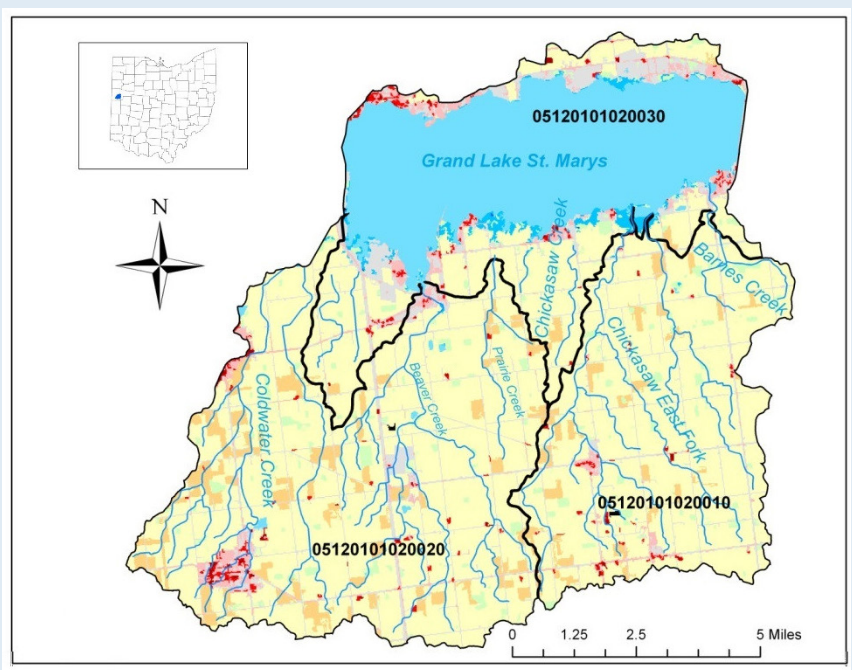


Figure CS5-1. Grand Lake St. Marys watershed

Monitoring and Sampling

Data from eleven water column monitoring sites were used in the assessment, with the five lake sites (shown in Figure CS5-2) sampled every two weeks after alum treatment. Samples at these five sites were always collected at 0.5 m from the surface, while some sampling events also included samples at the bottom of the water column. Samples were analyzed for TP, soluble reactive phosphorus, alkalinity, and chlorophyll. The Ohio Environmental Protection Agency (OEPA) also conducted routine sampling of tributaries, with sample analysis including TP (Tetra Tech, Inc. 2013).

Figure CS5-2. Tributary and lake sampling stations

Mass-Balance Approach

The mass-balance approach helped estimate internal TP loading before and after alum treatment. This approach consisted of five basic steps: (1) Estimating the water budget for GLSM; (2) Developing a basic P budget for the same time period as the water budget (May 2010 through May 2011 prior to any alum addition); (3) Predicting GLSM mean TP concentrations using a P mass balance model for which input values are based on available monitoring data for inflows and outflows; (4) Comparing estimated GLSM mean TP concentrations with measured TP concentrations; and (5) Adjusting the rates of P sedimentation and release of P into the water column (internal loading) to match predicted with measured TP concentrations in GLSM (Tetra Tech, Inc. 2013).

Water budget

A water budget for GLSM was determined at a two-week time step. Change in lake storage was determined using the following equation:

> Change in GLSM lake storage = Inflow (creek and WWTP inputs) + Precipitation – Outflow (water treatment plant withdrawal, groundwater loss, outlets) – Evaporation + Groundwater

The only tributary for which flow data were collected continuously was Chickasaw Creek where USGS has a gaging station (see Figure CS5-2). Wastewater treatment plant (WWTP) flow volumes were obtained from WWTP records and removed from the creek flow volumes so that loads from the four WWTPs in the watershed to GLSM could be calculated separately. Flow volumes from ungaged tributaries and areas draining directly to the lake were estimated by multiplying the adjusted Chickasaw Creek flow (minus WWTP) by the ratio between the other contributing drainage and Chickasaw Creek drainage areas. If creeks were observed to be dry, the flow was assumed to be zero for that period (Tetra Tech, Inc. 2013).

Precipitation records were obtained from a nearby weather station and multiplied by the surface area of the lake to get a volume of direct inflow from precipitation. Monthly mean pan evaporation rates were taken from the Hydrologic Atlas for Ohio (Harstine 1991; after Farnsworth and Thompson 1982).

Groundwater inflow was negligible and the rate for groundwater loss was assumed based on productivity of the underlying aquifer. This rate was adjusted such that there was more loss or recharge during the drier months when there was no outflow. Daily WWTP withdrawals were obtained from plant records. GLSM has two spillways, neither of which is continuously gaged. Lake level data were used to determine when losses would occur over the spillways and two instantaneous flow measurements were used to check estimated flows over the west spillway which is the major outflow. Outflow over the east spillway was assumed to be 10 percent of the west spillway outflow based on communication with local experts (Tetra Tech, Inc. 2013).

Total Phosphorus mass-balance model

A TP mass balance model was developed using the same two-week time step as used for the water budget (Perkins et al. 1997; Tetra Tech, Inc. 2013). Mass was estimated for two-week periods by multiplying the estimated flow volume and mean TP concentration. The principal use of the mass-balance model was to estimate changes in internal P loading for GLSM based on input of measured and estimated values for other terms in the model. Model calibration was based on matching predicted with measured lake TP concentration (Tetra Tech, Inc. 2013).

The following model was used to predict whole lake TP concentrations:

$$dTP/dt = W_{ext} + W_{int} – W_s – W_{out},$$

where W_{ext} is external loading, W_{int} is internal loading, W_s is loss to sediments, and W_{out} is loss through the lake outlet. Predicted whole-lake TP concentrations were compared to observed whole lake mean TP concentrations determined from monitoring at the five lake sites (Figure CS5-2).

Tributary TP concentrations were based on samples collected by OEPA during its routine monitoring. An average of all tributary TP concentrations was used for the ungaged portion of the basin. The TP concentration in direct precipitation was assumed to be 20 µg/L based on an average areal loading rate at Lake Erie from 1996 to 2002 (Dolan and McGunagle 2005). Concentration data for WWTPs were obtained from OEPA where available, and a concentration of 2 mg/L based on an OEPA analysis was assumed otherwise.

Assuming complete mixing, all but one outflow TP concentration was set equal to the whole lake average TP concentration predicted by the model. The actual measured TP concentration of the outflow, 210 µg/L, was used in the model for a single, very large storm event. Sedimentation rates (loss of TP to sediments) and sediment release rates (internal loading) of TP were adjusted in the model to reflect alum applications and to improve the relationship between predicted and measured lake TP concentrations (Tetra Tech, Inc. 2013).

Results

The phosphorus mass balance model was used to determine whole-lake mean TP concentrations based on external loading, internal loading, TP sedimentation, and TP loss through outflows. Whole-lake mean TP concentrations predicted by the 2012 model were compared to observed concentrations as collected and analyzed by OEPA. Sedimentation rates were adjusted to fit the predicted to measured TP concentrations in the lake (Figure CS5-3). With the 2012 model thus calibrated, results were compared with those from 2010 and 2011 to determine if changes in internal TP loading had occurred as a result of alum treatments (Tetra Tech, Inc. 2013).

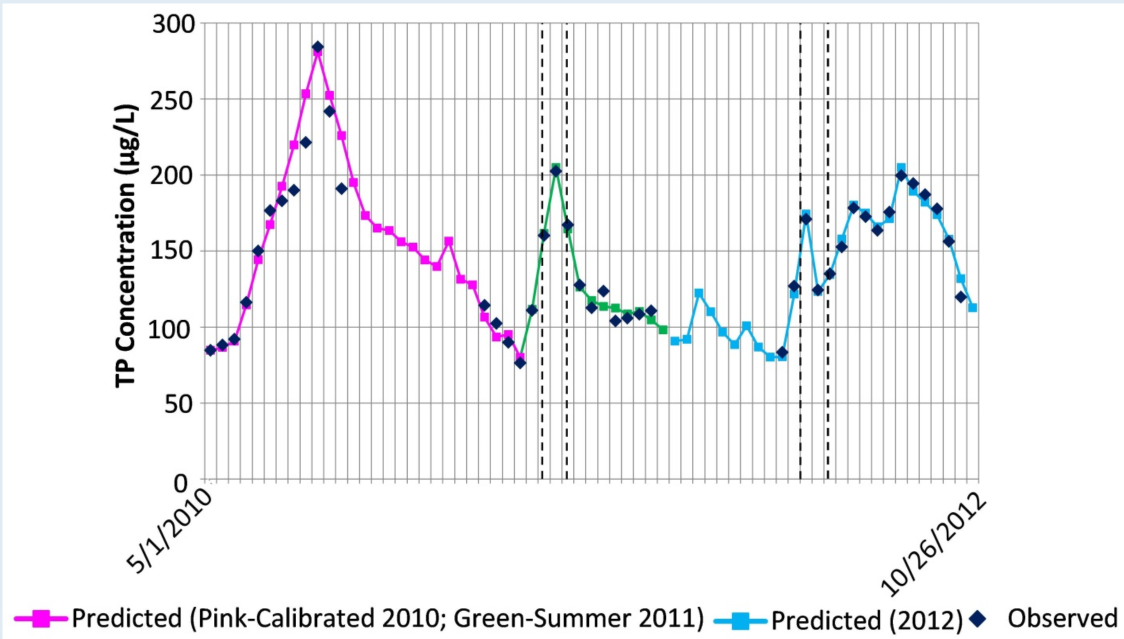


Figure CS5-3. GLSM predicted vs. observed TP concentrations from May 2010 through October 2012 (Adjustments made to internal loading estimates to match predicted November 2011 – October 2012 values to observed TP concentrations)

Table CS5-1 shows that gross summer internal TP loading to GLSM declined steadily from 2010 to 2012. The mass-balance modeling showed that average summer internal loading rate decreased from 4.0 mg/m2 per day before alum treatment to 1.8 mg/m2 per day after the two alum treatments, even though the combined 2011 and 2012 treatments totaled only 70 percent of the recommended treatment for the lake (Tetra Tech, Inc. 2013).

Table CS5-1. Comparison of internal TP loading in GLSM (2010–2012)

	2010	2011	2012
Total Gross Summer Internal TP Load (kg)	26,470	16,487	11,374
Average Summer Internal Loading Rate (SRR) (mg/m2-day)	4.0	2.4	1.8

(Tetra Tech, Inc. 2013)

References

Dolan, D.M. and K.P. McGunagle. 2005. Lake Erie total phosphorus loading analysis and update: 1996-2002. Journal of Great Lakes Research 31 (Suppl. 2):11-22. <http://www.cee.mtu.edu/~nurban/classes/ce5508/2007/Readings/dolan05.pdf>. Accessed November 2013.

Farnsworth, R.K. and E.S. Thompson. 1982. Mean Monthly, Seasonal, and Annual Pan Evaporation for the United States. NOAA Technical Report NWS 34. National Oceanic and Atmospheric Administration, National Weather Service, 82 pp. <http://www.nws.noaa.gov/oh/hdsc/PMP_related_studies/TR34.pdf> Accessed November 2013.

Harstine, L.J. 1991. Hydrologic Atlas for Ohio: Average Annual Precipitation, Temperature, Streamflow, and Water Loss for a 50-Year Period, 1931-1980. Ohio Department of Natural Resource Division of Water, Ground Water Resources Section. Water Inventory Report No. 28.

Hoorman, J., T.Hone, T.Sudman Jr., T.Dirksen, J. Iles, and K.R. Islam. 2008. Agricultural impacts on lake and stream water quality in Grand Lake St. Marys, Western Ohio. Water, Air, and Soil Pollution. 193: 309-322.

ODNR. 2013. Grand Lake St. Marys State Park. Ohio Department of Natural Resources. <http://parks.ohiodnr.gov/grandlakestmarys>. Accessed November 2013.

Perkins, W.W., E.B. Welch, J. Frodge, and T. Hubbard. 1997. A zero degree of freedom total phosphorus model: Application to Lake Sammamish, Washington, Lake and Reserv. Manage. 13:131-141.

Tetra Tech, Inc. 2013. *Preliminary Assessment of Effectiveness of the 2012 Alum Application—Grand Lake St. Marys.* <http://www.lakeimprovement.com/sites/default/files/GLSM%20Alum%20Report%200220 2013.pdf>. Accessed November 2013.

7.6.3 Estimate Land Treatment Needs

A watershed project must set land treatment goals based on estimates of pollutant reductions needed and the BMPs available to accomplish those reductions. Where aquatic habitat improvement is needed, the project's plan must also be based on an assessment of the change in habitat parameters (e.g., water temperature, cobble embeddedness, flow characteristics as well as pollutant loadings) needed to support aquatic life. Various approaches to determining land and in-stream treatment needs to restore and protect aquatic habitat have been documented (e.g., OWEB 1999, Rosgen 1997). Obviously, the BMPs selected must be those capable of addressing the pollutants and sources identified in the planning process. Setting goals for the level and extent of BMP implementation is necessary, but is an inexact science, partially because of the largely voluntary (and hence poorly predictable) nature of land treatment programs, and partially because it is difficult to predict water quality response to BMP implementation at the watershed level. See USEPA (2008) for a comprehensive discussion of watershed project planning.

Where local data on BMP performance exist (e.g., a documented 45 percent reduction in suspended sediment load through a water and sediment control structure or a 25 percent reduction in runoff phosphorus concentration from fields in conservation tillage), they can be applied to estimate pollutant reductions anticipated from different levels of implementation. Where locally-validated data do not exist, there is ample information in published literature (e.g., Simpson and Weammert 2009, USDA-NRCS 2012). Planners should use caution when applying performance data from other studies due to potential local site differences.

It should be noted that published BMP efficiencies do not generally account for interactions in multiple practice systems or address pollutant transport or delivery processes beyond the edge of field or BMP site scale. Modeling, e.g. the Soil Water Assessment Tool (SWAT), may be a better method for estimating treatment needs because some models account for routing of BMP effects through a watershed. Simple pollutant load estimation tools such as USEPA's STEPL (Spreadsheet Tool for Estimating Pollutant Load) can be used to provide general estimates of load reductions achievable via various BMP implementation options, but STEPL, for example, addresses a limited set of pollutants and simulates a limited set of BMPs.

7.6.4 Estimate Minimum Detectable Change

One critical step in watershed project planning is to use the data that have already been collected to evaluate the Minimum Detectable Change (MDC), the smallest monitored change in a pollutant concentration or load over a given period of time required to be considered statistically significant. Understanding of the MDC will assist in planning both land treatment and water quality monitoring design and will support predictions of project success. See section 3.4.2 for details.

The basic concept in the calculation of MDC is simple: variability in water quality measurements is examined to estimate the magnitude of changes in water quality needed to detect significant differences over time. The MDC is a function of pollutant variability, sampling frequency, length of monitoring time, explanatory variables or covariates (e.g., season, meteorological, and hydrologic variables) used in the analyses which 'adjust' or 'explain' some of the variability in the measured data, magnitude and structure of the autocorrelation, and statistical techniques and the significance level used to analyze the data. In general, MDC decreases with an increase in the number of samples and/or duration of sampling in a monitoring program.

The MDC for a system can be estimated from data collected within the same system during the planning or the pre-BMP project phase or from data collected in a similar system, such as an adjacent watershed.

As noted above, MDC is influenced by the statistical trend test selected. For the MDC estimate to be valid, the required assumptions must be met. Independent and identically distributed residuals are requirements for both parametric and nonparametric trend tests. Normality is an additional assumption placed on most parametric trend tests. However, parametric tests for step or linear trends are fairly robust and therefore do not require 'ideally' normal data to provide valid results.

The standard error on the trend estimate, and therefore, the MDC estimate, will be minimized if the form of the expected water quality trend is correctly represented in the statistical trend model. For example, if BMP implementation occurs in a short period of time after a pre-BMP period, a trend model using a step change would be appropriate. MDC in this case is an extension of the Least Significant Difference (LSD) concept (Snedecor and Cochran 1989). If the BMPs are implemented over a longer period of time, a linear or ramp trend would be more appropriate. Calculation of the MDC is discussed in detail in Spooner et al. (2011a) and illustrated in section 3.4.2. The reader is advised to consult that publication to calculate and apply the MDC analysis.

MDC provides an excellent feedback to whether the planned BMPs (type and location, acres served) will result in an amount of change in pollutant concentration or loads that can be statistically documented. Results of the MDC analysis can also be applied to the design of a long-term monitoring program (e.g., sampling frequency, monitoring duration). Decisions about data analysis such as the use of covariates to reduce effective variability and thereby reduce MDC can be made, or MDC calculations can be used to better understand the potential and limitations of an ongoing monitoring effort. Note that the MDC technique is applicable to water quality monitoring data collected under a range of monitoring designs including single fixed stations and paired watersheds. MDC analysis can be performed on datasets that include either pre- and post-implementation data or just limited pre-implementation data that watershed projects have in the planning phase.

7.6.5 Locate Monitoring Stations

The general location of monitoring stations is described for each monitoring design in section 2.4. Analysis of pre-project data, in conjunction with monitoring objectives, can provide insight into optimum location of monitoring stations to be used in watershed project effectiveness evaluation. Section 3.3 provides a discussion on how site characteristics, access, and logistics influence decisions on locating monitoring stations. Spatial analysis of land use and management data, including understanding of relationships between land use and management patterns and water quality (see section 7.5.2.3) can be used to inform monitoring site selection. Inferences on critical source areas (section 7.5.2.4) should also be used to guide station location. Subwatersheds showing very high and very low NO_2+NO_3-N concentrations in Figure 7-19, for example, might be selected for monitoring as treatment and control watersheds, respectively.

7.7 Data Analysis for Assessing Individual BMP Effectiveness

The availability of BMPs that perform a known water quality function is fundamental to NPS watershed projects. Many practices have a long history (e.g., buffers, conservation tillage for erosion control, grassed waterways) and their efficiency in reducing NPS pollutants is well-documented by research, although highly variable depending on site, management, and other factors. The performance of other

BMPs, such as novel practices or practices not common locally, may not be fully understood. In such cases, and in cases where specific assurance that BMPs will perform adequately in local circumstances is required, the effectiveness of individual BMPs may be assessed through monitoring.

Common monitoring designs for assessing BMP effectiveness include:

- Plot studies
- Input/output at the BMP practice scale
- Above/below at the site scale
- Paired watershed at the edge-of-field scale

Data analysis for above/below and paired-watershed BMP monitoring is essentially the same as for these designs at the watershed project level (see section 7.8). This section will focus on discussion of data analysis for plot studies and for BMP input/output studies.

7.7.1 Analysis of Plot Study Data

Controlled, replicated plot or field studies are effective for testing specific practices of undocumented effectiveness or evaluating the effectiveness of a BMP program or system at a farm or watershed scale (USEPA 1997b). To some extent, plots represent microcosms of an area where a full-scale BMP might be applied, where inputs, management, and outputs can be controlled and measured to a degree that would be extremely challenging at full scale. Most importantly, because plots are small (often less than 100 m^2), it is possible to test different levels of treatment and replicate treatments in the same experiment, thus potentially capturing enough variability to have some statistical confidence in the outcome.

As discussed in section 2.4.2.2, there are a variety of plot study designs, including factorial experiments, Latin Squares, and complete and incomplete block designs. Approaches to analyzing data from these various options differ to some degree, but most follow three basic steps:

- Test to see if there are significant differences among the treatments
- Test to find which treatments are significantly different
- Determine the magnitude of differences

Statistical approaches discussed in this section focus on one- and two-factor designs (generally Randomized Complete Block, RCB). Readers should consult statistics textbooks and other resources for information on procedures to analyze data from the more complicated designs such as Latin Squares and incomplete block designs.

Data from simple plot studies are usually analyzed using ANOVA (parametric) or the Kruskal-Wallis test (nonparametric). These procedures allow the determination of significant differences in group means for pollutant concentration or load coming from plots. When a plot study is conducted for a single precipitation/runoff event (either natural or simulated rainfall), the groups tested would be the replicate plots for each type or level of treatment, plus control plots. For a plot study conducted over a series of events, the groups tested could be data from replicate groups within individual events or mean concentration or total load over the entire series of events, depending on the study design. Note that the ANOVA and Kruskal-Wallis procedures only document that one or more group means differ significantly from the other groups. To determine which of the group means are significantly different, use a multiple comparison test such as Tukey's or the Least Significant Difference tests (Snedecor and Cochran 1989, USEPA 1997b). Applications of the Least Significant Difference and Tukey's tests are illustrated in section 4.6.1 (pages 4-55 to 4-56) and 4.6.4 (pages 4-63 to 4-64), respectively, of the 1997 guidance (USEPA 1997b).

The ANOVA procedure can also be used where there is more than one factor or explanatory variable (e.g., plot, slope), whereas the Kruskal-Wallis test handles only one factor. The Friedman nonparametric test is recommended for more than one factor. Application of these tests is described and illustrated in section 4.6 (pages 4-52 to 4-64) of the 1997 guidance (USEPA 1997b).

One-factor comparisons using ANOVA assume random samples, independent observations, and normal distributions for each group, as well as the same variance across groups. Group sample sizes can differ, however. An illustrative example application of the Kruskal-Wallis test for one-factor comparisons is included in the 1997 guidance (USEPA 1997b), pages 4-56 to 4-58.

Two-factor comparisons using ANOVA depend on whether the factors interact. An example of an interaction is the relationship between crop yield and precipitation, both of which can independently influence soil nitrate levels; greater yields remove more nitrate from the soil profile and greater precipitation moves more nitrate through the soil profile. Yield, however, is also influenced by precipitation (e.g., drought or excessively wet soil conditions), so there is an interaction between the two factors. The plot study analysis from Vermont (see Example 7.7-1) illustrates consideration of interactions.

Both the scope of inferences that can be made and the F statistic calculation differ for fixed effect models (e.g., rainfall simulation studies in which rainfall rates are not randomly selected) versus models using randomly selected or combinations of randomly selected and fixed factors. Readers are recommended to section 4.6.2 (pages 4-58 to 4-61) of the 1997 guidance (USEPA 1997b) for an illustrative example and a discussion of these and other important considerations when applying ANOVA to two-factor comparisons. If the data are log-transformed prior to ANOVA, the treatment effects are then interpreted as multiplicative (rather than additive) in the original units. An alternative approach is to rank-transform the data prior to ANOVA, resulting in a comparison of the medians of the data in the original units (see pages 4-61 of the 1997 guidance for details).

Once a statistically significant difference has been demonstrated and the different group means have been identified, it is possible to explore the magnitude of such differences. Methods for two random samples, two paired samples, or a single sample versus a reference (e.g., criterion for a WQS) are described in section 4.5.3 (pages 4-51 to 4-52) of the 1997 guidance. It is important to take the extra step of determining confidence intervals for difference estimates.

In addition to using statistical tests to document differences among treatment groups, plot data can be evaluated by direct comparison of event mean concentration (EMC) or event load (or areal load) among treatments. For plot studies evaluating practice performance over a series of events, a cumulative export plot (where the sum of cumulative mean export from each group is plotted sequentially over the study) will illustrate the behavior of treatment groups in different events. It must be cautioned that data and quantitative inferences about practice performance from plots are usually very difficult to extrapolate to field or watershed scale because physical processes like runoff velocity are not well-represented in very small areas.

Example 7.7-1. Plot Study Analysis: Bacteria Runoff from Manure Application in Vermont

Objective

Evaluate several practical methods for controlling *E. coli* in runoff from manure application sites.

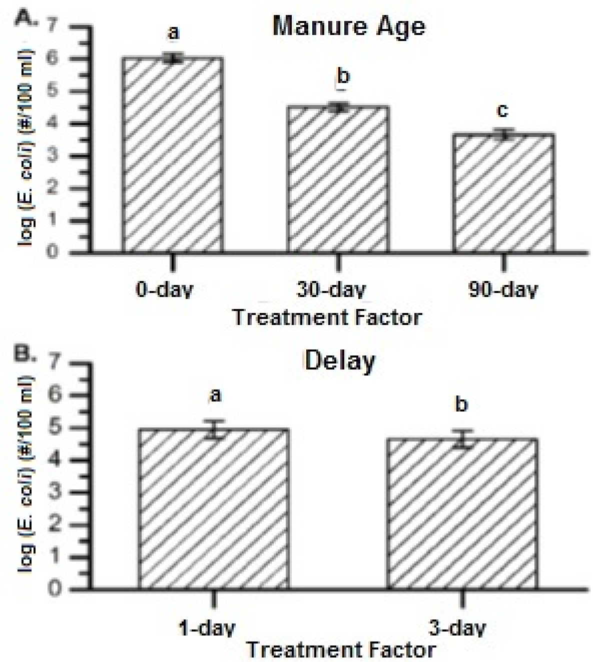
Specific objectives included: (1) determine the effect of manure storage time on *E. coli* losses in runoff from hay and corn land receiving liquid dairy manure; (2) determine the effect of manure incorporation on *E. coli* losses from corn land receiving manure; (3) determine the effect of vegetation height on *E. coli* losses in runoff from hay land; and (4) determine the effect of delay between manure application and rainfall on *E. coli* losses in runoff from hay land and corn land.

Monitoring Design

Two runoff experiments were conducted at separate hay land and corn land sites. For each experiment, 40 1.5- by 3-m plots were created, representing a factorial design of 3 replicates for each treatment combination, 3 manure ages, 2 vegetation heights (for hay) or incorporated/unincorporated (for corn), 2 delay to rain durations, resulting in 3 x 3 x 2 x 2 (36) treatments, plus three control plots (no manure applied), and one extra plot reserved as a backup. Specific treatments were assigned to plots randomly. A rainfall simulator was used to generate runoff from the test plots by continuously and uniformly applying water at an intensity resembling natural rainfall. For each experiment, the first hour or first 19 L of runoff was collected from each plot.

ANOVA table for hay land runoff experiment. Results show significant manure age, delay to rain, and two interactions

Analysis of Variance					
Source	df	Sum of Squares	Mean of Squares	*F* Ratio	*P*
Model	7	34.8385	4.9769	37.135	<0.001
Error	26	3.4846	0.1340		
Total	33	38.32311			
Effects Tests					
Source	df	Sum of Squares		*F* Ratio	*P*
Manure Age	2	31.1188		116.096	<0.001
Vegetation Height	1	0.0673		0.502	0.485
Delay to Rain	1	0.602		4.494	0.044
Manure Ag x Vegetation Height	2	0.7427		2.771	0.081
Vegetation Height x Delay to Rain	1	1.2076		9.011	0.006



Levels of *E. coli* in hay land plot runoff by two treatment factors. Error bars represent ±1 standard deviation; bars labeled with different letter(s) differ significantly ($P \leq 0.1$).

Data Analysis

Statistical analysis of *E. coli* data was conducted on $\log_{10}$ transformed data to satisfy the assumptions of normality and equal variances. All statistical tests were performed using JMP software at an α of 0.1. The effect of treatment on levels of *E. coli* in runoff was evaluated by multi-factor analysis of variance (ANOVA). After an initial pass that included all treatment factors and all possible interactions, nonsignificant ($P > 0.1$) interactions were removed from the model and a final reduced-model ANOVA was conducted. Interpretations of treatment effects were based on the reduced model.

Source: Meals and Braun 2006

7.7.2 Analysis of BMP Input/Output Data

For some BMPs, such as agricultural water and sediment control basins or stormwater treatment devices, it is possible to assess practice effectiveness by directly monitoring input and output pollutant concentration and load. In either an agricultural or an urban setting, inflow and outflow variables such as flow volume, peak flow, EMC, or pollutant loads, are measured and the effectiveness of the BMP is calculated by comparing input vs. output.

Paired input and output data can be compared by testing for significant differences in group means using the parametric paired Student's t or the nonparametric Wilcoxon Rank Sum test. Comparison of random observations from two samples (e.g., input and output from a large constructed wetland for which it is not possible to collect paired samples due to uncertain or variable flow pathways or time of travel) can also be made with a t-test if equal variance is confirmed (e.g., F test); the Mann-Whitney test is the nonparametric alternative in this case. These tests are described and illustrated in detail in chapter 4 (pages 4-34 to 4-52) of the 1997 guidance (USEPA 1997b).

Once a statistically significant difference is confirmed, BMP efficiency can be reported in a number of ways, including:

- Efficiency ratio (percent reduction in flow, EMC, or load),
- Summation of loads (percent reduction in sum of all monitored loads)
- Regression of loads (reduction efficiency is expressed as the slope of a regression line for input load vs. output load)
- Efficiency of individual storm load reductions across all monitored events
- Percent removal relative to a water quality criterion

All of these methods are described and illustrated by Geosyntec and WWE (2009). It is recommended that more than one method is used wherever possible because the results may differ. For example, results from the summation of loads and efficiency ratio (e.g., EMC) methods may not agree because of differences in how the water budgets are represented (Erickson et al. 2010b).

The EMC is the total event load divided by the total runoff volume. It should be noted that, for large practices such as some constructed wetlands, the influent EMC (EMC_I) must be adjusted to account for rain that falls directly onto the practice (Erickson et al. 2010a). Long-term performance can be determined by calculating the average EMCs (AvgEMC) for both influent (input or $AvgEMC_I$) and effluent (output or $AvgEMC_O$) and using these values to calculate the percent reduction in concentration (Erickson et al. 2010b). The simple equation becomes:

$$Long-Term\ Efficiency = 100 \times \left(\frac{AvgEMC_I - AvgEMC_O}{AvgEMC_I}\right)$$

An alternative approach that can add statistical power is to pair the input and output EMCs for each storm and calculate the average of the differences as an estimate of pollutant reduction efficiency. A paired t-test can then be used to determine both the statistical significance of and confidence interval for the reduction. See section 4.2.1 (pages 4-11 to 4-14) of the 1997 guidance (USEPA 1997b) for additional information and an illustrative example of EMC calculations.

The percent reduction in the sum of all monitored loads is calculated using the summed loads for both the input (L_I) and output (L_O):

$$Percent\ Reduction = 100 \times \frac{(L_I - L_O)}{L_I}$$

Similar to the alternative proposed for EMCs, the average differences between paired input and output loads can also be used as an estimate of pollutant reduction efficiency.

Erickson et al. (2010b) illustrate a method for determining the uncertainty of long-term performance estimates that are based on either the EMC or summation of load method they describe. Required input is the number of storm events, the standard deviation of the performance data, and a Student's t value.

Using data from Erickson et al. (2010b), Figure 7-21 illustrates regression of effluent against influent event loads. It should be noted that in this example the y-intercept was not constrained to the origin as recommended[9] by Geosyntec and WWE (2009). The slope of the line indicates that effluent concentration is 37 percent of influent concentration above the baseline level (y intercept) of 0.01 kg TP. In other words, the BMP reduces the load by 63 percent (100-37), a number that agrees well with the 57.5 percent removal rate calculated by summation of loads (Erickson et al. 2010b). Regression analysis is illustrated and described at [CADDIS Volume 4: Data Analysis](#).

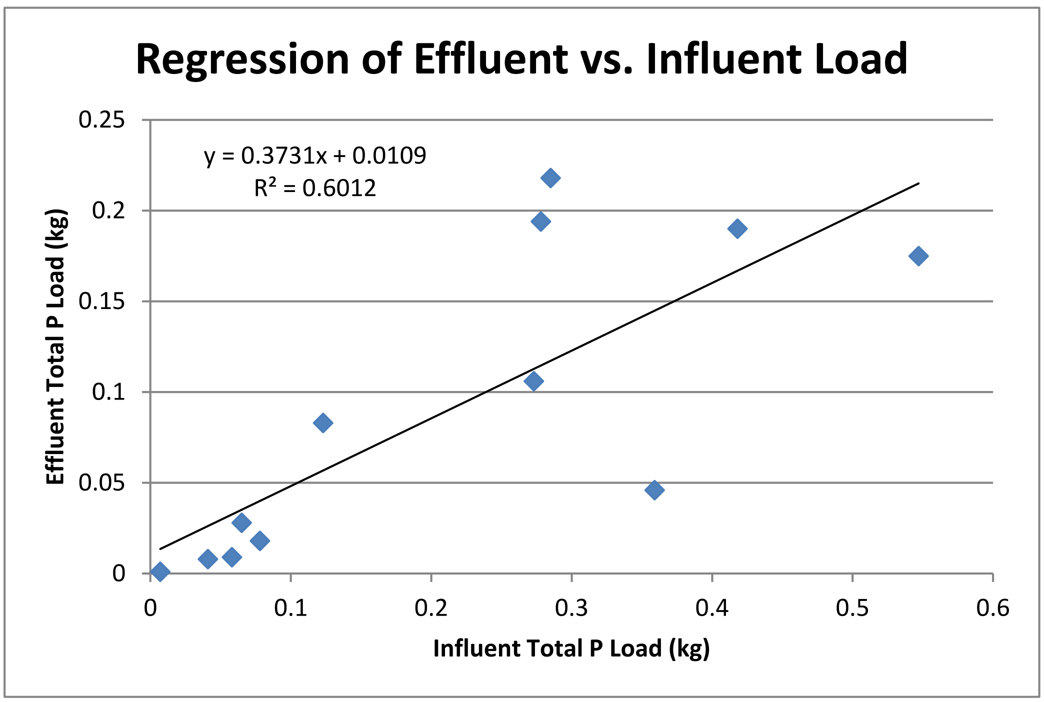


Figure 7-21. Regression of output versus input load (data from Erickson et al. 2010b)

[9] While specified in the definition of the regression of loads method, Geosyntec and WWE (2009) includes a comment suggesting that such a constraint "is questionable and in some cases could significantly misrepresent the data."

BMP efficiency evaluated by input/output monitoring is frequently reported as simply percent removal of a pollutant. In most cases, this is an inadequate basis for assessing BMP performance. Percent removal is primarily a function of input quality, and BMPs with a high apparent removal percentage may still have unacceptably high concentrations or loads in their output. Some BMPs with long retention times (e.g., constructed wetlands) show long-term performance that is not evident in comparing paired input-output samples because material from one event is not discharged until a subsequent event (i.e., the samples are not paired or matched). Finally, a simple percent removal calculation can be dominated by outliers that distort an average performance indicator.

Constructing an Effluent Probability Plot

The cumulative distribution function for the EMCs for the outflows and inflows can be created from the following steps:

- Calculate the EMC for each storm's outflows.
- Rank all EMCs for all storms from smallest to largest.
- Assign a 0 to 1 'probability' to the data based upon their ranked order. For example, if 10 storms were monitored, the ranked values would receive a 'probably ranking' value of 0.1, 0.2, … 1.0 for the lowest to highest EMC values.
- Plot the 'probability ranking' values on the Y-scale and the EMCs on the X-scale. The Y-scale should be plotted on a probability scale. Alternatively, the Y-axis could be expressed as the number of standard deviations (e.g., +/- 3). Because the EMCs are likely to follow a log-normal distribution, the X-axis should be a log scale.
- Repeat the procedure for the inflows and plot on the same graph.

For these and other reasons, USEPA and ASCE have recommended the Effluent Probability Method for evaluating input/output data from a BMP (Geosyntec and WWE 2009). In this procedure, a statistically significant difference between input and output EMC or load is verified (e.g., by Student's t Test). Then, a normal probability plot is constructed of input and output data that allows comparison of BMP performance over the full range of monitored conditions. For example, Figure 7-22 shows an effluent probability plot for chemical oxygen demand (COD) from an urban wet detention pond evaluation The plot shows that COD was poorly removed at low concentrations (<20 mg/L), but that removal increased substantially for higher concentrations.

The Effluent Probability Method is essentially a cumulative distribution function for the EMCs of the inflows and outflows. The cumulative distribution function depicts the probability of values being below a given EMC value or the EMC values that a percentage (e.g., 50 percent) of the data falls above.

The magnitude of the difference in EMC (or loads) from the inflow and outflows can be examined across the range of EMC values. The Kolmogorov–Smirnov test is based on cumulative distribution functions and can be used to determine if the two empirical distributions are significantly different (Snedecor and Cochran 1989).

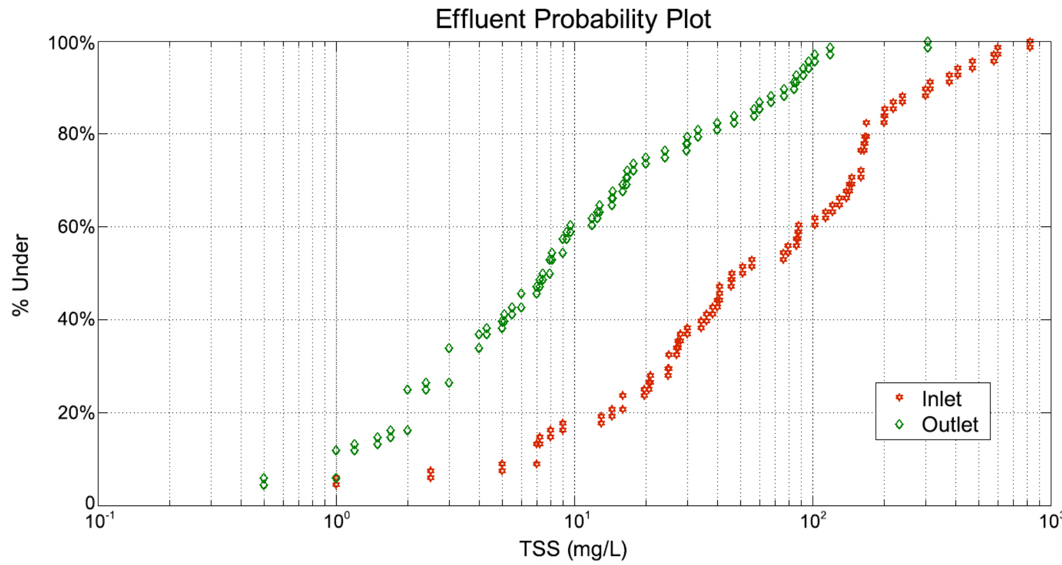


Figure 7-22. Effluent probability plot for input/output monitoring of a wet detention pond

$$Percent\ Removal\ vs\ Criterion = 100 \times \frac{(C_I - C_O)}{(C_I - C_C)}$$

Percent removal relative to a water quality criterion provides an indication of how well a BMP is performing compared to limits or expectations established for the local waterbody. Use of this method is recommended for specific event analysis, but not for a series of events (Geosyntec and WWE 2009). Calculation requires values for the criterion (C_C), input (C_I), and output (C_O), all expressed in the same units (concentration in this case):

For example, in a watershed with a target total N concentration of 0.75 mg/L, storm inlet and outlet concentrations of 3.6 mg/L N and 1.6 mg/L N, respectively, would yield a relative percent removal of 70 percent.

The reader is referred to *Urban Stormwater BMP Performance Monitoring* (Geosyntec and WWE 2009) for additional information on evaluating urban stormwater BMP performance through monitoring.

7.7.3 Analysis of BMP Above/Below Data

As noted earlier, BMP performance can be assessed using an above/below-before/after monitoring design, as long as the added area monitored by the downstream station is either entirely or predominantly influenced by the BMP. In such cases, analysis of monitoring data is done by the same approach as described in section 7.8.2.2. An example of this kind of above/below-before/after analysis of a single BMP can be found in the Otter Creek (WI) NNMP project, which assessed the effects of barnyard runoff control (see Example 7.7-2). This example illustrates application of the Hodges-Lehmann estimator described in section 4.5.3 of the 1997 guidance (USEPA 1997b).

Example 7.7-2. Above/Below-Before/After Analysis: Barnyard Runoff BMPs in Wisconsin

Monitoring Design

Sampling stations upstream and downstream of two investigated dairy barnyards were established in 1994/1995. At the upstream sampling stations, stream stage and precipitation were continuously monitored, and discrete water samples were collected automatically; at the downstream stations, only water quality samples were collected. Over the course of the study, 11 – 15 storm runoff periods were sampled at each of the sites. Continuous streamflow and instantaneous concentration data were used to estimate pollutant loads for individual storm-runoff periods.

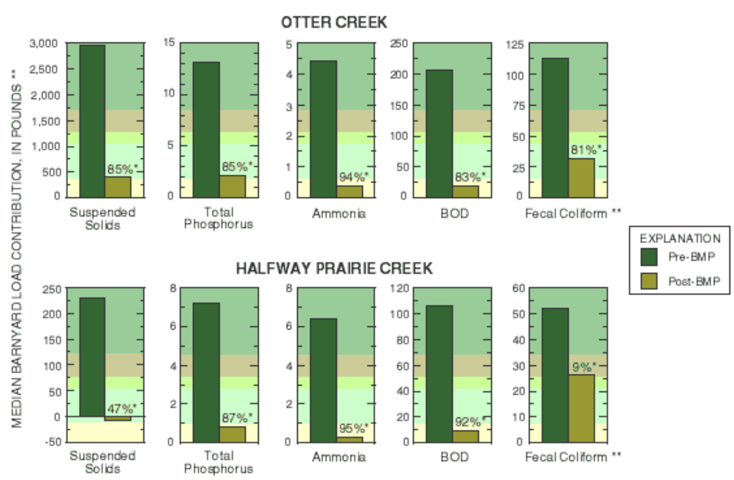
Pre-BMP Analysis

A critical aspect of obtaining useful conclusions for this study was the ability to document that downstream loads were significantly greater than upstream loads before the BMP systems were implemented. Results of t-Tests showed that, for the pre-BMP period at both creeks, downstream loads of total P, ammonia, BOD, and fecal coliform bacteria were significantly greater than upstream loads. At Otter Creek, pre-BMP downstream loads of total suspended solids also were significantly greater than those upstream. These significant differences indicated that each barnyard was an important contributor to the instream pollutant loads for the storm-runoff periods monitored.

Effects of Treatment

The difference between upstream and downstream constituent loads was computed for each pre- and post-BMP storm-runoff period. These differences were considered to be the load contributed by each barnyard. The bar graphs indicate that both barnyard BMP systems have reduced loads in the stream for each constituent. Each bar represents the median of all the differences between upstream and downstream constituent loads for both pre- and post-BMP storm-runoff periods. Although these medians could have been used to determine the percentage reduction achieved by each barnyard BMP system, it was decided that use of the Hodges-Lehmann estimator would be a more accurate approach (Helsel and Hirsch 2002). The Hodges-Lehman estimator is the median of all possible pairwise differences between pre- and post-BMP barnyard loads. This median difference was then divided by the pre-BMP median barnyard load for each constituent. The result was a percentage load reduction for each constituent.

The barnyard BMP system at Otter Creek reduced loads of total suspended solids by 85 percent, total P by 85 percent, ammonia by 94 percent, BOD by 83 percent, and microbial loads of fecal coliform bacteria by 81 percent; the respective loads at Halfway Prairie Creek have been reduced by 47, 87, 95, 92, and 9 percent.



* Percentage reduction is computed by dividing the Hodges-Lehmann estimator for pre- and post-BMP barnyard loads by the pre-BMP median barnyard load.
** Fecal Coliform microbial load in 10^{11} colonies.

Source: Stuntebeck and Bannerman 1998

7.7.4 Analysis of BMP Paired-Watershed Data

Some BMPs – especially agricultural BMPs that involve treatment of an entire field such as conservation tillage, cover crops, or nutrient management – can be evaluated using a paired-watershed design. In this case, monitoring takes place at the edge of field-sized watersheds, wherein one entire monitored field is designated to receive the BMP treatment. Automated samplers are required to collect storm event runoff. In the paired-watershed design, monitoring occurs during a calibration period in which both fields or subwatersheds have identical management. Then, after their pollutant responses to the same rainfall events are correlated, a treatment period occurs in which one of the subwatersheds receives the BMP treatment and the other remains in the 'controlled' management. Analysis of covariance (ANCOVA) is used to analyze the monitoring data from this type of study. See section 7.8.2.1 for details.

7.8 Data Analysis for Assessing Project Effectiveness

7.8.1 Recommended Watershed Monitoring Designs

Assessing the effectiveness of a watershed project where multiple BMPs are implemented in a land treatment program across a broad watershed area is a complex task with many sources of variability and uncertainty. Attributing changes in water quality documented through monitoring to land treatment, rather than to other causes such as drought or extreme weather, is another significant challenge. Monitoring designs (see chapter 2) recommended for assessing watershed project effectiveness are:

- Paired-watershed (link to section 2.4.2.3)
- Above/below-before/after (link to section 2.4.2.6)
- Nested-watershed (link to 2.4.2.3)
- Single watershed trend (link to section 2.4.2.5)

While not generally recommended because of cost and logistical constraints (see section 2.4.2.8), data analysis for multiple-watershed studies is also discussed here. These designs vary in their ability to evaluate watershed project effectiveness while controlling for sources of change other than land treatment; the designs also vary in the appropriate approach to data analysis. The paired-watershed design is generally considered to be the best design for this purpose because it strives for a controlled experiment to evaluate BMP effectiveness at a watershed scale, accounting for year-to-year variability in weather and streamflow through the use of a control watershed. Several common watershed project designs are excluded from the above list because they are not generally capable of reliably documenting water quality change and attributing the change to land treatment. Single watershed before/after and side-by-side watersheds, for example, cannot be recommended for watershed project effectiveness monitoring because they cannot be used directly to separate the effects of the BMPs from those of climate or watershed differences (e.g., soils, slope, land management) which may be the actual causes of the observed differences (see section 3.4). The single watershed before/after design can, however, be useful in comparing pollutant loads over time to determine if TMDL goals have been achieved (see section 7.9).

None of these designs will perform effectively, however, if all the requirements of the design are not met. In some cases, failure to meet a single criterion (e.g., unexpected treatment in the control watershed of a paired design, or changing analytical procedures during a long-term single-station study) may doom the effort.

Each of these designs is discussed in chapter 2; information relevant to data analysis procedures are provided in this section.

7.8.2 Recommended Statistical Approaches

The following sections recommend statistical approaches to analysis of data from recommended watershed monitoring designs. Additional details on specific statistical tests can be found in chapter 4 (Data Analysis) of the 1997 guidance (USEPA 1997b).

7.8.2.1 Paired Watershed

As described in chapter 2, the most effective practical design for evaluating watershed-level BMP effectiveness through monitoring is the paired-watershed design due to the presence of an experimental control for year-to-year hydrologic variability (Clausen and Spooner 1993). The paired-watershed design has been discussed in section 2.4.2.3. The basic design involves two watersheds (a control, where no BMPs are to be implemented, and a treatment watershed where land treatment will be applied) and two periods (a pre-treatment or calibration period, and a treatment period). Analysis of paired data (i.e., frequently collected chemical or physical data) from treatment vs. control areas should show a statistically significant correlation and result in a strong linear regression model (usually using log-transformed data) that changes from the pre-treatment to post-treatment period. In the case of biological monitoring (e.g., sampling twice per year), relationships between treatment and control watersheds should change in a more qualitative manner from pre- to post-treatment periods. For example, treatment and control watersheds may both be of "poor" quality in the pre-treatment (or pre-BMP) period, whereas the treatment watershed improves to "good" quality while the control watershed remains at "poor" quality during the post-treatment period. Additional considerations for paired-watershed designs with more than one treatment watershed are discussed at the end of this section.

Additional Information on ANCOVA

- USEPA. 1997b. *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* Chapter 4;
- Clausen and Spooner. 1993. Paired Watershed Study Design. 841-F-93-009;
- Grabow et al. 1999. Detecting Water Quality Changes Before and After BMP Implementation: Use of SAS for Statistical Analysis; and
- Grabow et al. 1998. Detecting Water Quality Changes Before and After BMP Implementation: Use of a Spreadsheet for Statistical Analysis of Paired Watershed, Upstream/Downstream and Before/After Monitoring Designs.

See section 4.8 of the 1997 guidance (USEPA 1997b) for details and an example including a method for determining if enough calibration data has been collected to warrant advancing to the BMP treatment period. Failure to establish a statistically valid pre-treatment correlation will doom the evaluation design.

7.8.2.1.1 Analysis of Covariance (ANCOVA) Procedure – Paired-Watershed Analysis

The Analysis of Covariance (ANCOVA) procedure is used to analyze data from a paired-watershed study (Clausen and Spooner 1993, Wilm 1949, Clifford et al. 1986, Meals 2001). ANCOVA combines the features of ANOVA with regression (Snedecor and Cochran 1989) and is an appropriate statistical technique to use in analysis of watershed designs that compare pre- and post-BMP periods using treatment and control watershed measurements. When applied to the analysis of paired-watershed data, ANCOVA is used both (a) to compare pre- and post-BMP regression equations between water quality measurement values (e.g., sediment concentration) for the treatment and control watersheds and (b) to test for differences in the average value (e.g., of sediment concentration) for the treatment watershed between the two time periods after adjusting measured values for covariates such as flow. Covariates are added to the analysis to decrease the residual error and give a more precise comparison between covariate-adjusted mean values.

There are three basic steps to performing ANCOVA:

1. Obtain paired observations
2. Select the proper form of linear model
3. Calculate the adjusted means (LS-means) and their confidence intervals

Paired observations could represent observations collected on the same date, the same time period for composite samples, or from the same storm event. Weekly flow-weighted composite samples taken at the outlet of both control and study watersheds would satisfy this requirement.

The second step is to select the proper form of the model. There are two basic statistical models here for paired-watershed studies:

- The change in treatment watershed concentration with change in control watershed concentration (i.e., the slope of the linear relationship between paired samples) remains constant through both the calibration and treatment periods.
- The slope of the relationship changes from calibration to treatment period.

ANCOVA for paired-watershed studies is illustrated by Figure 7-23 where pollutant concentration (or load) pairs are plotted with the treatment basin values on the Y-axis and the control basin values on the X-axis. The slopes of the pollutant concentrations plotted for both periods are tested to determine if they are significantly different (see B in Figure 7-23) or if the same slope can be assumed (see A in Figure 7-23). A change in slope and/or mean value indicates that pollutant concentrations for the treatment watershed exhibited different patterns, or magnitude, after BMPs were applied as compared to the calibration period. For example, in both A and B of Figure 7-23 the same concentration in the control watershed corresponds to a lower concentration in the treatment watershed in the post- (treatment) versus the pre-BMP (calibration) period, indicating beneficial effects from the BMPs. In the case of B, both the mean and the slope are reduced in the treatment period. The adjusted mean concentrations (LS-means) for the calibration and treatment periods are also compared for differences as described above under "ANCOVA Procedure."

The best statistical model for a particular dataset is determined with a test for homogeneity of slopes (i.e., same or different slopes) using the 'full analysis of covariance model' that allows for separate regression lines (i.e., different slopes and intercepts, Figure 7-23B) for the calibration and treatment periods (i.e., the groups) for the regression of the treatment watershed variable (Y) on the control watershed variable (X):

$$Y_{ij} = b_{0i} + \sum_{i=1}^{k} b_{1i}\left(X_{ij}\right) + e_{ij} \qquad \text{("Full statistical model" for different slopes)}$$

Where:

Y_{ij} = the j^{th} observation for Y in period i (e.g., pollutant concentration or load from treatment watershed)
b_{0i} = the intercept (B_0) for period i
b_{1i} = the regression coefficient (B_1) of Y on X for period i
X_{ij} = the jth observation for X in period i (e.g., pollutant concentration or load from control watershed paired with same sample time as Y_{ij})
k = number of time periods (with 'calibration' and 'treatment' periods, k=2)
e_{ij} = the residuals or experimental error for the j^{th} observation for Y in period i. Note: if the data are weekly, biweekly, or monthly, this error series is likely autocorrelated with Autoregressive, Lag 1 or AR(1) and depicted as V_{ij} or V_t. A statistical model that allows for this autocorrelated error structure should be used (e.g., PROC AUTOREG in SAS software (SAS Institute 2016d) or use a correction for the standard error on the test of LS-means (See section 7.3.6)

The F-Test for the homogeneity of slopes is used to see if the best model requires separate slopes for each period or the same (pooled) slope (Clausen and Spooner 1993). The best model will have the lowest residual sum of squares (SSE). The F-statistic for testing the homogeneity of slopes is:

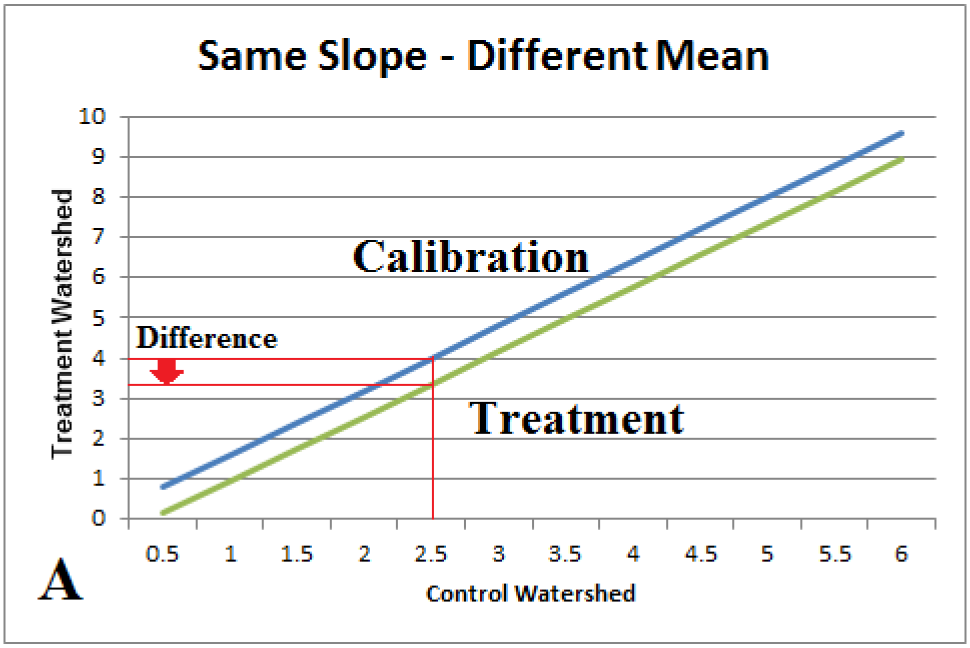
$$\text{F statistic} = \left[\frac{(\text{SSE}_\text{R} - \text{SSE}_\text{F})}{(\text{k} - 1)}\right] / \text{MSE}_\text{F}$$

Where:

SSE_R = Residual sum of squares for the reduced model with a common (pooled) slope (see below)
SSE_F = Residual sum of squares for the full model which allows for separate slopes for the calibration and treatment periods
k = number of groups (calibration + treatment periods = 2 in this case)
MSE_F = Mean square error from the full model

This F-statistic is compared to an F distribution with (k-1) and (N-2k) degrees of freedom (d.f.), where k is the number of groups and N is the total sample size (i.e., the total number of paired samples used in the analysis). See Example 7.8-1 below for examples of how to test if the slopes are different using an 'interaction' term in the statistical software programs.

If there is no evidence for separate slopes, then a "reduced model" with the same slopes assumed for each group (based on pooled data) should be used (see Figure 7-23A). If the interaction term is significant, then the "full model" is the correct model and the significance of the difference between all possible pairs can be obtained (see Figure 7-23B).



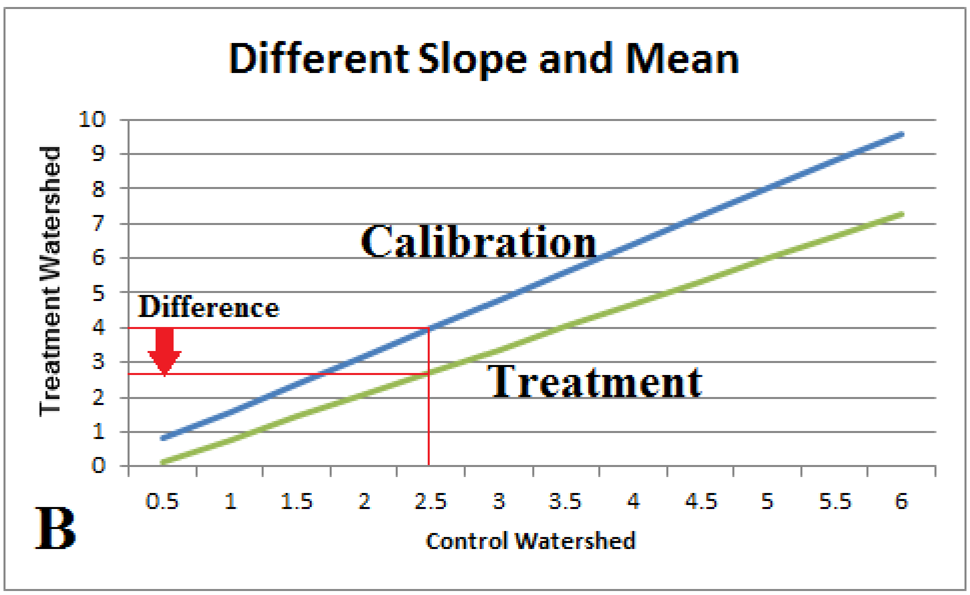


Figure 7-23. Conceptualized regression plots for paired-watershed data. The red line indicates the comparison of the treatment watershed from the calibration vs. treatment periods evaluated at the LSMEANS value of 2.5 (the mean of all sampled values in the control watershed over the entire sampling duration (both treatment and calibration period).

Example 7.8-1. Software Examples for the Statistical Analyses using Analysis of Covariance (ANCOVA) for the Paired-Watershed Study

Statistical software packages may vary in how they address ANCOVA. A few examples are given below. NOTE: We will provide a sample dataset (e.g., Walnut Creek, IA) and results for this example so users can test their own techniques and software.

A. SAS Software, assuming no autocorrelation

The SAS (SAS Institute 2010) program statements that generate a covariance model with unique slopes for each group ("full model", different slopes) are:

```
PROC GLM; CLASS PERIOD;
MODEL Y = X PERIOD PERIOD*X/ SOLUTION;
LSMEANS PERIOD /PDIFF;
```

Where the user inputs the variable names used for their project data for:

Y = Name of variable which contains the treatment watershed values (e.g., concentration/load)

X = Name of variable which contains the control watershed values (e.g., concentration/load)

PERIOD = calibration or treatment period

PERIOD*X = the "interaction" term that allows for different slopes for each PERIOD

The other terms are part of the SAS program software syntax. SOLUTION is optional but generates the regression equation for each PERIOD. The LSMEANS SAS statement generates the LS-means for each PERIOD. The PDIFF option produces significance tests to compare the LS-means for each PERIOD for statistically significant differences.

If there is no evidence for separate slopes (i.e., the PERIOD*X interaction term in the SAS output is not significant), then a "reduced model" with the same slopes assumed for each group (based on pooled data) should be used. If the interaction term is significant, then the "full model" is the correct model and the significance of the difference between all possible pairs can be obtained from the PDIFF option in the LSMEANS statement above.

The SAS program statements that generate a covariance model with common slope but unique intercepts for each period ("reduced model") are:

```
PROC GLM; CLASS PERIOD;
MODEL Y = PERIOD X/ SOLUTION SS1 SS3;
LSMEANS PERIOD /PDIFF;
```

NOTE regarding data setup:

The input data set has columns for each of the variables: Y, X, PERIOD, and DATE. Although DATE is not used in this software example, it is useful to match the values in each row for Y, X, and PERIOD to the correct sample collection date so that the Y and X values are correctly paired up. For the PROC GLM software procedure, PERIOD can be "0" and "1" or "Pre" and "Post" or any other numeric or character value desired. But, be aware that internal to SAS, "0" and "1" values will be generated based upon the alphabetical order – something to consider when interpreting the solutions for the regression line equations for each time period.

Example 7.8-1. Continued

B. SAS Software, data set with autoregressive, lag 1, AR(1) autocorrelation

The SAS (SAS Institute 2010) program statements that generate a covariance model with *unique slopes for each group ("full model", different slopes)* and accommodate an AR(1) error structure are:

```
PROC AUTOREG;
MODEL Y = X PER PER_INTER/NLAG=1 DWPROB;
```

Where the user inputs the variable names used for their project data for:

Y = Name of variable which contains the treatment watershed values (e.g., concentration/load)

X = Name of variable which contains the control watershed values (e.g., concentration/load)

PER = calibration or treatment period ("0" for pre-BMP period values; "1" for post-BMP values)

PER_INTER = the "interaction" term that allows for different slopes for each period. This is a numeric variable whose values are created by multiplying the values of X and PER for each observation

The other terms are part of the SAS program software syntax. NLAG=1 indicates a lag 1 error structure (PROC AUTOREG assumes an autoregressive error structure).

If there is no evidence for separate slopes (i.e., the PER_INTER interaction term in the SAS output is not significant), then a "reduced model" with the same slopes assumed for each group (based on pooled data) should be used. If the interaction term is significant, then the "full model" is the correct model.

The SAS program statements that generate a covariance model *with common slope but unique intercepts for each period ("reduced model")* are:

```
PROC AUTOREG;
MODEL Y = X PER /NLAG=1 DWPROB;
```

NOTE regarding data setup:

The data setup is similar to the PROC GLM software example in A above, except there is no CLASS option in PROC AUTOREG. Numeric input variables needs to be created for all input variables (e.g., 0 and 1 for pre- and post- BMP periods). Since this model includes is a time series error structure, the data must be sorted by date order and have equal spaced time intervals. PROC AUTOREG can correctly handle missing values. In such cases, a data record for the date should be included, but with missing values (indicated by a "." for the missing data input values.

When the reduced model with common slopes is used, the following equation (Snedecor and Cochran (1989) should be used to describe the linear regression for each time period, i, which would have the same slope, but be allowed to have different intercepts:

$$Y_{ij} = b_{0i} + b_1(X_{ij}) + e_{ij} \qquad \text{("Reduced model" for same slopes)}$$

Where:

Y_{ij} = the jth observation for Y in period i (e.g., treatment watershed concentration or load)
b_{0i} = the intercept for period i
b_1 = the regression coefficient of Y on X pooled over all periods
X_{ij} = the jth observation for X in period i (e.g., control watershed concentration or load)
e_{ij} = the residual or experimental error for the jth observation for Y in period i (V_{t} for autocorrelated error series)

Note that this version of the covariance model forces the slope of the regression of Y on X to be the same for each group, but allows the intercept to be unique (i.e., the regression lines representing each group are parallel).

Example 7.8-1. Continued

C. JMP Software, data set with no autocorrelation

Steps: Analyze => Fit Model => Select "Y' Variable, Add variables to the Model Effects ("X" and "PERIOD, highlight PERIOD and X variables in Select Colum and then select 'Cross' in Model Effects to include interaction term=>Run

NOTE regarding data setup:

The input data set has columns for each of the variables: Y, X, PERIOD, and DATE. Although DATE is not used in this software example, it is useful to match the values in each row for Y, X, and PERIOD to the correct sample collection date so that the Y and X values are correctly paired up. For the PROC GLM software procedure, PERIOD can be "0" and "1" or "Pre" and "Post" or any other numeric or character value desired. But, be aware that internal to SAS, "0" and "1" values will be generated based upon the alphabetical order – something to consider when interpreting the solutions for the regression line equations for each time period.

Note: if data has autocorrelated, autoregression, order 1 or AR(1) error series, the standard error on the differences between the LS-means can be adjusted and then the corrected significant differences can be determined by:

$$std.dev._{corrected} = std.dev._{uncorrected}\sqrt{\frac{1+\rho}{1-\rho}}$$

Where ρ = autocorrelation coefficient at lag 1
Std. dev = standard error on the differences of the LS-means

D. MiniTab Software, data set with no autocorrelation

Steps: Stat > ANOVA > General Linear Model. In the responses, model, and random factors dialogue boxes, enter "Y", "X PERIOD X*PERIOD", and "PERIOD", respectively. The user can choose whether to use adjusted or sequential sum of squares under the options button and pairwise comparisons can be chosen from the comparisons button. Pressing OK button runs the general linear model.

Reference: Minitab (2016)

Lastly, calculation of the adjusted means and their confidence intervals can be performed. After the correct model is determined ("Full" or "Reduced" model), then the adjusted LS-means[10] which correct for the bias in X between periods can be calculated. The LS-mean of each period (i.e., calibration and treatment periods in this case) is the period mean for Y adjusted to an overall common value of X. In other words, the LS-means are the calibration and treatment period regression values for the treated watershed evaluated at the mean of all the control watershed values over both time periods (e.g., mean of all the X values). Operationally, inserting the mean of all X values into the regression equations for the calibration and treatment periods will yield the LS-mean values for each period, respectively. An F-test of the adjusted LS-means then determines if there is sufficient evidence to conclude that the adjusted LS-mean for the treatment period is different from the adjusted LS-mean for the calibration period. The SAS program performs this F-test on the "Period" variable in Example 7.8-1.

[10] LS-means (least square means) are used in ANCOVA as a better comparison of average values between periods as compared to arithmetic means. LS-means are estimated values that are evaluated at the average value of the specified covariate(s) such as the control watershed values in the paired-watershed study design.

Caution must be used when interpreting the results for the comparisons of adjusted means when individual slopes are used. When the slopes are not parallel, the comparisons of adjusted means may not be the most meaningful question. One may be more interested in the behavior over the entire range of X. In this case a graphical presentation may be most appropriate.

For samples collected daily, weekly, biweekly, or monthly, autocorrelation may be significant. In these cases, autocorrelation can be addressed by using a software regression program that incorporates the autocorrelation in the error term, for example PROC AUTOREG by SAS (SAS Institute 2016d); see Example 7.8-1.

7.8.2.1.2 Multivariate ANCOVA-Paired Watershed with Explanatory Variables

Note that the above analysis employed a basic univariate ANCOVA model that included only data on the pollutant variable of interest (e.g., concentration or loads) from the control and treatment watersheds. The New York NNPSMP project demonstrated the successful use of a multivariate ANCOVA technique that included hydrologic variables (e.g., instantaneous peak flow rate, event flow volume, and average event flow rate) in the model (Bishop et al. 2005). The project found that including the flow covariates explained 80 to 90 percent of observed variability in annual and seasonal event P loads, an improvement of 16 to 50 percent versus a simpler univariate model. In addition, inclusion of covariates reduced the minimum detectable treatment effect by 11 to 53 percent versus the univariate model, a result that indicates potential cost savings through reduced sample size requirements. It is important to note that the inclusion of additional covariates (i.e., those in addition to the variable of interest in the control watershed) is prefaced upon the assumption that they are not affected by BMP implementation. In this example, testing indicated no influence of BMPs on farm runoff volume, event peak flow, or average event flow.

In the case of a paired-watershed study, explanatory variables (covariates) would be added to the statistical model. The full model which allows for different slopes for each time period and covariate is:

$$\mathrm{Y_{ij}} = \mathrm{b_{0i}} + \sum_{i=1}^{k} \mathrm{b_{1i}}\left(\mathrm{X_{1ij}}\right) + \sum_{c=2}^{d+1} \mathrm{b_{ci}}\left(\mathrm{X_{cij}}\right) + \mathrm{e_{ij}}$$

Where:

Y_{ij} = the jth observation for Y in period i (e.g., pollutant concentration or load from treatment watershed)
b_{0i} = the intercept (b_0) for period i
b_{1i} = the regression coefficient (b_1) of Y on X_1 for period i
b_{ci} = the regression coefficient (b_c) for covariate X_c for period i
k = number of time periods (with ‘calibration’ and ‘treatment’ periods, k=2)
X_{1ij} = the jth observation for X_1 in period i (X_1 is the pollutant concentration or load from control watershed paired with same sample time as Y_{ij})
d = number of explanatory variables in addition to the control watershed variable. For example, if only flow was used as a covariate, d=1 and the explanatory variable for flow would be X_2.
X_{cij} = the jth observation for X_c covariate in period i
e_{ij} = the residuals or experimental error for the jth observation for Y in period i (V_{ij} for autocorrelated error structure)

As discussed above, a test for the homogeneity of slopes (by including interaction terms) would be performed to see if a full or reduced model is the best choice, followed by calculation of adjusted means and their confidence intervals to see if a significant difference exists between the two periods.

While the focus above has been on a basic paired-watershed study design consisting of two watersheds (control and treatment) and two periods (calibration and treatment), ANCOVA is a powerful tool that can also be applied to paired-watershed studies with multiple control and treatment watersheds and more than two periods, as well as to above-below studies that have two or more time periods.

7.8.2.1.3 Multiple Paired Watersheds

Both the Jordan Cove (CT) and Lake Champlain Basin (VT) NNMP projects included three watersheds in their paired-watershed designs. The Jordan Cove project included a previously developed drainage area as a control, and two newly developed drainage areas, one following traditional subdivision requirements and another using low-impact development BMPs (Clausen 2007). The Vermont project employed a three-way paired design including one control watershed and two treatment watersheds receiving similar BMP systems at different intensities (Meals 2001). For both studies, the two treatment watersheds were separately compared versus the control watershed using ANCOVA.

Changes versus the control watershed for the Jordan Cove project were represented by the percent change in flow, concentration, and export (Clausen 2007). These calculations were made by comparing mean predicted values (P) from the calibration regression equations to observed values (O) using the equation:

$$\%Change = \frac{(O - P)}{P} \times 100$$

Meals (2001) performed a series of analyses to examine the results of the Lake Champlain Basin study. Where full ANCOVA models were used, the calibration and treatment period regression lines intersected, suggesting, for example, that TP concentrations in one of the treatment watersheds decreased in the high range, but not in the lower range (Figure 7-24). The importance of this observation is that the higher range is where active runoff conditions occur, indicating that the BMPs may have been performing as expected.

Calculations similar to those performed for the Jordan Cove project were performed to estimate the magnitude of change (i.e., %Change), but two additional analyses were carried out to estimate this change from different perspectives:

- Breakpoint analysis for intersecting or crossed regression lines, and
- Assessment of predicted-without-treatment versus observed-with-treatment.

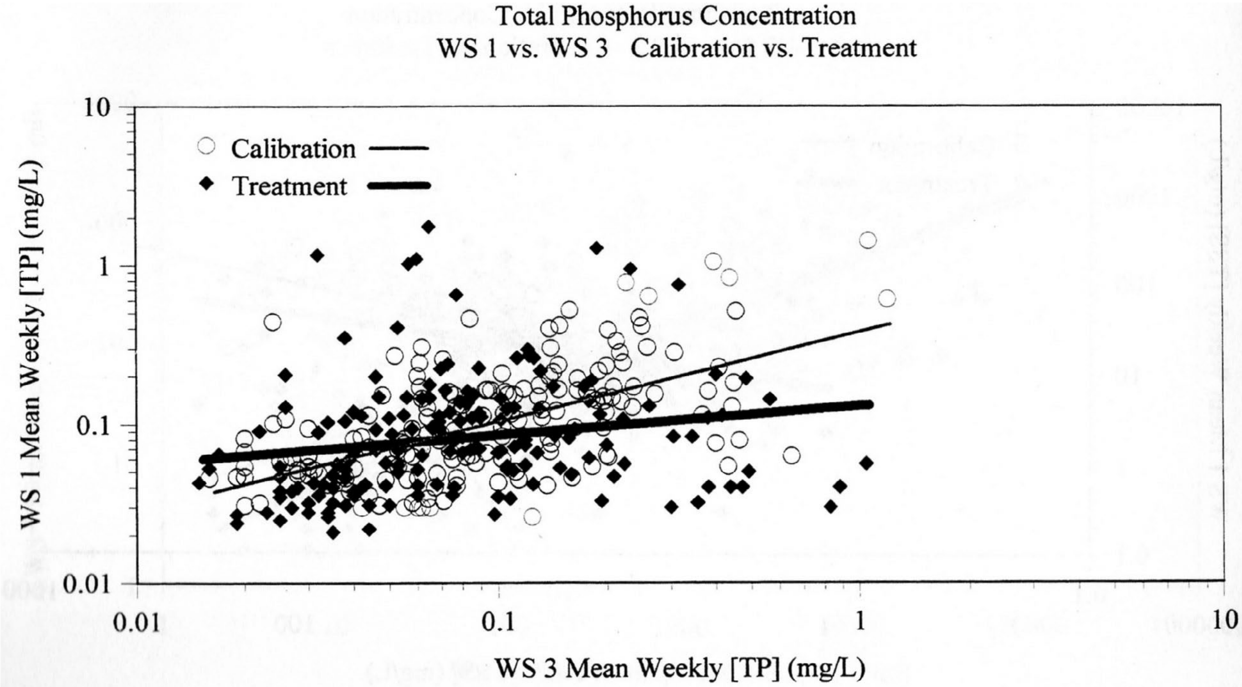


Figure 7-24. Example of intersecting regression lines (Meals 2001)

For the former analysis, the point where the regression lines crossed (the "breakpoint") was used in conjunction with the cumulative frequency of the breakpoint value in the control watershed to derive the proportion of time or conditions at which concentration or load reductions did or did not occur in the treatment watershed (Meals 2001). For example, the breakpoint in Figure 7-24 occurs at 0.055 mg/L in the control watershed (WS 3), a value for which the cumulative frequency for the entire project period was 0.32, or 32 percent. This is interpreted to mean that TP levels in the treatment watershed (WS 1) were not reduced 32 percent of the time when the concentration in the control watershed was less than 0.055 mg/L. Conversely, TP levels were reduced 68 percent of the time when control watershed concentrations exceeded 0.055 mg/L. This compares with an ANCOVA result that TP concentrations were reduced 15 percent in the treatment watershed.

The latter analysis was intended to assess the net treatment response regarding pollutant export over the full range of project conditions (Meals 2001). In this analysis, all weekly values for the treatment period in the control watershed were input to the calibration period regression for each treatment watershed to estimate what the pollutant export would have been for the hydrologic conditions of the treatment period under pre-treatment management, a what-if scenario. In other words, it is an estimate of the difference between measured loads for the treatment period and what those loads would have been if the BMPs had not been implemented.

7.8.2.1.4 Multiple Time Periods within a Paired-Watershed Study

Small watershed projects will generally have a period before BMP implementation, a period during BMP implementation, and a period after BMP implementation. The implementation and post-implementation periods are often lumped into the same period for data analysis, but this can complicate interpretation of results if the BMPs are not fully functional throughout the post-BMP period. Where feasible, it may be most appropriate to separate true implementation, and in some cases maturation of living BMPs, from post-implementation, to establish a better test of BMP or project effectiveness. There is also a very real possibility that BMP implementation will occur in phases, creating the potential for more than two or three periods of interest. For example, in the Waukegan River NNMP project, the state Water Survey designed biotechnical and other practices to resist high velocity runoff while increasing riparian habitat for stream fisheries within the stream channel (White et al. 2011). However, as the project progressed it became clear that insufficient pool depth and the lack of pools and riffles were important impairments yet to be addressed. As a result, pool-and-riffle sequences were later added to the restoration program, creating a two-phase implementation effort. Still, however, project scientists concluded that there is a remaining need to address sewage and stormwater management problems and take steps to increase implementation of alternative conservation practices that infiltrate and treat stormwater. Were the monitoring program to be continued, these could be considered additional BMP implementation phases.

Taken to the extreme, each year could also be considered its own period or group and the groups tested for differences, but this is not recommended[11]. In some cases, BMPs may have different effects depending on the season of the year, so including a seasonal covariate(s) may be appropriate. The New York NNMP project identified four seasons that reflect seasonal variation in both source activities and hydrologic runoff processes (Bishop et al. 2005). ANCOVA was performed separately on both seasonal and full-year datasets. Despite the wide range of possibilities, time periods for the types of projects envisioned by this guidance will largely be drawn from the following set of options:

- pre-BMP or calibration,
- BMP implementation (may be subdivided by growth stage if it involves vegetative BMPs), and
- post-BMP implementation (which may include BMP implementation as well).

Where multiple phases of BMPs are to be implemented, however, there could be a separate pre-BMP implementation and post-BMP implementation for each phase. It is important to identify and plan for these phases at the beginning of the monitoring project. Adjustments may be warranted later, however, because the implementation of BMPs may be more gradual or sporadic than anticipated during the planning phases of a study, and some BMPs, like forested buffers, may take longer than expected to reach critical growth stages.

For example, in a 15-year project monitoring the effectiveness of a riparian forest buffer in an agricultural watershed, it was expected that it would take several years for the planted seedlings to have a measureable influence on water quality (Newbold et al. 2009). To account for this, the calibration period was taken to be the first five years (1992-1996) of monitoring, a period during which the seedlings became established but remained too small to affect stream nutrient concentrations. Regression analysis was used to detect gradual change and one-way ANOVA was performed on the differences between paired samples, with year treated as the main effect.

7.8.2.1.5 Other Statistical Approaches for Paired-Watershed Analyses

Paired watersheds can also be analyzed with other statistical techniques. For example, some authors have used the differences between sample pairs taken at each watershed for each sampling date (Carpenter et al. 1989; Bernstein and Zalinski 1983; MacKenzie et al. 1987; and Palmer and MacKenzie 1985) for input into t-test or intervention analysis. Hornbeck et al. (1970), Hibbert (1969), and Meals (1987) calculated a

---

[11] It is feasible that a 2-year study could include one year each of pre-BMP and post-BMP monitoring, but this would be highly unusual and not, in fact, recommended. A similar situation would be a 3-year study with a pre-BMP, BMP-implementation, and post-BMP year.

linear regression equation relating the observations from the two watersheds for the calibration period. Observations from the treated watershed in the treatment period were compared to predicted values from the calibration period regression. If the deviations exceeded the 95 percent confidence intervals placed about the calibration regression, the treatment was thought to be significant (Hornbeck et al. 1970).

7.8.2.2 Above/Below – Before/After

An above/below-before/after watershed design monitors a water resource (e.g., a stream) above and below the drainage area in which land treatment is applied for multiple years before and after BMP implementation (see section 2.4.2.6). Consistency of sampling regime at both stations over time is essential. Hydrologic explanatory variables (e.g., covariates) such as stream flow must also be monitored to permit correction for changes in these conditions.

7.8.2.2.1 Comparing Means and Differences between Means

Two principal approaches can be taken to statistical analysis of data from this monitoring design. Both approaches are illustrated by the projects in Examples 7.8-2–7.8-5. In the first approach, mean upstream and downstream pollutant concentrations and/or loads can be compared (e.g., with the Student's t or Wilcoxon Rank Sum tests) prior to the application of BMPs to evaluate statistically significant differences between group means. The purpose of this analysis is to confirm and quantify the pre-treatment ("before") pollutant contribution of the untreated downstream area. This analysis is then repeated for the "after" data to document the changes in pollutant contribution of the treated downstream area. Differences between upstream and downstream conditions from the before to the after condition can be evaluated simply by examining the percent reductions in concentration or load or by conducting a group means test of the differences between upstream and downstream concentrations or loads from the before to the after period. A significant decrease in this upstream/downstream difference in the "after" period, for example, would suggest a significant effect of treatment. In addition to quantitative statistical tests, it is also possible to visualize differences between above/below and before/after using comparative boxplots, bar graphs, or other graphical techniques (see section 7.3.2).

A more statistically powerful approach would be to use the paired Student's t-test to test the differences between the downstream and upstream sample values in the pre-BMP period. In the post-BMP period, a Student's t-test can be applied to the average downstream-upstream differences in the pre- vs. post-BMP periods. Other explanatory variables can be added (e.g., stream discharge) by using an ANCOVA statistical approach.

Differences between above and below stations were examined as part of the analyses performed for the Otter Creek (WI) watershed project (Stuntebeck 1995). This project also incorporated innovative sampling procedures to maximize the potential for distinguishing between upstream and downstream water quality, including programming water quality samplers to be activated by precipitation so that time-integrated samples were collected initially before stage-triggered samples were collected. This allowed sampling of barnyard runoff in the stream before stage increased, thereby isolating runoff from sources upstream. It also allowed sampling during small storms where barnyard runoff occurred in the absence of substantial upstream contributions. In addition, investigators collected concurrent samples from both the above and below sites via computer linkage to aid data interpretation. Paired Student's t-tests were used to determine that the pre-BMP average of the differences between downstream and upstream event-mean concentrations was different from zero at the 95 percent confidence level. An MDC analysis revealed that the average downstream post-BMP event-mean concentrations of TP would need to decrease by at least

50 percent for the change to be considered statistically significant at the 95 percent confidence level. In the final analysis, the Hodges-Lehmann estimator was used to determine that the barnyard BMP system at Otter Creek reduced loads of suspended solids by 85 percent, TP by 85 percent, ammonia by 94 percent, BOD by 83 percent, and microbial loads of fecal coliform bacteria by 81 percent (Stuntebeck and Bannerman 1998; See Example 7.7-2). The nonparametric Hodges-Lehmann estimator is the median of all possible pairwise differences between pre- and post-BMP barnyard loads (see section 4.5.3 of the 1997 guidance (USEPA 1997b) for a discussion of the Hodges-Lehmann estimator). This median difference was divided by the pre-BMP median load for each constituent to determine percentage load reductions.

7.8.2.2.2 ANCOVA

A second approach for analysis of the above/below-before/after design involves the application of ANCOVA. The statistical analysis approach is the same as with the paired-watershed study (see section 7.8.2.1) In this case, a significant linear regression relationship for a water quality variable (e.g., weekly mean total P concentration, weekly suspended sediment load) between the upstream and downstream stations is obtained during the "before" period. The upstream station is considered to be the "control" watershed. This regression relationship is then compared to a similar relationship during the "after" period and significant difference between the two regression models indicates the effect of treatment. Note that the analysis can include explanatory variables (e.g., covariates) like precipitation or flow in a multiple regression model that may explain more of the variability in the water quality variable than a simpler model.

Example 7.8-2. Above/Below-Before/After Design - Long Creek, NC NNPSMP

A number of successful projects have used multiple approaches to analyzing their data. For example, data from an above/below-before/after study of livestock exclusion as part of the Long Creek (NC) NNPSMP project were first log-transformed and then analyzed using t-tests, two-way ANOVA, and ANCOVA (Line et al. 2000). While the specific questions addressed by each method differ somewhat, the results all supported the conclusion that livestock exclusion and establishment of riparian vegetation reduced mean weekly loads of TSS, TKN, and TP.

Example 7.8-3. Above/Below-Before/After Design (biological data) - Waukegan River, IL NNPSMP

The Waukegan River (IL) NNPSMP project illustrates the application of the above/below design for biological monitoring. In this project, the South Branch was divided into an upstream untreated reference site designated as station S2 and a severely eroding downstream treated area designated as station S1 (Spooner et al. 2011b). At each location fish, macoinvertebrates, and habitat were sampled during the spring, summer, and fall seasons. Sampling was also conducted at stations N1 and N2 on the North Branch for reference. Qualitative analysis of biological data collected through 2006 indicated that the number of fish species and abundance in the South Branch had improved after the construction of lunkers and rock grade control structures. The IBI rose sharply from a limited aquatic resource into the moderate category after construction. Sites on both the South and North Branches where lunkers and Newbury Weirs were applied averaged higher IBI scores and fish population with more fish species than the untreated control at S2 or the N2 bank armored site from 1996 through 2006.

Example 7.8-4. Above/Below-Before/After Design with Flow as an Explanatory Variable - Pequea and Mill Creek Watershed, PA NNPSMP

A Pennsylvania study of the effects of streambank fencing on surface-water quality, near-stream ground water, and benthic macroinvertebrates employed both a paired-watershed and above/below-before/after design (Galeone et al. 2006). Data for this Section 319 NNMPMS project were collected from 1993 to 2001, with the calibration period from October 1993 through mid-July 1997. Streambank fencing was installed from May 1997 through July 1997. The above/below-before/after design featured two sites above fence installation (T-3 and T-4) and two sites located to show the effects of fencing (T-1 and T-2); T1 and T2 were paired with T3 and T4, respectively, for data analysis. Both low-flow and storm-flow samples were collected and analyzed for nutrients, suspended sediment, and fecal streptococcus (only low-flow samples). Explanatory data collected during the study included precipitation, inorganic and organic nutrient applications, and the number of cows.

Figure 7-25 illustrates the major data preparation steps and statistical procedures used by the project to analyze the chemical/physical data. Low-flow, storm-flow, pre-treatment, and post-treatment data were separated as a preliminary step. Concentrations were flow adjusted using a LOcally WEighted Scatterplot Smoothing (LOWESS) procedure (Helsel and Hirsch 2002). Statistical tests were performed on both original and flow-weighted data to determine if factoring out the variability caused by flow affected the results.

After the above steps were completed, the project applied the nonparametric rank-sum test (see Mann-Whitney test and Wilcoxon Rank Sum test on pages 4-50 of the 1997 guidance, USEPA 1997b) to determine if data for any one site significantly changed from the pre-treatment to the post-treatment period. In addition, the nonparametric Kruskal-Wallis test (see pages 4-56 of the 1997 guidance) was carried out to determine if there were significant differences between any of the sites, considering pre-treatment and post-treatment data separately. Where significant differences were found, the Tukey multiple-comparison test (see Multiple Comparisons on pages 4-63 of the 1997 guidance) was used to identify which sites were significantly different. The nonparametric signed-rank test (see Wilcoxon Signed Ranks test on pages 4-42 of the 1997 guidance) was used to determine if there were significant differences (i.e., not zero) between paired observations (e.g., matched samples from above/below sites). Finally, ANCOVA (see section 4.8 of the 1997 guidance and section 7.8.2.1 for detailed discussions of the ANCOVA procedure) was applied to determine the effects of streambank fencing using a procedure highlighted by Grabow et al. (1999). ANCOVA was performed on concentrations and loads for both low-flow and storm-flow samples. Loads were analyzed in two ways, as actual measured loads and as weighted loads adjusted with a factor determined by dividing the annual mean discharge for each water year by the mean discharge for the entire period for each station.

The procedures used by Galeone et al. (2006) demonstrated improvements relative to control or untreated sites in surface-water quality (nutrients and suspended sediment) during the post-treatment period at T-1, but T-2 showed reductions only in suspended sediment. N species at T-1 were reduced by 18 percent (dissolved nitrate) to 36 percent (dissolved ammonia); yields of total P dropped by 14 percent. Conversely, T-2 had increases in N species of 10 percent (dissolved ammonia) to 43 percent (total ammonia plus organic N), and a 51-percent increase in total P load. The average reduction in suspended-sediment load for the treated sites was about 40 percent. Two factors were evident at T-2 that helped to overshadow any positive effects of fencing on nutrient yields. One was the increased concentration of dissolved P in shallow ground water (also monitored). In addition, cattle excretions at the low-cost, in-stream cattle crossings appeared to increase concentrations of dissolved ammonia plus organic N and dissolved P. See chapter 3 Case Study #1 for a discussion of how the benthic macroinvertebrate data were analyzed.

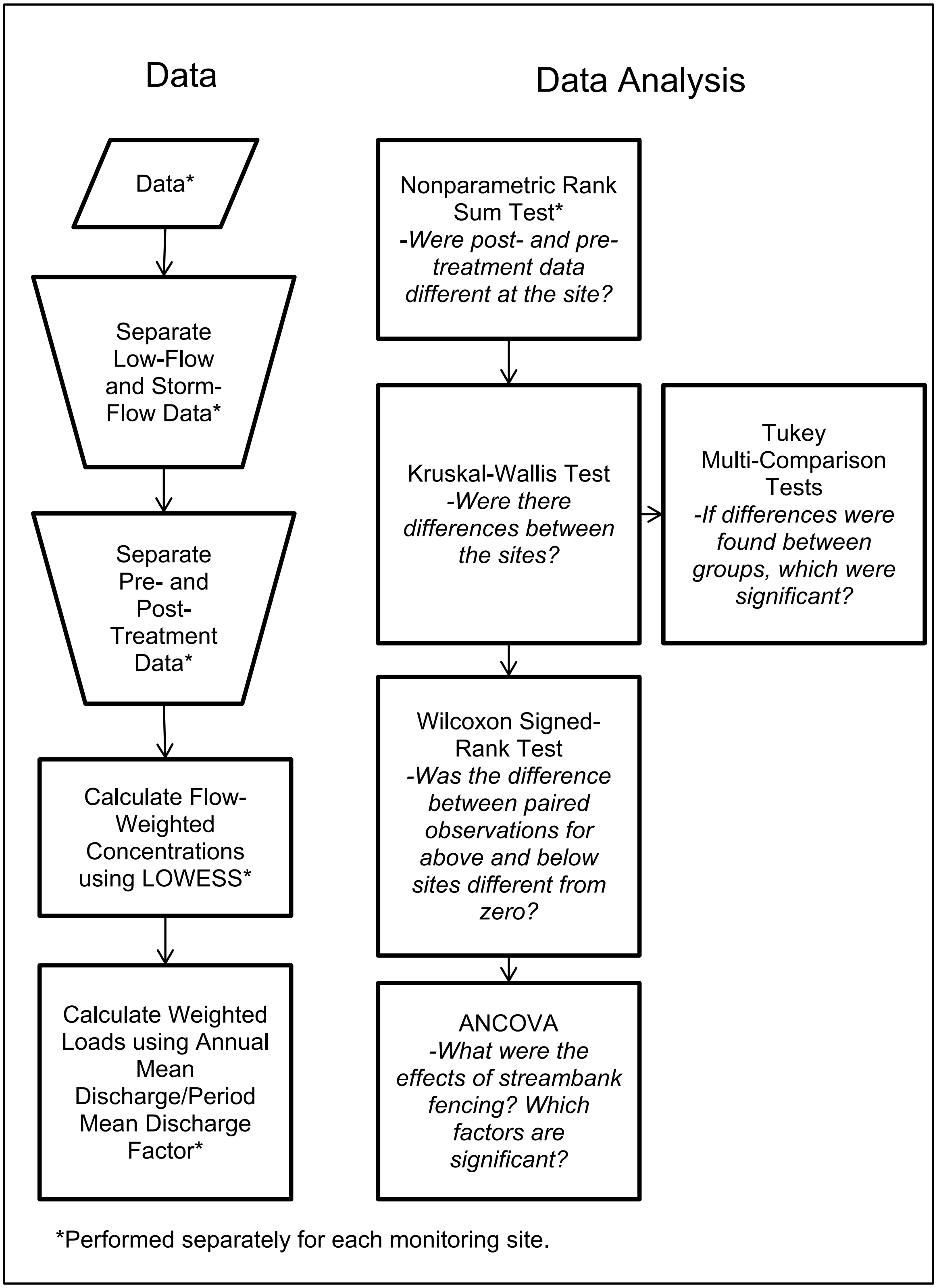


Figure 7-25. Basic data preparation and analysis procedure for above/below-before/after study in Pennsylvania (Galeone et al. 2006)

Example 7.8-5. Above/Below-Before/After Design with Upstream Concentration and Flow as Explanatory Variables - Walnut Creek, IA NNPSMP

In some cases, projects are forced to develop alternative plans for data analysis due to unforeseen circumstances they cannot control. The Walnut Creek (IA) NNMP project, for example, began as a ten-year paired-watershed study that also included an above/below-before/after design and three subwatershed single-station designs within each of the paired watersheds (Schilling and Spooner 2006). The primary purpose of the project was to evaluate the response of stream nitrate concentrations to conversion of row crops to native prairie. The normal approach of analyzing project data (both for the paired-watershed and above/below-before/after designs) using ANCOVA was compromised by two facts: prairie conversion began before the calibration period was completed, and conversion to prairie was gradual instead of rapid. Based on the guidelines and experiences of others (Spooner et al. 1987, Grabow et al. 1998 and 1999), multiple linear regression analysis on all ten monitoring sites was selected as an alternative approach to project evaluation (see Example 7.8-7 for the general form of equation used). Treatment in this case was modeled as time with covariates such as upstream concentration used to factor out hydrologic variability. For the downstream site in the treatment watershed, a model using month (for seasonality), upstream nitrate concentration, and downstream nitrate concentration in the control watershed provided the best fit to the data. For all other sites, month and the log of baseflow discharge from the same or a different site were used as covariates in the best-fit regression model. All tests resulted in detection of significant trends in nitrate concentrations, with the downstream treatment site trend indicating nitrate reductions due to conversion to prairie (the treatment). A negative coefficient on the time variable (-0.119 mg $l^{-1}yr^{-1}$) indicated a nitrate reduction of 1.2 mg l^{-1} over 10 years at this site. It was also found that in the control site, where land was unexpectedly converted from grassland to row crops, nitrate concentrations increased during the project period.

If the errors (e.g., residuals) in the statistical models are autocorrelated, a statistical software procedure should be used that incorporates the autocorrelation structure into the model. For example, PROC AUTOREG of the SAS software (SAS Institute 2010) is useful with autoregressive autocorrelation typical of weekly, biweekly, and monthly series. Alternatively, a correction of the standard deviation of the slope estimate and revised confidence intervals can be used with the correction given in section 7.3.6.

It should be cautioned that changes in pollutant concentrations or loads measured at a downstream station (either before or after land treatment) versus upstream may be difficult to detect if incoming concentrations or loads at the upstream station are high and the contribution of the additional area draining to the downstream station is small. Conversely, if the upstream contribution is very low compared to that of the treated area, a change or difference due to treatment may be difficult to attribute to BMPs because of dilution. If the upstream pollutant inputs do not respond similarly to hydraulic changes (e.g., rainfall), then the design effectively becomes a single watershed design. The Walnut Creek (IA) NNPSMP project provides an example of the former case where annual mean nitrate concentrations ranged from 10.0 to 12.7 mg/L at the upstream site and 6.8 to 9.5 mg/L at the site below the treatment area (Schilling and Spooner 2006). The treatment in this case was conversion of row crops to native prairie, and the study design (paired watersheds and above/below-before/after) was compromised by the fact that land conversion began before pre-treatment conditions could be established. See Example 7.8-5 for a discussion of how data from this project were analyzed using multiple linear regression, a technique typically applied to single watershed trend designs.

7.8.2.3 Nested Watershed

As described in section 2.4.2.3, it is preferred that the nested subwatershed is used as the control watershed[12] and is located above the remainder of the watershed where treatment occurs (Hewlett and Pienaar 1973). However, a valid nested design can also entail the treatment watershed in a small headwater subbasin; the control being the much larger watershed outlet. This design requires calibration (before) and treatment (after) periods similar to the paired-watershed design.

Analysis of data from a nested watershed design can be done using the same ANCOVA procedure described in section 7.8.2.1 for the paired-watershed design. In the case of nested watersheds, the paired data represent observations collected on the same date, time period, or storm at both the nested and main watershed stations. As noted above, data from the nested watershed should represent the control watershed, while data from the main watershed outlet represent the treatment watershed.

7.8.2.4 Single Watershed Trend Station

As noted in section 2.4.2.5, monitoring at a single watershed outlet is not a strong design for documenting the effectiveness of watershed land treatment on water quality. Without the ability to control for the effects of varying weather and hydrology, it is difficult to attribute any observed changes in water quality to the land treatment program. However, because the coupling of budget limitations and accountability requirements often leads to single-station designs, the unfortunate fact that some paired-watershed and other superior designs fail due to unforeseen circumstances, and the simple reality that some NPS watershed programs must rely on watershed outlet monitoring conducted by another party (e.g., a state long-term surveillance program or a USGS network station), it is useful to discuss how best to analyze data from such stations to assess the effects of a watershed project. In addition, experience has shown that projects with failed paired-watershed or above/below-before/after designs may resort to trend analysis as the best option for analyzing project data (see Example 7.8-6).

Long-term water quality data may show a *monotonic* trend (a continuous change, consistent in direction, either increasing or decreasing) or a *step* trend (an abrupt shift up or down). Trend analysis may be the best — or perhaps only — approach to documenting response to treatment in situations where water quality data are collected only at a single watershed outlet station or where land treatment was widespread, gradual, and inadequately documented. Data from long-term, fixed-station monitoring programs where gradual responses such as those due to incremental BMP implementation or continual urbanization are likely to occur are more aptly evaluated with monotonic trend analyses that correlate the response variable (i.e., pollutant concentration or load) with time or other independent variables. These types of analyses are useful in situations where vegetative BMPs like the riparian buffers implemented in the Stroud Preserve NNPSMP project (Newbold et al. 2008) must mature, resulting in gradual effects expressed over time. Analysis of step trends, on the other hand, is most appropriate when the change in response to BMP implementation is rapid and abrupt (e.g., when a municipal stormwater management regulation is enforced) and the timing of that change is known and well-documented. Biological data can also be evaluated with either monotonic or step-trend tests. A potential limitation is that most biological programs will only sample once a year and the time to acquire sufficient samples to detect a meaningful trend might be longer than what is practical.

[12] A reverse situation, where the downstream subwatershed area is the control is possible in theory, but all effort would need to be made to ensure that upstream contributions to constituents measured at the downstream control area are minimized.

Example 7.8-6. Single Trend Watershed with Covariates - Sycamore Creek, MI NNPSMP

This project planned a paired-watershed study with two treatment watersheds (Willow Creek and Marshall Drain) and one control watershed (Haines Drain), but implementation of no-till and continuous cover in the control watershed compromised the study (Suppnick 1999). Each watershed was then analyzed independently, with regression analysis ultimately successful in linking reductions in TSS (95 percent confidence level) and TP (90 percent confidence level) loads to the percentage of land in no-till in the Willow Creek watershed (Grabow 1999, Suppnick 1999). Following is a summary of the steps taken to establish the TSS relationship for Willow Creek (Grabow 1999):

1. Regression analysis on sediment yield versus storm discharge and/or peak flow to reduce the analysis to water quality change over time independent of hydrologic variability. All variables were log-transformed.
2. Two methods were then used to answer the question of whether there was a water quality trend over time.
    a. Regression equation incorporating elapsed time and explanatory variables. This addresses the question of whether there has been a change in water quality over time while simultaneously accounting for hydrologic variability.
    b. Regression of residuals[1] from regression on the water quality variable and explanatory variables versus elapsed time. This addresses the question of whether there has been a water quality change over time after adjusting for hydrologic variability.
3. Correlation of land use change to water quality change via multiple linear regression analysis. Terms incorporated in the regression model were percent of land in no-till, percent of land in continuous cover, storm discharge, and peak flow.

Step 1 yielded correlation between TSS load (kg/storm) and both storm discharge (mm) and peak flow (liters/second). Discharge and peak flow were tested for collinearity which was found to be not an issue (see Box 7.8-1).

Step 2 analyses indicated statistically significant trends in TSS and TP in Willow Creek watershed. Method “a” used the following basic equation:

$$log[TSS] = \beta_0 + \beta_1 log[Q] + \beta_2 log\left[Q_p\right] + \beta_3 t$$

Where TSS is the TSS storm load (kg), Q is the total storm discharge, Q_p is the peak stream discharge, t is elapsed time in days, and the β terms are regression parameter estimates. A significant negative value for β_3 indicated a reduction in TSS load over time. Insertion of average log values of total storm discharge and peak discharge, and setting the beginning and ending days (1 and 2,629 for t_{begin} and t_{end} in this case) would then yield the average change in loadings from the first to last data of data collection.

[1]Residuals are the differences between actual and predicted values: Actual-Predicted.

Example 7.8-6. (continued)

Sycamore Creek, MI NNPSMP

Method "b" of Step 2 used the following equation:

$$TSSres = \beta_0 + \beta_1 t$$

Where TSSres is the residuals (log kg/storm) from the regression in Step 1 and t is again elapsed time. In this approach, a statistically significant value for β_1 would indicate a change in the relationship between TSS and the explanatory variables (total and peak discharge), suggesting an impact due to land use change. The value $\beta_1 \times t_{end}$ would then estimate the change in loading (in log units) over the data collection period. The average change in loading is determined by then plugging the average values for log [Q] and log[Q_p] into the regression equation used in Step 1.

In this case, method "a" indicated a 60 percent reduction in TSS load, whereas method "b" estimated a 59 percent reduction.

With a statistically significant reduction in TSS load now documented, Step 3 explored the linkage between that reduction and land use change by adding the percentage of land in no-till (NoTill) and the percentage of land in continuous cover (ContCov) as additional terms in the multiple linear regression used for method "a" in Step 2. Statistically significant regression parameters β_3 and/or β_4 in the following equation would indicate correlation between log[TSS] and the percentage of land in no-till and/or continuous cover.

$$log[TSS] = \beta_0 + \beta_1 log[Q] + \beta_2 log\left[Q_p\right] + \beta_3 NoTill + \beta_4 ContCov + \beta_5 t$$

A statistically significant value of -0.01969 was found for β_3, but β_4 was insignificant, suggesting that for every percent increase in the percentage of land under no-till, the TSS load (as log kg) would decrease by 0.01969 log units. Regression estimates based on average storm discharge and peak flow were then used in conjunction with first-year and last-year values of no-till percentages to estimate a TSS load reduction of 52 percent, with a 95 percent confidence interval of 18-72 percent. This agreed well with the estimates of 59 and 60 percent reduction from Step 2.

Combining the results from the above analyses by Grabow (1999) with additional project information, it was concluded that it is very likely that streambank stabilization also contributed to the reduction in TSS observed in Willow Creek (Suppnick 1999).

Box 7.8-1. Collinearity

What is Collinearity?

Collinearity in multiple regression analysis occurs when there is a linear relationship between two explanatory (x) variables. Although this does not impact the reliability of the overall model, it does create great uncertainty regarding the model coefficients. There are ways to address collinearity, including recognizing the ambiguity in the interpretation of regression coefficients (USF n.d.) or simply removing one of the variables from the regression model (Martz 2013).

Various statistics programs have tests for collinearity (or multicollinearity), including the Variance Inflation Factor (VIF), Tolerance (1/VIF), and the Condition Index (SAS 2016a and 2016c, USF n.d.). Guidelines vary, but VIF values greater than 5 to 10, Tolerance values close to 0, and Condition Index values greater than 15 to 30 indicate problems with collinearity. See Belsley et al. (1980) for additional details.

Several statistical trend analysis techniques will be mentioned in this section; the topic of trend analysis is covered in more detail in *Tech Notes 6: Statistical Analysis for Monotonic Trends* (Meals et al. 2011). Before proceeding, it is important to recognize some limitations of trend analysis. First, trend analysis is most effective with long periods of record; general guidelines are ≥5 years of monthly data for monotonic trends and ≥2 years of monthly data before and after a step trend (Hirsch 1988). Short monitoring periods and small sample sizes make documentation of trends difficult, and it must be recognized that - especially over the short term - some increasing or decreasing patterns in water quality are not trends. A snapshot of water quality data over a few months may suggest a trend, but examination of a full year may show this "trend" to be part of a regular cycle associated with temperature, precipitation, or cultural practices. Autocorrelation may also be mistaken for a trend, especially over a short time period. Changes in sampling schedules, field methods, or laboratory practices can cause shifts in data that could be erroneously interpreted as step trends.

Perhaps most importantly, statistical trend analysis can help to identify trends and estimate the rate of change, but will not provide much insight into attributing a trend to a particular cause (e.g., land treatment). Interpreting the cause of a trend requires knowledge of the watershed and a deliberate study design (see section 7.8.1).

Before proceeding to numerical analysis, it is useful to examine time series plots for visual evidence of a trend. Visualization of trends in noisy data can be clarified by various data smoothing techniques. Plotting moving averages or medians, for example, instead of raw data points, reduces apparent variation and may reveal general tendencies. Spreadsheets can display a moving-average trend line in time-series scatterplots with adjustable averaging periods. A more complex smoothing algorithm, such as *LOWESS* (*LO*cally *We*ighted *S*catterplot *S*moothing), can reveal patterns in very large datasets that would be difficult to resolve by eye (see Helsel and Hirsch 2002). Most pollutant concentrations and loads in surface waters show strong seasonal patterns. Seasonal variations in precipitation and flow are often main drivers of these patterns, but seasonal changes in land management and use may also play a role. See section 4.3 of the 1997 guidance (USEPA 1997b) for additional information on seasonality.

Some techniques to address seasonality beyond controlling for the effects of flow covariates are often necessary for water quality trend analysis. For example, the relationship between concentration and discharge may not be consistent over time, perhaps due to seasonal variations in BMP implementation. The relationship (or slope) can be allowed to change between time periods by the use of interaction terms between the time periods and discharge in an analysis of covariance (ANCOVA) statistical model. An alternative that might develop more traction with experiences is to consider a weighted regressions on time, discharge and season (WRTDS) proposed by Hirsch et al. (2010) (see section 7.9.2 for more information on WRTDS).

When multiple explanatory variables are included in the trend models, it is common that these variables will be related to each other (collinearity) and/or a few data points may have a lot of 'influence' over the regression results (Belsley et al. 1980). Regression analysis performed with various software programs will provide leverage plots as part of the output to help identify these data features.

7.8.2.4.1 Monotonic Trends

Table 7-8 lists some monotonic trend tests available for different circumstances, including adjustments for a covariate and the presence of seasonality. The tests are further divided into parametric, nonparametric, and mixed types. Regression tests require that the expected value of the dependent variable is a linear function of each independent variable, the effects of the independent variables are additive, the errors in the model are independent (e.g., no correlation between consecutive errors in the case of time series data), and the errors exhibit both normality and constant variance. Nonparametric tests require only constant variance and independence. Parametric trend tests (see Examples 7.8-7 and 7.8-8) are considered more powerful and/or sensitive to detect significant trends than are nonparametric tests (see Example 7.8-9), especially with a small sample number. However, unless the assumptions for parametric statistics are met, it is generally advisable to use a nonparametric test (Lettenmaier 1976, Hirsch et al. 1991, Thas et al. 1998).

Table 7-8. Classification of tests for monotonic (nonparametric) or linear (parametric) trend (adapted from Helsel and Hirsch 2002)

	Type of Test	Not Adjusted for covariate (X)	Adjusted for covariate (X)
No Seasonality	Parametric	Linear regression of Y on t	Multiple linear regression of Y on X and t
	Mixed	-	Mann-Kendall on residuals from regression of Y on X
	Nonparametric	Mann-Kendall	Mann-Kendall on residuals from LOWESS of Y on X
Seasonality	Parametric	Linear regression of Y on t and periodic functions or indicator X's for months	Multiple linear regression of Y on X, t, and periodic functions or indicator X's for months
	Mixed	Regression of deseasonalized Y on t	Seasonal Kendall on residuals from regression of Y on X
	Nonparametric	Seasonal Kendall on Y	Seasonal Kendall on residuals from LOWESS of Y on X
Other Explanatory variables or covariates (e.g., stream discharge)	Parametric	Linear regression of Y on t and covariates (X)	Multiple linear regression of Y on t, X covariates
	Mixed	Regression of deseasonalized Y on X	Seasonal Kendall on residuals from regression of Y on X
	Nonparametric	Seasonal Kendall on Y	Seasonal Kendall on residuals from LOWESS of Y on X

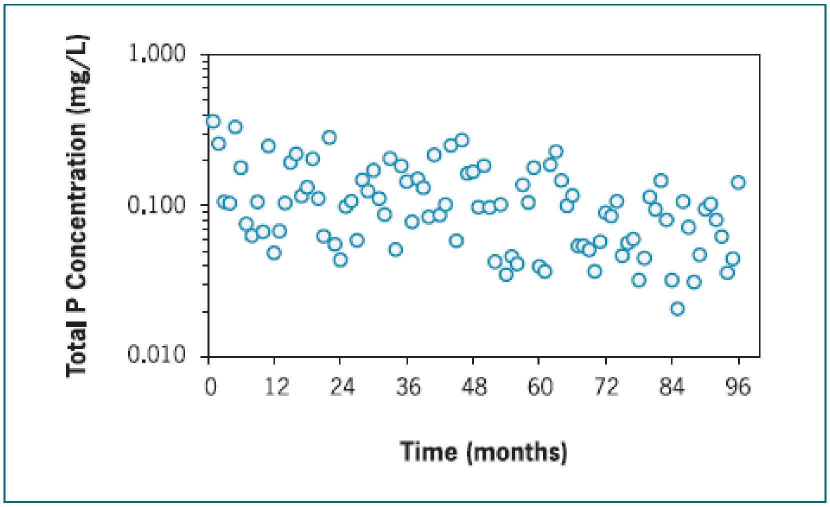
Y = dependent variable of interest; X = covariate; t = time

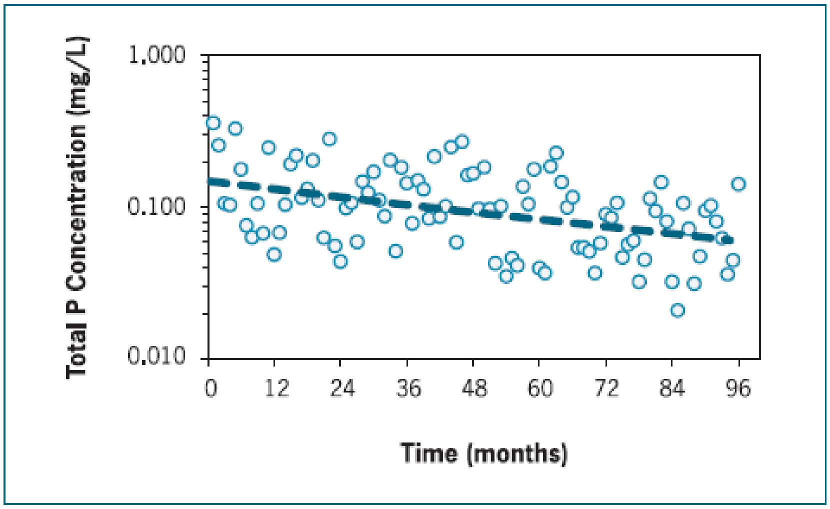
Refer to *Tech Notes 6: Statistical Analysis for Monotonic Trends* (Meals et al. 2011) for details on the tests listed in Table 7.8-1. Chapter 4 (pages 4-86 through 4-89) of the 1997 guidance (USEPA 1997b) also discusses the computation of Mann-Kendall and Seasonal Kendall statistics.

If the trend model has autocorrelated errors, a statistical model that incorporates the autoregessive errors should be employed. Alternatively, a correction of the standard error of the slope that is given in section 7.3.6 can be used to calculate the correct confidence interval of the slope on t (time, date) to determine if it is significantly different from zero (e.g., evidence of a trend over time) in the pollutant concentration or load.

Example 7.8-7. Simple Linear Regression - Samsonville Brook in Vermont

- Eight years of monthly TP concentration data from Samsonville Brook in Vermont
- Data satisfy assumptions for regression after log transformation: normal distribution, constant variance, independence (low autocorrelation)





Simple linear regression (using Excel® or any basic statistical package)

Log[TP] = -0.8285-0.00414(Time)
$r^2 = 0.18$, F = 21.268 P ≤ 0.001

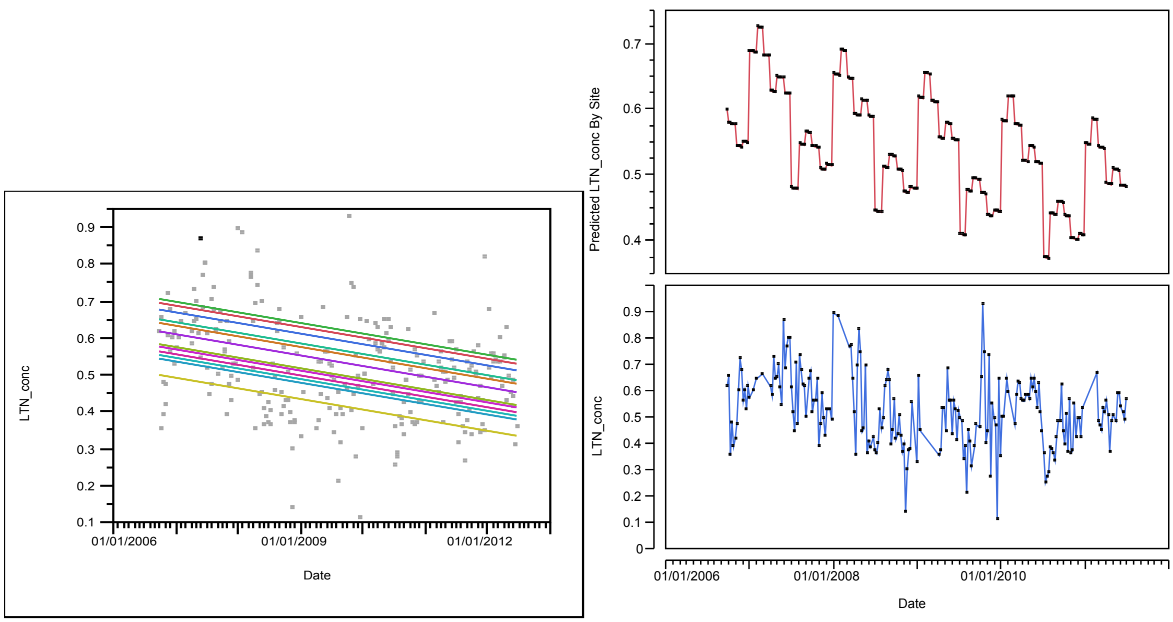
Rate of change:
Slope of log-transformed date = -0.00414
$(10^{-0.00414} - 1)\times100$ = -0.95%/month or about -11%/year

This result suggests that TP concentrations have decreased significantly over the period at a rate of approximately 11 percent per year.

Note: Data used in this example are taken from the Vermont NNMP project, *Lake Champlain Basin agricultural watersheds section 319 national monitoring program project, 1993 – 2001* (Meals 2001).

Example 7.8-8. Linear Regression with Monthly Seasons as a Covariate - Corsica River, MD NNPSMP

A significant trend was detected in a small watershed within the Corsica River Basin, Maryland, using times series analysis that adjusted for autocorrelation as well as monthly (seasonal) differences for log transformed, flow-weighted total nitrogen (TN) concentrations. In this example, monthly indicator variables were used to adjust for seasonality in an ANOVA regression model. See section 7.3.6 for details on adjustments for autocorrelation and seasonality.



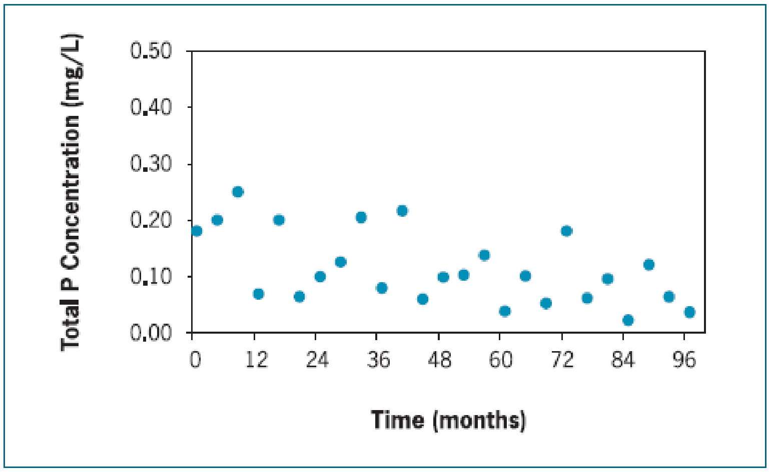
By addressing seasonality in the regression model with monthly indicator variables, most of the regression degrees of freedom were preserved, a more powerful approach than if each month was evaluated separately. Each line in the plot on the left represents the trend line (log transformed, ffflow-weighted TN concentration) for a single month (i.e., January, February … December). The trend slopes for each month were assumed to be the same, but the intercept was allowed to vary, enabling the differences in concentration due to season to be removed from the test for trends and therefore making it easier to isolate and detect trends due to other factors (e.g., BMPs).

The bottom right graph shows the raw data. The noise due to seasonal differences and other factors makes it difficult to pick out any trends. The top right graph shows the predicted value from the seasonal regression model with the indicator variables. A downward trend is apparent and it is also clear from this graph that the highest TN concentration is found in February, followed by January, March, May, April, June, Sept, August, October, November, December, and July (lowest).

Example 7.8-9. Mann-Kendall Procedure – Single Trend Watershed - Samsonville Brook in Vermont.

The data from Samsonville Brook in Vermont:

- Eight years of quarterly mean TP concentration data
- Data satisfy assumptions for constant variance and independence, but are not normally distributed without transformation



Month (n=25)	[TP] mg/L
1	**0.180**
5	**0.200**
9	**0.250**
13	**0.068**
17	**0.201**
21	**0.063**
25	**0.099**
29	**0.125**
33	**0.205**
37	**0.078**
41	**0.216**
45	**0.059**
49	**0.098**
53	**0.102**
57	**0.137**
61	**0.037**
65	**0.100**
69	**0.051**
73	**0.180**
77	**0.060**
81	**0.095**
85	**0.021**
89	**0.120**
93	**0.063**
97	**0.035**

The Mann-Kendall trend test for this example may be evaluated in two ways. First, in a manual calculation, use the formulas below. The value of S (sum of the signs of differences between all combinations of observations) can be determined either manually or by using a spreadsheet to compare combinations, create dummy variables (-1, 0, and +1), and sum for S.

Mann-Kendall $S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign\left(y_j - y_i\right) = -106$

$\tau = \frac{S}{n(n-1)/2} = \frac{-106}{300} = -0.353$ **(decreasing trend)**

Calculating Z_S as $(S\pm1)/\sigma_S$ where

$$\sigma_s = \sqrt{\left(\frac{n}{18}\right) \times (n-1) \times (2n+5)} = 42.817$$

$Z = \frac{-105}{42.817} = -2.454$ **(USEPA 1997a)**

This Z statistic is significant at P=0.014, indicating a significant trend, i.e., we are 98.6 percent confident there is a decreasing trend in TP. See USEPA (1997a) for the calculation of σ_S when there are ties among the data.

To estimate the rate of change, use the Sen slope estimator:

$\beta_1 = median\left(\frac{y_j - y_i}{x_j - x_i}\right)$ **211 individual slopes -000945 to +0.00766**

Median slope = -0.0011 mg/L/month = -0.013 mg/L/yr

This result suggests that TP concentration decreased significantly over the period at a rate of about 0.013 mg/L/yr.

Note: Data used in this example are taken from the Vermont NNMP project, Lake Champlain Basin agricultural watersheds section 319 national monitoring program project, 1993 – 2001 (Meals 2001).

7.8.2.4.2 Step Trends

Monotonic trend analysis may not be appropriate for all situations. Other statistical tests for discrete changes (step trends) should be applied where a known discrete event (like BMP implementation over a short period) has occurred. Testing for differences between the "before" and "after" conditions is done using two-sample procedures such as the Student's t test and ANCOVA (parametric techniques with and without covariates) and nonparametric alternatives such as the rank-sum test, Mann-Whitney test, and the Hodges-Lehmann estimator of step trend magnitude (Helsel and Hirsch 2002, Walker 1994). Application of the Mann-Whitney/Wilcoxon's rank sum test and the Hodges-Lehmann estimator are illustrated in sections 4.5.2 and 4.5.3, respectively, of the 1997 guidance (USEPA 1997b). A key principle in step trend analysis is that the hypothesized timing of the step change must be selected in advance (i.e., define the pre- and post- periods before conducting statistical tests). Knowledge of watershed management activities and examination of data plots will be helpful in identifying a potential step in time.

For example, the Mann-Whitney test was used to associate changes in P management practices with a decrease in annual median soluble reactive P concentration from a 9-ha grassland catchment in Northern Ireland (Smith et al. 2003). Weekly samples were collected from 1989 through 2000, with the change in P management instituted in 1998. A comparison of data from 1997 with data from 2000 indicated that the change from whole-farm to site-specific P management reduced SRP concentrations significantly.

If the trend model has autocorrelated errors, a statistical model that incorporates the autoregessive errors should be employed. Alternatively, a correction of the standard error of the slope that is given in section 7.3.6 can be used to calculate the correct confidence interval of the step change (difference) between time periods to determine if it is significantly different from zero (e.g., evidence of a step change) in the pollutant concentration or load.

7.8.2.5 Multiple Watersheds

In the simplest case of a multiple watershed design, where monitored watersheds fall into two groups, treated and untreated, data may be analyzed by Student's t test or the non-parametric Wilcoxon Rank-Sum test. Such an analysis would test the (null) hypothesis that there was no significant difference in mean pollutant concentration or load between the treated and untreated watershed groups. Where monitored watersheds occur in more than two groups (e.g., untreated, treatment A, treatment B, etc.), significant differences in group means can be evaluated using ANOVA or the Kruskal-Wallis test. For example, Clausen and Brooks (1983) assessed mining impacts on MN peat lands using a multiple watershed design. Results – analyzed by ANOVA for normally distributed variables and otherwise by nonparametric Kruskal-Wallis and Chi-Square tests – documented significant impacts of peat mining on water quality. Lewis (2006) describes application of fixed-effect and mixed-effect (i.e., includes random effects) regression models to multiple-watershed studies involving logging. A 13-watershed study involving 3 controls, 5 clear-cuts, and 5 partial cuts was carried out over sixteen years with monitoring of storm volumes during four years before cutting, three years of logging, and nine years[13] of post-logging. The best fit was obtained when the proportion harvested, antecedent wetness, regrowth, and spatial autocorrelation were all incorporated into the model. This study design and analytic approach allows the prediction of streamflow response to harvesting in other watersheds considered part of the same population of watersheds included in the study.

[13] Three years of post-cut monitoring at seven stations and nine years at six stations.

7.8.3 Linking Water Quality Trends to Land Treatment

A central objective of many NPS watershed projects is to determine not only if water quality changes can be documented but also if water quality changes can be associated with changes in land treatment. Such documentation is necessary to help build an information base to support continued improvement in preventing and solving water quality problems. It is also needed in many cases to justify expenditure on clean-up efforts.

For a range of reasons, including budgets and programmatic constraints, watershed project monitoring efforts are almost never designed to satisfy the rigorous criteria for establishing true cause and effect relationships (see Box 7.8-2). Rather, project effectiveness monitoring designs are generally intended to measure improvements in water quality and, hopefully, relate that improvement to activities undertaken to influence water quality. A plausible argument that what was done on the ground improved water quality is often the best that can be hoped for and that is usually not a simple task at the watershed level. The ability to control for factors other than land treatment (e.g., weather, hydrology, land use change) is a key ingredient in making such a plausible argument.

Box 7.8-2. Cause-effect requirements (Mosteller and Tukey 1977).

A cause-effect relationship must satisfy the following criteria:

- *Consistency* - the direction and degree of the relationship between the measured variables (such as TP loads and acres treated with nutrient management) holds in each data set.
- *Responsiveness* - as one variable changes in a known manner, the other variable changes similarly. For example, as the amount of land treatment increases, further reduction of pollutant delivery to the water resource is documented.
- *Mechanistic* - the observed water quality change is that which is expected based on the known or hypothesized physical processes involved in the installed BMPs.

Control refers to eliminating or accounting for all factors that may affect the response to the treatment so that the treatment effect can be isolated. In a laboratory experiment, control is usually obtained by subjecting the entire system to the same conditions, varying only the treatment variable and selecting replicates at random to assure that unmeasured sources of variability do not affect the interpretation. Such an approach is rarely if ever possible for monitoring projects in watersheds dominated by nonpoint sources. Instead, we hope to show an association between change in water quality and change in land use or management by selecting a project design that includes monitoring for important explanatory variables (covariates) and applying appropriate statistical tools to include and adjust for these covariates in the analysis. By factoring explanatory variables into trend analyses, we remove some of the noise in the data to uncover water quality trends that are closer to those that would have been measured had no changes in climatic or other explanatory variables occurred over time. When performing statistical analyses with both water quality and land treatment data, it is important to note that it is not necessary to summarize the water quality data on the same (less frequent) time scale as the land treatment data. Rather, land treatment data can be incorporated within a trend analysis, for example, as repeating explanatory variables. That is, the values of land treatment and land use are treated as X variables in a statistical trend model. Because land management data are usually taken less frequently than water quality data, the land management information for a given X variable can be repeated for the time range of water quality samples that is represented by the land management value.

Although association by itself is not sufficient to infer causal relationships, it can contribute to a plausible argument that pollution control activities have resulted in environmental improvement. Thus, knowledge of land management and land treatment in the watershed is essential to demonstrate an association between changes on the land and changes in water quality. For example, section 7.8.2.2 described how the Sycamore Creek (MI) NNMP project used multiple linear regression to link log[TSS] load to the percentage of land under no-till cropping (Grabow 1999). Additional explanatory variables included the logs of total storm discharge and peak stream discharge.

Data on both the temporal progress and spatial extent of land treatment and other watershed land use/management activities should be used to build an association between land treatment and observed water quality. For example, on a temporal scale, land treatment and management data can be analyzed and linked to water quality in these ways:

Define monitoring periods: Documentation of BMP implementation can be used to define critical project periods, like pre- and post-treatment periods in before/after and paired-watershed designs or to establish a hypothesis on the timing of a step trend.

Explain observed water quality: Knowledge of not only BMP implementation history but also dates of tillage, manure or agrichemical applications, street sweeping, and other watershed management activities can be extremely useful in qualitatively explaining observed water quality patterns, especially extreme or unusual values.

Quantify the level of treatment: Quantitative expressions of land treatment can become the independent variable in an analysis of correlation between land management and water quality. Analyze land treatment data collected in the watershed monitoring program to form such variables as:

- Number or percent of watershed animal units under animal waste management
- Acres or percent of cropland in cover crops
- Acres or percent of cropland under conservation tillage
- Annual manure or fertilizer application rate and extent
- Extent and capacity of stormwater infiltration practices

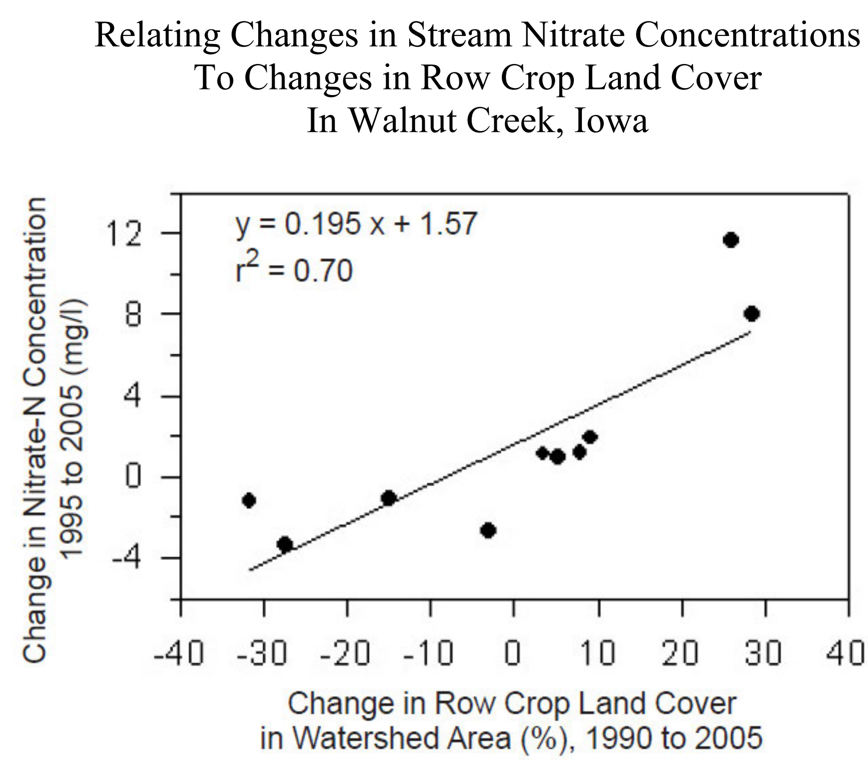
Such variables can be tested for correlation with mean total P concentration, annual suspended sediment load, or other annual water quality variables.

Document areas receiving BMPs: Use knowledge of locations of land treatment to:

- Select appropriate watersheds for analysis in a multiple watershed design
- Confirm conditions in above/below and nested watershed designs
- Document the integrity of the control and treatment watersheds in a paired-watershed design

Relate land treatment to critical source areas: A comparison of critical pollutant sources to locations that received treatment can assist in evaluating effectiveness of land treatment efforts and establish expectations for how much of the NPS problem the land treatment program potentially addresses.

The Walnut Creek (IA) NNPSMP project, for example, monitored stream NO_3-N concentrations and tracked conversion of row crop land to restored prairie vegetation (Schilling and Spooner 2006). By linking the two monitored variables, the project was able to suggest a clear association between restoring native prairie and reducing stream nitrate levels (see Figure 7-26).



Relationship between change in percentage of land cover in row crops and change in stream NO_3-N concentrations in Walnut Creek, IA

Figure 7-26. Linking stream nitrate concentration to land cover (Schilling and Spooner 2006)

7.9 Load Estimation

Determination of pollutant load is a key objective for many NPS monitoring projects. The mass of nutrients delivered to a lake or estuary drives the productivity of the waterbody. The annual suspended sediment load transported by a river is usually a more meaningful indicator of soil loss in the watershed than is a suspended sediment concentration. The foundation of water resource management embodied in the TMDL concept lies in assessment of the maximum pollutant load a waterbody can accept before becoming impaired and in the measurement of changes in pollutant loads in response to implementation of management measures.

Estimation of pollutant load through monitoring is a complex task that requires accurate measurement of both pollutant concentration and water flow and careful calculation, often based on a statistical approach. It is imperative that an NPS monitoring program be designed for good load estimation at the start. This section addresses important considerations and procedures for developing good pollutant load estimates in NPS monitoring projects. Much of the material is taken from an extensive monograph written by Dr. R. Peter Richards, of Heidelberg University, *Estimation of Pollutant Loads in Rivers and Streams: A Guidance Document for NPS Programs*. The reader is encouraged to consult that document and its associated tools for additional information on load estimation. Much of this information is also summarized in Meals et al. (2013).

7.9.1 General Considerations

7.9.1.1 Definitions

Load may be defined as the mass of a substance that passes a particular point of a river (such as a monitoring station on a watershed outlet) in a specified amount of time (e.g., daily, annually). Mathematically, load is essentially the product of water discharge and the concentration of a substance in the water. Flux is a term that describes the loading rate, i.e., the instantaneous rate at which the load passes a point in the river. Water discharge is defined as the volume of water that passes a cross-section of a river in a specified amount of time, while flow refers to the discharge rate, the instantaneous rate at which water passes a point. Refer to Meals and Dressing (2008) for guidance on appropriate ways to estimate or measure surface water flow for purposes associated with NPS watershed projects.

> **Basic Terms**
>
> **Flux** – instantaneous loading rate (e.g., kg/sec)
>
> **Flow rate** – instantaneous rate of water passage (e.g., L/sec)
>
> **Discharge** – quantity of water passing a specified point (e.g., m^3)
>
> **Load** – mass of substance passing a specified point (e.g., metric tons)

If we could directly and continuously measure the flux of a pollutant, the results might look like the plot in Figure 7-27. The load transported over the entire period of time in the graph would simply be equal to the shaded area under the curve.

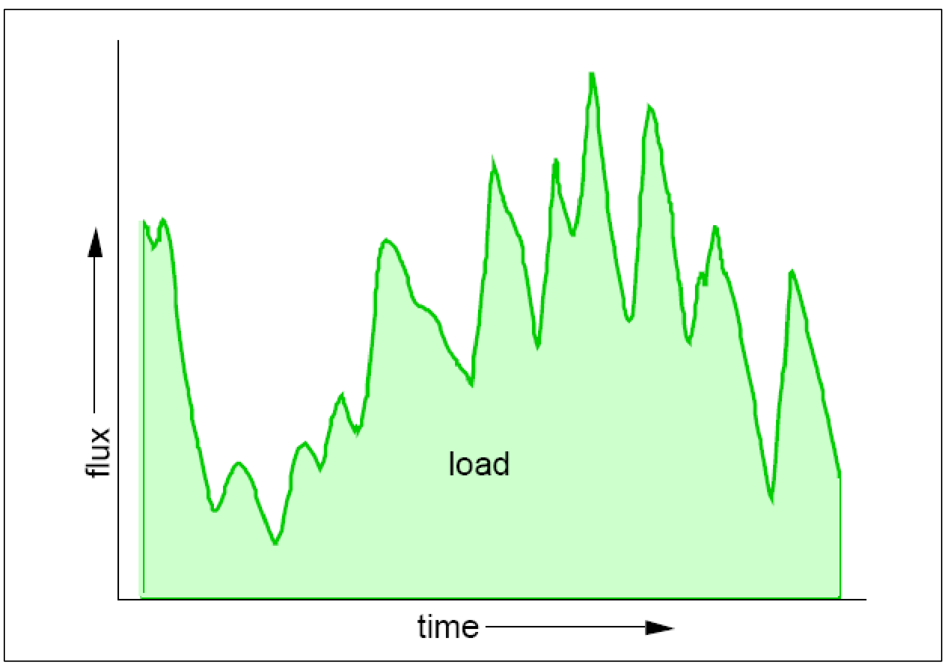


Figure 7-27. Imaginary plot of pollutant flux over time at a monitoring station (Richards 1998)

However, we cannot measure flux directly, so we calculate it as product of instantaneous concentration and instantaneous flow:

$$Load = k \int_t c(t)q(t)dt$$

where c is concentration and q is flow, both a function of time (t), and k is a unit conversion factor. Because we must take a series of discrete samples to measure concentration, the load estimate becomes the sum of a set of *n* products of concentration (c), flow (q), and the time interval (Δt) over which the concentration and flow measurements apply:

$$Load = k \sum_{i=1}^{n} c_i q_i \Delta t$$

The main monitoring challenge becomes how best to take the discrete samples to give the most accurate estimate of load. Note that the total load is the load over the timeframe of interest (e.g., one year) determined by summing a series of unit loads (individual calculations of load as the product of concentration, flow, and time over smaller, more homogeneous time spans). The central problem is to obtain good measures of concentration and flow during each time interval; calculation of total load by summing unit loads is simple arithmetic.

7.9.1.2 Issues of Variability

Both flow and concentration vary considerably over time, especially in NPS situations. Accurate load estimation becomes an exercise in both how many samples to take and when to take them to account for this variability.

Sampling frequency has a major influence on the accuracy of load estimation, as shown in Figure 7-28. The top panel shows daily suspended solids load (calculated as the products of daily total suspended solids (TSS) concentration and mean daily discharge measured at a continuously recording USGS station) for the Sandusky River in Ohio. The middle panel represents load calculated using weekly TSS samples and mean weekly discharge; the lower panel shows load calculated from monthly TSS samples and mean monthly discharge data. Clearly, very different pictures of suspended solids load emerge from different sampling frequencies, as decreasing sampling frequencies tend to miss more and more short-term but important events with high flow or high TSS concentrations.

Because in NPS situations most flux occurs during periods of high discharge (e.g., ~80 – 90 percent of annual load may be delivered in ~10 – 20 percent of time), choosing *when* to sample can be as important as how often to sample. The top panel in Figure 7-29 shows a plot of daily suspended solids load derived from weekly sampling superimposed on daily flux data; the bottom panel shows daily loads derived from monthly and quarterly sampling on top of the same daily flux data. Weekly samples give a reasonably good visual fit over the daily flux pattern. The monthly series gives only a very crude representation of the daily flux, but it is somewhat better than expected, because it happens to include the peaks of two of the four major storms for the year. A monthly series based on dates about 10 days later than these would have included practically no storm observations, and would have seriously underestimated the suspended solids load. Quarterly samples result in a poor fit on the actual daily flux pattern.

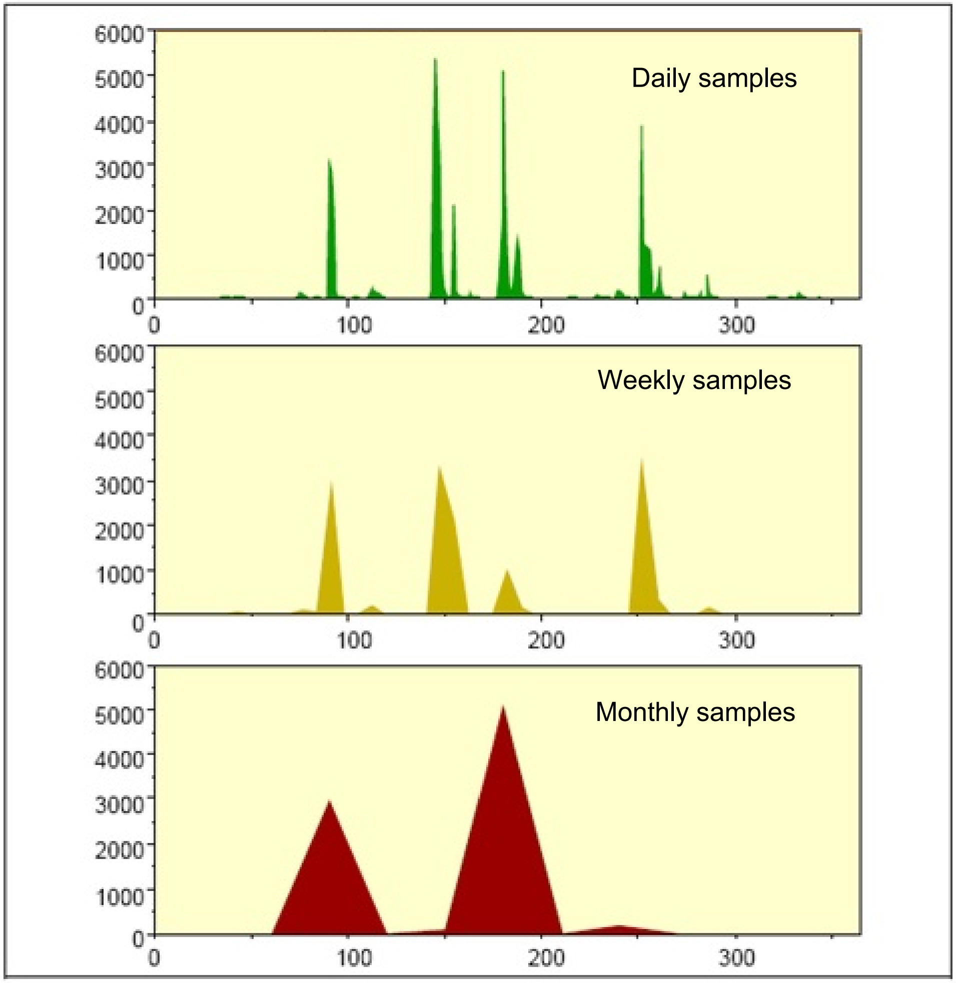


Figure 7-28. Plot of suspended solids loads for the Sandusky River, water year 1985 (Richards 1998). *Top,* daily TSS samples; *Middle*, weekly samples; *Bottom,* monthly samples. Weekly and monthly sample values were drawn from actual daily sample data series. Flux is on y-axis, time is on x-axis, and area under curve is load estimate.

The key point here is that many samples are typically needed to accurately and reliably capture the true load pattern. Quarterly observations are generally inadequate, monthly observations will probably not yield reliable load estimates, and even weekly observations may not be satisfactory, especially if very accurate load estimates are required to achieve project objectives.

7.9.1.3 Practical Load Estimation

Ideally, the most accurate approach to estimating pollutant load would be to sample very frequently and capture all the variability. Flow is relatively straightforward to measure continuously (see Meals and Dressing 2008), but concentration is expensive to measure and in most cases impossible to measure continuously. It is therefore critically important to choose a sampling interval that will yield a suitable characterization of concentration.

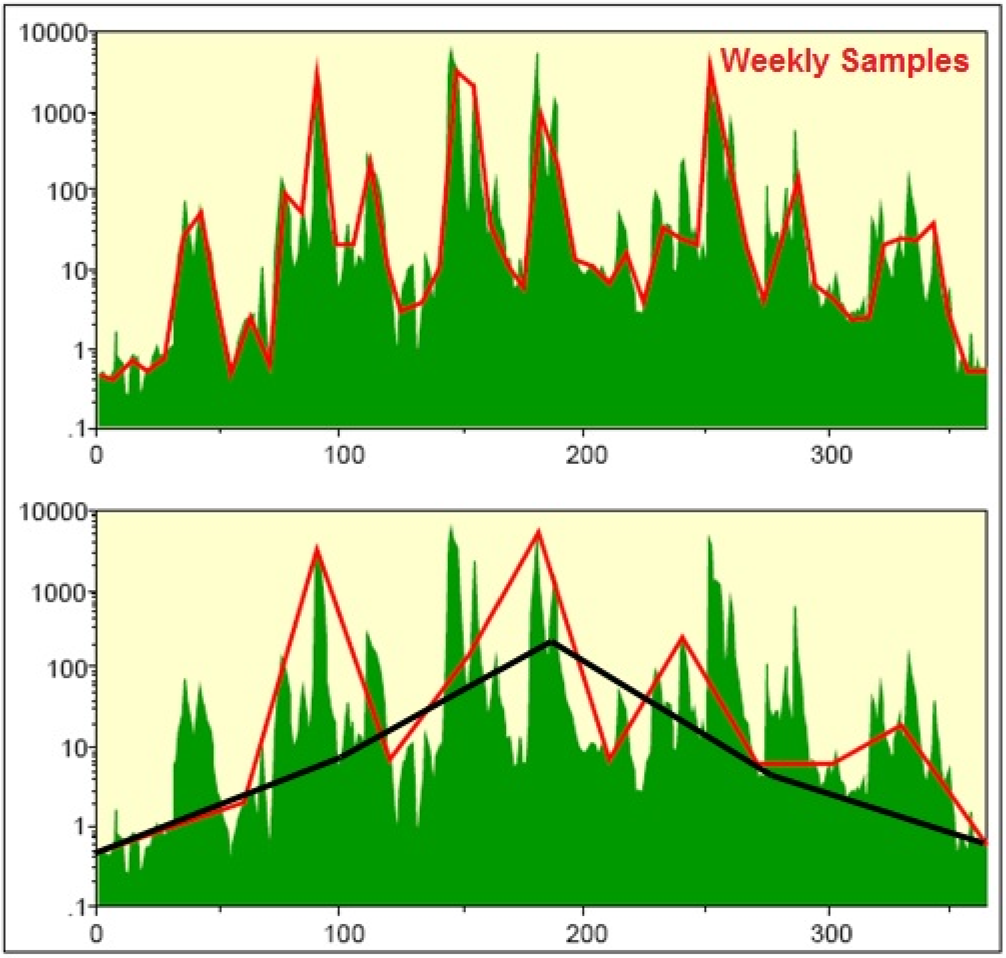


Figure 7-29. Weekly (red line in top panel), monthly (red line), and quarterly (black line in bottom panel) suspended solids load time series superimposed on a daily load time series (Richards 1998). Log of flux is on y-axis, time is on x-axis, and area under curve is load estimate.

There are three important considerations involved in sampling for good load estimation: sample type, sampling frequency, and sample distribution in time. Grab samples represent a concentration only at a single point in time and the selection of grab sampling interval must be made in consideration of the issues of variability discussed above. Integrated samples (composite samples made up of many individual grab samples) are frequently used in NPS monitoring. Time-integrated or time-proportional samples are either taken at a constant rate over the time period or are composed of subsamples taken at a fixed frequency. Time-integrated samples are poorly suited for load estimation because they are taken without regard to changes in flow (and concentration) that may occur during the integration period and are usually biased toward the low flows that occur most often. Flow-proportional samples (where a sample is collected for every n units of flow that pass the station), on the other hand, are ideally suited for load estimation, and in principle should provide a precise and accurate load estimate if the entire time interval is properly sampled. However, collecting flow-proportional samples is technically challenging and may not be suitable for all purposes. Also, even though a flow-proportional sample over a time span (e.g., a week) is a good summation of the variability of that week, ability to see what happened within that week (e.g., a transient spike in concentration) is lost. Flow-proportional sampling is also not compatible with some monitoring demands, such as monitoring for ambient concentrations that are highest at low flow or for documenting exceedance of critical values (e.g., a water quality standard).

Sampling frequency determines the number of unit load estimates that can be computed and summed for an estimate of total load. Using more unit loads increases the probability of capturing variability across the year and not missing an important event (see Figure 7-29); in general, the accuracy and precision of a load estimate increases as sampling frequency increases. Over a sufficiently short interval between samples, a sampling program will probably not miss a sudden peak in flux. If, for example, unit loads are calculated by multiplying the average concentration for the time unit by the discharge over the same time unit, the annual load that is the sum of four quarterly unit loads will be considerably less accurate than the annual load that is the sum of twelve monthly loads. Note that this example does not mean that an annual load calculated from 12 monthly loads is sufficiently accurate for all purposes.

There is a practical limit to the benefits of increasing sampling frequency, however, due to the fact that water quality data tend to be autocorrelated (see section 7.3.6). The concentration or flux at a certain point today is related to the concentration or flux at the same point yesterday and, perhaps to a lesser extent, to the concentration or flux at that spot last week. Because of this autocorrelation, beyond some point, increasing sampling frequency will accomplish little in the way of generating new information. This is usually not a problem for monitoring programs, but can be a concern, however, when electronic sensors are used to collect data nearly continuously.

Consideration of the basic sampling frequency – n samples per year – does not address the more complex issue of timing. The choice of *when* to collect concentration samples is critical. Most NPS water quality data have a strong seasonal component as well as a strong association with other variable factors such as precipitation, streamflow, or watershed management activities such as tillage or fertilizer application. Selecting when to collect samples for concentration determination is essentially equivalent to selecting when the unit loads that go into an annual load estimate are determined. That choice must consider the fundamental characteristics of the system being monitored. In northern climates, spring snowmelt is often the dominant export event of the year; sampling during that period may need to be more intensive than during midsummer in order to capture the most important peak flows and concentrations. In southern regions, intensive summer storms often generate the majority of annual pollutant load; intensive summer monitoring may be required to obtain good load estimates. For many agricultural pesticides, sampling may need to be focused on the brief period immediately after application when most losses tend to occur. Issues of random sampling, stratified random sampling, and other sampling regimes should be considered. Simple random sampling may be inappropriate for accurate load estimation if, as is likely, the resulting schedule is biased toward low flow conditions. Stratified random sampling – division of the sampling effort or the sample set into two or more parts which are different from each other but relatively homogeneous within – could be a better strategy. In cases where there is a conflict between the number of observations a program can afford and the number needed to obtain an accurate and reliable load estimate, it may be possible to use flow as the basis for selecting the interval between concentration observations. For example, planning to collect samples every x thousand ft^3 of discharge would automatically emphasize high flux conditions while economizing on sampling during baseflow conditions. Sampling levels following this strategy could be based on an annual average flow, recognizing that the number of samples per year would vary.

7.9.1.4 Planning for Load Estimation

Both discharge and concentration data are needed to calculate pollutant loads, but monitoring programs designed for load estimation will usually generate more flow than concentration data. This leaves three basic choices for practical load estimation:

1. Find a way to estimate un-measured concentrations to go with the flows observed at times when chemical samples were not taken;
2. Throw out most of the flow data and calculate the load using the concentration data and just those flows observed at the same time the samples were taken; and
3. Do something in between - find some way to use the more detailed knowledge of flow to adjust the load estimated from matched pairs of concentration and flow.

The second approach is usually unsatisfactory because the frequency of chemical observations is likely to be inadequate to give a reliable load estimate when simple summation is used. Thus almost all effective load estimation approaches are variants of approaches 1 or 3.

Unfortunately, the decision to calculate loads is sometimes made after the data are collected, often using data collected for other purposes. At that point, little can be done to compensate for a data set that contains too few observations of concentration, discharge, or both, collected using an inappropriate sampling design. Many programs choose monthly or quarterly sampling with no better rationale than convenience and tradition. A simulation study for some Great Lakes tributaries revealed that data from a monthly sampling program, combined with a simple load estimation procedure, gave load estimates which were biased low by 35 percent or more half of the time (Richards and Holloway 1987).

To avoid such problems, the sampling regime needed for load estimation must be established in the initial monitoring design, based on quantitative statements of the precision required for the load estimate. The resources necessary to carry out the sampling program must be known and budgeted for from the beginning.

The following steps are recommended to plan a monitoring effort for load estimation:

- Determine whether the project goals require knowledge of load, or if goals can be met using concentration data alone. In many cases, especially when trend detection is the goal, concentration data may be easier to work with and be more accurate than crudely estimated load data.
- If load estimates are required, determine the accuracy and precision needed based on the uses to which they will be put. This is especially critical when the purpose of monitoring is to look for a change in load. It is foolish to attempt to document a 25 percent load reduction from a watershed program with a monitoring design that gives load estimates ±50 percent of the true load (see Spooner et al. 2011a).
- Decide which approach will be used to calculate the loads based on known or expected attributes of the data.
- Use the precision goals to calculate the sampling requirements for the monitoring program. Sampling requirements include both the total number of samples and, possibly, the distribution of the samples with respect to some auxiliary variable such as flow or season.
- Calculate the loads based on the samples obtained after the first full year of monitoring, and compare the precision estimates (of both flow measurement and the sampling program) with the initial goals of the program. Adjust the sampling program if the estimated precision deviates substantially from the goals.

It is possible that funding or other limitations may prevent a monitoring program from collecting the data required for acceptable load estimation. In such a case, the question must be asked: is a biased, highly uncertain load estimate preferable to no load estimate at all? Sometimes the correct answer will be no.

7.9.2 Approaches to Load Estimation

Several distinct technical approaches to load estimation are discussed below. The reader is encouraged to consult Richards (1998) for details and examples of these calculations. Do not estimate annual loads based on simple multiplication of an annual average concentration and average discharge as load estimates will be biased low for positively correlated parameters such as suspended sediment and total phosphorus.

7.9.2.1 Numeric Integration

The simplest approach is numeric integration, where the total load is given by

$$Load = \sum_{i=1}^{n} c_i q_i t_i$$

where c_i is the concentration in the ith sample, q_i is the corresponding flow, and t_i is the time interval represented by the ith sample, calculated as:

$$\frac{1}{2}(t_{i+1} - t_{i-1})$$

It is not required that t_i be the same for each sample.

The question becomes how fine to slice the pie – few slices will miss much variability, many slices will capture variability but at a higher cost and monitoring effort. Numeric integration is only satisfactory if the sampling frequency is high - often on the order of 100 samples per year or more, and samples must be distributed so that all major runoff events are captured. Selection of sampling frequency and distribution over the year is critical – sampling must focus on times when highest fluxes occur, i.e., periods of high discharge.

As noted above, flow-proportional sampling is a special case of mechanical rather than mathematical integration that assumes that one or more samples can be obtained that cover the entire period of interest, each representing a known discharge and each with a concentration that is in proportion to the load that passed the sampling point during the sample's accumulation. If this assumption is met, the load for each sample is easily calculated as the discharge times the concentration, and the total load for the year is derived by summation. In principle, this is a very efficient and cost-effective method of obtaining a total load.

7.9.2.2 Regression

When, as is often the case with NPS-dominated systems, a strong relationship exists between flow and concentration, using regression to estimate load from continuous flow and intermittent concentration data can be highly effective. In this approach, a regression relationship is developed between concentration and flow based on the days for which concentration data exist. Usually, these data are based on grab samples for concentration and mean daily flow for the sampling day (see Example 7.9-1). This relationship may involve simple or multiple regression analysis using covariates like precipitation. In most applications, both concentration and flow are typically log-transformed to create a dataset suited for regression analysis (see section 7.3.2 and Meals and Dressing 2005) for basic information on data transformations). The regression relationship may be based entirely on the current year's samples, or it may be based on samples gathered in previous years, or both. This method requires that there be a strong linear association between flow and concentration that does not change appreciably over the period of interest. If BMP implementation is expected to affect the relationship between flow and concentration, such relationships must be tracked carefully - if BMPs change the relationship, the concentration estimation procedure must be corrected.

Once the regression relationship is established, the regression equation is used to estimate concentrations for each day on which a sample was not taken, based on the mean daily flow for the day. The total load is calculated as the sum of the daily loads that are obtained by multiplying the measured or estimated daily concentration by the total daily discharge.

The goal of chemical sampling under this approach is to accurately characterize the relationship between flow and concentration. The monitoring program should be designed to obtain samples over the entire range of expected flow rates. If seasonal differences in the flow/concentration relationship are likely, the entire range of flows should be sampled in each season. In some cases, separate seasonal flow-concentration regressions may need to be developed and used to estimate seasonal loads. Examples of such flow-concentration regressions are shown in Figure 7-30 and example 7.9-1.

This approach is especially applicable to situations where continuous flow data already exist, e.g., from an ongoing USGS hydrologic station. Grab samples can be collected as needed and then associated with the appropriate flow observations. Economy is another significant advantage of this approach. After an initial intensive sampling period to develop the regression, it may be possible to maintain the regression model with ~20 samples a year for concentration, focusing on high-flow or critical season events. Software exists to calculate and manage this approach, e.g. Flux32 (Walker 1990, Soballe 2014). Flux32 is an interactive program designed for use in estimating the loadings of nutrients or other water quality components passing a tributary sampling station over a given period of time. Data requirements include (a) grab-sample nutrient concentrations, typically measured at a weekly to monthly frequency for a period of at least 1 year, (b) corresponding flow measurements (instantaneous or daily mean values), and (c) a complete flow record (mean daily flows) for the period of interest. Using six calculation techniques, Flux32 maps the flow/concentration relationship developed from the sample record onto the entire flow record to calculate total mass discharge and associated error statistics. An option to stratify the data into groups based upon flow, date, and/or season is also included. The USGS program LOADEST is also available and is widely used to estimate loads together with an estimate of precision using the regression approach. LOADEST includes an adjusted maximum likelihood estimation method that can be used for censored data sets and a least absolute deviation method to use when the regression residuals are not normally distributed. A web-based version of LOADEST program is available at https://engineering.purdue.edu/~ldc/LOADEST/. Another USGS load estimation calculation tool – FLUXMASTER – has been used in the SPARROW (SPAtially Referenced Regressions On Watershed attributes) watershed modeling technique to compute unbiased detrended estimates of long-term mean flux, and to provide an estimate of the associated standard error (Schwarz et al. 2006). These models include seasonal and temporal terms in their formulation that can improve the estimate of load; however, care is needed to ensure the model form is correct by reviewing the diagnostic plots.

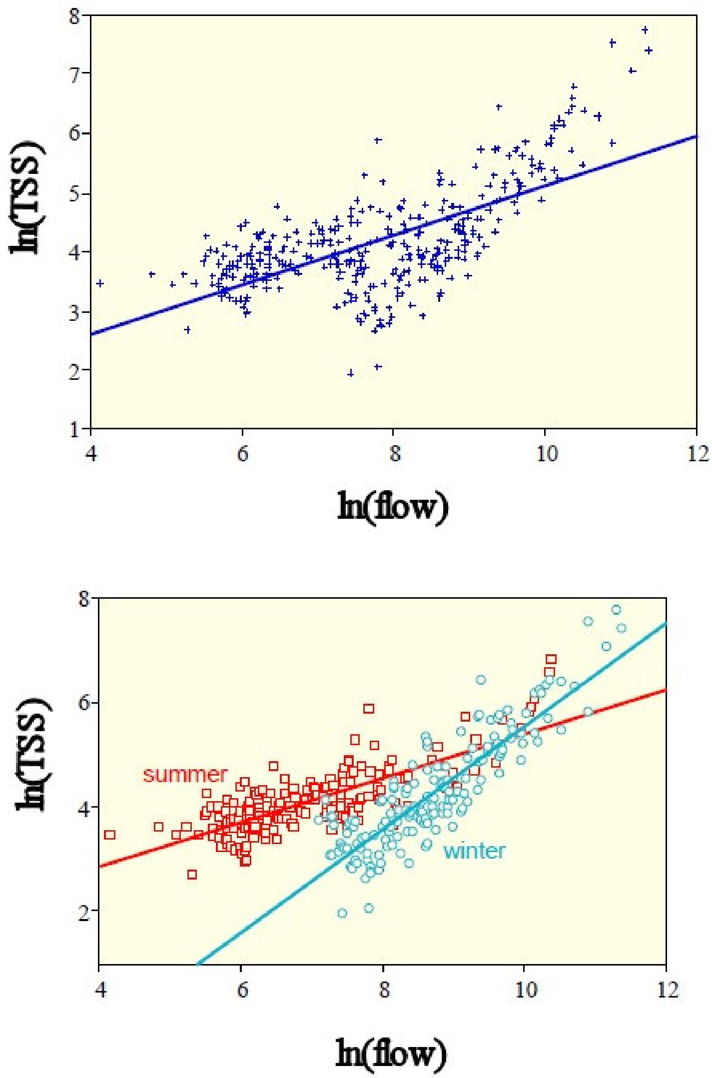


Figure 7-30. Flow-concentration regressions from the Maumee River, Ohio (Richards 1998). ***Top panel,*** **regression relationship between log of total suspended solids concentration and log of flow for the 1991 water year dataset;** ***Bottom panel*****, plot of same data divided into two groups based on time of year. Within each season, the regression model is stronger, has lower error, and provides a more accurate load estimate.**

Example 7.9-1. Mill Creek Watershed, PA NNPSMP

In this project, loads per unit area of nutrients and suspended sediment were estimated by combining the non-storm (i.e., low flow) and storm-flow loads (Galeone et al. 2006). Low-flow and storm-flow loads were computed using a multiple regression technique that included explanatory variables such as discharge, season, and time to estimate concentrations (and subsequently loads). Regressions were developed separately for low-flow and storm-flow periods, and for both low flow and storm flow, separate models were generated for the pre- and post-treatment periods for each site. Models were selected on the basis of the highest adjusted R^2 and residuals plots to detect trends, and all F-values had to exceed the value for the F distribution for the appropriate degrees of freedom and an alpha equal to 0.05.

Continuous discharge data for the four sites was first separated into low-flow and storm-flow periods using site-specific criteria defining a storm event. Sampled storms were reviewed to determine the typical rate of stage-height increase that initiated storm sampling. The recession and subsequent completion of storm sampling was also reviewed to determine the typical endpoint of storm sampling at each of the four sites. This information was used with 5- or 15-minute stage data to manually separate storm-flow discharge data from low-flow data.

For low-flow periods, a subset of the grab-sample data was used to develop the relation between constituent concentrations and explanatory variables. Prior to using the grab-sample data, the cumulative frequency distribution for each site was determined using the continuous discharge data for the entire period of record. Grab samples collected at flows above the 97th percentile were deleted prior to load analysis. With these higher flows deleted, the relation between constituent concentrations and explanatory variables was developed. The low-flow constituent concentrations were estimated on a daily basis using the daily-mean discharge data for low-flow periods. The estimated concentrations were multiplied by the daily-mean discharge to estimate daily loads.

Storm-flow loads for nutrients and suspended sediment were estimated by use of the mean discharge and mean constituent concentration for sampled storms. The mean discharge-concentration relation developed for sampled storms using regression analysis was used to predict the concentrations for unsampled storms. The mean discharge was calculated for unsampled storms using the 5- or 15-minute continuous-stage data for the sites. This mean discharge was applied to the predicted concentration to estimate constituent loads for unsampled storms. Increases in stage caused by snowmelt events were analyzed separately by subsetting the storm events sampled during snowmelts and using these regression relations to estimate loads for non-sampled snowmelt events. The percentage of the storms sampled at each site was somewhat dependent on the location of the surface-water site, ranging from about 50-60 percent at outlet sites and 35-45 percent at upstream sites where flashiness was greater and defined storms more frequent.

Constituent loads for each continuous surface-water site were estimated by summing the low-flow and storm-flow loads. The annual load data for the constituents were divided by the basin drainage areas to determine constituent yields. The percentage of the total yield in storm-flow was determined by summing the sampled and unsampled storm yields and dividing by the total yield. The remaining yield was attributed to low-flow periods. Data also were separated into pre- and post-treatment periods.

There are a few potential disadvantages to this approach. First, as mentioned earlier, potential changes or trends in the concentration-flow relationship – sometimes a goal of watershed projects – must be tracked. If the relationship changes a new regression model must be constructed. Second, the monitoring program must be managed to effectively capture the entire range of flows/conditions that occur; the use of data from fixed-interval time-based sampling is not appropriate for this purpose because of bias toward low flow conditions.

Hirsch et al. (2010) propose a weighted regression on time, discharge, and season (WRTDS) method that addresses some of these shortcomings. Principally, the WRTDS method relies on the same function regression structure as LOADEST; however, the fitted coefficients are allowed to vary with time. For example, the amplitude of the seasonal cycle could be relatively large in some periods of the record and then dampen to smaller cycles in other portions of the record. This is achieved through using a weighted regression that "windows in" on a portion of the record in time, flow and season. It is noteworthy that the researchers recommend that this method is primarily developed for data sets with more than 200 samples collected over 20 years. Like other flow adjustment tools there is a requirement of flow stationarity, that is, there isn't a basis for expecting a change in flow over time such as a new reservoir whether that change is observed over the entire year or just during a portion of the year. Extended dry or wet periods are simply an expected part of the long term record. WRTDS is generally intended for gradual changes that might be expected with NPS projects or sites that represent the cumulative effect of multiple point sources, and less for abrupt changes. WRTDS has been built into Exploration and Graphics for RivEr Trends (EGRET): An R-package for the analysis of long-term changes in water quality and streamflow. User guidance is available at https://github.com/USGS-R/EGRET/wiki although more current releases are available through R (R Core Team 2013). The WRTDS method was applied to eight monitoring sites on the Mississippi River investigating nitrate (Sprague et al. 2011) and compared to the more traditionally recommended ESTIMATOR by Moyer et al. (2012) in an evaluation using data from the Chesapeake Bay.

7.9.2.3 Ratio Estimators

The concept of ratio estimators is a powerful statistical tool for estimating pollutant load from continuous flow data and intermittent concentration data. Ratio estimators assume that there is a positive linear relationship between load and flow that passes through the origin. On days when chemistry samples are taken, the daily load is calculated as the product of grab-sampled concentration and mean daily flow, and the mean of these loads over the year is also calculated. The mean daily load is then adjusted by multiplying it by a flow ratio, which is derived by dividing the average flow for the year as a whole by the average flow for the days on which chemical samples were taken. A bias correction factor is included in the calculation, to compensate for the effects of correlation between discharge and load. The adjusted mean daily load is multiplied by 365 to obtain the annual load.

When used in a stratified mode (e.g., for distinct seasons), the same process is applied within each stratum, and the stratum load is calculated by multiplying the mean daily load for the stratum by the number of days in the stratum. The stratum loads are then summed to obtain the total annual load. The Beale Ratio Estimator is one technique, with an example provided by Richards (1998). Several formulas are available to calculate the number of samples (random or within strata) required to obtain a load estimate of acceptable accuracy based on known variance of the system. Stratification may improve the precision and accuracy of the load estimate by allocating more of the sampling effort to the aspects which are of greatest interest or which are most difficult to characterize because of great variability such as high flow seasons.

7.9.2.4 Comparison of Load Estimation Approaches

Although strongly driven by available resources, the monitoring program design (that should have included consideration of load estimation issues from the beginning), and the natural system itself, the choice of load estimation approach can make an enormous difference in the resulting load estimate.

In an analysis of total suspended solids data from the Maumee River in water year 1991, Richards (1998) demonstrated that different methods of load estimation applied to different datasets can result in substantially different estimates of pollutant load. Richards (1998) found that loads were often underestimated with the Beale Ratio Estimator and regression techniques, attributing this finding to missed high flow/TSS events and/or the estimation methods being biased toward low flow conditions. Notably, the Beale Ratio Estimator gave a load estimate closer to the true load (estimated through numeric integration) than did the regression method. For the full daily dataset, the single flow-concentration regression over the entire year appeared to seriously underestimate suspended solids load; while separating the data into summer and winter seasons improved the fit and the accuracy of the load estimate. In a summary of findings, Harmel et al. (2006) reported that the USGS regression method could result in annual constituent loads to within 10 percent of true loads in larger watersheds but no less than 30 percent for smaller watersheds.

Harmel and King (2005) and Harmel et al. (2006) concluded that flow-proportional, composite sampling was the most effective method to obtain high quality data for estimating loads from small agricultural watersheds. They concluded that composite sampling extended the sampler capacity with little effect on error, noting that intensive sampling strategies could achieve errors less than 10 percent. In their study, smaller sampling intervals should be used for constituents such as sediment which varies more during the course of a rainfall event in comparison to other constituents which vary less during a rainfall event.

Dolan et al. (1981) evaluated total phosphorus loadings to Lake Michigan from Grand River in 1976-77. They found that the Beale ratio estimator performed better than regression or other simplified calculations. Quilbé et al. (2006) evaluated a 1989-1995 nutrient and sediment data set from the Beaurivage River (Québec, Canada). They chose to estimate loadings with a Beale ratio estimator because they found that the correlation between flow and various water quality parameters was too weak to develop regression equations while noting that regression techniques would have been preferred if good correlations were found. Marsh and Waters (2009) also found few cases with strong correlations in their evaluation of 31 storm events in Queensland. They concluded that there was no clear best technique, but noted that the ratio methods were more robust and regression techniques worked well when there was a “tight” correlation. Using hourly model output, Zamyadi et al. (2007) found that the Beale ratio did not perform well in comparison to averaging and interpolation procedures.

Taking the above literature into account, this guidance recommends that numeric integration be used when the full time series of water quality and flow data are available as in the case of flow-proportional composited samples. Regression approaches are appropriate for incomplete water quality records if good correlations between water quality and flow exist, with the Beale ratio recommended otherwise. It is important to take into account stratification by flow regime, season, and other covariates for both regression and the Beale ratio.

7.9.3 Load Duration Curves

A particularly useful diagnostic tool for load estimation data is the load duration curve. Simply stated, a duration curve is a graph representing the percentage of time during which the value of a given parameter (e.g., flow, concentration, or load) is equaled or exceeded. A load duration curve is therefore a cumulative frequency plot of mean daily flows, concentrations, or daily loads over a period of record, with values plotted from their highest value to lowest without regard to chronological order. For each flow, concentration, or load value, the curve displays the corresponding percent of time (0 to 100) that the value was met or exceeded over the specified time – the flow, concentration, or load duration interval.

Extremely high values are rarely exceeded and have low duration interval values; very low values are often exceeded and have high duration interval values.

The process of using load duration curves generally begins with the development of a flow duration curve, using existing historical flow data (e.g., from a USGS gage), typically using mean daily discharge values. A basic flow duration curve runs from high to low along the x-axis, as illustrated in Figure 7-31. The x-axis represents the duration or percent of time, as in a cumulative frequency distribution. The y-axis represents the flow value (e.g., ft^3/sec (cfs)) associated with that percent of time. Figure 7-31 illustrates that the highest observed flow for the period of record was about 5,400 cfs, while the lowest flow was 6 cfs. The median flow – the flow exceeded 50 percent of the time – was about 200 cfs.

In the next step, a load duration curve is created from the flow duration curve by multiplying each of the flow values by the applicable numeric water quality target (usually a water quality criterion) and a unit conversion factor, then plotting the results as for the flow duration curve. The x-axis remains as the flow duration interval, and the y-axis depicts the load rather than the flow. This curve represents the allowable load (e.g., the TMDL) at each flow condition over the full range of observed flow. An example is shown in Figure 7-32 for the same site as shown in the flow duration curve, using a target of 0.05 mg/L total P. Then, observed P load observations associated with the flow intervals are plotted along the same axes. Points located above the curve represent times when the actual loading is exceeding the target load, while those plotting below the curve represent times when the actual loading is less than the target load.

A key feature of load duration curve analysis is that the pattern of loads – and impairment – can be easily visualized over the full range of flow conditions. Because flow variations usually correspond to seasonal patterns, this feature can address the requirement that TMDLs account for seasonal variations. The pattern of observed loads exceeding target loads can be examined to see if impairments occur only at high flows, only during low flows, or across the entire range of flow conditions. A common way to look at a load duration curve is by dividing it into zones representing, for example: high flows (0-10 percent flow duration interval), moist conditions (10-40 percent), mid-range flows (40-60 percent), dry conditions (60-90 percent), and low flows (90-100 percent). Data may also be grouped by season (e.g., spring runoff versus summer base flow). Sometimes, analysis of the load duration curve can provide insight on the source of pollutant loads. Measured loads that plot above the curve during flow duration intervals above 80 percent (low flow conditions), for example, may suggest point sources that discharge continuously during dry weather. Conversely, measured loads that plot above the curve during flow duration intervals of about 10 to 70 percent tend to reflect wet weather contributions by NPS such as erosion, washoff, and streambank erosion. Figure 7-32 illustrates that allowable total P loads in the Sevier River were exceeded during all flow intervals, and that P concentrations were independent of flow.

It should be noted that an individual load duration curve applies only to the point in the stream where the data were collected. A load duration curve developed at a watershed outlet station (e.g., for a TMDL) applies only to loads observed at that point. If significant pollution sources exist upstream, a single load duration analysis at the watershed outlet can underestimate the extent of impairment in upstream segments. For this reason, it is usually wise to develop multiple load duration curves throughout the watershed to address the spatial distribution of impairments. Such an exercise can also be useful in targeting land treatment to critical watershed source areas.

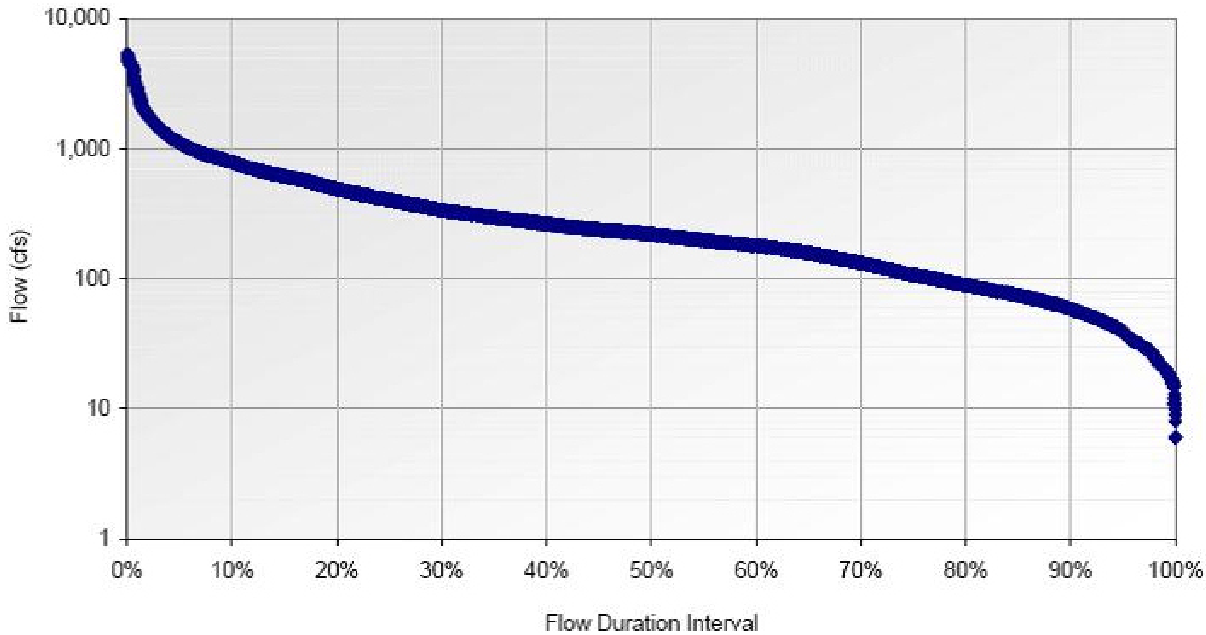


Figure 7-31. Flow duration curve for the Sevier River near Gunnison, UT, covering the period January 1977 through September 2002

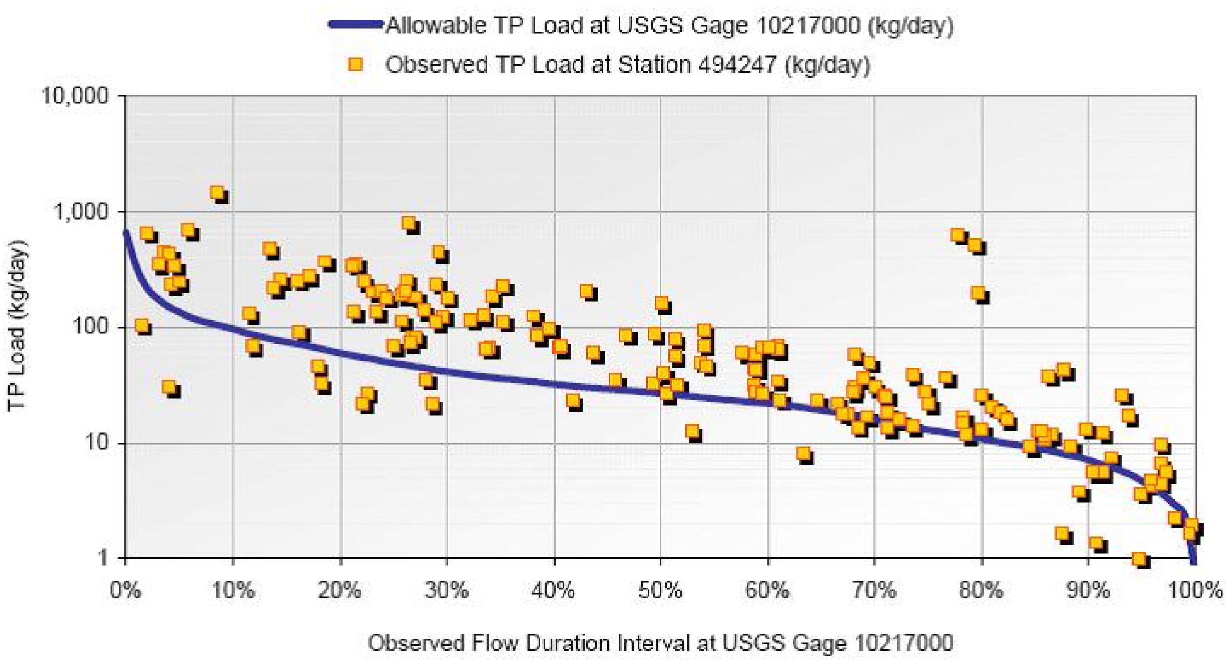


Figure 7-32. Load duration curve for the Sevier River near Gunnison, UT, January 1977 through September 2002. Blue line represents allowable total P load calculated as the product of each observed flow duration interval and the target total P concentration of 0.05 mg/L. Yellow points represent observed total P loads at the same flow duration intervals.

For more detailed discussion of load duration curves, particularly their application to the TMDL process, refer to:

- USEPA. 2007. *An Approach for Using Load Duration Curves in the Development of TMDLs*

7.9.4 Assessing Load Reductions

The same statistical tools recommended for flow and concentration data in section 7.8.2 and elsewhere in this chapter can be used to analyze program effectiveness with regard to load reductions. For example, loads might be estimated on a weekly basis using numeric integration and flow-proportional, composite sample data. Under a paired-watershed approach, the weekly-paired loads would be grouped as pre- and post-treatment and analyzed using ANCOVA.

For comparisons of annual loadings, the analyst will have limited data to perform analyses (i.e., one annual loading value per site-year) and will be generally limited to reporting simple change in loading and drawing anecdotal comparisons to the control watershed. Normalizing the loadings based on watershed size, annual rainfall, and other covariates might prove helpful.

Depending on the watershed and the types of installed BMPs, it is also appropriate to compare storm loadings from individual storms before and after BMP implementation in a single watershed. The particular challenge here is to control for other covariates and select/analyze storms of a certain size (e.g., rainfall between 2.5-5.0 cm) and occurring at key times during the year (e.g., within 6 weeks of spring planting). This type of analysis might also be limited to drawing simple comparisons due to sample size.

7.10 Statistical Software

Modern computers and software packages make it simple to perform the statistical analyses described in this chapter. Most standard spreadsheet programs include basic statistical functions and graphing capabilities, but more sophisticated and powerful statistical software packages might be needed for advanced analyses such as ANCOVA or cluster analysis. An extensive list and comparison of statistical software packages is available at Wikipedia. Practical Statistics, a web site maintained by Dennis Helsel, provides a more environmental-centric review of low-cost software tools. Table 7-9 lists some examples and websites to visit for more information about the many statistical packages available.

Table 7-9. Sampling of available statistics software packages

Package Name	Web Site URL
Analyse-It (add in for MS Excel)	http://www.analyse-it.com
DataDesk	http://www.datadesk.com
JMP	http://www.jmp.com/en_gb/software.html
Mathematica	http://www.wolfram.com/mathematica/
MATLAB	http://www.mathworks.com/products/matlab/
MINITAB	https://www.minitab.com/en-us/
R	https://www.r-project.org/
SAS/Stat, SAS/Insight	http://www.sas.com/technologies/analytics/statistics/index.html
SPSS	http://www.spss.com/spss/
SYSTAT	http://www.systat.com/products/Systat/
WINKS	http://www.texasoft.com/

7.11 References

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons, New York.

Bernstein, B.B. and J. Zalinski. 1983. An optimum sampling design and power tests for environmental biologists. *Journal of Environmental Management* 16(1):35-43.

Bishop, P.L., W.D. Hively, J.R. Stedinger, M.R. Rafferty, J.L. Lojpersberger, and J.A. Bloomfield. 2005. Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. *Journal of Environmental Quality* 34:1087–1101.

Box, G.E.P. and D.R. Cox. 1964. An analysis of transformations - series B (methodological). *Journal of the Royal Statistical Society* 26(2):211-252.

Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis: Forecasting And Control*. Revised Edition. Holden-Day, Oakland, CA.

Carpenter, S.R., T.M. Frost, D. Heisey, and T.K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology* 70(4):1142-1152.

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

Clausen, J.C. 2007. *Jordan Cove Watershed Project Final Report*. University of Connecticut, College of Agriculture and Natural Resources, Department of Natural Resources Management and Engineering. Accessed January 8, 2016. http://jordancove.uconn.edu/jordan_cove/publications/final_report.pdf.

Clausen, J.C. and K.N. Brooks. 1983. Quality of runoff from Minnesota peatlands: II. a method for assessing mining impacts. *Water Resources Bulletin* 19(5):769-772.

Clausen, J.C. and J. Spooner. 1993. *Paired Watershed Study Design*. 841-F-93-009. Prepared for S. Dressing, U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 12, 2016.

Cleveland, W.S. 1993. *Visualizing Data.* AT&T Bell Laboratories/Hobart Press, Murray Hill, NJ/Summit, NJ.

Clifford, R., Jr., J.W. Wilkinson, and N.L. Clesceri. 1986. Statistical Assessment of a Limnological Data Set. In *Statistical Aspects of Water Quality Monitoring, Proceedings of the Workshop Held at the Canada Centre for Inland Waters*, October 7-10, 1985, Volume 27 of Developments in Water Science series, ed. A.H. El-Shaarawi and R.E. Kwiatkowski. Elsevier Publishers, New York. pp. 363-380.

Dolan, D.M., A.K. Yui and R.D. Geist. 1981. Evaluation of river load estimation methods for total phosphorus. *Journal of Great Lakes Research* 7(3):207-214.

Dolan, D.M. and K.P. McGunagle. 2005. Lake Erie total phosphorus loading analysis and update: 1996-2002. *Journal of Great Lakes Research* 31(Supplement 2):11-22. Accessed March 24, 2016. http://www.cee.mtu.edu/~nurban/classes/ce5508/2007/Readings/dolan05.pdf.

Drake, D. 1999. *Multivariate Analysis of Fish and Environmental Factors in the Grande Ronde Basin of Northeastern Oregon*. Oregon Department of Environmental Quality, Biomonitoring Section, Laboratory Division, Portland. Accessed March 24, 2016. http://www.deq.state.or.us/lab/techrpts/docs/Bio012.pdf.

Elliott, A.C. 2012. *Descriptive Statistics Using Microsoft Excel*. Texa Soft Mission Technologies, Cedar Hill, TX. Accessed March 24, 2016. http://www.stattutorials.com/EXCEL/EXCEL-DESCRIPTIVE-STATISTICS.html.

Erickson, A.J., P.T. Weiss, J.S. Gulliver, R.M. Hozalski. 2010a. Analysis of Individual Storm Events, In *Stormwater Treatment: Assessment and Maintenance* ed. J.S. Gulliver and A.J. Erickson, University of Minnesota, St. Anthony Falls Laboratory. Minneapolis, MN. Accessed March 24, 2016. http://stormwaterbook.safl.umn.edu/

Erickson, A.J., P.T. Weiss, J.S. Gulliver, and R.M. Hozalski. 2010b. Analysis of Long-Term Performance. In *Stormwater Treatment: Assessment and Maintenance*. ed. J.S. Gulliver, A.J. Erickson, and P.T. Weiss. University of Minnesota, St. Anthony Falls Laboratory. Minneapolis, MN. Accessed March 24, 2016. http://stormwaterbook.safl.umn.edu/content/analysis-long-term-performance.

Farnsworth, R.K. and E.S. Thompson. 1982. *Mean Monthly, Seasonal, and Annual Pan Evaporation for the United States*. Technical Report NWS 34. National Oceanic and Atmospheric Administration, National Weather Service. Accessed March 24, 2016. http://www.nws.noaa.gov/oh/hdsc/PMP_related_studies/TR34.pdf.

Fuller, W.A. 1976. *Introduction to Statistical Time Series*. John Wiley & Sons, Inc. New York.

Galeone, D.G., R.A. Brightbill, D.J. Low, and D.L. O'Brien. 2006. *Effects of Streambank Fencing of Pastureland on Benthic Macroinvertebrates and the Quality ofSurface Water and Shallow Ground Water in the Big Spring Run Basin of Mill Creek Watershed, Lancaster County, Pennsylvania, 1993-2001*. Scientific Investigations Report 2006-5141. U. S. Geological Survey, Reston, VA. Accessed March 24, 2016. http://pubs.usgs.gov/sir/2006/5141/.

Geosyntec and WWE (Wright Water Engineers, Inc.). 2009. *Urban Stormwater BMP Performance Monitoring*. Prepared for U.S. Environmental Protection Agency, Water Environment Research Foundation, Federal Highway Administration, and Environmental and Water Resources Institute of the American Society of Civil Engineers, by Geosyntec Consultants and Wright Water Engineers, Inc., Washington, DC. Accessed March 24, 2016. http://www.bmpdatabase.org/Docs/2009%20Stormwater%20BMP%20Monitoring%20Manual.pdf.

Gilliom, R.J., R.M. Hirsch, and E.J. Gilroy. 1984. Effect of censoring trace-level water-quality data on trend-detection. *Environmental Science and Technology* 18(7):530-535.

Grabow. G.L. 1999. *Summary of Analyses Performed for Sycamore Creek Section 319 NNMP Project*. North Carolina State University, NCSU Water Quality Group, Raleigh, NC. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1998. Detecting water quality changes before and after BMP implementation: use of a spreadsheet for statistical analysis. *NWQEP Notes* 92:1-

9. North Carolina State University Cooperative Extension, Raleigh. Accessed March 24, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/92.pdf.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1999. Detecting water quality changes before and after BMP implementation: use of SAS for statistical analysis. *NWQEP Notes* 93:1-11. North Carolina State University Cooperative Extension, Raleigh. Accessed March 24, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/93.pdf.

Harmel, R.D. and K.W. King. 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. *Transactions of the American Society of Agricultural and Biological Engineers* 48(5): 1713-1721.

Harmel, R.D., K.W. King, B.E. Haggard, D.G. Wren, and J.M. Sheridan. 2006. Practical guidance for discharge and water quality collection in small watersheds. *Transactions of the American Society of Agricultural and Biological Engineers* 49(4):937-948.

Harstine, L.J. 1991. *Hydrologic Atlas for Ohio: Average Annual Precipitation, Temperature, Streamflow, and Water Loss for a 50-Year Period, 1931-1980*. Water Inventory Report No. 28. Ohio Department of Natural Resource, Division of Water, Ground Water Resources Section.

Helsel, D.R. 2012. *Statistics for Censored Environmental Data Using Minitab and R*. 2nd ed. Wiley and Sons, New York.

Helsel, D.R. and T.A. Cohn. 1988. Estimation of descriptive statistics for multiply censored water quality data. *Water Resources Research* 24(12):1997-2004.

Helsel, D.R., and R.M. Hirsch. 2002. Statistical Methods in Water Resources. Book 4, Chapter A3 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 10, 2016. http://pubs.usgs.gov/twri/twri4a3/.

Hewlett, J.D. and L. Pienaar. 1973. Design and Analysis of the Catchment Experiment, In *Proceedings of a Symposium on Use of Small Watersheds in Determining Effects of Forest Land Use on Water Quality*, ed. E. H. White, University of Kentucky, Lexington, KY, May 22-23, 1973. Accessed January 26, 2016. http://coweeta.uga.edu/publications/hewlett_73_catchment.pdf.

Hibbert, A.R. 1969. Water yield changes after converting a forested catchment to grass. *Water Resources Research* 5(3):634-640.

Hirsch, R.M. 1988. Statistical methods and sampling design for estimating step trends in surface water quality. *Water Resources Research* 24:493-503.

Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20(6):727-732.

Hirsch, R.M., R.B. Alexander, and R.A. Smith.1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research* 27:803-813.

Hirsch, R.M. D.L. Moyer, and S.A. Archfield. 2010. Weighted regressions on time, discharge and season (WRTDS), with an application to Chesapeake Bay river inputs. *Journal of the American Water Resources Association* 46(5):857-880.

Hoorman, J., T.Hone, T.Sudman Jr., T.Dirksen, J. Iles, and K.R. Islam. 2008. Agricultural impacts on lake and stream water quality in Grand Lake St. Marys, Western Ohio. *Water, Air, and Soil Pollution* 193: 309-322.

Hornbeck, J.W., R.S. Pierce, and C.A. Federer. 1970. Streamflow changes after forest clearing in New England. *Water Resources Research* 6(4):1124-1132.

Jambu, M. 1991. *Exploratory and Multivariate Data Analysis.* Academic Press, Inc., Boston.

Kilpatrick, F.A. and J.F. Wilson, Jr. 1989. Measurement of Time of Travel in Streams by Dye Tracing. Book 3, Chapter A9 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed March 24, 2016. http://pubs.usgs.gov/twri/twri3-a9/pdf/twri_3-A9.pdf.

Lettenmaier. D.P. 1976. Detection of trends in water quality data from records with dependent observations. *Water Resources Research* 12:1037-1046.

Lettenmaier, D.P. 1978. Design considerations for ambient stream quality monitoring. *Water Resources Bulletin* 14(4):884-902.

Lewis, J. 2006. *Fixed and Mixed-Effects Models for Multi-Watershed Experiments*. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, Arcata, CA. Accessed March 24, 2016. http://www.fs.fed.us/psw/publications/4351/Lewis06.pdf.

Line, D.E., W.A. Harman, G.D. Jennings, E.J. Thompson, and D.L. Osmond. 2000. Nonpoint-source pollutant load reductions associated with livestock exclusion. *Journal of Environmental Quality* 29: 1882–1890.

Loftis, J.C. and R.C. Ward. 1980a. Sampling frequency selection for regulatory water quality monitoring. *Water Resources Bulletin* 16:501-507.

Loftis, J.C. and R.C.Ward, 1980b. Water quality monitoring – some practical sampling frequency considerations. *Environmental Management* 4:521-526.

MacKenzie, M.C., R.N. Palmer, and S.P. Millard. 1987. Analysis of statistical monitoring network design. *Journal of Water Resources Planning and Management* 113(5):599-615.

Marsh, N. and D. Waters. 2009. Comparison of Load Estimation Methods and Their Associated Error. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, ed. R.S. Anderssen, R.D. Braddock and L.T.H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 3322-3328. Accessed March 24, 2016. http://mssanz.org.au/modsim09/I4/marsh_I4.pdf.

Martz, E. 2013. *Enough Is Enough! Handling Multicollinearity in Regression Analysis*. The Minitab Blog, Minitab Inc. Accessed March 24, 2016. http://blog.minitab.com/blog/understanding-statistics/handling-multicollinearity-in-regression-analysis.

Matalas, N.C. and W.B. Langbein. 1962. Information content of the mean. *Journal of Geophysical Research* 67(9):3441-3448.

Meals, D.W. 1987. Detecting changes in water quality in the LaPlatte River watershed following implementation of BMPs. *Lake and Reservoir Management* 3(1):185-194.

Meals, D.W. 2001. *Lake Champlain Basin Agricultural Watersheds Section 319 National Monitoring Program Project, Final Project Report: May, 1994-September, 2000*. Vermont Department of Environmental Conservation, Waterbury, VT.

Meals, D.W. and S.A. Dressing. 2005. *Monitoring Data – Exploring Your Data, the First Step*, Tech Notes #1, July 2005. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W. and D.C. Braun. 2006. Demonstration of methods to reduce E.coli runoff from dairy manure application sites. *Journal of Environmental Quality* 35:1088-1100.

Meals, D.W. and S.A. Dressing. 2008. *Surface Water Flow Measurement for Water Quality Monitoring Projects*, Tech Notes #3, March 2008. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W., J. Spooner, S.A. Dressing, and J.B. Harcum. 2011. *Statistical Analysis for Monotonic Trends*, Tech Notes #6, September 2011. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Meals, D.W., R.P. Richards, and S.A. Dressing. 2013. *Pollutant Load Estimation forWater Quality Monitoring Projects.* Tech Notes #8. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Minitab. 2016. Minitab 17. Minitab Inc., State College, PA. Accessed January 22, 2016. http://www.minitab.com/en-US/products/minitab/.

Mosteller, F. and J.W. Tukey. 1977. *Data Analysis and Regression: Second Course in Statistics*. Addison-Wesley Pub. Co., Reading, MA.

Moyer, D.L., R.M. Hirsch, and K.E. Hyer. 2012. *Comparison of Two Regression-Based Approaches for Determining Nutrient and Sediment Fluxes and Trends in the Chesapeake Bay Watershed*. Scientific Investigations Report 2012-5244. U.S. Geological Survey, Reston, VA. Accessed March 24, 2016. http://pubs.usgs.gov/sir/2012/5244/.

NCDENR (North Carolina Department of Environment and Natural Resources). 2016. *Sanitary Survey*. North Carolina Department of Environment and Natural Resources, Division of Marine Fisheries, Raleigh, NC. Accessed March 24, 2016. http://portal.ncdenr.org/web/mf/sanitary-survey.

Newbold, J.D., S. Herbert, B.W. Sweeney, and P. Kiry. 2009. Water quality functions of a 15-year-old riparian forest buffer system. *NWQEP Notes* 130:1-9. North Carolina State University Cooperative Extension, Raleigh. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/issues/notes130.pdf.

Newbold, J.D., S. Herbert, and B.W. Sweeney. 2009. *Mitigation of Nonpoint Pollution by a Riparian Forest Buffer in an Agricultural Watershed of the Mid-Atlantic Piedmont: Stroud Preserve Watersheds National Monitoring Project Final Report*. Stroud Water Research Center, Avondale PA. Accessed March 24, 2016. http://www.stroudcenter.org/research/projects/StroudPreserve/StroudNMPFinalReport2009.pdf.

ODNR (Ohio Department of Natural Resources). 2013. *Grand Lake St. Marys State Park*. Ohio Department of Natural Resources. Accessed March 24, 2016. http://parks.ohiodnr.gov/grandlakestmarys.

OWEB (Oregon Watershed Enhancement Board). 1999. *Oregon Aquatic Habitat Restoration and Enhancement Guide*. Oregon Watershed Enhancement Board, Salem, OR. Accessed April 25, 2016. http://www.oregon.gov/OWEB/docs/pubs/habguide99-complete.pdf.

Palmer, R.N. and M.C. MacKenzie. 1985. Optimization of water quality monitoring networks. *Journal of Water Resources Planning and Management* 111(4):478-493.

Perkins, W.W., E.B. Welch, J. Frodge, and T. Hubbard. 1997. A zero degree of freedom total phosphorus model: application to Lake Sammamish, Washington. *Lake and Reservoir Management* 13:131-141.

Porter, P.S., R.C. Ward, and H.F. Bell. 1988. The detection limit, water quality monitoring data are plagued with levels of chemicals that are too low to be measured precisely. *Environmental Science and Technology* 22:856-861.

Primrose, N.L. 2003. *Report on Nutrient Synoptic Surveys in the Corsica River Watershed, Queen Annes County, Maryland, April 2003*. Maryland Department of Natural Resources, Watershed Services, Annapolis, MD.

Quilbé, R., A.N. Rousseau, M. Duchemin, A. Poulin, G. Gangbazo, J. Villeneuve. 2006. Selecting a calculating method to estimate sediment and nutrient loads in streams: application to the Beaurivage River (Québec, Canada). *Journal of Hydrology* 326(2006):295-310.

R Core Team. 2013. *The R Project for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Accessed April 25, 2016. http://www.R-project.org/.

Richards, R.P. 1998. *Estimation of Pollutant Loads in Rivers and Streams: A Guidance Document for NPS Programs*. Prepared for U.S. EPA Region VIII, by Heidelberg University, Water Quality Laboratory, Tiffin, OH. Accessed February 5, 2016. http://141.139.110.110/sites/default/files/jfuller/images/Load_Est1.pdf.

Richards, R.P. and J. Holloway. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research* 23:1939-1948.

Roseboom, D., T. Hill, J. Rodsater, J. Beardsley, and L. Duong. 1999. *Evaluation of Sediment Delivery to Lake Pittsfield After Best Management Practice Implementation-National Watershed Monitoring Project*. Illinois Environmental Protection Agency, Springfield, IL.

Rosgen, D.L. 1997. A Geomorphological Approach to Restoration of Incised Rivers. In *Proceedings of the Conference on Management of Landscapes Disturbed by Channel Incision*, ed. S.S.Y. Wang, E.J. Langendoen and F.D. Shields, Jr. 1997. Accessed March 24, 2016.

http://www.wildlandhydrology.com/assets/A_Geomorphological_Approach_to_Restoration_of_Incised_Rivers.pdf.

SAS Institute. 1985. *SAS® User's Guide; Statistics. Version 5 Edition*. SAS Institute Inc., Cary, North Carolina 27513.

SAS Institute. 2010. *SAS® Version 9.2, SAS/ETS*. SAS Institute Inc., Cary, NC.

SAS Institute. 2016a. *Collinearity Diagnostics, SAS 9.2 Documentation, SAS/STAT(R) 9.22 User's Guide*. SAS Institute, Inc., Cary, NC. Accessed March 24, 2016. http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_reg_sect038.htm.

SAS Institute. 2016b. *JMP® Version 9.0.2, 2010*. SAS Institute Inc., Cary, NC. Accessed March 24, 2016. http://www.jmp.com/software/.

SAS Institute. 2016c. *Multiple Linear Regression, SAS 9.2 Documentation.* SAS Institute, Inc., Cary, NC. Accessed March 24, 2016. http://support.sas.com/documentation/cdl/en/anlystug/58352/HTML/default/viewer.htm#chap11_sect3.htm.

SAS Institute. 2016d. *The Autoreg Procedure, SAS/ETS(R) 9.2 User's Guide*. SAS Institute, Inc., Cary, NC. Accessed March 31, 2016. http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#autoreg_toc.htm.

Schilling, K.E., and J. Spooner. 2006. Effects of watershed-scale land use change on stream nitrate concentrations. *Journal of Environmental Quality* 35:2132–2145.

Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith. 2006. *The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation*. Techniques and Methods 6-B3. U.S. Geological Survey, Reston, VA. Accessed March 24, 2016. http://pubs.usgs.gov/tm/2006/tm6b3/.

Simpson, T. and S. Weammert. 2009. *Developing Best Management Practice Definitions and Effectiveness Estimates for Nitrogen, Phosphorus, and Sediment in the Chesapeake Bay Watershed*. Final Report. University of Maryland Mid-Atlantic Water Program. Accessed March 24, 2016. http://archive.chesapeakebay.net/pubs/BMP_ASSESSMENT_REPORT.pdf.

Smith, R.V., S.D. Lennox, and J.S. Bailey 2003. Halting the upward trend in soluble phosphorus transported from a grassland catchment. *Journal of Environmental Quality* 32:2334–2340.

Snedecor, G.W. and W.G. Cochran. 1989. *Statistical Methods*. 8th Edition. Iowa State University Press, Ames, IA.

Soballe, D.M. 2014. Flux32. Developed in conjunction with Minnesota Pollution Control Agency, by U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS. Accessed March 24, 2016.

Spooner, J., C.J. Jamieson, R.P. Maas, and M.D. Smolen. 1987. Determining statistically significant changes in water pollutant concentrations. *Lake Reservoir Management* 3:195–201.

Spooner, J., S.A. Dressing, and D.W. Meals. 2011a. *Minimum detectable change analysis.* Tech Notes #7. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Spooner, J., L.A. Szpir, D.E. Line, D. Meals, G.L Grabow, D.L. Osmond and C. Smith. 2011b. *2011 Summary Report: Section 319 National Monitoring Program Projects, National Nonpoint Source Watershed Project Studies*. North Carolina State University, Biological and Agricultural Engineering Department, NCSU Water Quality Group, Raleigh, NC. Accessed March 15, 2016. http://www.bae.ncsu.edu/programs/extension/wqg/319monitoring/toc.html.

Sprague, L.A., R.M. Hirsch, and B.T. Aulenbach. 2011. Nitrate in the Mississippi River and its tributaries, 1980 to 2008: are we making progress? *Environmental Science and Technology* 45 (17):7209–7216.

Statistics Solutions. 2016. *Correlation (Pearson, Kendall, Spearman)*. Statistics Solutions, Clearwater, FL. Accessed March 24, 2016. http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/.

Stuntebeck, T.D. 1995. *Evaluating Barnyard Best Management Practices in Wisconsin Using Upstream-Downstream Monitoring*. Fact Sheet FS-221-95. U.S. Geological Survey, Madison, WI. Accessed February 5, 2016. http://pubs.usgs.gov/fs/1995/fs221-95/.

Stuntebeck, T.D. and R.T. Bannerman. 1998. *Effectiveness of Barnyard Best Management Practices in Wisconsin*. Fact Sheet FS-051-98. U.S. Geological Survey, Madison, WI. Accessed March 24, 2016. http://pubs.er.usgs.gov/publication/fs05198.

Suppnick, J. 1999. *Water Chemistry Trend Monitoring in Sycamore Creek and Haines Drain, Ingham County, Michigan 1990-1997*. Staff report MI/DEQ/SWQD-99-085. Michigan Department of Environmental Quality, Surface Water Quality Division. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Tetra Tech. 2013. *Preliminary Assessment of Effectiveness of the 2012 Alum Application—Grand Lake St. Marys*. Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. http://www.lakeimprovement.com/sites/default/files/GLSM%20Alum%20Report%2002202013.pdf.

Thas O., L. Van Vooren, and J.P. Ottoy. 1998. Nonparametric test performance for trends in water quality with sampling design applications. *Journal of the American Water Resources Association* 34(2):347-357.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Co., Reading, MA.

USDA-NRCS (U.S. Department of Agriculture-Natural Resources Conservation Service). 2012. *Assessment of the Effects of Conservation Practices on Cultivated Cropland in the Upper Mississippi River Basin*. U.S. Department of Agriculture, Natural Resources Conservation Service. Accessed March 24, 2016. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?&cid=nrcs143_014161.

USEPA (U.S. Environmental Protection Agency). 1997a. *Linear Regression for Nonpoint Source Pollution Analysis*. EPA-841-B-97-007. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 24, 2016.

USEPA (U.S. Environmental Protection Agency). 1997b. *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls*. EPA 841-B-96-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-guidance-determining-effectiveness-nonpoint.

USEPA (U.S. Environmental Protection Agency). 2007. *An Approach for Using Load Duration Curves in the Development of TMDLs*. EPA 841-B-07-006. U.S. Environmental Protection Agency, Office of Wetlands, Oceans and Watersheds, Watershed Branch, Washington, DC. Accessed March 24, 2016. https://www.epa.gov/tmdl/approach-using-load-duration-curves-development-tmdls.

USEPA (U.S. Environmental Protection Agency). 2008. *Handbook for Developing Watershed Plans to Restore and Protect Our Waters.* EPA 841-B-08-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 24, 2016. http://www.epa.gov/polluted-runoff-nonpoint-source-pollution/handbook-developing-watershed-plans-restore-and-protect.

USEPA (U.S. Environmental Protection Agency). 2010. *Causal Analysis/Diagnosis Decision Information System (CADDIS)*. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Accessed March 24, 2016. http://www.epa.gov/caddis.

USEPA (U.S. Environmental Protection Agency). 2013. Information on impaired waters and total maximum daily loads. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed March 24, 2016. http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/index.cfm.

USF (University of South Florida). n.d. *Collinearity*. University of South Florida , College of Arts and Sciences, Tampa, FL. Accessed March 24, 2016. http://faculty.cas.usf.edu/mbrannick/regression/Collinearity.html.

Vollenweider, R.A. 1976. Advances in defining critical loading levels for phosphorus in lake eutrophication. *Memorie dell'Istituto italiano di idrobiologia dott. Marco De Marchi* 33:53-83.

Vollenweider, R.A., and J. Kerekes. 1982. *Eutrophication of Waters: Monitoring, Assessment and Control*. Organization for Economic Co-Operation and Development, Paris.

Walker, J.F. 1994. Statistical techniques for assessing water quality effects of BMPs. *Journal of Irrigation and Drainage Engineering* 120(2):334-347.

Walker, W.W. 1990. *Flux Stream Load Computations. DOS Version 4.4*. Prepared for U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS.

Walker, W.W. 1999. *Simplified Procedures for Eutrophication Assessment and Prediction: User Manual*. Prepared for U.S. Army Corps of Engineers, Water Operations Technical Support Program, Instruction Report W-96-2, September 1996 (Updated April 1999). Accessed March 24, 2016. http://www.wwwalker.net/bathtub/Flux_Profile_Bathtub_DOS_1999.pdf.

White, W., J. Beardsley, and S. Tomkins. 2011. *Waukegan River Illinois National Nonpoint Source Monitoring Program Project*. Contract Report 2011-01. , University of Illinois at Urbana-

Champaign, Institute of Natural Resource Sustainability, Illinois State Water Survey, Champaign, Illinois. Accessed March 24, 2016. http://www.isws.illinois.edu/pubdoc/CR/ISWSCR2011-01.pdf.

Whitfield, P.H. 1983. Evaluation of water quality sampling locations on the Yukon River. *Water Resources Bulletin* 19(1):115-121.

Whitfield, P.H. and P.F. Woods. 1984. Intervention analysis of water quality records. *Water Resources Bulletin* 20(5):657-668.

Wilm, H.G. 1949. How long should experimental watersheds be calibrated? *Transactions of the American Geophysical Union* 30(2):272-278.

Zamyadi, A., J. Gallichand, and M. Duchemin. 2007. Comparison of methods for estimating sediment and nitrogen loads from a small agricultural watershed. *Canadian Biosystems Engineering* 49:1.27-1.36.

8 Quality Assurance and Quality Control

By S.A. Lanberg, J.G. O'Donnell, and J.B. Harcum

8.1 Introduction

Quality assurance and quality control (QA/QC) are commonly thought of as procedures used in the laboratory to ensure that all analytical measurements made are accurate. Yet QA/QC extends beyond the laboratory and includes a wide range of issues that nonpoint source (NPS) managers consider when addressing the challenges of developing a monitoring program (see chapters 2 and 3). When considered independently from monitoring program design, QA/QC may seem burdensome. Yet the purpose of QA/QC is the same as a well-intentioned NPS manager, which is to ensure that the monitoring data generated are complete, accurate, and suitable for the intended purpose. By integrating certain QA/QC aspects with monitoring program design, NPS managers can reduce repetition and ultimately reduce total costs by developing a more efficient monitoring design.

The remainder of this section defines QA/QC, discusses their value in NPS monitoring programs, and explains EPA's policy on these topics. Section 8.2 provides an overview of the Data Quality Objectives (DQO) process. EPA recommends that organizations use the DQO process to systematically plan their monitoring programs. Typically, written QA/QC documentation takes the form of a quality assurance project plan (QAPP). As discussed in section 8.3, a QAPP details the technical activities and QA/QC procedures that should be implemented to ensure the data meet the specified standards.

The QAPP should identify who will be involved in the project and their responsibilities; the nature of the study or monitoring program; the questions to be addressed or decisions to be made based on the data collected; where, how, and when samples will be taken and analyzed; the requirements for data quality; the specific activities and procedures to be performed to obtain the requisite level of quality (including QC checks and oversight); how the data will be managed, analyzed, and checked to ensure that they meet the project goals; and how the data will be reported. The QAPP should be implemented and maintained throughout a project.

Sections 8.4, 8.5, and 8.6 provide more specific information for preparing QAPPs with respect to field operations, laboratory operations, and data and reporting requirements, respectively. Although there are many commonalities, QAPP development to support modeling and secondary data usage is beyond the scope of this chapter. The reader is referred to CREM (2009) and USEPA (2002b) for guidance on the development and application of environmental models and related QAPPs. EPA also provides guidance about the evaluation of existing (secondary) data quality (USEPA 2012) and information needed to develop QAPPs for secondary data projects (USEPA 2008a).

8.1.1 Definitions of Quality Assurance and Quality Control

8.1.1.1 Quality assurance:

An integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the client (USEPA 2001c).

8.1.1.2 Quality control:

The overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer; operational techniques and activities that are used to fulfill requirements for quality (USEPA 2001c).

In a laboratory setting, QC procedures include the regular inspection of equipment to ensure it is operating properly and the collection and analysis of blank, duplicate, and spiked samples and standard reference materials to ensure the accuracy and precision of analyses. 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 8-1 lists some common activities that fall under the heading of QA/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.

Table 8-1. Common QA/QC activities

QA Activities
• Organization of the project into component parts • Assignment of roles and responsibilities to project staff • Determine the number of QC 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 • Auditing of field and laboratory operations • Maintenance of accurate and complete records of all project activities • Training of personnel to ensure consistency of sample collection techniques and equipment use
QC Activities
• Collection of duplicate samples for analysis • Analysis of blank, duplicate, and spike samples • Regular inspection and calibration of analytical equipment • Regular inspection of reagents and water for contamination • Regular inspection of refrigerators, ovens, etc. for proper operation • Regular evaluation of data against QC objectives

Adapted from Drouse et al., 1986, and Erickson et al., 1991.

8.1.2 Importance of QA/QC Programs

While it is desirable to stay below 10 percent, development and implementation of a QA/QC program can require up to 10 to 20 percent of project resources (Cross-Smiecinski and Stetzenback 1994). This cost, however, can be recaptured in lower overall costs of a well-planned and executed project. Likely problems are anticipated and accounted for before they arise, eliminating the need to resample, reanalyze data, or revisit portions of the project to determine where an error was introduced. A QAPP can serve as a foundation for documenting standard operating procedures for all project activities, ensuring that project tasks are conducted consistently by all personnel and can support training for new personnel as the project moves forward. During a project, QA/QC information can provide essential feedback to ongoing project management. Most importantly, a QA/QC program helps ensure that project data are of known accuracy and precision, that errors are minimized, and that all critical project activities are conducted consistently. As long as the QA/QC procedures are followed, the data and information collected by the project will be adequate to support technical conclusions and choices from among alternative courses of action. These conclusions and actions will be defensible based on quality of the data and information collected. In short, 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).

8.1.3 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 methods, 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 unsuited for decision making. QA/QC practices should be integral parts 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 2001b).

EPA Order CIO 2105.0 (formerly EPA Order 5360.1 A2), EPA's *Policy and Program Requirements for the Mandatory Agency-wide Quality System* (USEPA 2000b), provides requirements for the conduct of quality management practices, including QA/QC activities, for all environmental data collection and environmental technology programs performed by or for EPA. The *EPA Quality Manual for Environmental Programs* (USEPA 2000a) provides program requirements for implementing EPA's mandatory quality system. In accordance with EPA Order CIO 2105.0, EPA requires that environmental programs be supported by a quality system that complies with the quality system standard developed by the American National Standard ANSI/ASQC E4-1994, *Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs* (ANSI/ASQC 1994). The ANSI/ASQC E4-1994 quality system standard was later updated as ANSI/ASQ E4-2004, *Quality Systems for Environmental Data and Technology Programs - Requirements with Guidance for Use* (ANSI/ASQ 2004).

EPA's mandatory agency-wide Quality System Policy (EPA Policy CIO 2106.0) 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 2008b). This policy is now based on the quality system standard developed by the American National Standards Institute and the American Society of Quality Control (ANSI/ASQ 2004). 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/QC operations applied to environmental programs (USEPA 2001b). In addition, each non-EPA organization must have an approved QAPP that covers each monitoring or measurement activity associated with a project (Erickson et al. 1991, USEPA 1983, 2008b). Additional implementation guidance is provided in *EPA Quality Manual for Environmental Programs* (USEPA 2000a).

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) in order to meet the project's water quality and land treatment objectives. See section 2.1 for a discussion of appropriate objectives for NPS monitoring projects.

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. Although a QA/QC officer oversees major aspects of QAPP implementation, 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 data collection 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 2001c).

8.2 Data Quality Objectives

When monitoring data are being used to assess water quality and the effects of land-based activities on water quality or the effectiveness of best management practices, EPA recommends that states, tribes, and non-governmental organizations (NGOs) consider using the systematic planning tool called the Data Quality Objectives (DQO) Process. The DQO process should be part of project planning and development of a proposed monitoring strategy.

The DQO process is used to establish performance or acceptance criteria that serve as the basis for designing a plan for collecting data of sufficient quality and quantity to support the objectives of a study. The DQO process consists of seven iterative steps (USEPA 2006):

1) **State the problem:** define the problem that necessitates the study; identify the planning team, examine budget, schedule.

2) **Identify the goal of the monitoring program:** state how monitoring data will be used in meeting objectives and solving the problem, identify study questions, define alternative outcomes.
3) **Identify information inputs:** identify data and information needed to answer questions.
4) **Define the boundaries of the study:** specify the target population and characteristics of interest, define spatial and temporal limits, scale of inference.
5) **Develop the analytic approach:** define the parameters of interest, specify the type of inference, and develop the logic for drawing conclusions from findings.
6) **Specify performance or acceptance criteria:** develop performance criteria for new data being collected or acceptance criteria for existing data being considered for use.
7) **Develop the plan for obtaining data:** select the resource-effective monitoring plan that meets the performance criteria.

Several iterations of the process might be required to specify the DQOs for a monitoring program. Because DQOs are continually reviewed during data collection activities, any needed corrective action can be planned and executed to minimize problems before they become significant. General guidance and examples of planning for monitoring programs are also provided in related guidance (USEPA 2003a).

8.2.1 The Data Quality Objectives Process

The 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 2006).

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.

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 on Systematic Planning Using the Data Quality Objectives Process* (USEPA 2006). This reference contains a case study example of the DQO process. A computer program, *Data Quality Objectives Decision Error Feasibility Trials* (USEPA 2001a), 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.

8.2.1.1 (1) State the problem

In this first step, concisely describe the problem to be studied . 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.

8.2.1.2 (2) Identify the goal of the monitoring program

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” or an objective that will link the principal study question to one or more possible actions that should solve the problem. Example NPS monitoring program objectives might be to “determine the sources of bacteria causing the water quality standard violation in Duck Creek” or “determine the effects of land treatment program xyz on phosphorus loads to Lake Eutrophy.” Results from the monitoring program would then support management decisions to take action, modify an action, or take no action.

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.

8.2.1.3 (3) Identify information inputs

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.

8.2.1.4 (4) Define the boundaries of the study

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 variable, 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 (because 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.

8.2.1.5 (5) Develop the analytic approach

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 septic systems are contributing to water quality standard violations, then failing septic systems will be remediated; 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.

8.2.1.6 (6) Specify performance or acceptance criteria

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 an impact when there is none [a false

positive, or type I error] or deciding there is no impact when there is [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 points above and below the gray region that reflect the tolerable probability for the occurrence of decision errors.

8.2.1.7 (7) Develop the plan for obtaining data

Evaluate information from the previous steps and generate alternative data collection designs. Some aspects of this may be considered informally during the project planning process, and less attention can be given to some alternatives. 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. As resources dictate, it may be necessary to reduce or restate 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.

8.2.2 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 2006). The QAPP presents the policies, organization, and objectives of the data collection effort and explains how particular QA/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 2006). 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 2006).

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 2001c, 2006). 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 2006).

8.3 Elements of A Quality Assurance Project Plan

QAPPs must be prepared in accordance with EPA Requirements for Quality Assurance Project Plans (USEPA, 2001b) and Guidance for Quality Assurance Project Plans (USEPA 2002a). EPA requires that four types of elements be discussed in a Quality Assurance Project Plan (QAPP): Project Management, Measurement and Acquisition, Assessment and Oversight, and Data Validation and Usability. These elements are listed in Table 8-2. For complete descriptions and requirements, see USEPA (2001b). 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 in Table 8-2 should always be addressed in the QAPP, unless otherwise directed by the overseeing or sponsoring EPA organization(s). Both laboratory and field operations should be included. The types, quantity, and quality of environmental data collected for each project could be quite different. The level of detail in each QAPP will vary according to the nature of the work being performed and the intended use of the data (USEPA 2001b). 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 data collection can begin.

Table 8-2. Elements required in an EPA Quality Assurance Project Plan. (USEPA, 2001b)

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
A8	Special Training/Certification
A9	Documents and Records
B1	Sampling Process Design (Experimental Design)
B2	Sampling Methods
B3	Sampling Handling and Custody
B4	Analytical Methods
B5	Quality Control
B6	Instrument/Equipment Testing, Inspection, Maintenance
B7	Instrument/Equipment Calibration and Frequency
B8	Inspection/Acceptance of Supplies and Consumables
B9	Non-direct Measurements
B10	Data Management
C1	Assessments and Response Actions
C2	Reports to Management
D1	Data Review, Verification, and Validation
D2	Verification and Validation Methods
D3	Reconciliation and User Requirements

Standard Operating Procedures (SOPs) must be provided or referenced in the QAPP such that they are available to all participants. An SOP typically presents in detail the method for a given technical operation, analysis, or action in sequential steps and it includes specific facilities, equipment, materials and methods, QA/ QC procedures, and other factors necessary to perform the operation, analysis, or action for the particular project. By following the SOP, the operation should be performed the same way every time. Activities typically include field sampling, laboratory analysis, software development, and database management. EPA presents examples of the format and content of SOPs (USEPA, 2007). The format and content requirements for an SOP are flexible because the content and level of detail in SOPs vary according to the nature of the procedure. SOPs should be revised when new equipment is used, when comments by personnel indicate that the directions are not clear, or when a problem occurs. Organizations should ensure that current SOPs are used. SOPs are critical in the training of new personnel during the conduct of a long-term project.

Definitions of selected data quality terms

Precision (reproducibility) is an expression of mutual agreement of multiple measurements of the same property (e.g., duplicate field samples or duplicate lab samples) conducted under 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.

Accuracy (bias) 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). Accuracy can be 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 is defined as the confidence with which one data set can be compared to another (Erickson et al., 1991). 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 is a measure of how representative the data obtained for each parameter are compared with the values of the same parameter within the population being measured. Because 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). 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).

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). 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).

8.4 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/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/QC for each of the phases separately. This will ensure that no details are omitted.

8.4.1 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, for recording site visit information and field data, should be prepared beforehand. Where applicable data sheets should include data quality reminders to 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.

8.4.2 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. A copy of the site entry permission letter or document should be taken to the site at the time of field visit.

8.4.3 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/QC. All measurement equipment must be routinely checked and calibrated to verify that it is operating properly and generating reliable results, 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.

8.4.4 Sample Collection

A Sampling Plan should be developed and approved prior to sampling. 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. 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.

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. 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.

8.4.5 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). Chain-of-custody seals must be applied to sample containers and shipping containers.

8.4.6 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.

8.5 Laboratory Operations

Laboratory operations should be conducted with great care and attention to detail. Often, an independent laboratory conducts sample analyses, so QA/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.

8.5.1 General Laboratory QA/QC

EPA recommends using an accredited laboratory with an established QA/QC policy to ensure that results will be defensible. The National Environmental Laboratory Accreditation Conference (NELAC) Institute provides accreditation of environmental testing laboratories. Numerous references are available on laboratory QA/QC procedures, and one or more should be consulted to gain an understanding of laboratory QA/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.

8.5.2 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).

8.5.3 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 (USGS), the USEPA, and the American Society for Testing and Materials (ASTM), or those published in *Standard Methods for Examination of Water and Wastewater* (Rice et al., 2012). A compendium of methods for environmental analysis is maintained by the National Environmental Methods Index (NEMI), supported by both USGS and USEPA. 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).

8.5.4 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).

8.5.5 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).

8.5.6 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.

8.6 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. See sections 3.9 and 3.10 for additional details on data management, reporting, and presentation.

8.6.1 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.

8.6.2 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 2001c):

- 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 should be evaluated relative to the project requirements.
- 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 (see section 3.4.2).
- 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 (see section 7.8.2.4).

8.6.3 Documentation, Record Keeping, and Data Management

- 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

One aspect that merits further discussion is documentation and management of data, from the collection process through the data analysis. Data management activities include documenting the nature of the data and subsequent analyses so that the data from different sites are comparable. Data management also includes handling and storing both hard copies and electronic files containing field and laboratory data. A data management system that addresses project needs should be selected at the beginning of the monitoring program (see section 3.9). It is also important to understand and comply with applicable state agency and/or grant policies and standards regarding data collection and generation.

Some grants might require local NPS and water resources managers to add their data to EPA's storage and retrieval (STORET) database (https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange). STORET contains raw biological, chemical, and physical data on surface water and ground water collected by federal, state, and local agencies; tribes; volunteer groups; academics; and others. Each sampling result in STORET is accompanied by information on where the sample was taken (latitude, longitude, state, county, hydrologic unit code, and brief site identification), when the sample was gathered, the medium sampled (e.g., water, sediment, fish tissue), and the name of the organization that sponsored the monitoring. Staff working with the database should have expertise and training in the software and in the procedures for data transport, file transfer, and system maintenance.

The operation of the data management system should include QA oversight and QC procedures. If changes in hardware or software become necessary during the course of the project, the data manager should obtain the most appropriate equipment and test it to verify that the equipment can perform the necessary jobs. Appropriate user instructions and system documentation should be available to all staff using the database system. Developing spreadsheet, database, and other software applications involves performing QC reviews of input data to ensure the validity of computed data.

8.6.4 Report Preparation

The original project description should include a schedule and 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.

8.7 Geospatial Data

Projects should incorporate procedures for documenting geospatial data appropriately. Geospatial Information System (GIS) data can vary from relatively simple site locations to complex with many overlapping contextual boundaries. For example, the development of a watershed implementation plan may involve analyzing water samples from industrial dischargers, developing water quality models, creation of new geospatial data, or even updating existing geospatial data. Use of geospatial data from external sources may require the development of a secondary data QAPP. QAPPs also apply to geospatial data (USEPA 2001b), but should vary with the complexity of the project (see Table 8-3). The project planning phase should determine the scope and complexity of the project that will inform the complexity of the QAPP (USEPA, 2003b).

Table 8-3. Continuum of Geospatial Projects with Differing Intended Uses

Purpose of Project	Typical Quality Assurance Issues	Level of QA
Regulatory compliance Litigation Congressional testimony	Legal defensibility of data sources Compliance with laws and regulatory mandates applicable to data gathering Legal defensibility of methodology	
…	…	
Regulatory development Spatial data development (Agency infrastructure development)	Compliance with regulatory guidelines Existing data obtained under suitable QA program Audits and data reviews	
…	…	
Trends monitoring (non-regulatory) Reporting guidelines (e.g., Clean Water Act) "Proof of principle"	Use of accepted data-gathering methods Use of accepted models/analysis techniques Use of standardized geospatial data models Compliance with reporting guidelines	
…	…	
Screening analyses Hypothesis testing Data display	QA planning and documentation as appropriate Use of accepted data sources Peer review of products	

Source: USEPA, 2003b.

8.7.1 Performance Criteria for a Geospatial Data Project

Projects with geospatial components will likely follow the same DQO process described in section 8.2 of this chapter. In decision-making programs taking the form of the DQO process, data quality to achieve a desired level of confidence in the decision takes a number of typical forms as listed below (USEPA 2003b):

- A description of the resolution and accuracy needed in input data sources
- Statements regarding the speed of applications programs written to perform data processing (e.g., sampling at least "n" points in "m" minutes)
- Criteria for choosing among several existing data sources for a particular geospatial theme (e.g., land use); geospatial data needs are often expressed in terms of using the "best available" data, but different criteria—such as scale, content, time period represented, quality, and format—

may need to be assessed to decide which are the "best available" (when more than one is available) to use on the project

- Specifications regarding the accuracy needs of coordinates collected from GPS receivers
- Criteria for aerial photography or satellite imagery geo-referencing quality, such as specifications as to how closely these data sources need to match spatially with ground-based reference points or coordinates
- Criteria for minimum overall match rate, tolerances including whether or not spatial offsets are to be supplied in the resulting coordinates procedures, and if so, the offset factor in address matching
- Topology, label errors, attribute accuracy, overlaps and gaps, and other processing quality indicators for map digitizing
- Criteria to be met in ground-truthing classified satellite imagery

8.7.2 Spatial Data Quality Indicators for Geospatial Data

The most comprehensive way to track the quality and applicability of a geospatial data set is through the use of metadata. EPA requires that appropriate metadata accompany every data set, in accordance with Federal Geographic Data Committee standards (FGDC 1998). There are five components applicable to the Federal Geographic Data Committee metadata requirements (FGDC 1998, USEPA 2003b):

- **Accuracy – positional:** The closeness of the locations of the geospatial features to their true position.
- **Accuracy – attribute:** The closeness of attribute values (characteristics at the location) to their true values.
- **Completeness:** The degree to which the entity objects and their attributes in a data set represent all entity instances of the abstract universe (defined by what is specified by the project's data use in systematic planning). It is in the metadata where the user may define the abstract universe with criteria for selecting features to include in the data set. The information is relevant to any user who wishes to independently replicate geospatial procedures. Missing, or incomplete data can affect logical consistency needed for correct processing of data by software.
- **Logical consistency:** The data in any spatial data set is logically consistent when it complies with the structural characteristics of the data model and is compatible with attribute constraints defined for the system.
- **Lineage:** The description of the origin and processing history of a data set.

8.8 References

Air National Guard. 1993. *Draft Final Remedial Investigation/Feasibility Study Work Plan*. Prepared for Wyoming Air National Guard 153rd Tactical Airlift Group, Cheyenne Municipal Airport, Cheyenne, Wyoming, by Aepco, Inc. and Tetra Tech, Inc.

ANSI/ASQ (American National Standards Institute/American Society for Quality). 1994. *Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs.* American Society for Quality Control (ASQC). ASQC Quality Press, Milwaukee, Wisconsin.

ANSI/ASQ (American National Standards Institute/American Society for Quality). 2004. *American National Standard, Quality systems for environmental data and technology programs – Requirements with guidance for use*. American Society for Quality (ASQ). ASQ Quality Press. Milwaukee, Wisconsin.

Coffey, S.W., J. Spooner, and M.D. Smolen. 1993. *The Nonpoint Source Manager's Guide to Water Quality and Land Treatment Monitoring*. North Carolina State University, Department of Biological and Agricultural Engineering, NCSU Water Quality Group, Raleigh, NC.

CREM (Council for Regulatory Environmental Modeling). 2009. *Guidance on the Development, Evaluation, and Application of Environmental Models.* EPA/100/K-09/003. U.S. Environmental Protection Agency, Council for Regulatory Environmental Modeling, Washington, DC.

Cross-Smiecinski, A., and L.D. Stetzenback. 1994. *Quality Planning for the Life Science Researcher: Meeting Quality Assurance Requirements*. CRC Press, Boca Raton, Florida.

Drouse, S.K., D.C. Hillman, J.L. Engles, L.W. Creelman and S.J. Simon. 1986. *National Surface Water Survey. National Stream Survey (Phase 1 - Pilot, Mid-Atlantic Phase 1 Southeast Screening, and Episodes Pilot) Quality Assurance Plan*. EPA/600/4-86/044. NTIS No. PB87-145819. Prepared for U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada, by Lockheed Engineering and Management Services Co., Inc., Las Vegas.

Erickson, H.E., M. Morrison, J. Kern, L. Hughes, J. Malcolm and K. Thornton. 1991. *Watershed Manipulation Project: Quality Assurance Implementation Plan for 1986-1989*. EPA/600/3-91/008. NTIS No. PB91-148395. Prepared for Corvallis Environmental Research Laboratory, Oregon, by NSI Technology Services Corporation, Corvallis, OR.

FGDC (Federal Geographic Data Committee). 1998. *Content Standard for Digital Geospatial Metadata*. FGDC-STD-001-1998. Federal Geographic Data Committee, Washington, DC.

Knapton, J.R., and D.A. Nimick. 1991. *Quality Assurance for Water-Quality Activities of the U.S. Geological Survey in Montana*. Open File Report 91-216. U.S. Geological Survey, Helena, Montana.

Pritt, J.W., and J.W. Raese, ed. 1992. *Quality Assurance/Quality Control Manual*. Open File Report 92-495. U.S. Geological Survey, National Water Quality Laboratory, Reston, Virginia.

Rice, E.W., R.B. Baird, A.D. Eaton, and L.S. Clesceri, ed. 2012. *Standard Methods for the Examination of Water and Wastewater*. 22nd ed. American Public Health Association, American Water Works

Association, Water Environment Federation. American Public Health Association Publications, Washington, DC.

Spreizer, G.M., T.J. Calabrese and R.S. Weidner. 1992. Assessing the Usability of Historical Water Quality Data for Current and Future Applications. In *Current Practices in Ground Water and Vadose Zone Investigations*, ASTM STP 1118. Ed. D.M. Nielsen and M.N. Sara, pp. 377-390. American Society for Testing and Materials, Philadelphia, Pennsylvania.

USEPA (U.S. Environmental Protection Agency). 1983. *Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans*. EPA-600/4-83-004. QAMS-005/80. U.S. Environmental Protection Agency, Office of Monitoring Systems and Quality Assurance, Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2000a. *EPA Quality Manual for Environmental Programs*. CIO 2105-P-01-0. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2000b. *Policy and Program Requirements for the Mandatory Agency-wide Quality System*. CIO 2105.0. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001a. *EPA Data Quality Objectives Decision Error Feasibility Trials Software (DEFT) – USER'S GUIDE, EPA QA/G-4D*. EPA/240/B-01/007. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001b. *EPA Requirements for Quality Assurance Project Plans, EPA QA/R-5*. EPA 240/B-01/003. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001c. *EPA Requirements for Quality Management Plans, EPA QA/R-2*. EPA/240/B-01/002. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002a. *Guidance for Quality Assurance Project Plans,* EPA QA/G-5. EPA 240/R-02/009. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002b. *Guidance for Quality Assurance Project Plans for Modeling, EPA QA/G-5M*. EPA/240/R-02/007. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2003a. *Elements of a State Water Monitoring and Assessment Program*. EPA 841-B-03-003. U.S. Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans and Watersheds, Assessment and Watershed Protection Division, Washington, DC. Accessed January 28, 2016.

USEPA (U.S. Environmental Protection Agency). 2003b. *Guidance for Geospatial Data Quality Assurance Project Plans, EPA QA/G-5G.* EPA/240/R-03/003. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2006. *EPA Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G-4*. EPA/240/B-06/001. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2007. *Guidance for Preparing Standard Operating Procedures (SOPs), EPA QA/G-6.* EPA 600/B-07/001. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2008a. *NRMRL QAPP Requirements for Secondary Data Projects*. U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Cincinnati, OH.

USEPA (U.S. Environmental Protection Agency). 2008b. *U.S. Environmental Protection Agency Quality Policy*. CIO 2106.0. U.S. Environmental Protection Agency, Office of Environmental Information. Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2012. *Guidance for Evaluating and Documenting the Quality of Existing Scientific and Technical Information Addendum to: A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*. U.S. Environmental Protection Agency, Science and Technology Policy Council, Washington, DC.

9 Monitoring Costs

By S.A. Dressing and D.W. Meals

9.1 Introduction

Monitoring plans must be designed to help achieve watershed project or program goals. This could be a relatively simple task with an unlimited budget, but perhaps the most frequently cited problems for those who design and carry out water quality monitoring programs are the limitations and unpredictability of funding. Although cost should not be the defining factor in the design of monitoring plans, it must be considered from the start. Both "cheap" monitoring programs that are inadequate to achieve project objectives or great monitoring programs that are discontinued because funding disappears are worse than no monitoring at all because much or all the money spent is essentially wasted.

While funding can almost never be guaranteed over the course of a multi-year monitoring effort, careful cost analysis at the beginning can help design a monitoring plan that will meet objectives and fit within a cost range that can be sustained until the project ends. In some cases, project budgets might be insufficient to carry out meaningful monitoring; in such cases, monitoring should not be done. In all other cases, project staff must seek a balance that provides the ability to achieve monitoring objectives that are supportive of project or program goals at an affordable cost.

Although an exact monitoring budget will be highly specific to the setting of a particular project, monitoring costs can be estimated reasonably well as part of project planning. Even a very good cost estimate, however, will miss the mark on category specific costs. For example, sampling trips may take more or less time than anticipated, equipment costs can change drastically if equipment is washed away or needed equipment suddenly becomes available from a discontinued monitoring effort, or data analysis and reporting requirements change under new management or because of unexpected findings or additional requests for information. While the total budget allotted to a monitoring project may not change, projects should maintain flexibility to shift resources within a budget to ensure that project objectives are met with maximum cost efficiency.

In this chapter, potential monitoring costs for the types of monitoring described in this guidance document are illustrated using a spreadsheet tool that has been developed to estimate monitoring costs for nonpoint source watershed projects (Dressing 2012, 2014). Two user-editable versions of the spreadsheet can be downloaded at this site: (https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed). The master spreadsheet allows users to determine every detail in their cost estimation, whereas the simplified spreadsheet includes default assumptions for monitoring designs, sampling types, and parameters, as well as basic algorithms to allow users to generate cost estimates with as little input as possible. See Appendix 9-1 for additional details on the cost estimation spreadsheets.

9.2 Monitoring Cost Items and Categories

A complete accounting of monitoring costs begins with watershed characterization and development of a QAPP (see chapter 9) and ends with data analysis (see chapter 8) and reporting. Costs incurred by monitoring efforts can include monitoring site selection, construction of monitoring stations, installation and setup of monitoring equipment, sample collection, laboratory analysis, and the ultimate removal of monitoring sheds at the conclusion of the project (see chapters 2 and 3 for details on monitoring designs). Some monitoring efforts include the cost of contracts or grants for monitoring support.

Specific cost items can be grouped and summarized in many (and often overlapping) ways, including the categories shown in Table 9-1. These categories are basically people and things, whereas the categories shown in Table 9-2 are organized by project phases and key project elements. Some costs are incurred once during a project (e.g., site establishment) while others are recurring (e.g., sampling site visits), so annual costs often vary, particularly for the first and last years of a project.

Table 9-1. Costs grouped by type of item or activity

Cost Category	Items Included In Category
Labor	All labor costs (inclusion of fringe benefits optional).
Installed Structures	Materials and labor costs.
Other Site Establishment Costs	One-time fee, electricity connection, setup, etc.
Purchased Equipment	All purchased monitoring equipment.
Rental Equipment	All rented monitoring equipment.
Monitoring Supplies	All startup and annual monitoring supplies.
Office Equipment	All purchased office equipment.
Office Supplies	All startup and annual office supplies.
Travel/Vehicles	All use of vehicles for travel, construction, sample pickup, etc.
Laboratory Analysis	Annual laboratory analysis.
Data Purchases	Maps, data, satellite & aerial photography.
Printing/Media	Printing and other report output media (e.g., CD, web).
Electricity/Fuel	All fuel and power costs for operating sites.
Site Service and Repair	All service, repair, and replacements of sites and equipment.
Annual Site Fees	All annual fees for site access. Does not include initial fee.
Contracts	All non-itemized contracts costs.
Grants	All non-itemized grants costs.

Table 9-2. Costs grouped by project phase or element

Cost Category	Items Included in Category
One-Time Costs	
Proposal and QAPP	Cost for development of proposal and QAPP or equivalent document (added to Year 1 cost).
Watershed Characterization	Cost for characterization of watershed to aid monitoring design (added to Year 1 cost). Includes windshield surveys and analysis of existing data and maps.
Site Establishment	Includes one-time costs for setting up station, including purchase of equipment that remains at site. Site selection, preparation, and excavation costs are all included.
Portable Sampling Equipment and Startup Supplies Costs	Includes one-time costs for all portable sampling equipment or instruments that are taken to the site for use and then taken away for use at another site or time. Equipment includes such items as kick nets, pH meters, etc. Also includes one-time cost for initial purchase of supplies such as pipettes, vials, and bottles.

Cost Category	Items Included in Category
One-Time Office Equipment and Startup Supplies Costs	Includes computer hardware and software and related items.
Station Demolition and Site Restoration	Includes all costs associated with tearing down the station and restoring the site at the end of the project.
First-Year Report	This is the cost for data analysis and writing, printing, and distribution of the first-year report. Data analysis and reporting can be combined or kept separate.
Final Report	This is the cost for data analysis and writing, printing, and distribution of the final report. Data analysis and reporting can be combined or kept separate.
Annual Costs	
Access Fees	Any fees paid to landowners for allowing access to the site.
Sampling Trips to Sites	Includes labor, vehicle use, and other equipment (e.g., boat) costs for site visits.
Volunteer Training	Annual cost to train volunteers or others collecting data for the project.
Sample Analysis	Cost for laboratory analysis of samples. Includes travel to and from laboratory if done in addition to sampling trip travel. Can include costs for shipping samples to laboratories as “Other” cost.
Annual Data Analysis and Reports	This is the cost for annual analysis of project data and annual or more frequent reporting in years other than the first and last year. Includes labor and materials. Data analysis and reporting can be combined or kept separate.
Site Operation and Maintenance	Includes service/repair/replacement of equipment and structures, electric and fuel bills (e.g., for heating), and annual cost to establish and update stage/discharge relationship.
Supplies and Rental Equipment	This cost is primarily for consumable supplies (e.g., sample preservative), but can include sample bottles and other items. Also includes rental equipment and office supplies.
Land Use Tracking	Labor, travel, and services (e.g., aerial photography or data purchase) needed to track land use and land treatment.
Total Cost of Monitoring	Total cost of monitoring for the entire project period.

9.3 Cost Estimation Examples

The cost spreadsheets have been used to estimate costs for a wide variety of monitoring designs and applications. The cost estimates highlighted here were developed for three different purposes. First, the master spreadsheet was used to provide a range of estimates for a diverse set of monitoring options, with estimated costs generated for eight different monitoring scenarios covering a wide range of timeframes (see section 9.3.1). The ten cost estimates summarized in section 9.3.2 cover various monitoring approaches relevant to assessing the watershed-scale water quality impacts of programs such as USDA’s National Water Quality Initiative (NWQI). Finally, the simplified spreadsheet was used to estimate costs for 60 basic, 5-year monitoring scenarios that are summarized in section 9.3.3. It is important to note that assumptions regarding the need and cost for labor, equipment, monitoring parameters, sampling frequency, and sampling duration are all important determinants of the final cost estimates, so costs are presented in this section more for a comparative analysis than as accurate estimates for any specific monitoring type or effort. The examples are particularly useful to contemplate trade-offs among cost categories and to evaluate where cost-effectiveness can be improved, e.g., offsetting high labor costs with the purchase of automated equipment.

9.3.1 Cost Estimates for a Diverse Range of Monitoring Options

Cost estimates for the following eight monitoring scenarios are presented in this section.

1. Synoptic Survey
2. TMDL – Water Quality Standards
3 TMDL – Loads
4. Paired-Watershed - Loads
5. Long-term Single Station - Biomonitoring
6. Above/Below BMP Effectiveness - Biomonitoring
7. Input/Output Urban Low Impact Development (LID) Effectiveness
8. Photo-Point Monitoring

These eight scenarios were chosen to represent a wide range of monitoring approaches, addressing both problem assessment and project evaluation, using chemical, physical, and biological (Barbour et al. 1999) monitoring methods. With the exception of 1-year synoptic surveys (Scenario 1), costs are estimated for 1, 2, 5, and 8 years. A more detailed comparison of Scenarios 2-8 is based on five-year cost estimates. See Appendix 9-2 for additional details on these eight scenarios.

9.3.1.1 Discussion

Table 9-3 summarizes the total costs for each scenario for 1, 2, 5, and 8 years. Cost totals are taken from the base scenarios in which all equipment is purchased and all monitoring is stand-alone; that is, there are no cost savings assumed for monitoring activities that may be combined with other activities (e.g., another monitoring effort in the same area) to save on travel or labor. It should be no surprise that biological (Scenarios 5 and 6) and photo-point (Scenario 8) monitoring are the least expensive monitoring approaches in this analysis. Sampling frequency (2x/year) for biological and photo-point monitoring is far less than is assumed for water quality monitoring and load estimation, and laboratory and equipment costs are generally lower as well.

While total cost provides the best measure for comparing the costs for alternative monitoring designs, the breakout of costs by category gives a better picture of where cost savings can be found within each monitoring design. For example, labor accounted for the greatest share of total costs in all five-year scenarios, ranging from 68 percent for Scenario 7 (urban LID) to 90 percent for quantitative photo-point monitoring (Figure 9-1). Labor accounted for only 45 percent of the total cost for the 1-year synoptic survey.

Equipment costs ranged from 2 percent for Scenario 2 (TMDL water quality standards) to 12 percent of total 5-year costs for qualitative photo-point monitoring. About 45 percent of the 1-year budget for synoptic surveys was devoted to equipment. Laboratory analysis costs accounted for 16 percent of total 5-year costs for Scenario 7 (urban LID), 9 percent of the 1-year cost for a synoptic survey, and 5 percent of the 5-year cost for Scenario 2 (TMDL water quality standards), but were responsible for less than 1 percent of costs for all other scenarios.

Vehicle (mileage) costs ranged from 1 percent for Scenario 1 and quantitative photo-point monitoring to 10 percent of total 5-year costs for Scenario 2. Both Scenario 3 and Scenario 7 had 5-year budgets in which vehicle costs accounted for 9 percent of the total cost.

It is important to ensure that whatever monitoring approach is used will provide the type, quality, and quantity of information necessary to meet project monitoring objectives. Despite its lower cost, photo-point monitoring will usually not be appropriate as a stand-alone monitoring approach for tracking progress in achieving a TMDL. Likewise, biological monitoring cannot be used to estimate pollutant loads. On the other hand, weekly grab sampling for water chemistry may be wasteful if monitoring is intended to track attainment of aquatic life support, and photo-point monitoring could be appropriate for a trash TMDL such as that established for the Anacostia River (MDOE and DDOE 2010).

Table 9-3. Summary of scenario costs for diverse range of monitoring options

Scenario	Total Cost ($1,000)			
	1 Year	2 Years	5 Years	8 Years
1. Synoptic Survey	30	n/a	n/a	n/a
2. TMDL WQS	47	90	215	339
3. TMDL Loads	62	107	238	368
4. Paired-Watershed Load	93	158	348	537
5. Long-Term Biological	16	26	53	80
6. Above/Below BMP Effectiveness - Biological	17	28	58	88
7. Input/Output Urban LID Effectiveness	68	115	252	388
8. Photo-Point Monitoring – Qualitative Analysis	8	11	19	26
8. Photo-Point Monitoring – Quantitative Analysis	25	39	75	111

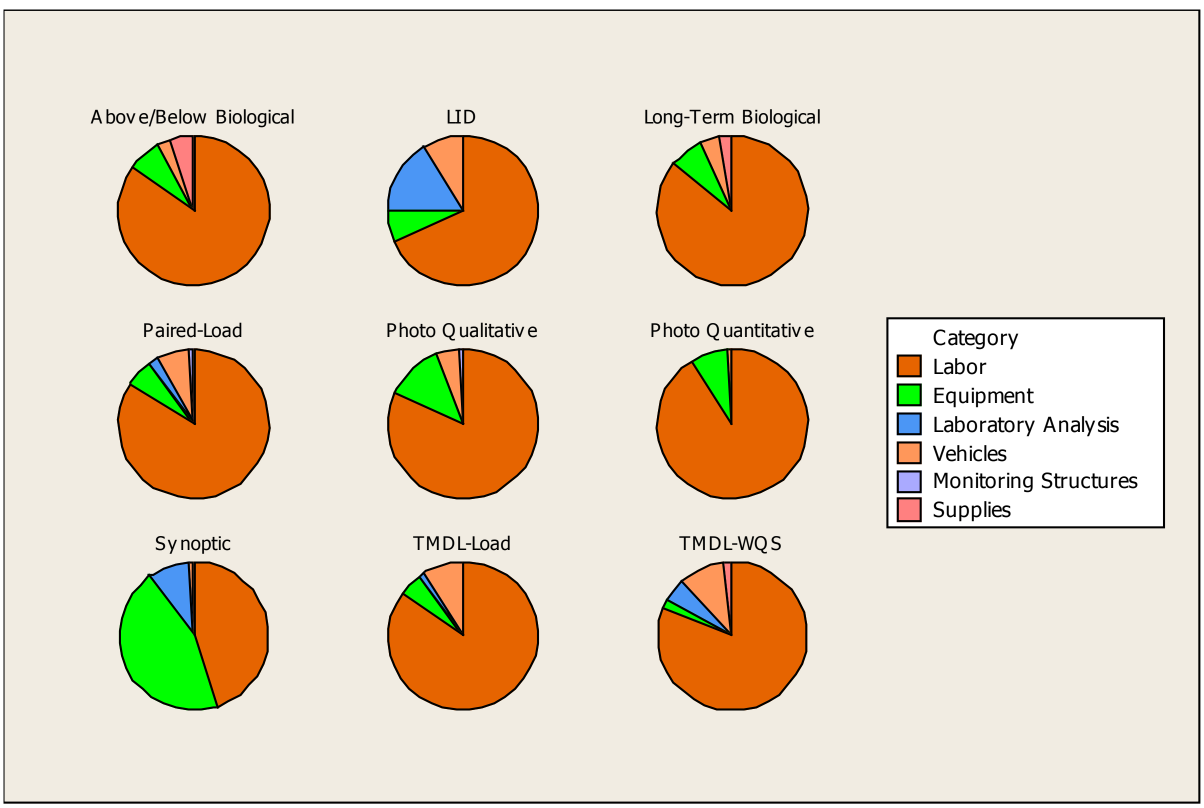


Figure 9-1. Breakout of costs for diverse range of monitoring options

9.3.2 Cost Estimates for Watershed-Scale Evaluation of Agricultural BMP Implementation

This cost analysis was performed to explore different options for planning and assessing the water quality impacts of watershed-scale implementation of agricultural BMPs. The setting assumed for the cost scenarios is a 12-digit HUC watershed covering 10,117 ha (25,000 ac), primarily in agricultural use. Monitoring is performed in perennial streams with the exception of the paired-watershed scenario that assumes intermittent flow.

Cost estimates were generated for a total of 84 scenarios, including synoptic surveys, compliance monitoring, soil testing, multiple-watershed monitoring, and paired-watershed, trend, and above/below monitoring. Cost estimates were developed for three different driving distances to the watershed to illustrate how that factor influences costs, particularly the labor share of total costs. Three timeframes were considered (three, five, and seven years) for all but synoptic surveys which were assumed to be completed within one year.

For simplicity, all labor was assumed to be performed by contractors, but this may not be affordable in many situations. Pay rates assumed (including fringe and overhead) and basic job functions are summarized in Table 9-4. Rates for government or university employees and volunteers would clearly differ, and contractor rates would vary depending on location.

Additional assumptions about number of sampling sites, monitoring frequency, monitoring variables, and various other aspects of the monitoring designs are documented in Appendix 9-3.

Table 9-4. Labor costs assumed for watershed-scale evaluation scenarios

Pay Level	Rate ($/hr)[1]	Job Functions
4	130	Monitoring design, statistical analysis, oversight, etc.
3	80	Lead field person for monitoring, data collection, bulk of writing
2	56	Field technician, lab tech, etc.
1	34	Secretarial and support staff

[1]Includes fringe and overhead.

9.3.2.1 Discussion

Results for 5-year monitoring efforts are summarized in Figure 9-2. Not shown in this figure are 1-year synoptic surveys which had the lowest cost, ranging from $12,000 to $18,000 depending on distance traveled to the watershed. The low cost of synoptic surveys compared to the cost of other scenarios indicates that they can be a very good investment for generating additional information to support final decisions on both the land treatment plan and long-term monitoring design.

Compliance monitoring is also relatively inexpensive as defined in these scenarios, ranging from $21,000 to $55,000 for 5-year efforts depending on distance traveled. The cost for a soil testing program ranges from $32,000 to $50,000 for five years with a far smaller influence of distance traveled on total cost compared to compliance monitoring. This is because soil testing requires a large amount of time collecting samples at the site, whereas sampling for compliance monitoring is relatively quick once the site is reached.

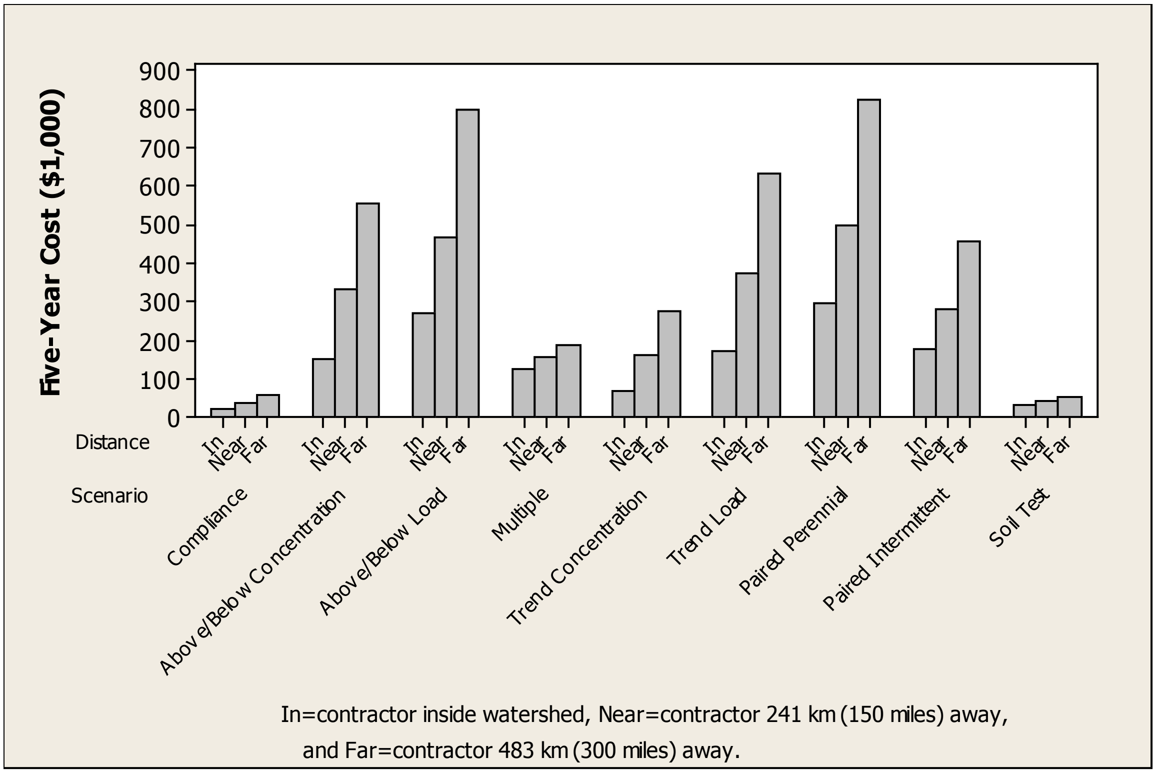


Figure 9-2. Cost estimates for watershed-scale assessment of agricultural BMP projects

Trend monitoring costs can range from $68,000 to $275,000 for a 5-year effort with grab sampling to a range of $172,000 to $630,000 for a 5-year effort with automated sampling and pollutant load estimation. Although not shown in Figure 9-3, this analysis presents an interesting choice between a 7-year grab sampling effort ($92,000-$382,000) and a 3-year load estimation effort ($112,000-$391,000) for trend analysis. This cost information coupled with an MDC analysis (see section 9.4) could lead to cost-effective solutions to monitoring needs.

The cost of above/below monitoring ranges from $152,000 to $553,000 for a 5-year grab sampling effort to $268,000 to $799,000 for a 5-year load estimation effort. Costs for above/below monitoring designs are roughly twice the cost of the parallel trend monitoring designs for grab sampling, but can be much less than double the cost for load estimation. For example, comparing 5-year costs for above/below with trend concentration monitoring shows that the “near” cost for above/below ($329,000) is about twice the “near” cost for the trend design ($159,000). However, the 5-year cost for above/below load monitoring ($466,000) is far less than double the cost for trend load monitoring ($371,000). The different patterns are largely explained by the costs for site establishment and automated sampling equipment for load estimation.

Paired-watershed monitoring (loads) are found to be similar to above/below monitoring in this analysis. Costs ranged from $176,000 to $455,000 for a 5-year effort on an intermittent stream to $294,000 to $824,000 for 5 years on a perennial stream. The major difference between paired-watershed and above/below monitoring costs is the travel between watersheds and larger area involved in land use/treatment tracking for paired-watershed monitoring.

The cost of monitoring 20 subwatersheds in a multiple-watershed design is estimated to range from $125,000 to $185,000 for a 5-year effort. Grab sampling is assumed for multiple-watershed monitoring scenarios in this analysis.

For scenarios assuming 5 years of monitoring and the "near" distance (monitoring team 241 km from watershed), labor consumes 72 to 86 percent of total cost estimates. The proportion of total costs devoted to labor often changes with project duration, however, as illustrated in Figure 9-3. In this comparison, the labor share of cost decreases with increasing monitoring duration for soil testing (assuming 20 sites), but increases for a paired study measuring loads on a perennial stream. The different trends result primarily from differences in first-year costs. The paired design assumes significant labor and equipment (~equal) costs for site establishment and purchased equipment, while the soil testing design assumes substantial labor cost to select sites via desktop analysis. It should be noted that for both scenarios total labor costs increase over time, whereas equipment, site selection, and site establishment costs are incurred in the first year only.

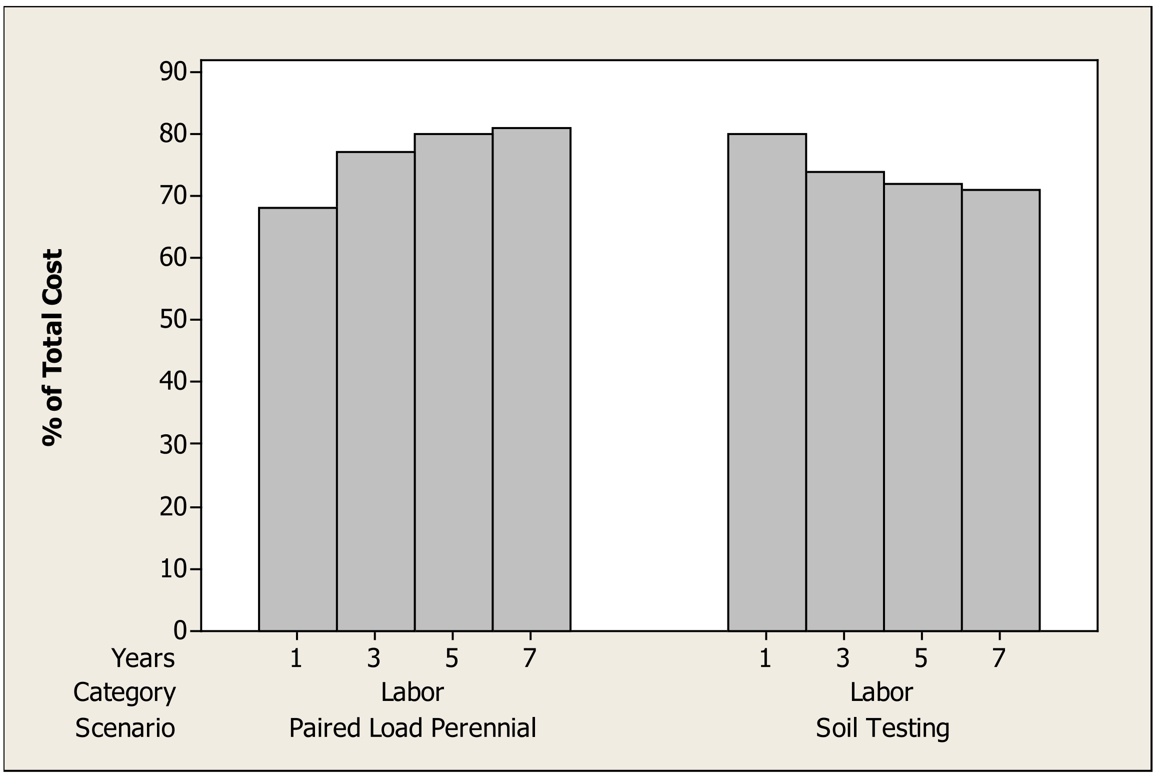


Figure 9-3. Comparison of labor cost category percentage over time

9.3.3 Cost Estimates for Five-Year Trend and Above/Below Monitoring

Cost estimates were generated for 160 scenarios that address two different designs (trend and above/below); four different monitoring variable sets (nutrient and sediment grab samples – [NSC], nutrient and sediment loads – [NSL], biological/habitat with kick net – [BioK], and sondes for nutrients and turbidity – [SNT]); four watershed sizes (202, 2023, 10117, and 20234 ha)[1]; and five different

[1] 500; 5,000; 25,000; and 50,000 acres

distances to the watershed (0, 40, 80, 121, and 161 km)[2]. All scenarios assume 5 years of monitoring, while it was assumed that sampling frequency was 2 and 26 times per year for biological and all other monitoring, respectively. In addition to sample collection and analysis, total monitoring costs also include watershed characterization, site establishment, land use/treatment tracking, data analysis, and reporting. This cost analysis was designed to test application of the simplified spreadsheet to the designs most commonly used by NPS watershed projects. See Appendix 9-4 for additional details.

Additional scenarios using all combinations of the following conditions were run to illustrate how assumptions on salary and equipment affect total cost estimates:

- Labor cost of $0 and salary adjustment of factors of 0.5, 0.7, and 1 (baseline).
- Purchase of all equipment (baseline) and equipment cost of $0.

These scenarios were run for a 2,023-ha (5,000 ac) watershed where the monitoring team was 80 km (50 mi) from the watershed, parameters that best represent the median total costs for each design and variables set.

9.3.3.1 Discussion

Figure 9-4 summarizes the results from this analysis. The box plots on the top show clearly that load estimation (NSL) is the most expensive approach when compared to concentration monitoring with grab samples (NSC) and the use of sondes for nutrients and turbidity (SNT). Biological monitoring (BioK) is the cheapest option overall, but sampling is only done twice per year versus the assumed 26 times per year for the other three options. Above/below monitoring is more expensive than trend monitoring for all variable sets because there are twice as many stations. The cost, however, is less than double because of efficiencies in labor, travel, analysis and other cost categories. It should be noted that paired designs would have costs similar to those for the above/below design.

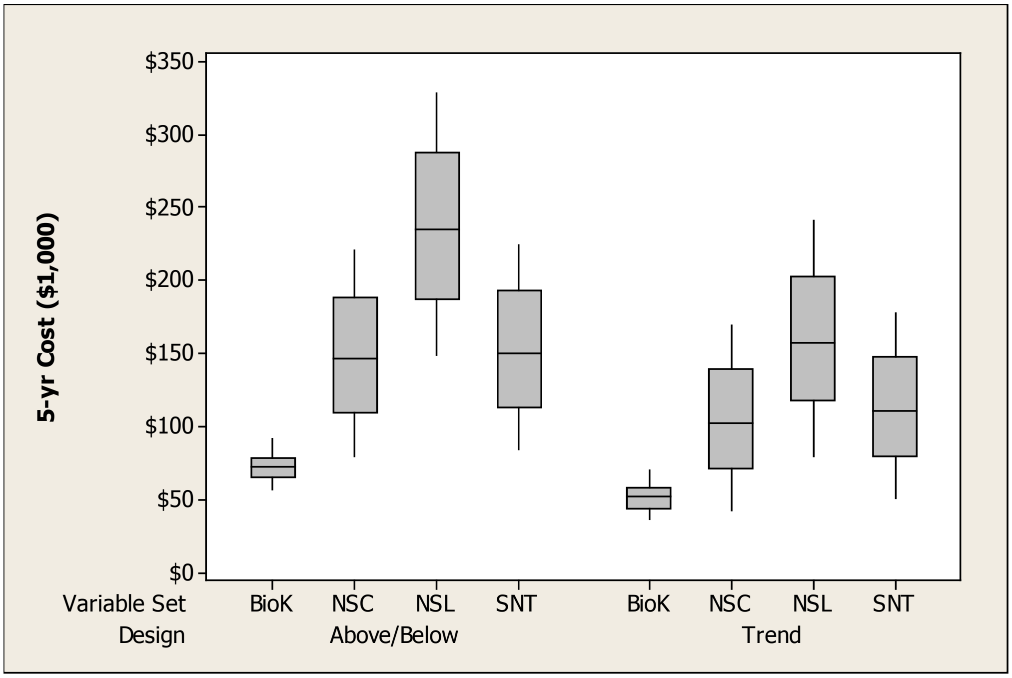
When costs are reduced to cost per sampling trip to each monitoring site (bottom of Figure 9-4), biological monitoring is by far the most expensive approach of the scenarios considered. This is due primarily to the fact that only 2 samples are collected each year versus 26 samples per year for the other scenarios. Load monitoring is more expensive than both concentration and sonde monitoring. This figure also points out the cost efficiency of above/below versus trend monitoring when using a biological approach; the extra site is relatively inexpensive. Readers should keep in mind that, as described above, total costs include more than just sample collection and analysis.

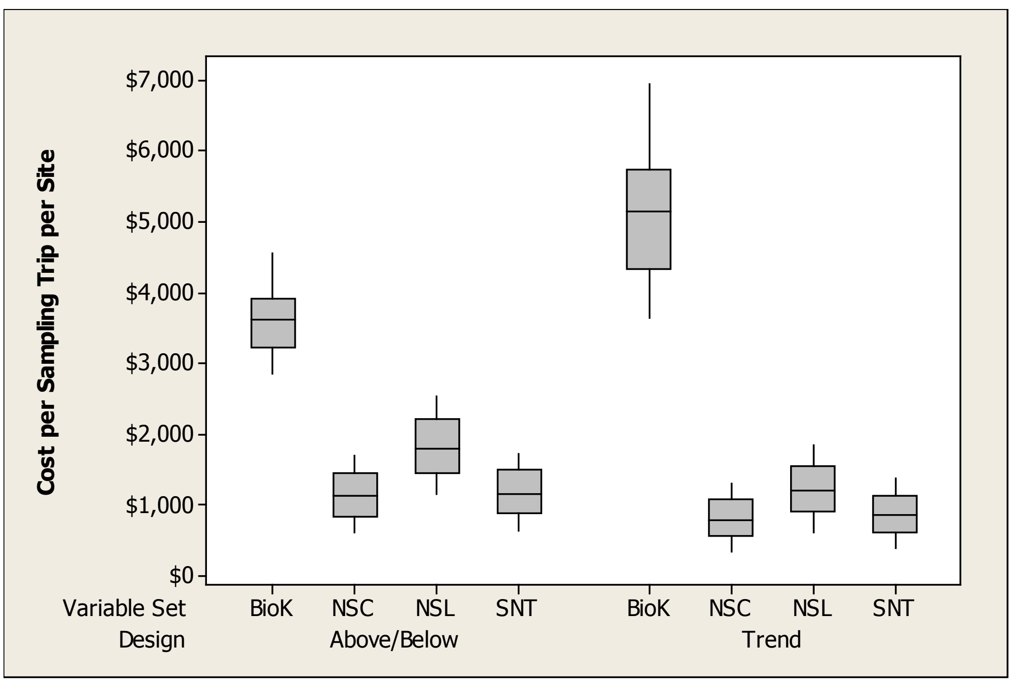
In all cases examined here, labor accounted for the largest share of costs, ranging from 63% to 84 percent of total cost (66 percent to 85 percent if labor for analysis of biological samples is included). Competitive contractor rates were assumed for labor, but the importance of labor costs can vary greatly because monitoring efforts may use far less expensive staff (e.g., volunteers) or assume that labor is not an additional cost because in-house staff are used.

Labor generally accounted for a larger share of total costs for scenarios that required less equipment, ranging from 63 percent to 74 percent for biological (74-85 percent including analysis of biological samples) and 74 percent to 84 percent for nutrient/sediment concentration monitoring. A slightly lesser share of total cost was devoted to labor in cases where sondes were assumed (66-82 percent) or loads were estimated with continuous flow measurement and automatic sampling (67-81 percent). Despite the

[2] 0, 25, 50, 75, and 100 miles

greater importance of labor in costs for biological monitoring, Figure 9-5 illustrates that the dollar amount is still far less than for other monitoring options, whether labor for analysis of biological samples is included (BioK-A) or not (BioK).





BioK=Biological monitoring with kick net; NSC=Nutrient and sediment concentration; NSL=Nutrient and sediment load: SNT=sondes for nutrients and turbidity

Figure 9-4. Box plots summarizing cost estimates for five-year monitoring efforts

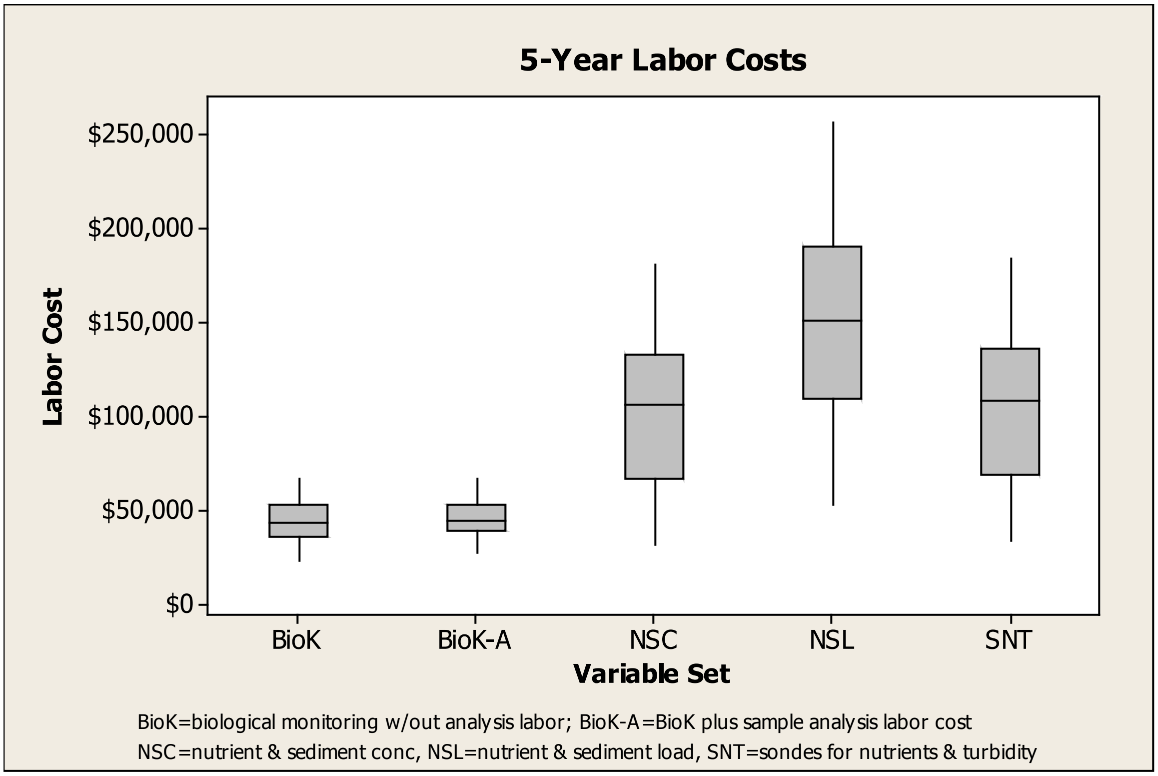


Figure 9-5. Box plots summarizing five-year labor costs

Equipment and supplies accounted for 6-27 percent of total costs for BioK, NSL, and SNT, but only a maximum of 2 percent for NSC. The large difference in importance of this cost category for biological monitoring versus grab sampling for nutrients and sediments (NSC) hinges largely on the vast differences in sampling frequencies (2x/yr vs. 26x/yr) and thus labor costs. The difference between NSC and NSL and SNT is due to the far larger reliance on purchased equipment for monitoring with sondes and measurement of loads. Sample analysis generally accounted for 2-25 percent of total costs for all scenarios.

Vehicle costs were typically well under 10% of total costs for these scenarios, and per diem costs were zero except in cases where watersheds were very large (10,117 or 20,234 ha) and monitoring teams were remote (121 or 61 km from the watershed). Overnight stays were associated with watershed characterization and land use/treatment tracking, not water quality monitoring. Each cost scenario assumes that the watershed will be characterized in the first year of monitoring, and that land use/treatment will be tracked twice per year every year.

Assumptions regarding salary and equipment costs have a substantial impact on total cost estimates as illustrated in Table 9-5. If pay rates are reduced to 70 percent of the default values, the total cost is reduced by 23-25 percent for all 64 scenarios[3] included in this analysis versus the baseline scenario of full pay rates (see Table 9-5) and purchase of all equipment. A reduction to 50 percent of default pay rates reduces the total cost by 38-42 percent. If labor costs are zeroed out, total costs are reduced by 68-84 percent. If pay rates are maintained at the default values and all equipment is assumed to be in hand with no purchases required, costs are reduced by 1-20 percent versus the baseline scenario. If equipment

[3] Two designs, 4 variable sets, 4 salary levels, 2 equipment cost levels (2x4x4x2=64).

purchases are assumed to not be needed and labor costs are reduced to 70 percent of the default values, total costs are reduced by 25-43 percent. If neither labor nor equipment costs are included, total cost is reduced by 81-98 percent of baseline cost, clearly illustrating the importance of assumptions on pay rates and equipment needs when estimating total monitoring program costs.

Table 9-5. Cost reductions due to lowering of labor and equipment costs

Salary Assumption	Equipment and Supplies	Cost Reduction vs. Base Scenario[1]	
		Range	Median
Full Cost	Purchase All	0[2]	0[2]
Reduced to 70%	Purchase All	23-25	24
Reduced to 50%	Purchase All	38-42	39
No cost for Labor	Purchase All	68-84	77
Full Cost	Zero cost	1-20	12
Reduced to 70%	Zero cost	25-43	35
Reduced to 50%	Zero cost	41-59	51
No cost for Labor	Zero cost	81-98	88

[1]Base scenario assumes full contractor salary levels and purchase of all equipment and supplies. All scenarios assume 5-yr monitoring in a 2,023-ha watershed 80 km from monitoring team.
[2]Base scenario of full pay rates (Table 9-4) and purchase of all equipment.

9.3.4 Major Conclusions from Cost Estimation Scenarios

The cost estimates provided in this section are intended to illustrate the importance of estimating the costs for *all* elements of monitoring for both the short- and long-term as part of establishing an effective and sustainable monitoring program that will meet watershed project monitoring objectives. Those who use either spreadsheet will find that they can tailor assumptions and add localized cost information to improve their estimation capabilities. With increasing experience, including making adjustments based on comparison of estimated versus actual costs, users should be able to improve the accuracy of their cost estimates over time. In all cases, but especially where budgets for monitoring are limited, accurate cost estimation is essential to assessing the potential for conducting a monitoring effort that will satisfy project objectives. Anything short of that is likely to be a waste of resources.

Because labor is such an important cost factor for all monitoring designs considered here, it provides the greatest opportunity for cost savings. These savings can be generated a number of ways, including:

- Using volunteers whenever possible. (Training costs may be incurred, however, and practical and legal limitations apply.)
- Using in-house labor. (This is not free and may involve diversion of labor from other projects or programs.)
- Negotiating contracts to ensure greater use of lower cost staff wherever appropriate.
- Using labor sources based within or near the watershed. (This will also reduce vehicle and lodging costs, but may limit options.)
- Piggybacking sampling trips with other duties to maximize benefits of travel time.

- Strategic use of in-house, volunteer, and contractor/grant labor to more efficiently match functions with capabilities and needs.
- Substituting higher initial cost equipment for some labor in long-term projects (e.g., telecommunications/data logging to reduce data collection trips).

In many cases the addition of non-instrumented monitoring sites to a watershed project can be relatively inexpensive because of the labor already invested in getting to the watershed for sampling events, as well as the labor needed to characterize the watershed, track land use and land treatment, analyze data, and develop reports. The incremental cost of adding monitoring stations should always be assessed in light of how they could contribute to achieving project monitoring objectives. For example, paired-watershed and above/below monitoring designs are inherently more powerful than single-station trend designs for evaluating the effectiveness of BMP implementation on a localized or watershed scale. The incremental cost of sampling two stations instead of one may support a stronger monitoring design that could yield results in a shorter time period, perhaps reducing overall costs in the end. In addition, findings may be more conclusive and the risk of failure reduced.

Equipment is never cheap, but the relatively low cost for equipment in most cost estimates developed here suggests that it may be cost-effective to use sophisticated equipment and instruments if they can offset higher personnel costs. Conversely, substituting labor for equipment (e.g., sending staff out to collect frequent observations vs. using a data logger) is not likely to be cost-effective. Finally, it is very important that equipment is maintained and operated in accordance with manufacturer recommendations to both obtain good data and to ensure that equipment is operable over its expected lifespan.

While this chapter did not focus on how total cost is affected by the selection of monitoring variables, it is clear that analysis of constituents such as pesticides and metals, as well as advanced methods such as microbial source tracking will cost more than *in situ* measurement of temperature or laboratory analysis of basic variables such as suspended sediment. Planners can use the spreadsheets to assess tradeoffs between adding more or different variables versus increasing sampling frequency or duration, or adding monitoring sites. Careful consideration of these and other design options should lead to better decisions regarding the makeup of a monitoring plan while both achieving monitoring objectives and staying within the budget.

9.4 Using Minimum Detectable Change to Guide Monitoring Decisions

As noted earlier, cost should not be the defining factor in the design of monitoring programs. Program designers must seek a balance that provides the ability to achieve monitoring objectives that are supportive of watershed project goals at an affordable cost. Monitoring design, for example, should be guided by the results of MDC analysis (see section 3.4.2) whenever possible. To illustrate this approach, cost estimates were developed for options considered in Example 1 (*A linear trend with autocorrelation and covariates or explanatory variables; Y values log-transformed*) of a technical note on MDC (Spooner et al. 2011). In the first scenario, weekly samples are collected for five years, resulting in an MDC of 15 percent, or an average of 3 percent change per year. By extending the monitoring period to 10 years, the MDC is increased to 20 percent, but with a lower average change of 2 percent per year required. Assuming that total P is the monitoring parameter of interest ($20 per sample analysis) the total cost (including a QAPP, reports, travel, etc.) for five years is estimated at $190,000, with 83 percent devoted to labor. A 10-year effort would cost $377,000. So, an additional $187,000 is needed to reduce the average annual change needed from 3 percent to 2 percent. This type of analysis would provide project managers with the cost information needed to determine whether they would prefer to enhance implementation of BMPs to achieve a faster rate of change or commit to a longer monitoring period to measure a slower rate of change.

The cost-benefit of adding explanatory variables can also be assessed through a combination of MDC analysis and cost estimation. For example, Spooner et al. (1987) demonstrated that adding salinity as a covariate in the Tillamook Bay, Oregon watershed study decreases the MDC for fecal coliform (yearly geometric concentration means) over an 11-year period of time (20 samples/yr; 14 sites) from 42 percent to 36 percent. For this same study, the MDC for fecal coliform decreases from 55 percent to 42 percent when doubling sampling frequency from 10 to 20 times per year over an 11-year study.

To estimate costs for the Tillamook Bay scenarios, it is assumed that there are 14 monitoring sites and fecal coliform is measured from one grab sample per site ($20/sample). Salinity is measured using a hand-held meter ($765). Sample size is increased by 10 percent for QA/QC. Sampling trips are assumed to involve 2 people for 8 hours each, including a 322-km (200 mi) round-trip to cover all 14 sites. The cost for a QAPP is assumed to be $1,400 and data analysis and reporting costs are $2,268 for the first and last years and $622 for the other nine years. The costs for watershed characterization, site establishment, and land use/treatment tracking are assumed to be zero.

These scenarios are summarized in Table 9-6. Adding salinity to the base scenario increases the 11-year cost by only $800 ($75/year) while improving the MDC by 8 percent from 55 percent (5 percent per year) to 47 percent (4.3% per year). Increasing sampling frequency nearly doubles the total 11-year cost while improving the MDC by 13 percent, from 55 percent to 42 percent (3.8 percent per year). Adding salinity measurement to the increased sampling frequency adds just $800 to the total 11-year cost, but reduces the overall MDC by an additional 6 percent to 36 percent (3.3 percent per year). Clearly, with or without an increase in sampling frequency, the additional $800 cost for salinity, while almost negligible, buys substantial additional sensitivity to detect a change in fecal coliform counts.

Table 9-6. Illustration of costs and MDC in response to changes in sampling program in Tillamook Bay, Oregon (Spooner et al. 1987)

Scenario	Sampling Program	Cost (11 years)	Cost Change[1]	MDC	MDC Change[1]
Base	10x/yr, FC	$182,600	--	55%	--
Add salinity	10x/yr, FC, salinity	$183,400	$800	47%	8%
Double frequency	20x/yr, FC	$347,400	$164,800	42%	13%
Double frequency, add salinity	20x/yr, FC, salinity	$348,200	$165,600	36%	19%

[1]Change versus Base scenario.

9.5 References

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*. 2nd ed. EPA/841-D-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 9, 2016. http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm.

Dressing, S.A. 2012. *Monitoring Cost Estimation Spreadsheet – Master Version*. Tetra Tech, Inc., Fairfax, VA. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

Dressing, S.A. 2014. *Monitoring Cost Estimation Spreadsheet – Simplified Version*. Tetra Tech, Inc., Fairfax, VA. Accessed April 29, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/monitoring-and-evaluating-nonpoint-source-watershed.

MDOE (Maryland Department of the Environment) and DDOE (District of Columbia Department of the Environment). 2010. *Total Maximum Daily Loads of Trash for the Anacostia River Watershed, Montgomery and Prince George's Counties, Maryland and the District of Columbia-Final*. Maryland Department of the Environment and District Department of the Environment-Natural Resources Administration. Accessed April 6, 2016.

NEMI (National Environmental Methods Index). 2006. *National Environmental Methods Index*. National Water Quality Monitoring Council. Accessed February 5, 2016. www.nemi.gov.

Spooner, J., R.P. Maas, M.D. Smolen, and C.A. Jamieson. 1987. Increasing the Sensitivity of Nonpoint Source Control Monitoring Programs. In *Symposium On Monitoring, Modeling, and Mediating Water Quality*, American Water Resources Association, Bethesda, Maryland, pp. 243-257.

Spooner, J., S.A. Dressing, and D.W. Meals. 2011. *Minimum detectable change analysis.* Tech Notes #7. Prepared for U.S. Environmental Protection Agency, by Tetra Tech, Inc., Fairfax, VA. Accessed March 24, 2016. https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes.

Appendix 9-1. Overview of Cost Estimation Spreadsheets

Both the master and simplified spreadsheets support cost estimation covering all items shown in Table 9A1-1 and Table 9A1-2. Various options exist for users to change spreadsheet costs (e.g., for labor or laboratory analysis) based on local information or experience, as well as assumptions regarding labor, equipment, and other requirements for monitoring designs of interest to the user. The master spreadsheet provides total flexibility in changing cost assumptions, whereas the simplified spreadsheet is designed to provide a set of default assumptions that facilitates development of cost estimates with minimal data entry. The master spreadsheet supports costing of virtually any monitoring design, while the simplified spreadsheet supports cost estimation for only above/below, paired, and trend monitoring designs.

Data entry requirements for the simplified worksheet are:

- Beginning year for monitoring (for inflation estimates).
- Monitoring design (above/below, paired, or trend – results in 1 or 2 sites).
- Watershed size and size of second watershed for paired design.
- Distance monitoring team is from watershed.
- Extra distance to drop samples off at laboratory.
- Average speed limit for drive to watershed.
- Average speed limit within watershed.
- Mileage rate paid for vehicles.
- Per diem rate (food and non-lodging expenses).
- Lodging rate (including taxes).
- Type of sampling (biological/habitat, grab, sondes, loads).
- Variable set (2 or 3 options per sampling type).
- Sampling frequency (same at each site).
- Duration of monitoring effort.

The simplified spreadsheet provides the sample type and variable set options shown in Table 9A1-1 (Note: codes are used in Appendix 9.4). Variable sets for these options are shown in Table 9A1-2 through Table 9A1-5. The number of units needed is calculated for each cost item based on the number of monitoring sites, sampling frequency, and monitoring plan duration. Note that inclusion of specific vendor products does not indicate EPA endorsement.

Table 9A1-1. Sample type and variable set options for simplified spreadsheet

Sample Type	Grab		Load		Biological/Habitat		Sondes	
	Variables	Code	Variables	Code	Variables	Code	Variables	Code
Variable Set Options	Nutrients, Sediment	NSC	Nutrients, Sediment	NSL	Biological Monitoring with Kick Net	BioK	Nutrients, Turbidity	SNT
	Bacteria, Nutrients, Sediment	BNSC	Bacteria, Nutrients, Sediment	BNSL	Biological Monitoring with D-Frame Dip Net	BioD	Nutrients, Turbidity, Metals	SNTM
	Metals, Sediment	MSC	Metals, Sediment	MSL				

Table 9A1-2. Grab sampling variable sets

Variable Set	Cost Items	
	Equipment and Supplies	Laboratory Analysis
Nutrients and Sediment (NSC)	Style A Staff Gage (13.5 ft), T-style post, and post driver	Total N using EPA Method 351.4
	Rain Gage (plastic)	Total P using EPA Method 365.4
	Cooler (54-quart) and ice for cooler	Suspended Sediment Concentration (USGS Method)
	Bottles-1000 ml wide mouth (HDPE, Box of 24)	
	Sulfuric Acid (10 N) Liter	
Bacteria, Nutrients, and Sediment (BNSC)	Same as above	Total N using EPA Method 351.4
		Total P using EPA Method 365.4
		Suspended Sediment Concentration (USGS Method)
		E. coli and total coliform via Micrology Labs Coliscan Easygel
Metals (Total and Dissolved) and Sediment (MSC)	Above items, *minus* sulfuric acid and *plus* the items below:	Suspended Sediment Concentration (USGS Method)
	Geopump Series 1 Peristaltic Pump AC/DC	Hardness EPA Method 130.2 - Titrimetry using EDTA
	Silicone Tubing, Size 24, 25'L (for use with peristaltic pumps)	Metals Scan (5 metals) using EPA Method 200.7 ($12/metal)
	12V Battery and Charger (for peristaltic pumps)	
	Solinst Model 860 Disposable Filters (0.45 µm) 1 filter	
	1:1 Nitric acid 500ml	

As shown below, the simplified spreadsheet allows users to apply labor adjustment factors (0 to 1.5 times default assumptions) to better simulate local labor costs. Inflation can also be factored into cost estimates. The base year assumed for inflation is 2012 because most costs in the spreadsheet are from that year. Users can also change default assumptions in the simplified spreadsheet to tailor them to local costs, but this requires a level of effort that mimics what is required for the master spreadsheet.

Salary Adjustment Factor:	1	
Inflation Rate (vs. 2012)	0.0	%

The simplified spreadsheet generates simple pie charts to show costs by category (see Figure 9A1-1). Total cost is also broken down as in Table 9A1-6. Total costs are given with and without inflation estimates. Annual costs are also generated by the simplified spreadsheet as shown in Table 9A1-7. The effect of inflation is illustrated by the change in costs between the years 2017 through 2021 which would all be the same without inflation.

Table 9A1-3. Load monitoring variable sets

Variable Set	Cost Items	
	Equipment and Supplies	Laboratory Analysis
Nutrients and Sediment (NSL)	USGS portable steel gage house (2'x3'x5' tall), connection to power grid, and surge protector	Total N using EPA Method 351.4
	Style A Staff Gage (13.5 ft), T-style post, and post driver	Total P using EPA Method 365.4
	Isco Model 6712FR Fiberglass Refrigerated Sampler, 2-bottle kit (7.5-liter polyethylene), 2 extra 7.5-liter polyethylene bottles for each site, intake line with strainer, battery-backed power pack, and Flowlink Software	Suspended Sediment Concentration (USGS Method)
	Isco 730 Bubbler Flow Module	
	Isco 581 RTD (rapid transfer device) for field retrieval of Model 6712FR data	
	Pygmy-type Current Meter w/ AquaCount data logger	
	Tipping Bucket Rain Gauge	
	HOBO Event Rainfall Logger (for tipping bucket rain gauge) and Boxcar Software	
	Cooler (54-quart) and ice for cooler	
	Sulfuric Acid (10 N) Liter	
Bacteria, Nutrients, and Sediment (BNSL)	Same as above	Total N using EPA Method 351.4
		Total P using EPA Method 365.4
		Suspended Sediment Concentration (USGS Method)
		E. coli and total coliform via Micrology Labs Coliscan Easygel ($18.50 for 10 tests)
Metals (Total and Dissolved) and Sediment (MSL)	Above items, *minus* sulfuric acid and *plus* the items below:	Suspended Sediment Concentration (USGS Method)
	Bottles-1000 ml wide mouth (HDPE, Box of 24)	Hardness EPA Method 130.2 - Titrimetry using EDTA
	Geopump Series 1 Peristaltic Pump AC/DC	Metals Scan (5 metals) using EPA Method 200.7 ($12/metal)
	Silicone Tubing, Size 24, 25'L (for use with peristaltic pumps)	
	12V Battery and Charger (for peristaltic pumps)	
	Solinst Model 860 Disposable Filters (0.45 μm) 1 filter	
	1:1 Nitric acid 500ml	

Table 9A1-4. Biological monitoring variable sets

Variable Set	Cost Items
Kick Net Option (BioK)	Style A Staff Gage (13.5 ft), T-style post, and post driver
	YSI 556 D.O., pH, conductivity, temperature meter with pH kit
	pH buffer, conductivity, and ORP calibration solutions for YSI 556
	Hach Model 2100Q Portable Turbidimeter with USB+Power Module for 2100Q (for data transfer to PC) and Gelex Secondary Standards Kit
	Silicone oil and portable turbidimeter sample cells for Hach Turbidimeter
	Pentax Option W30 waterproof digital camera
	Garmin eTrex 30 GPS
	Current meter outfit (Pygmy-type). Meter, headphones, and rod.
	Bottom kick net (500 µm mesh)
	Forceps (straight fine point)
	Sieve bucket
	First aid kit, 119-piece, economy
	STEARNS neoprene chest waders and fluorescent orange PVC gloves
	Bottles-1000 ml wide mouth (HDPE, Box of 24)
	Low plastic specimen jars and black molded caps
	Ice (cooler full)
	95% Ethanol (3.8 L)
D-Frame Dip Net Option (BioD)	Above items, *minus* bottom kick net and *plus* item below
	D-Frame dip net (500 µm mesh)

Table 9A1-5. Sondes monitoring variable sets

	Cost Items	
Variable Set	Equipment and Supplies	Laboratory Analysis
Nutrients and Turbidity Set (SNT)	Style A Staff Gage (13.5 ft), T-style post, and post driver	Total P using EPA Method 365.4
	Rain Gage (plastic)	
	Hydrolab DataSonde 5 - DS5 w/ built-in data logger, temperature sensor, and connecting cable (takes 10 sensors, measures up to 15 parameters simultaneously)	
	pH, polarographic DO, temperature (comes with unit), nitrate, self-cleaning turbidity, ammonia, chlorophyll a, and conductivity sensors for DS5	
	5-meter communication cable and battery pack for DS5	
	Bottles-1000 ml wide mouth (HDPE, Box of 24)	
	Cooler (54-quart) and ice for cooler	
	1:1 Nitric acid 500ml	
	Sulfuric Acid (10 N) Liter	

Variable Set	Cost Items	
	Equipment and Supplies	Laboratory Analysis
Nutrients, Turbidity, and Metals (Total and Dissolved) Set (SNTM)	Above items *plus* the items below:	Total P using EPA Method 365.4
	Geopump Series 1 Peristaltic Pump AC/DC	Hardness EPA Method 130.2 - Titrimetry using EDTA
	Silicone Tubing , Size 24, 25'L (for use with peristaltic pumps)	Metals Scan (5 metals) using EPA Method 200.7 ($12/metal)
	12V Battery and Charger (for peristaltic pumps)	
	Solinst Model 860 Disposable Filters (0.45 µm) 1 filter	

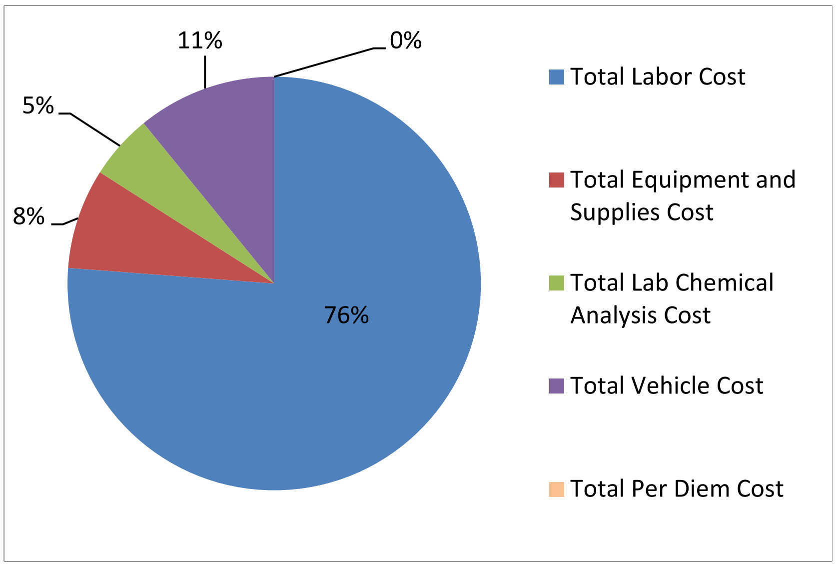


Figure 9A1-1. Pie chart from simplified spreadsheet

Table 9A1-6. Tabular output from simplified spreadsheet

Cost Category	Total Cost	% of Total
Labor	$205,167	72
Equipment and Supplies	$2,158	1
Sampling Analysis	$53,654	19
Vehicles	$22,921	8
Per Diem	$0	0
TOTAL COST	**$283,900**	**100**
Average Annual Cost	$40,557	
Total Cost with Inflation	*$325,887*	
Average Annual Cost with Inflation	*$46,555*	

Table 9A1-7. Annual costs from simplified spreadsheet

Inflation Rate: 2%			
Begin: 2016			
End: 2022			
Year	Inflation Factor Applied	Annual Cost without Inflation	Annual Inflation-Adjusted Cost
2016	1.08	$47,349	$51,253
2017	1.10	$39,279	$43,367
2018	1.13	$39,279	$44,234
2019	1.15	$39,279	$45,119
2020	1.17	$39,279	$46,021
2021	1.20	$39,279	$46,942
2022	1.22	$40,157	$48,951
	TOTAL	$283,900	$325,887

Appendix 9-2. Cost Estimates for a Diverse Range of Monitoring Options

As described in Appendix 9-1, a large number of assumptions must be made to estimate costs for various monitoring scenarios. Thus, while these cost estimates are intended to be informative, they are more or less relevant to any particular monitoring effort based on how well the assumptions match the realities of that specific situation. Cost estimates given here are more likely to be high than low because it is always assumed that contractors perform the monitoring (i.e., no use of in-house labor that was hired to do monitoring) and all monitoring equipment must either be leased or purchased.

Cost Scenarios and Assumptions

Cost estimates for the following eight monitoring scenarios are presented in this section.

1. Synoptic Survey
2. TMDL – Water Quality Standards
3. TMDL – Loads
4. Paired-Watershed – Loads
5. Long-term Single Station – Biomonitoring
6. Above/Below BMP Effectiveness – Biomonitoring
7. Input/Output Urban LID Effectiveness
8. Photo-Point Monitoring

These eight scenarios address both problem assessment and project evaluation, using chemical, physical, and biological (Barbour et al. 1999) monitoring methods. Five-year total costs are used for comparing Scenarios 2-8, but costs are also provided for 1, 2, and 8 years. The synoptic survey is considered a one-year effort.

The Watershed

The setting assumed for the cost scenarios is a 3,035 ha (7,500 ac) watershed, primarily in agricultural use with some urban influence. Monitoring is performed in perennial streams.

For the synoptic survey (Scenario 1) it is assumed that the nature and extent of water quality problems in the watershed are totally unknown. Thus, water chemistry sampling includes a wide range of variables. For Scenarios 2-7, the problems are assumed to be associated with sediment, nutrients, aquatic life use support, and cadmium toxicity. Stream channel restoration is the focus of Scenario 8.

Labor Costs

All monitoring is assumed to be performed by contractors; different pay rates would apply to government and university employees, and volunteers would work for free. Pay rates assumed (including fringe and overhead) and basic job functions are summarized in Table 9A2-1.

Table 9A2-1. Labor costs assumed for scenarios

Pay Level	Rate ($/hr)[1]	Job Functions
4	130	Monitoring design, statistical analysis, oversight, etc.
3	80	Lead field person for monitoring, data collection, bulk of writing
2	56	Field technician, lab tech, etc.
1	34	Secretarial and support staff

[1]Includes fringe and overhead.

Other Cost Assumptions

Monitoring proposals are assumed to be QAPPs (Quality Assurance Project Plans) prepared in 16 hours by a team that includes an expert and support staff at a cost of $1,400 for each scenario.

Transportation costs (vehicle and labor) include driving to and from the watershed, driving to monitoring sites within the watershed, and delivering samples to a laboratory for analysis. It is assumed that the watershed is 160 km (100 mi) from the base of those performing the monitoring. The sample analysis laboratory is assumed to be “on the way,” so no additional mileage is added for delivering samples to the laboratory.

Watershed characterization (windshield survey) costs are included only in Scenario 1. Monitoring site selection and establishment (as needed) costs are included in all scenarios. While it is a very important part of most NPS monitoring designs and is addressed by the spreadsheet, costs for meteorological monitoring were not included in these scenarios.

Analytical methods for water quality variables were obtained from various sources such as NEMI (http://www.nemi.gov/). Constraints associated with these methods (e.g., cooling samples to 4˚C for suspended sediment, and pre-acidification for hardness) are reflected in the cost estimates through, for example, the purchase of refrigerated samplers or the use of both pre-acidified and non-acidified sample containers.

For safety reasons, all sampling is assumed to be performed by teams of at least two people. In some cases, one or two additional people are added for a limited number of sampling trips. Larger teams are assumed necessary for QA/QC checks, stage-discharge calibration during a regularly scheduled sampling event, and scenarios where both intensive water chemistry and biological monitoring are performed. In all cases where continuous flow is measured, additional labor is assumed for stage-discharge calibration.

Scenario Description and Results

Scenario 1: Synoptic Survey

Under this scenario, a windshield survey is performed to characterize the watershed and select monitoring sites. It is assumed that the survey covers 512 km (320 mi) within an 8-hour day. Water quality monitoring at six sites (1.6 km [1 mi] apart from each other) is performed on two separate sampling dates to cover both high-flow and low-flow conditions. Each sampling run is assumed to require 400 km (250 mi) and a 12-hour day (1 hour per site, plus driving time) for a team of three in a single vehicle.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and jars, water quality sonde with 6 sensors (D.O., pH, temperature, conductivity, turbidity, and chlorophyll a), pygmy-type meter with data logger, kick net
- Sampling for all 6 sites: B.O.D., hardness, SSC, TP, TKN, NO_2+NO_3 –N, *E. coli* and total coliforms, biological monitoring, flow
- Sampling for 3 sites (targeted locations to keep costs down): grab sample for pesticides scan and metals scan (5 metals)

As shown in Table 9A2-2, the total cost for this one-year effort is estimated at $30,000. Equipment and labor each account for 45% of the total cost. Assuming that the contractor already has the basic monitoring equipment, however, the one-year total cost is reduced to just over $17,000.

Scenario 2: TMDL – Water Quality Standards

Scenario 2 envisions a TMDL under which water quality monitoring is performed at a single site to both track dissolved cadmium concentration (weekly grab samples) and assess aquatic life use support through biological monitoring.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles, multi-probe water quality meter for *in situ* D.O., pH, conductivity, and temperature measurements, kick net
- Sampling: cadmium and hardness, biological monitoring (2x/yr)

As shown in Table 9A2-2, the total cost for five years is about $214,900. Costs for one year, 2 years, and 8 years are estimated at $47,100, $90,300, and $339,400, respectively. Nearly 83% of the total cost is associated with sampling trips, with another 7% for analysis of samples for cadmium and hardness. Labor accounts for 81% of the total budget, and equipment account for only 2% of the total 5-year budget.

Scenario 3: TMDL – Pollutant Load

Under this scenario, weekly flow-weighted composite samples are taken for suspended sediment load estimation at a single site. Continuous discharge is measured with a bubbler water level sensor and a pygmy-type current meter is used for calibration.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample shed, refrigerated automatic sampler (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, surge protector, pygmy-type current meter and data logger
- Sampling: discharge and suspended sediment concentration

As shown in Table 9A2-2, the total five-year cost for this scenario is estimated at $237,500. Total costs for 1, 2, and 8 years are $61,800, $106,500, and $368,400, respectively. Sampling trips and labor account for 87% and 84% of the total cost, respectively.

Scenario 4: Paired-Watershed Loads

This scenario is in many ways a doubling of Scenario 3, but shared equipment (e.g., pygmy-type current meter) is not duplicated and incremental costs for analyzing and reporting on data from the second monitoring station are assumed to be half the cost for the first monitoring station. The watersheds are assumed to be 12.8 km (8 mi) apart. Weekly flow-weighted composite samples are taken for suspended sediment load estimation at each site using an automatic sampler. Continuous discharge is measured with a bubbler water level sensor and a pygmy-type current meter is used for calibration. Unlike for Scenario 3, tracking of land use and land treatment is included in the analysis, with the cost essentially twice that for Scenario 5.

Equipment and sampling assumptions for this scenario are:

- Equipment: 2 sample sheds, 2 refrigerated automatic samplers (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, 2 surge protectors, pygmy-type current meter and data logger
- Sampling: discharge and suspended sediment concentration

As shown in Table 9A2-2, this is the most expensive scenario considered here with a total five-year cost estimated at $347,800. Total costs for 1, 2, and 8 years are $93,400, $158,100, and $537,400, respectively. Sampling trips account for about three-quarters of the total cost. Site establishment cost is significant under this scenario, accounting for nearly 7% of the total cost, while sample analysis represents about 2% of the total cost. Labor is the largest cost category at 84% of the total cost.

Scenario 5: Long-Term Trend Monitoring-Biological

This scenario assumes long-term biological monitoring (2x/yr) at a single site. Stage is measured as a covariate, but discharge is not estimated. Land use and BMP implementation are tracked via two whole-watershed surveys per year.

Equipment and sampling assumptions for this scenario are:

- Equipment: multi-probe water quality meter for *in situ* D.O., pH, conductivity, and temperature measurements, staff gage, kick net, sample bags
- Sampling: biological monitoring (2x/yr)

The total cost for five years is estimated at $52,800, while the total costs for 1, 2, and 8 years are estimated at $16,100, $25,800, and $79,800, respectively. As shown in Table 9A2-2, land use tracking accounts for about 28% of the total five-year cost, while annual sampling trips consume 26% of the five-year budget. An additional 20% is used for data analysis and reporting. The largest cost category is labor at 85% of the total cost.

Scenario 6: Above/Below BMP Effectiveness Monitoring-Biological

This scenario assumes long-term biological monitoring (2x/yr) at two monitoring sites in an above/below design to evaluate individual BMP effectiveness. Stage at the time of sampling is measured as a covariate, but discharge is not estimated. Land use and BMP implementation are tracked via two partial-watershed surveys per year.

Equipment and sampling assumptions for this scenario are:

- Equipment: multi-probe water quality meter for *in situ* D.O., pH, conductivity, and temperature measurements, 2 staff gages, kick net, sample bags
- Sampling: biological monitoring (2x/yr)

As shown in Table 9A2-2, the five-year total cost is estimated at $58,000. One-year, two-year, and eight-year total costs are estimated at $17,200, $28,000, and $88,000, respectively. The total cost for this scenario nearly matches that for Scenario 5. Despite having two sites instead of one, annual sampling trips for Scenario 6 ($15,720) cost only slightly more than for Scenario 5 ($13,860). The time spent tracking land use/land treatment is substantially greater for Scenario 5 because the entire watershed is tracked versus only a portion of the watershed under the Scenario 6 above/below study. This difference explains the greater amount and percentage of the Scenario 5 budget devoted to land use tracking ($14,640, 28%) versus that for Scenario 6 ($11,800, 20%). Labor accounts for 85% of the five-year budget.

Scenario 7: Input/Output Urban LID Effectiveness

The analysis of inflow-outflow monitoring of urban LID practices assumes two monitoring stations, one storm event sampled per week at each station, discharge measurement, and analysis of both suspended sediment and five metals.

Equipment and sampling assumptions for this scenario are:

- Equipment: 2 small sample sheds, 2 refrigerated automatic samplers (with 2-bottle kit), data transfer device and software, 2 submersible pressure transducers with data logger, 2 V-notch weir boxes, 2 surge protectors
- Sampling: discharge, suspended sediment concentration, metals scan (5 metals)

As shown in Table 9A2-2, the five-year total cost for this scenario is estimated at $251,400, while estimated total costs for 1, 2, and 8 years are $68,000, $114,900, and $387,800. Costs for monitoring site establishment and equipment contribute to the high first-year cost of this study design. After five and eight years, however, the average annual costs drop to about $50,300 and $48,500, respectively. Annual sampling trips account for nearly 71% of the total five-year budget, while annual sample analysis accounts for 16%, and equipment and site establishment combine for just over 8%. Labor is the largest cost category at 68% of the total five-year budget.

Scenario 8: Photo-Point Monitoring

This scenario assumes repeat photography of a riparian zone restoration project using two photo points (see chapter 5). Each photo point has a single camera point. Cost estimates were developed for both qualitative and quantitative approaches, with digital image analysis assumed for the quantitative approach.

Equipment assumptions for qualitative photo-point monitoring are:

- 2 meter boards, digital camera with tripod, GPS unit, field computer, compass, level, sledge hammer, measuring tape, rebar, shovel, whiteboard, metric staff gage

As shown in Table 9A2-2, the five-year cost for qualitative photo-point monitoring is estimated to be about $18,600, with 81% of the cost devoted to labor. If it is assumed that the contractor already has the major equipment, the total cost for five years is reduced to about $16,300. Total costs for 1, 2, and 8 years are estimated at $8,100, $11,100, and $26,000, respectively. Annual sampling trips account for about 48% of the total five-year budget, while site establishment, portable sampling equipment, and startup supplies consume a combined 22% of the budget. Labor is the largest cost category at 81% of the total.

When considering photo-point as an add-on monitoring activity (e.g., the same individuals who perform biological monitoring or collect water chemistry samples also take the photos), the five-year cost is reduced to $8,500 due primarily to savings in labor and vehicle costs. Coupled with the assumption that the contractor already has the major equipment the 5-year cost drops to about $6,200.

Quantitative photo-point analysis requires image processing software, and labor requirements for data analysis are increased substantially. Because quantitative photo-point analysis has not been used to any measurable extent in watershed projects, the cost estimates provided here are highly uncertain. The total cost for five years is estimated at $74,900 with 90% of the cost for labor. Assuming the contractor has all major equipment and software, the 5-year cost is reduced to about $68,700. If quantitative photo-point monitoring is added to a water chemistry or biological monitoring program, the cost is estimated at just over $53,000 for five years. Coupled with the assumption that the contractor already has the major equipment and software the 5-year cost drops to $46,800.

Table 9A2-2. Total costs for eight diverse scenarios

	Scenario								
	1	2	3	4	5	6	7	8 Qualita-tive	8 Quantita-tive
Cost Phase or Element	1 Year	5 Years	5 Years	5 Years	5 Years	5 Years	5 Years	5 Years	5 Years
Proposal and QAPP	$1,400	$1,400	$1,400	$1,400	$1,400	$1,400	$1,400	$1,400	$2,440
Watershed Characterization	$1,858	$0	$0	$0	$0	$0	$0	$0	$0
Site Establishment	$0	$0	$11,409	$22,829	$2,110	$2,234	$17,598	$1,860	$1,860
Portable Sampling Equipment and Startup Supplies Costs	$13,332	$4,210	$4,803	$4,803	$3,595	$3,595	$3,056	$2,260	$6,241
One-Time Office Equipment and Startup Supplies Costs	$0	$0	$0	$0	$0	$0	$0	$0	$0
Station Demolition and Site Restoration	$0	$0	$808	$1,616	$0	$0	$136	$114	$114
First-Year Report	$3,610	$1,952	$1,692	$2,448	$1,952	$1,952	$2,250	$720	$11,676
Final Report	$0	$3,608	$2,976	$4,422	$2,656	$2,656	$3,336	$1,224	$10,980
Annual Access Fees	$0	$0	$0	$0	$0	$0	$0	$0	$0
Annual Sampling Trips to Sites	$6,008	$177,660	$205,660	$266,539	$13,860	$15,720	$177,840	$8,820	$13,960
Annual Volunteer Training	$0	$0	$0	$0	$0	$0	$0	$0	$0
Annual Sample Analysis	$3,774	$15,880	$2,860	$5,720	$4,960	$9,920	$40,040	$0	$0
Annual Data Analysis	$0	$2,268	$2,268	$3,402	$2,268	$2,268	$2,412	$342	$17,280
Annual Reports	$0	$3,588	$3,588	$5,316	$3,588	$3,588	$3,288	$1,818	$10,368
Annual Site Operation and Maintenance	$0	$0	$0	$0	$0	$0	$0	$0	$0
Annual Supplies and Rental Equipment	$59	$4,313	$0	$0	$1,795	$2,855	$0	$0	$0
Annual Land Use Tracking	$0	$0	$0	$29,280	$14,640	$11,800	$0	$0	$0
TOTAL	**$30,040**	**$214,879**	**$237,464**	**$347,775**	**$52,824**	**$57,988**	**$251,356**	**$18,558**	**$74,919**

Appendix 9-3. Cost Estimates for Watershed-Scale Evaluation of Agricultural BMP Implementation

As described in Appendix 9-1, a large number of assumptions must be made to estimate costs for various monitoring scenarios. Thus, while these cost estimates are intended to be informative, they are more or less relevant to any particular monitoring effort based on how well the assumptions match the realities of that specific situation. Cost estimates given here are more likely to be high than low because it is always assumed that contractors perform the monitoring (i.e., no use of in-house labor that was hired to do monitoring) and all monitoring equipment must either be leased or purchased.

Cost Scenarios

Cost estimates for the following seven monitoring scenarios are described in this section. Results of the cost analysis are summarized in Figure 9-2. One year is assumed for the synoptic survey, and costs for other scenarios are estimated for 3, 5, and 7 years.

1. Preliminary Synoptic Survey
2. Compliance Monitoring
3. Above/Below Monitoring (sub-scenarios for concentration and load: 3C, 3L)
4. Multiple-Watershed Monitoring
5. Trend Monitoring (sub-scenarios for concentration and load: 5C, 5L)
6. Paired-Watershed Monitoring (sub-scenarios for perennial and intermittent flows: 6P, 6I)
7. Soil Testing

The Watershed

The setting assumed for these cost scenarios is a 12-digit HUC watershed covering 10,117 ha (25,000 ac), primarily in agricultural use. Monitoring is performed in perennial streams with the exception of Scenario 6I which assumes intermittent flow. Scenario 1 assumes that the nature and extent of water quality problems in the watershed are totally unknown, so a wider range of monitoring variables is included. For Scenarios 2-7, the problems are assumed to be associated with nutrients from agricultural sources.

Labor Costs

Labor cost assumptions are the same as described in Appendix 9-2 (Table 9A2-1).

Driving Distances and Sampling Times

Transportation costs include driving to and from the watershed, driving to monitoring sites within the watershed, and delivering samples to a laboratory for analysis. To bracket a wide range of possibilities for transportation costs and sampling times, three one-way distances and associated drive times are assumed:

- "In" - Monitoring staff are within the watershed: distance and travel time are zero.
- "Near" - Monitoring staff are based 240 km (150 mi) from the watershed, with a one-way drive time of 2 hour and 45 minutes.
- "Far" - Monitoring staff are based 480 km (300 mi) from the watershed, with a one-way drive time of 5.5 hours.

Drive distances and times for sampling runs *within* (i.e., in addition to travel distance and times *to* the watershed) the watershed are assumed to be:

- Zero miles and time for trend monitoring (1 station)
- 25 km (16 mi) and 0.5 hours R/T for compliance and above/below monitoring (2 stations)
- 48 km (30 mi) and 0.75 hours R/T for paired-watershed monitoring (2 stations, 1 in nearby watershed 24 km [15 mi] away)
- 96 km (60 mi) and 2.5 hour R/T for multiple-watershed study (20 sub-watershed stations all within same watershed)
- 80 km (50 mi) and 2 hours R/T for a soil testing study (20 fields within the same watershed)

For all scenarios in which driving to the watershed is required, it is assumed that collected samples are dropped off at the laboratory in transit with no additional driving mileage. For scenarios in which the contractor is based in the watershed, 80 km (50 mi) is added for delivery of the samples to the nearest laboratory, except for Scenario 7 for which soil samples are assumed mailed to the laboratory.

It is assumed that contractors within the watershed will not incur lodging fees, while lodging is (generally) assumed for others when work days exceed 12 hours. Efforts were made to combine activities (e.g., site establishment and discharge observation) to reduce the need for overnight stays.

For safety reasons, all sampling is assumed to be performed by teams of at least two people. Two-person teams are assumed for grab sampling and 3 people are assumed necessary for runs including discharge observations. Periodic trips for QA/QC (e.g., 4 times per year for weekly sampling) by a QA/QC expert are also included.

The time required for grab sampling is assumed to be 0.5 hours per site, whereas sampling at sites with automatic sampling and discharge measurements is assumed to require 1.5 hours per site. Scenario 7 incorporates an assumption that 45 minutes is required to collect a composite soil sample for each 4-ha (10-acre) field that is monitored.

The cost of establishing a stage-discharge relationship is included for Scenarios 3L, 5L, 6P, and 6I. It is assumed that all monitoring is performed on wadeable streams, so time assumed for a discharge observation is set at 1.5 hours. Requirements for discharge observations on larger streams would be more expensive. Costs assume eight discharge observations per year, with 6 of these as separate trips and 2 as additional time during normal sampling runs. The driving distances and hours assumed necessary for discharge observations made within each study area as separate trips are summarized in Table 9A3-1.

Table 9A3-1. Driving and labor assumptions for discharge observations as stand-alone trips

Scenario	Discharge Observation		
	Number of Stations	Total Drive Distance Within Watershed[1]	Total Hours Within Watershed[1]
3L. Above/Below	2	25 km (16 mi)	3.50
5L. Trend (Load only)	1	0 km (0 mi)	1.50
6P. Paired	2	48 km (30 mi)	3.75
6I. Paired	2	48 km (30 mi)	3.75

[1]Does not include driving distance and time to arrive at watershed.

Table 9A3-2 summarizes assumptions regarding driving distances and time spent within (and between for paired-watershed design) each watershed for sampling runs. This does not include round-trip (R/T) travel to or from the watershed, nor does it include add-ons such as discharge observations.

Table 9A3-2. Sampling distances and times within watersheds

Scenario	No. of Sites	Travel Within Watershed		Travel Between Watersheds		Time at Each Site	Total PER Site	
		km	Hours	km	Hours	Hours	km	Hours
1. Synoptic	8	32	0.75	0	0	0.5	4	0.6
2. Compliance	2	25	0.5	0	0	0.5	12.5	0.75
3C. Above/Below (Conc.)	2	25	0.5	0	0	0.5	12.5	0.75
3L. Above/Below (Load)	2	25	0.5	0	0	1.5	12.5	1.75
4. Multiple	20	96	2.5	0	0	0.5	4.8	0.625
5C. Trend (Conc.)	1	0	0	0	0	0.5	0	0.5
5L. Trend (Load)	1	0	0	0	0	1.5	0	1.5
6P. Paired (Perennial)	2	0	0	48	1	1.5	24	2
6I. Paired (Intermittent)	2	0	0	48	1	1.5	24	2
7. Soil Test	20	80	2	0	0	0.75	4	0.85

Quality Assurance Project Plans (QAPPs)

Monitoring proposals are assumed to be QAPPs prepared in 16 hours by a team that includes an expert and support staff at a cost of $1,400 for each scenario.

Watershed characterization

Watershed characterization costs apply only to Scenario 1, including a windshield survey (240 km, 8 hours) and a review of available data and maps. For all other scenarios it is assumed that the watershed has been suitably characterized for development of the monitoring program.

9.5.1.1 Site selection and establishment

For Scenarios 1 and 2 (synoptic and compliance) site selection is assumed to be a desktop exercise, requiring two staff for four hours each. It is assumed that site selection for Scenario 7 involves more time because information must be gathered to find 20 fields via a random selection process. Two staff for 20 hours each is assumed for this effort, with any additional labor provided by cooperators within the watershed. For Scenarios 3-6 it is assumed that three staff each devote 2 hours of paper investigation to each monitoring site prior to traveling to the watersheds for field investigation.

Field costs for site selection include travel (to and within the watersheds) and labor. Costs assumed for field work under Scenarios 3-6 are summarized in Table 9A3-3. It is assumed that an additional person is needed for site selection that involves installation of a sampling shed and for Scenario 4 because 20 subwatersheds must be selected.

Monitoring site establishment (as needed) costs are included in Scenarios 3 through 6, with greater cost for sites with continuous discharge measurement and automated samplers. Major materials and equipment assumed for stations at which continuous flow is measured are summarized in Table 9A3-4. A tipping rain gauge, data logger, and software are purchased for Scenarios 3L, 5L, 6P, and 6I. Plastic rain gauges are purchased for Scenario 3C and 5C, while available local precipitation records are used for all other scenarios.

Table 9A3-3. Field work costs for site selection

Scenario [# stations]	1-Way Distance from Base	Travel and Site Investigation and Selection		Number of Staff / Number of Vehicles	Number of Overnight Stays[1]
	km/vehicle	km/vehicle	Hours/person		
3C. Above/Below (conc.) [2]	0	50	5	2/1	0
	240	530	10.5	2/1	0
	480	1,010	16	2/1	1
3L. Above/Below (load) [2]	0	50	5	3/1	0
	240	530	10.5	3/1	0
	480	1,010	16	3/1	1
4. Multiple Watershed [20]	0	322	48	3/1	3
	241	804	53.5	3/1	4
	483	1,287	59	3/1	4
5C. Trend (conc.) [1]	0	10	2.5	2/1	0
	240	490	8	2/1	0
	480	970	13.5	2/1	1
5L. Trend (load) [1]	0	10	2.5	3/1	0
	240	490	8	3/1	0
	480	970	13.5	3/1	1
6P. Paired (perennial) [2]	0	40	5	3/1	0
	240	520	10.5	3/1	0
	480	1,000	16	3/1	1
6I. Paired (intermittent) [2]	0	40	5	3/1	0
	240	520	10.5	3/1	0
	480	1,000	16	3/1	1

[1]Except where the contractor is based within the watershed, overnight lodging was assumed as needed to keep the length of work days reasonable (generally 12 hours or less).

Table 9A3-4. Major equipment and materials costs for stations measuring continuous discharge

Cost Item	Unit Cost
Build sampling shed (labor and materials)	$2,000
Connection to power grid	$800
Staff gage, post, and post driver	$154
Automatic sampler with bubble flow module, battery backup, 2-bottle kit, data transfer device, and software	$10,530
Pygmy-type current meter w/data logger	$2,015

Six hours is added per station (18 person-hours) in cases where a monitoring shed is installed for automatic sampling equipment. Table 9A3-5 summarizes travel and labor assumptions for site establishment field work for Scenarios 3 and 6.

Table 9A3-5. Site establishment costs for sites designed for load estimation

	2-Way Travel to Site[1]		Shed Construction and Setup		Total Without Discharge Observation	Hours Added for Discharge Observation[2]	Total	
Scenario [# stations]	km / Vehicle	Hours / Person	Hours / Person	# Staff / # Vehicles	Hours / Person	Hours / Person	Hours / Person	# Nights
3L. Above / Below [2]	0	0	12	3/1	12	0	12	0
	480	5.5	12	3/1	17.5	3	20.5	1
	960	11	12	3/1	23	3	26	2
5L. Trend [1]	0	0	6	3/1	6	0	6	0
	480	5.5	6	3/1	11.5	1.5	13	0
	960	11	6	3/1	17	1.5	18.5	1
6P. Paired [2]	48	1	12	3/1	13	0	13	0
	528	6.5	12	3/1	18.5	3	21.5	1
	1,008	12	12	3/1	24	3	27	2
6I. Paired [2]	48	1	12	3/1	13	0	13	0
	528	6.5	12	3/1	18.5	3	21.5	1
	1,008	12	12	3/1	24	3	27	2

[1]Paired watersheds are assumed to be 24 km apart. Above/below sites are assumed to be less than 1 km apart.
[2]Hours were added to perform a discharge observation at each site where long-distance travel was involved and pollutant load estimation is planned.

Site Demolition and Restoration

Site demolition and restoration is only required for sites with sampling sheds. It is assumed that 3 people are needed for this activity, each working 3 hours at each monitoring station. Assumptions are summarized in Table 9A3-6.

Table 9A3-6. Site demolition and restoration costs

Scenario [# stations]	2-Way Travel to Site[1]		Site Demolition and Restoration	# Staff / # Vehicles	Total	# Nights
	km / Vehicle	Hours / Person	Hours / Person		Hours / Person	
3L. Above / Below [2]	0	0	6	3/1	6	0
	480	5.5	6	3/1	11.5	0
	960	11	6	3/1	17	1
5L. Trend [1]	0	0	3	3/1	3	0
	480	5.5	3	3/1	8.5	0
	960	11	3	3/1	14	1
6P. Paired [2]	48	1	6	3/1	7	0
	528	6	6	3/1	12	0
	1,008	11.5	6	3/1	17.5	1
6I. Paired [2]	48	1	6	3/1	7	0
	528	6	6	3/1	12	0
	1,008	11.5	6	3/1	17.5	1

[1]Paired watersheds are assumed to be 24 km apart. Above/below sites are assumed to be less than 1 km apart.

Sample Analysis

Analytical methods for water quality variables included in the spreadsheet were obtained from various sources such as NEMI (2006). Constraints associated with these methods (e.g., cooling samples to 4˚C for suspended sediment, and pre-acidification for hardness) are reflected in the cost estimates through, for example, the purchase of refrigerated samplers and sample preservatives.

Sample analysis for total P assumes EPA Method 365.4 (NEMI 2006) at a cost of $21 per sample. Soil samples under Scenario 7 are analyzed for soil P (Mehlich 3), textural class, and organic matter, at a total cost of $26 per sample. Soil samples are assumed to be sent by ground shipment to the laboratory.

The number of samples analyzed is increased by 10% for QA/QC.

Land Use/Treatment Tracking

Tracking of BMP implementation is assumed to occur twice per year under Scenarios 3, 5, and 6. The baseline assumption for tracking effort within a 12-digit HUC is 240 km (150 mi) driving and 8 hours R/T each time, with variations across scenarios due to differing monitoring scales and specifics. Travel distances and times to the watershed are added as appropriate.

For Scenario 4 it is assumed that a cooperator (e.g., NRCS) provides the data for the 20 subwatersheds on an annual basis; additional observations can be made during the 30-minute visits for grab sampling in each of the 405-ha (1,000-acre) subwatersheds. Under Scenario 7, annual data on organic and inorganic nutrient application rates and crop yields per field are assumed to be provided by a cooperator. The resulting assumptions are summarized in Table 9A3-7.

Table 9A3-7. Driving and labor assumptions for land use/treatment tracking

Scenario	Land Use/Treatment Tracking			Comments
	Drive Distance Within Watershed[1]	Hours Including Drive Time	Frequency (#/Yr)	
1. Synoptic	n/a	n/a	n/a	Not done.
2. Compliance	n/a	n/a	n/a	Not done.
3. Above/Below	128 km (80 mi)	6	2	Only part of watershed is tracked.
4. Multiple	n/a	n/a	n/a	Not done. Data provided by a cooperating agency.
5. Trend	240 km (150 mi)	8	2	Baseline assumption.
6. Paired	290 km (180 mi)	9	2	Tracking intensity varies by source and location.
7. Soil Test	n/a	n/a	n/a	Not done. Data provided by a cooperating agency.

[1]Does not include driving distance and time to arrive at watershed.

Supplies

Cost estimates include the purchase of ice for each sampling event and annual purchases of 1-liter HDPE bottles and sample preservative.

Data Analysis and Reports

Data analysis and reporting costs are set higher for the first and last years compared to the "middle" years. For example, involvement of higher paid staff is greater in the first and final years because of the challenges faced in developing data management and analysis procedures and rules. It is assumed that lower level staff can play a greater role in the middle years with oversight from senior staff.

The cost for analysis and reporting is greater for projects estimating pollutant loads versus those simply collecting concentration data. Synoptic surveys (Scenario 1) and compliance (Scenario 2) monitoring efforts are assumed to require less time than other scenarios because of greater simplicity. Data analysis and reporting for multiple-watershed studies (Scenario 4) is assumed to be the most time consuming despite less frequent sampling than found in Scenarios 3 and 6 because information is obtained from 20 subwatersheds. More hours are assumed for data analysis than for reporting in all cases for Scenario 7 because reports are assumed to be short and more straight-forward. Table 9A3-8 summarizes assumptions for data analysis and reporting.

Table 9A3-8. Labor assumptions for data analysis and reporting

Scenario	First-Year Report		Middle-Year Reports		Final-Year Report	
	Data Analysis (Hours)	Report Preparation (Hours)	Data Analysis (Hours)	Report Preparation (Hours)	Data Analysis (Hours)	Report Preparation (Hours)
1. Synoptic	12	18	n/a	n/a	n/a	n/a
2. Compliance	10	12	7	8	10	12
3C. Above/Below	15	22	12	12	17	22
3L. Above/Below	20	28	14	14	22	26
4. Multiple	36	34	26	24	38	32

Scenario	First-Year Report		Middle-Year Reports		Final-Year Report	
	Data Analysis (Hours)	Report Preparation (Hours)	Data Analysis (Hours)	Report Preparation (Hours)	Data Analysis (Hours)	Report Preparation (Hours)
5C. Trend	14	18	11	9	19	18
5L. Trend	16	26	13	12	20	23
6P. Paired	20	28	14	14	22	26
6I. Paired	20	28	14	14	22	26
7. Soil Test	16	12	13	10	20	11

Scenario Summaries

Scenario 1: Preliminary Synoptic Survey

Under this scenario, grab sampling is performed at 8 sites on two trips (low and high flow conditions). A team of 3 people conducts a windshield survey to characterize the watershed, but subsequent land use/land treatment tracking is not performed. Meteorological and flow data are assumed to be obtained as part of the desktop analysis of the watershed.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler
- Sampling: TP, SSC, B.O.D., *E. coli*, total coliform, discharge, and suspended sediment concentration

Scenario 2: Compliance Monitoring

Under this scenario, grab sampling (4x/yr) is performed at 2 sites. Land use/land treatment tracking is not performed.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler
- Sampling: TP

Scenario 3: Above/Below Monitoring

This scenario has two options. Land use/land treatment tracking (e.g., type and number of practices, acres treated) is performed 2x/yr for both options via windshield survey and collection of data from cooperators (e.g., USDA, Soil and Water Conservation District); emphasis is placed on the area between the above and below stations.

3C. Concentration Option: Weekly grab samples are collected at 2 sites.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, 2 plastic rain gages, 2 staff gages
- Sampling: TP, stage

3L. Load Option: Weekly flow-proportional composite samples are collected at 2 sites.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, 2 sample sheds, 2 tipping bucket rain gages and data logger, 2 staff gages, 2 refrigerated automatic samplers (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, 2 surge protectors, pygmy-type current meter and data logger
- Sampling: TP, continuous flow

Scenario 4: Multiple-Watershed Monitoring

Under this scenario there are 10 small watersheds each (n=20) with/without BMPs. Water quality sampling occurs six times per year. It is assumed that land use/land treatment tracking is performed 2x/yr by a cooperator, with additional observations made during water quality sampling runs.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, 20 staff gages
- Sampling: TP, stage

Scenario 5: Trend Monitoring

This scenario has one monitoring site and two options. Land use/land treatment tracking is performed 2x/yr for both options via windshield survey and collection of data from cooperators (e.g., USDA, Soil and Water Conservation District). Data are collected on the nature, extent, and timing of BMP implementation – as well as operation and maintenance after implementation.

5C. Concentration Option: Twice-monthly grab samples.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, plastic rain gage, staff gage
- Sampling: TP, stage, precipitation

5L. Load Option: Weekly flow-proportional composite samples.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, sample shed, tipping bucket rain gage and data logger, staff gage, refrigerated automatic sampler (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, surge protector, pygmy-type current meter and data logger
- Sampling: TP, continuous flow, precipitation

Scenario 6: Paired-Watershed Monitoring

This scenario has two monitoring sites (treated and untreated) and two options that address load estimation for a perennial and intermittent stream setting. For continuously flowing streams, a single weekly composite sample is collected for analysis. For intermittent streams, a flow-proportional composite sample is collected during each of 20 runoff events each year. Land use/land treatment tracking is performed 2x/yr in both watersheds for both options via windshield survey and collection of data from cooperators (e.g., USDA, Soil and Water Conservation District).

6I. Intermittent Stream Option: Twenty runoff events sampled per year at each site.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, 2 sample sheds, 2 tipping bucket rain gages and data logger, 2 staff gages, 2 refrigerated automatic samplers (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, 2 surge protectors, pygmy-type current meter and data logger
- Sampling: TP, continuous flow, precipitation

6P. Perennial Stream Option: Weekly flow-proportional composite samples at each site.

Equipment and sampling assumptions for this scenario are:

- Equipment: sample bottles and cooler, 2 sample sheds, 2 tipping bucket rain gages and data logger, 2 staff gages, 2 refrigerated automatic samplers (with bubble flow module, battery backup, 2-bottle kit), data transfer device and software, 2 surge protectors, pygmy-type current meter and data logger
- Sampling: TP, continuous flow, precipitation

Scenario 7: Soil Testing

This scenario involves random selection of 20 agricultural fields for annual soil sampling. Ten fields are beginning to adopt nutrient management, and the other ten are conventionally managed. Local precipitation records are used in lieu of on-site collection of precipitation data. Annual data on nutrient application and crop yields are provided by a cooperator (e.g., the landowner, USDA).

Equipment and sampling assumptions for this scenario are:

- Equipment: 2 soil probes, 2 buckets, and a supply of bags and ties for soil samples
- Sampling: soil P, textural class (covariate), and organic matter (covariate)

Appendix 9-4. Cost Estimates for Five-Year Trend and Above/Below Monitoring

As described in Appendix 9-1, a large number of assumptions must be made to estimate costs for various monitoring scenarios. Thus, while these cost estimates are intended to be informative, they are more or less relevant to any particular monitoring effort based on how well the assumptions match the realities of that specific situation. Cost estimates given here are more likely to be high than low because it is always assumed that contractors perform the monitoring (i.e., no use of in-house labor that was hired to do monitoring) and all monitoring equipment must either be leased or purchased.

The basic scenarios (n=160) assumed for this analysis are summarized in Table 9A4-1. The trend design assumes one monitoring site and the above/below design assumes two monitoring sites. All monitoring is assumed to continue for five years. Tracking of land use and land treatment is assumed to occur twice per year, with costs identical for all scenarios. The five-year costs for this tracking range from about $100 to $16,500 for all scenarios. Costs vary considerably based on the size of and distance to the watershed.

Table 9A4-1. Factors used in creating cost estimation scenarios

<table>
<tr><th>Scenario</th><th>Monitoring Variables Set (and source)</th><th>Sampling Frequency (times/year)</th><th>Monitoring Designs</th><th>Watershed Sizes (ha)</th><th>Distances to Watershed[1] (km)</th></tr>
<tr><td>Biological</td><td>BioK (Table 9A1-4)</td><td>2</td><td rowspan="4">Trend and Above/Below</td><td rowspan="4">202
2,023
10,117
20,234</td><td rowspan="4">0
40
80
121
161</td></tr>
<tr><td>Nutrient and Sediment Concentration</td><td>NSC (Table 9A1-2)</td><td rowspan="3">26</td></tr>
<tr><td>Nutrient and Sediment Load</td><td>NSL (Table 9A1-3)</td></tr>
<tr><td>Sondes for Nutrients and Turbidity</td><td>SNT (Table 9A1-5)</td></tr>
</table>

[1]Distance sampling team must travel to reach the watershed or nearest watershed being monitored.

Labor costs for these estimates use the same rates shown in Table 9A2-1. All scenarios include a mix of fixed labor assumptions (e.g., QAPP development cost is $1,400 for all[4] scenarios) and variable labor assumptions that are based on the monitoring design and watershed size. For example, watershed characterization costs vary depending on design and watershed size as illustrated in Table 9A4-2. A simple algorithm in the simplified spreadsheet estimates travel distances and drive times based on watershed size, affecting both labor and vehicle costs for watershed characterization.

[4] "All" scenarios refers to the base scenarios for which pay rates are those found in Table 9A2-1.

Table 9A4-2. Watershed characterization costs as function of design and watershed size

Design	Watershed Size (ha)			
	202	2,023	10,177	20,234
Trend	$1,516	$1,780	$2,952	$4,790
Above/Below	$1,888	$2,152	$3,324	$5,162

Distance to watershed is assumed = 80km.

Labor and vehicle requirements for sampling vary depending upon design, watershed size, and monitoring variables set. The variability of labor costs for data analysis and report development is illustrated in Table 9A4-3. These costs reflect the assumption that biological data require more time for analysis (at species level) than chemical/physical data collected using the other variable sets. Estimation and analysis of pollutant loads, likewise, is assumed to be more time-consuming than for either sonde or concentration data. Spreadsheet users, of course, can change these assumptions.

Table 9A4-3. Variability of costs for data analysis and reporting

Design	Variable Set	Samples/Year	5-Year Labor Cost for Data Analysis and Reporting
Trend	BioK	2	$15,889
	NSC	26	$10,051
	NSL	26	$16,271
	SNT	26	$11,899
Above/Below	BioK	2	$27,068
	NSC	26	$16,177
	NSL	26	$27,047
	SNT	26	$19,873

Assumes 2,023-ha watershed and 50 mile distance.

QA/QC is addressed in a number of ways. For sample analysis, sample size is increased by 10% to account for replicates. In addition, a QA/QC officer is assumed to join the sampling team once per year, and stage-discharge relationships are checked 8 times per year.

The results of running 160 scenarios for these above/below and trend monitoring designs are discussed in section 9.3.3 and summarized in Figure 9-4. Paired designs would have costs similar to those for the above/below design.

Additional cost estimates were run using a salary adjustment factor to see how this would affect total costs. Salaries were adjusted across the board by reducing them to 70%, 50%, and 0% of those in Table 9A2-1. Similarly, a rough assessment of the effects of equipment costs on total costs was performed by estimating costs where all or no equipment was purchased. These two equipment scenarios were also combined with the four salary options (0%, 50%, 70%, and 100% of the rates in Table 9A2-1) to explore the impacts of both adjustments on total costs. The results of these analyses are presented in section 9.3.3 and summarized in Table 9A3-3.