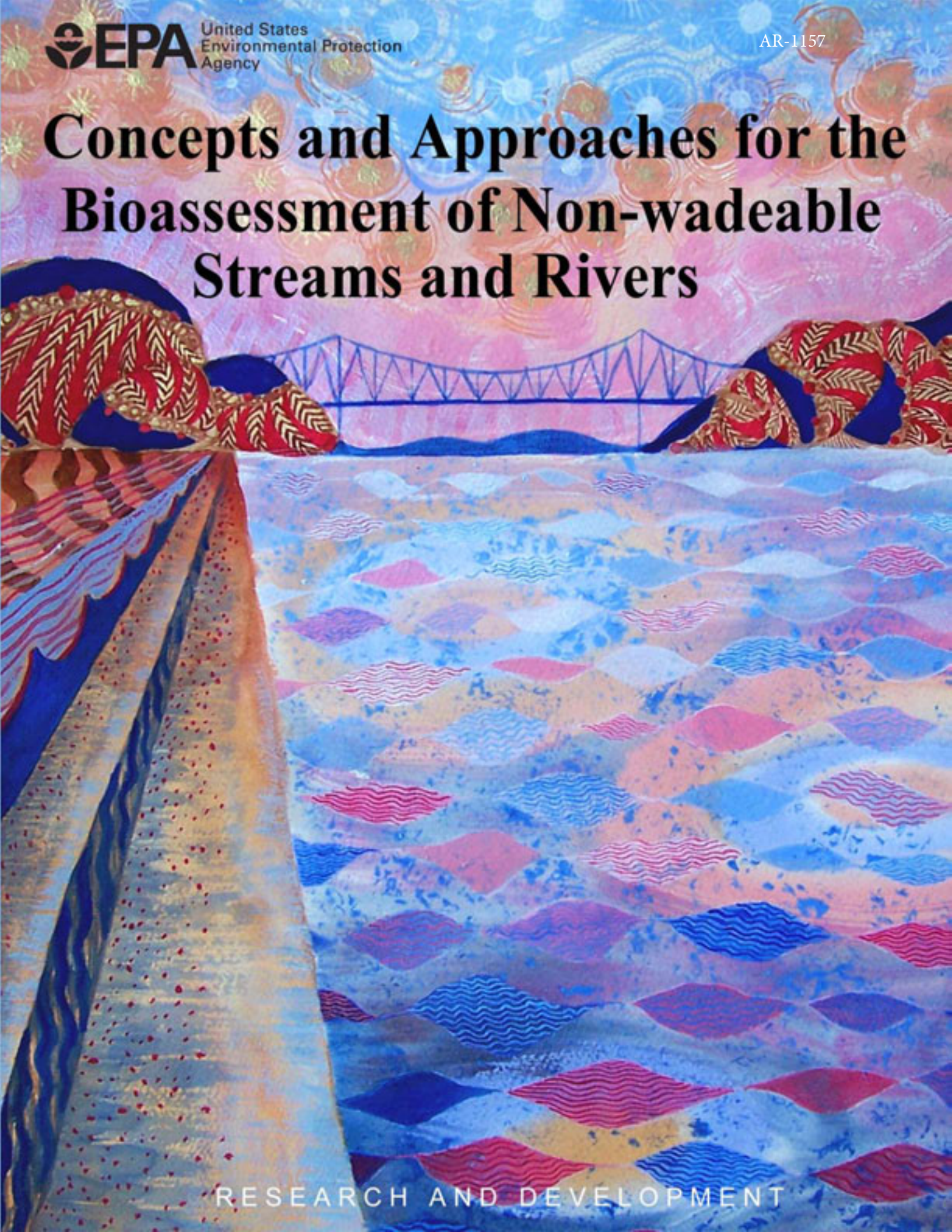
EPA United States Environmental Protection Agency

Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers



RESEARCH AND DEVELOPMENT

EPA/600/R-06/127
September 2006
www.epa.gov

Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers

Joseph E. Flotemersch[1], James B. Stribling[2], and Michael J. Paul[2]

[1]US Environmental Protection Agency
National Exposure Research Laboratory
26 W. Martin Luther King Dr., MS 642
Cincinnati, Ohio 45268

[2]Tetra Tech, Inc.
400 Red Brook Blvd., Suite 200
Owings Mills, Maryland 21117-5159

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

Notice

The U.S. Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to the Agency's peer and administrative review and approved for publication. The document was prepared at the EPA National Exposure Research Laboratory, Ecological Exposure Research Division, with support under contract 68-C-02-108 to Tetra Tech, Inc. Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy. Mention of trade names and commercial products does not constitute endorsement or recommendation for use.

The correct citation for this document is:

Flotemersch, J. E., J. B. Stribling, and M. J. Paul. 2006. Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.

This entire document can be downloaded from the website of the USEPA, Office of Research and Development, National Exposure Research Laboratory, Ecological Exposure Research Division:

http://www.epa.gov/eerd/rivers/

Foreword

"Downstream, that's where we are all headed. Out in the flow, we all feel it, sometimes turbulent, sometimes gentle flowing, but always moving…downstream."
Richard Russo
Empire Falls

Human history is replete with stories, songs, and pictures documenting the extraordinary importance and value of rivers to our lives. Almost every situation in life is reflected in rivers. Rivers are among the most animated of inanimate objects; among the liveliest of non-living natural things. Rivers offer us solace in times of trouble and instill awe in the power they can generate. They can provide a gentle place to rest, float, and think, but yet are persistent enough to erode even the hardest canyons the earth can muster.

Rivers touch all parts of the natural environment and nearly all aspects of human life and culture. They often act as centers of organization within landscapes. Their roles in providing natural resources such as fish and clean water are well known as are their roles in providing transportation, energy, diffusion of wastes and recreation (Naiman and Bilby 1998). However, as a consequence of this close relationship, the integrity of rivers is often challenged. Fish ecologist and essayist Peter Moyle has been quoted as saying "no matter how bad things are on land, you'll find that they're worse in nearby rivers." (Shepard 2001).

Sociocultural evolutionists have postulated that the United States has evolved from a pre-industrial society, to an industrial society, into what Daniel Bell (1973) coined a post-industrial society. Such societal changes are generally accompanied by changes in what society values. In the 1960's of the United States, this included an increased interest in the well-being and sustainability of our natural resources. ***This document is intended to provide support to those concerned with the well-being and sustainability of large rivers.***

There is a Chinese proverb that states *the mark of a successful man is that he has spent an entire day on the bank of a river without feeling guilty about it.* While this quote most likely speaks to the man's freedom from the need to earn money, the closest most of us will ever come is to enjoy earning our living on the banks of a river. This document represents an opportunity to do just that.

The US Environmental Protection Agency through its Office of Research and Development (Cincinnati, OH) and Regional Methods Initiative funded much of the research described herein and subsequent production of this document. The US Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington, DC provided additional funding for document production. Tetra Tech, Inc. provided primary technical support. Much appreciation is extended to the extensive list of reviewers who provided thoughtful and detailed critique of earlier drafts, and assisted our efforts to push the document toward scientific peer review standards. Any shortcomings, however, remain the responsibility of the authors.

Preface

In 1998, the National Exposure Research Laboratory (NERL) of the US Environmental Protection Agency's Office of Research and Development (USEPA-ORD) received funding to develop standardized protocols for the bioassessment of large (i.e., non-wadeable) streams and rivers. The request came from scientists in USEPA's regional offices who recognized that states and tribes need these protocols (which we term Large River Bioassessment Protocols or LR-BP), to meet their monitoring and enforcement objectives. In response, we conducted several years of research and development to adapt existing or devise new protocols, specific to the ecological and logistical demands of these large, flowing systems. We systematically compared alternative approaches and documented their performance characteristics, collaborating extensively with regional, State and Tribal scientists to ensure that the protocols were both technically feasible and economically practical.

We originally conceived of this document as a compilation of these research results. At the request of the user community, however, we have expanded it to present a comprehensive technical framework for the bioassessment of non-wadeable streams and rivers. While presentation of the LR-BP remains our main focus, several other bioassessment approaches exist that vary in purpose and technical approach. Therefore, the document is structured to show the technical relationship of the LR-BP to other protocols, and to assist the user in the selection of those that best allow programmatic management objectives to be met. We realize that in some cases protocols will need to be modified; to support these cases, the document provides information to assist the reader in determining the performance characteristics of the modified protocol.

In several locations in the document, specific programs have been highlighted to provide examples of how program elements might be more fully developed. Highlights are not intended to indicate endorsement or recommendation of these programs, nor should they be used as a stand-alone reference for field application. For more information on field applications, please consult the cited materials on these programs:

> Kaufmann, P. R. 2000. Physical habitat characterization - non-wadeable rivers. Chapter 6 in J. M. Lazorchak, B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). Environmental monitoring and assessment program - surface waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams. US Environmental Protection Agency, Cincinnati, OH. http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwatr/field/nonws1.html
>
> Moulton, S. R., II, J. G. Kennen, R. M. Goldstein, J. A. Hambrook. 2002. Revised protocols for sampling algal, invertebrate, and fish communities in the National Water-Quality Assessment program, U.S. Geological Survey Open-File Report 02-150. http://water.usgs.gov/nawqa/protocols/OFR02-150/index.html

Merritt, R. W., J. D. Allan, K. W. Cummins, K. J. Wessell, and J. G. O. Wilhelm. 2003. Qualitative biological and habitat survey protocols for Michigan's non-wadeable rivers. Submitted to the Michigan Department of Environmental Quality, Lansing, MI.

Ohio River Valley Water Sanitation Commission (ORSANCO). http://www.orsanco.org

For the most recent field operations material from the US Environmental Protection Agency's, Environmental Monitoring and Assessment Program, please consult:

Angradi, T. R. (editor). 2006. Environmental Monitoring and Assessment Program: Great River Ecosystems, Field Operations Manual. EPA/620/R-06/002. U.S. Environmental Protection Agency, Washington, D.C. http://www.epa.gov/emap/greatriver/fom.html

Peck, D. V., D. K. Averill, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, M. R. Cappaert, T. Magee, and P. A. Monaco. In press. Environmental Monitoring and Assessment Program - Surface Waters Western Pilot Study: Field Operations Manual for Non-Wadeable Rivers and Streams. EPA 620/R-0?/xxx. US Environmental Protection Agency, Washington, DC.

Acknowledgments

Contributions to the Development of this Document

Victor de Vlaming
School of Veterinary Medicine
University of California
Davis, Davis, CA

Kevin Goding
School of Veterinary Medicine
University of California
Davis, Davis, CA

Kenneth Lubinski
US Geological Survey
Upper Midwest Environmental Center
La Crosse, WI

Technical Support

Brad Autrey
US Environmental Protection Agency
National Exposure Research Laboratory
Ecological Exposure Research Division
Cincinnati, OH

Abby Markowitz
Tetra Tech, Inc.
Owings Mills, MD

Brenda Decker
Tetra Tech, Inc.
Owings Mills, MD

Primary Review

J. David Allan
The University of Michigan
School of Natural Resources and Environment
Ann Arbor, MI

Randy Bruins
US Environmental Protection Agency
National Exposure Research Laboratory
Ecological Exposure Research Division
Cincinnati, OH

John J. Hutchens, Jr.
Coastal Carolina University
Department of Biology
Conway, SC

Additional Review

Ted Angradi
US Environmental Protection Agency
National Health & Environmental
Effects Laboratory
Mid-Continent Ecology Division
Duluth, MN

Brad Autrey
US Environmental Protection Agency
National Exposure Research Laboratory
Ecological Exposure Research Division
Cincinnati, OH

Mike Beiser
Mississippi Department of
Environmental Quality
Compliance and Enforcement
Monitoring Section
Pearl, MS

Mike Bilger
EcoAnalysts, Inc.
Selinsgrove, PA

Karen Blocksom
US Environmental Protection Agency
National Exposure Research Laboratory
Ecological Exposure Research Division
Cincinnati, OH

Frank Borsuk
US Environmental Protection Agency
Region 3
Wheeling, WV

John Brumley
Kentucky Division of Water
Watershed Management Branch
Frankfort, KY

Erich Emery
The Ohio River Valley Water Sanitation
Commission (ORSANCO)
Cincinnati, OH

Steven W. Foster
US Army Corps of Engineers
Huntington District
Apple Grove, WV

Jennifer Greenwood
University of Tennessee at Martin
Biological Sciences
Martin, TN

Gretchen Hayslip
US Environmental Protection Agency
Region 10
Seattle, WA

Evan Hornig
US Environmental Protection Agency
Office of Water
Office of Science and Technology
Health & Ecological Criteria Division
Washington, D.C.

Brian Hill
US Environmental Protection Agency
National Health & Environmental
Effects Laboratory
Mid-Continent Ecology Division
Duluth, MN

Bob Hughes
Oregon State University
Department of Fisheries & Wildlife
Corvallis, OR

Donald Jackson
Mississippi State University
Department of Wildlife and Fisheries
Mississippi State, MS

Phil Kaufmann
US Environmental Protection Agency
National Health and Environmental
Effects Laboratory
Western Ecology Division
Corvallis, OR

Phil Larsen
US Environmental Protection Agency
National Health and Environmental Effects Laboratory
Western Ecology Division
Corvallis, OR

Gary Lester
President, EcoAnalysts, Inc.
Moscow ID

Robert L. Limbeck
Delaware River Basin Commission
West Trenton, NJ

Maggie Passmore
US Environmental Protection Agency
Region 3
Wheeling, WV

Rodney Pierce
Kentucky Division of Water
Watershed Management Branch
Frankfort, KY

Barry C. Poulton
US Geological Survey
Columbia Environmental Research Center
River Studies Station
Columbia, MO

Louis Reynolds
US Environmental Protection Agency
Region 3
Wheeling, WV

Greg Pond
US Environmental Protection Agency
Region 3
Wheeling, WV

Nancy Schuldt
Fond du Lac Band of Lake Superior Chippewa Reservation
Water Projects Coordinator
Cloquet, MN

John Van Sickle
US Environmental Protection Agency
National Health and Environmental Effects Laboratory
Western Ecology Division
Corvallis, OR

Blaine Snyder
Tetra Tech, Inc.
Owings Mills, MD

Jeff Thomas
The Ohio River Valley Water Sanitation Commission (ORSANCO)
Cincinnati, OH

Chris Yoder
Midwest Biodiversity Institute
Center for Applied Bioassessment & Biocriteria
Columbus, OH

Matt Wooten
Northern Kentucky Sanitation District #1
Watershed Management
Ft. Wright, KY

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Tables (continued)

List of Figures

List of Figures (continued)

List of Figures (continued)

Acronyms

ALU	Aquatic Life Use
USACE	US Army Corps of Engineers
ATtiLA	Analytical Tools Interface for Landscape Assessments
BCG	Biological Condition Gradient
BMP	Best Management Practice
BOD	Biological Oxygen Demand
BPJ	Best Professional Judgment
CAFO	Confined Animal Feeding Operation
CDG	Catchment Disturbance Gradient
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CSO	Combined Sewer Overflows
CWA	Clean Water Act
DEQ	Department of Environmental Quality
DQO	Data Quality Objectives
EMAP	Environmental Monitoring and Assessment Program
EPT	Ephemeroptera, Plecoptera, Trichoptera
FWPCA	Federal Water Pollution Control Act
GIS	Geographic Information Systems
GLEI	Great Lakes Environmental Indicators
HDG	Human Disturbance Gradient
HUC	Hydrologic Unit Code
IBI	Index of Biological/Biotic Integrity
ICI	Invertebrate Community Index
IDEQ	Idaho Department of Environmental Quality
ITFM	Intergovernmental Task Force on Monitoring Water Quality
LDM	Linear Discriminant Model
LR-BP	Large River Bioassessment Protocol
LWD	Large Woody Debris
MDCB	Methods and Data Comparability Board
MQO	Measurement Quality Objectives
NAWQA	National Water-Quality Assessment Program

NLCD	National Land Cover Database (NLCD)
NPDES	National Pollutant Discharge Elimination System
NWHI	Non-Wadeable Habitat Index
NWQMC	National Water Quality Monitoring Council
OHEPA	Ohio Environmental Protection Agency
ORSANCO	Ohio River Valley Water Sanitation Commission
ONRW	Outstanding Natural Resource Waters
PBMS	Performance Based Methods Systems
POTW	Publicly-Owned Treatment Works
QHEI	Qualitative Habitat Evaluation Index
RBP	Rapid Bioassessment Protocols
RIVPACS	River Invertebrate Prediction and Classification System
TALU	Tiered Aquatic Life Use
TMDL	Total Maximum Daily Load
SCI	Stream Condition Index
STP	Sewage Treatment Plants
UAA	Use Attainability Analyses
USEPA	United States Environmental Protection Agency
WLAs	Waste Load Allocations
WQS	Water Quality Standards
WWTP	Wastewater Treatment Plant

Chapter 1.0 Introduction

1.1 Purpose of the Document

The target readership of this document is primarily program managers and technical staff involved in the development and implementation of biological monitoring programs for non-wadeable streams and rivers. The document is intended to assist users in establishing or refining protocols, including the specific methods related to field sampling; laboratory sample processing; taxonomy; data entry, management, and analysis; and final assessment and reporting. It also reviews and provides information on development of monitoring designs to address certain types of environmental questions, and approaches for documenting and reporting data quality and performance characteristics for large river biological monitoring. The approaches presented are not intended to replace existing program components but may in some cases be useful for refining them. Throughout the document, "large rivers" is used as short-form for "non-wadeable streams and rivers," which are defined as lotic systems more effectively and safely sampled with boat-based field methods than with wading techniques.

The principal purposes of this document are to:

- Serve as a framework for the development of bioassessment programs and biocriteria for large rivers, as needed by water quality management agencies for establishing Water Quality Standards (WQS), determining attainment or nonattainment of designated uses, evaluating effectiveness of mitigation or restoration activities, and to contribute to the Total Maximum Daily Load (TMDL) process;
- Provide information that can be used to enhance existing river assessment programs, including elevation of the scientific/technical foundation;
- Provide the essential technical elements for routine biological monitoring and assessment programs;
- Foster clear communication among agencies and other entities for mainstem rivers crossing jurisdictional boundaries; and
- Describe how assessment procedures and strategies can be tailored for different types of rivers.

1.2 Transitioning from Streams to Rivers

Streams and smaller rivers that are considered "wadeable" (Section 1.2.3) are abundant in terms of number and total length, and relatively easy to sample compared to large rivers. As a result, efforts to develop appropriate sampling protocols for the bioassessment of lotic ecosystems have been focused primarily on smaller systems (e.g., Barbour et al. 1999). As these methods become increasingly refined and accepted, a growing number of government agencies are developing sampling protocols for large rivers (Humphries et al. 1998). Realizing that this may be a relatively new area of responsibility for many, a brief overview of key ecological concepts relating to the topic is warranted.

1.2.1 Key Ecological Concepts About Large Rivers

Scientific knowledge of river ecosystems has expanded greatly over the last three decades (e.g., Johnson et al. 1995, Lorenz et al. 1997, Ward 1998, Tockner and Stanford 2002). However, there remains a need to test current assumptions with data. The following concepts of river ecosystem structures, functions, and controlling factors are generally well-accepted today by river ecologists. Future monitoring of our Nation's large rivers will probably support many of these assumptions, while some may prove incomplete. In any event, ongoing and upcoming work will provide an opportunity to develop a better understanding of this class of ecosystems.

The ecological condition of large rivers is affected by drivers (e.g., climate, geology) and stressors that exist at multiple spatial scales (Frissell et al. 1986, Lubinski 1993, Naiman 1998, Ward et al. 2001, Wiens 2002). Drivers that operate at larger spatial scales tend to exert their control over longer temporal scales and cycles (Naiman 1998, Poff and Ward 1990). Within a basin, as rivers increase in size in the downstream direction, predictable gradients occur in the forces that shape the river, control the substrate, and provide organic material (Huet 1959, Vannote et al. 1980). In response to these natural forces, rivers are ever changing as they advance downstream (Ward 1998, Fausch et al. 2002).

Rivers tend to be located at lower elevations than smaller streams within the same basin. They also often have shallower elevation gradients, trap more sediment, and have longer retention times than their upstream tributaries. These conditions, with the exception of localized areas where the channel is constricted, generally result in substrates dominated by finer particles.

Under natural conditions, river discharge increases with downstream distance. The predictability of the flow regime of a large river is typically greater than that of its smaller, flashier tributaries (Johnson et al. 1995). Under natural conditions, the primary sources of energy in a large river (i.e., detritus, fine particulate organic material, and attached bacteria) are usually allochthonous (i.e., carried downstream by tributaries), except where water clarity allows development of substantial plant biomass. The River Continuum Concept (Vannote et al. 1980) holds that local photosynthesis in large rivers is limited by turbidity. However, the presence of dams, floodplains with large backwaters, or large amounts of woody debris in a large river reach can reset energy processes to conditions more like those that occur in moderate-sized streams (Ward and Stanford 1983, Junk et al. 1989, Thorp and DeLong 1994, Bayley 1995). Under these conditions, autochthonous (instream) energy production through photosynthesis and invertebrate production each increase.

Large rivers frequently exhibit distinctive reach or microhabitat characteristics that are attractive to individual or groups of species (Stalnaker et al. 1989, Montgomery and Buffington 1998, Ward 1998). Reach distinctions frequently are reflected in different riparian vegetative patterns, community types, and habitat (Lubinski 1993). Microhabitat associations are often observed during specific life history stages, seasons, or discharge ranges. An especially important characteristic of large rivers is that conditions in their microhabitats change widely with river discharge (Reash 1999). Population changes in response to year-to-year variations in discharge are considered to be an important contributor to riverine biodiversity (Galat et al. 1998, Knutson and Klass 1998).

The flora and fauna of large rivers are adapted to and controlled in large part by these physical, chemical, and hydrologic conditions. It is important to note, however, that large-scale distribution patterns of many species, both terrestrial and aquatic, still reflect zoogeographic patterns established by land-forming processes (e.g., glaciation) that occurred many thousands of years ago. Large rivers, in the context of either their tributary networks or even broader spatial scales, function as landscape corridors (Lubinski and Theiling 1999). The landscape corridor function of large rivers is of special value to migratory birds and fishes, especially for birds with ranges extending beyond the basin itself.

In large rivers with substantial floodplains, annual flood pulses of allochthonous material from the floodplain have been identified as perhaps the most important hydrologic feature governing year-to-year changes in ecosystem productivity, and possibly biological diversity (Junk et al. 1989, Ward 1989, Welcomme 1985). Over-bank flooding onto floodplains facilitates the lateral exchange of nutrients, organic matter, and organisms between the main channel and associated floodplains (Benke and Meyer 1988, Meyer 1990, Sparks et al. 1990). This in turn increases the biological activity of the river ecosystem (Bayley 1989, Junk et al. 1989, Meyer 1990) and expands the physical habitat available for fishes and aquatic invertebrates (Welcomme 1989). During periods of floodplain inundation, fish forage mainly on terrestrial organisms (Reimer 1991). Some organisms (e.g., burrowing crayfish, [Crustacea:Decapoda]) considered aquatic actually live in seasonally dry floodplains and actively enter the aquatic environment during flood conditions, comprising a significant portion of the diet of some riverine fish species (Flotemersch and Jackson 2003). Floodplain interactions contribute to increased food intake and growth rates in most river fishes (Lowe-McConnell 1975, Welcomme 1985), and may account for up to 75% of annual growth (Welcomme 1985).

Today, most large rivers have been altered by a variety of human activities (Welcomme 1985, Dynesius and Nilsson 1994, Galat and Frazier 1996). Humans have altered the physical templates of rivers, the hydraulic dynamics of their channels and tributary networks, and the land-use characteristics of their basins to an extent that has had a large, but complex, impact on the biota (Bayley 1995). Even so, efforts have been made to predict how riverine assemblages might respond to imposed changes (Ward and Stanford 1983, 1995). In such disturbed systems, management requires restoration of altered system features to desired levels of quality (i.e., to support designated uses) and the conservation of river features that still exhibit desirable conditions (National Research Council 1992).

1.2.2 Bioassessment and Rivers

The aquatic life of streams and rivers (fish, insects, plants, shellfish, amphibians, etc.) integrates the cumulative effects of multiple stressors generated by both point source and non-point source (NPS) pollution. Bioassessments, consisting of surveys and other direct measures of aquatic life, are the most effective way to measure the aggregate impact of these stressors on waterbodies. Bioassessments allow evaluation of the biological integrity of a waterbody, where biological integrity is:

> *The ability to support and maintain a balanced, integrated, and adaptive community with a biological diversity, composition, and functional organization*

comparable to those of natural aquatic ecosystems in the region (Frey 1977, Karr and Dudley 1981, Karr et al. 1986).

In recent years, this subject has been comprehensively addressed and interested readers should consult the large amount of existing literature (e.g., Plafkin et al. 1989, USEPA 1990, USEPA 1992, Davis and Simon 1995, Barbour et al. 1999, USEPA 2005).

All streams and rivers are susceptible to cumulative impacts from all upstream anthropogenic disturbances including chemical and organic pollution, dams, channelization, overharvest, invasive species, and land use. The greater the distance between a reach and its headwaters, the more these disturbances accumulate, so that large rivers are often the most ubiquitously disturbed type of lotic system. As a result, determining undisturbed conditions for large rivers is difficult. The fact that the natural structure and function of larger order streams are fundamentally different from those of smaller, wadeable systems (Vannote et al. 1980, Minshall et al. 1983, Junk et al. 1989, Sedell et al. 1989) highlights the need for bioassessment methods tailored to the special circumstances that large rivers present. For example, there are physical habitat conditions that are unique or of increased significance to large rivers, such as backwater habitat (Sheaffer and Nickum 1986, Scott and Nielsen 1989), islands (Thorp 1992), woody snags (Lehtinen et al. 1997), and floodplains (Petts 1996, Benke 2001). Because these areas serve as additional physical habitat in large rivers, they influence the dynamics of the biological community. However, although science recognizes the importance of these unique habitats to the overall condition of a river system, relatively few attempts have been made to incorporate habitat condition into an overall assessment of river condition (Poulton et al. 2003, Gutreuter et al. 1995).

The size of large rivers makes expense, logistics, and safety important issues that need to be incorporated into ecologically-sound sampling methods. For an adequate assessment of large rivers, the length of channel that must be sampled to capture the diversity of organisms and habitats is greater than that for smaller, wadeable streams. Many wadeable stream techniques are also not feasible or relevant to large river systems. These complications have led many river assessment programs to: 1) omit biological assessment of large rivers, 2) simply apply wadeable methods to wadeable areas of larger rivers, or 3) drop certain assessment parameters that are more difficult to measure in large rivers, such as benthic macroinvertebrates, and base assessments only on fish sampling and visual habitat assessments. None of these solutions allow for a comprehensive, and scientifically defensible evaluation of the condition of our Nation's large river systems, and therefore, will not provide the information needed to determine appropriate Aquatic Life Uses (ALUs) for the Clean Water Act (CWA).

1.2.3 Resource Typology

No consensus has been reached on what criteria should be used to differentiate between wadeable and non-wadeable (i.e., large river) systems. There is no clear geographic point along rivers that consistently discriminates when they become non-wadeable. Rather, there is a zone of gradual transition between wadeable and non-wadeable conditions before a river becomes predominately non-wadeable (Figure 1-1). As a result, criteria for defining large rivers will likely vary across the country.

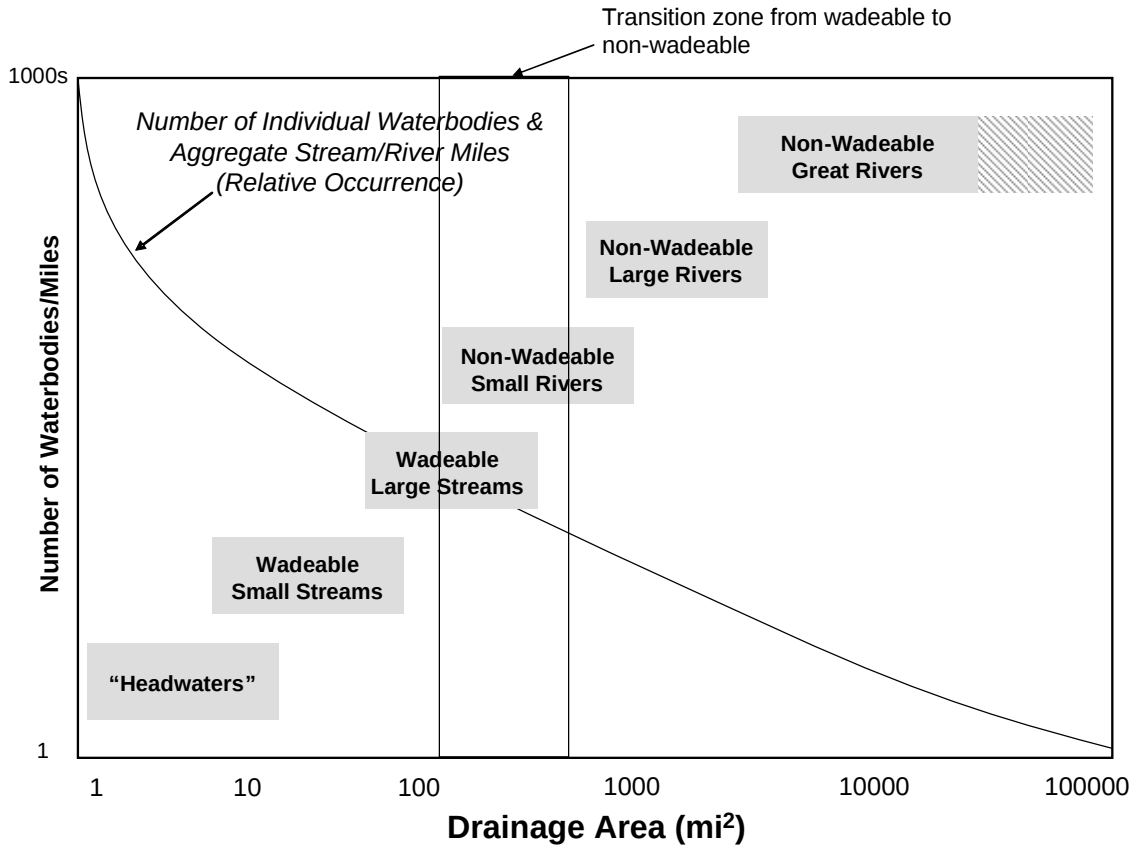


Figure 1-1. The delineation between wadeable and non-wadeable streams is not discrete, but rather a gradual transition (after C. Yoder, personal communication).

Some of the more common designations in use include a priori stream designations such as stream order (e.g., >4th order) (Strahler 1957) and drainage area (e.g., >5000 km^2). Using Strahler order, Leopold et al. (1964) estimated that there are approximately 5000 rivers of 5th through 7th order, and 50 of 8th through 10th order in North America. However, use of Strahler order alone has not proven to be a reliable stand-alone predictor of whether a river is wadeable or non-wadeable and, hence, whether a wadeable or non-wadeable sampling approach will be required for collection of a representative sample. In a recent paper by Wilhelm et al. (2005), the problem of defining the resource is discussed and quoted herein:

> A non-wadeable or large river can be defined as a reach where the investigator cannot wade along its length (Meador et al. 1993) or from bank to bank (Edsall et al. 1997). However, the progression from small to large river is continuous, and even the demarcation between wadeable and non-wadeable is an indistinct boundary, because the status of a single location can change between wet and dry months or years. It is desirable to establish guidelines that can be applied prior to visiting a site and used to define the sampling universe of large rivers for a region. Large rivers have been defined as those that exceed a drainage area of 1600 km2 (Ohio EPA 1989); an average depth of

1 m (Stalnaker et al. 1989); a width of 50 m (Simonson et al. 1994); or a river order of six or greater (Vannote et al. 1980, Sheehan and Rasmussen 1999). In contrast, Reash (1999) set a much higher threshold by defining a large river as one with a drainage area greater than 20,000 km^2.

An alternative to strict, a priori order or area designations is for field crews to make the designation after arriving on site. Others have adopted a multi-criteria approach. For example, Wilhelm et al. (2005) defined non-wadeable rivers of Michigan as those that equaled or exceeded a river order of five, drainage area of 1600 km^2, mainstem length of 100 km, and mean annual discharge of 15 m^3/s. Another example is provided by the Idaho Department of Environmental Quality (Grafe 2002) where several criteria are used for designating a system as wadeable or non-wadeable. Criteria considered include average width at baseflow, average depth at baseflow, average greatest depth, site discharge, mean annual site discharge, and site drainage area. After a review of the strengths and weaknesses of each parameter, three were selected, each of which is scored and then averaged. They are: 1) stream order on a 1:100,000-scale map, equal to or greater than 5 = 1 point, 2) average wetted width at base flow greater than or equal to 15 meters = 1 point, and 3) average depth at base flow greater than or equal to 0.4 meters = 1 point. If the average of the scores in the three categories is greater than or equal to 1.7, it is classified as non-wadeable. If the average is less than 1.7, it is considered a wadeable stream. Additional criteria are used to delineate between medium and large rivers. However, the same protocol is used in each.

Two other characteristics that can be used to classify sites as non-wadeable are whether they are boatable or raftable and whether riverine species predominate. This would include sites that have lowhead dams, small hydroelectric facilities, or navigational dams, yet retain the generalized form and function of a flowing river ecosystem. It would exclude sites that function as reservoirs (e.g., publicly owned reservoirs, reservoirs managed for flood control or water supply), which are better assessed with protocols designed specifically for lentic systems (e.g., USEPA 1998). An example of an application following the non-wadeable logic is Lyons et al. (2001) which defined rivers in Wisconsin as lotic systems having at least 3 km of contiguous river channel too deep to be sampled using wadeable techniques.

As an alternative, a conceptual classification that combines features of the above approaches with physical and biological attributes of the system can be used. As summarized by the River Continuum Concept (RCC) (Vannote et al. 1980), lotic systems present a longitudinal gradient of physical conditions including width, depth, velocity, flow volume, and temperature. For example, proceeding downstream, river systems become broader, canopy cover decreases, and water temperatures increase. In response to these changes, stream segments are progressively influenced less by adjacent and more by upstream contributions of materials. This conceptualization, however, must be tempered with the realization that many rivers receive significant inputs from the seasonal coupling with their adjacent floodplains as well as connections with adjoining off-channel waterbodies (e.g., wetlands, oxbows) (Junk et al. 1989, Meyer 1990). This gradient of conditions is likewise reflected in the aquatic communities (e.g., algae, benthic macroinvertebrates, and fish) that have adapted to the physical conditions of a given reach along the system (Flotemersch and Jackson 2003, 2005).

In response to this gradient of changes, the methods used to sample the biotic communities must likewise change. For fish sampling, this means a progression in gear from backpack electrofishers in fully wadeable stream reaches, to the use of tote barges in deeper wadeable waters, to boat- or raft-based electrofishers. For macroinvertebrate sampling, the downstream progression from wadeable to non-wadeable reaches generally entails a shift from sampling the available habitat of the full channel to sampling in shoreline areas with dip-nets or artificial substrates. For sampling of algae, there may be a need to switch from an assessment based completely on periphytic diatoms to one including phytoplankton and soft algae. It should be noted that at some sites, transitional zones may be encountered that are composed of both wadeable and non-wadeable sections, and thus may require a hybrid approach to meet specific study objectives.

Integrating these additional attributes of systems, a conceptual classification can be constructed to serve as a guide for site classification and assessment approaches. Descriptive characteristics could include drainage area, Strahler order, functional features, narrative definitions (i.e., ability to sample), or other discriminatory characteristics useful for a particular region. An advantage of this approach is that by paralleling the conceptual framework of the RCC, conceptual classification can be used to place a site or reach in context within a larger watershed or landscape and thus help define and focus bioassessment and monitoring activities and restoration goals. One drawback of this approach is that a categorical framework is being applied to a system that exists along a continuum. Consequently, some sites may not fit neatly into a single category. In such cases, additional information may be required or a weight-of-evidence approach employed.

Here we present a prototype classification that includes classes for large and great rivers using some of these characteristics (Table 1-1). Because of geographic differences that exist among river systems, modifications to the table will certainly be required to ensure broad applicability. As is evident in the provided example, the chart may be very general, exhibit much overlap, and vary greatly by region. The “Functional features” presented assume the systems being discussed are in undisturbed condition.

1.3 Overview of the Large River Bioassessment Protocols

As stated in the Preface, the impetus for the development of the Large River Bioassessment Protocols (LR-BP) was a need expressed by Regional scientists of the USEPA to develop standardized protocols specifically designed for the bioassessment of large rivers by States and Tribes. Criteria established for the final protocols were that they:

- permit the sampling of one or more sites per day;
- be scientifically defensible and statistically robust;
- be suitable for incorporation into routine monitoring programs;
- have the capacity of addressing often multiple and simultaneous objectives of agencies; and
- produce assessments acceptable to State, Tribal, and National programs with a reasonable level of effort.

TABLE 1-1. Prototype site classification approach for streams and rivers.

Bioassessment protocol class	Drainage area (range, km^2)[a]	Strahler order	Functional features (ecological)	Narrative definition (sampleability)
A. Headwater streams (Intermittent, Ephemeral)	< 3 km	0 – 3rd	**Habitat**: Riparian shading/canopy is heavy in forested streams but may be light along those draining desert, grassland, and agricultural fields. Debris dams common in forested headwater streams. Substrate type will vary depending upon geology and gradient. Bed material of high gradient streams will be dominated by cobble, boulder and bedrock, whereas finer substrates commonly dominate low gradient channels. The length of habitat units tend to be small relative to channel width; therefore, the distances between alternating units is short, particularly for high-gradient (>20%) channels that have step-pool formations. Headwater channels also have high length:width and width:depth ratios, such that a high proportion of water flowing through these streams is in direct contact with the stream bed and banks.[b] **Typical biotic assemblages**: *Algae*: Primarily benthic diatoms; some blue-green and green algae; mosses and liverworts common. *Benthic macroinvertebrates*: Shredders and predators (forested), collectors, scrapers, and predators (grassland and desert), endemic species commonly associated with spring-fed streams. *Fish*: Few (e.g., *Semotilus atromaculatus*, *Salvelinus fontinalis*) to none. *Amphibians*: Salamanders (e.g., Plethodontidae) and frogs (e.g., *Ascaphus* spp., *Rana clamitans*); salamanders are frequently the top stream predators.	All habitats are accessible for sampling; however modified methods may be required for the shallow and low-flow conditions. Summer sampling may be limited due to naturally intermittent streams, where channels may be completely dry or surface water is limited to isolated pools.

[a] There is overlap between estimated ranges of drainage areas.

[b] This generalized description pertains to systems with relatively undisturbed riparian vegetation. Disturbed reaches may have characteristics more typical of larger systems.

TABLE 1-1. Continued.

Bioassessment protocol class	Drainage area (range, km^2)[a]	Strahler order	Functional features (Ecological)	Narrative definition (Sampleability)
B. Wadeable streams and rivers	<1 – 700	1st – 3rd, or 4th	**Habitat**: Riparian shading/canopy cover may be heavy in forested streams. Channel dominated by stable substrates. Energy sources mainly from outside of stream (allochthonous); thus coarse particulate organic material (CPOM) contributions are significant. For desert streams, perennial water will persist in most seasons, but the water may disappear underground into the porous, sandy stream bottom. Pools may persist. The stream is open to direct sunlight. **Typical biotic assemblages:** *Algae*: Periphyton in desert streams: filamentous green algae. *Benthic macroinvertebrates*: Shredders/ collectors, in desert stream mainly fine particulate organic material (FPOM) - feeding gathering-collectors. *Fish*: Surface and water column feeders, generalized invertebrate feeders, and benthic invertebrate feeders.	River reaches where sampling of multiple habitats can be accomplished using simple wadeable techniques.
C. Transitional streams and rivers	500 – 1000	3rd – 5th	**Habitat:** Riparian shading significant in forested streams, but openings in canopy cover increasing. Channel dominated by stable substrates with increasing occurrence of unstable substrates. Unique habitats exist that host fauna from adjoining upstream and downstream segments. Transition in importance of energy sources from CPOM to FPOM. **Typical biotic assemblages:** *Algae*: Periphyton. *Benthic macroinvertebrates*: Collectors/ grazers. *Fish*: Benthic invertebrate and generalized invertebrate feeders common.	Contains both wadeable and non-wadeable segments with a mosaic of habitat types that shift in quantity and quality in response to prevailing flow conditions. Sampling often requires a combination of methods developed for wadeable streams and large rivers.

[a] There is overlap between estimated ranges of drainage areas as well as among orders.

TABLE 1-1. Continued.

Bioassessment protocol class	Drainage area (range, km^2)[a]	Strahler order	Functional features (Ecological)	Narrative definition (Sampleability)
D. Non-wadeable streams and rivers	800 – 40,000	4^{th} – 8^{th}	**Habitat**: Importance of riparian shading is minimal, even in forested streams, and stream surface area mostly unshaded. Left and right banks increasingly divergent in character but not functionally independent. Influences on stream reaches affect both banks but maybe to differing degrees. Occurrence of unstable substrates artificially high in impounded reaches. Importance of FPOM > CPOM. Most desert streams are heavily diverted in lower reaches and therefore may only have intermittent flow and no non-wadeable reaches, or are only non-wadeable during certain times of the year. **Typical biotic assemblages:** *Algae*: Periphyton more prevalent in free-flowing reaches. Increasing importance of phytoplankton where water retention time is sufficient for development. Especially true immediately upstream of dams and other heavily impounded sections. Macrophytes infrequent but increasing in incidence. *Benthic macroinvertebrates*: Collectors with appearance of great river species, more so with increasing impoundment levels. Mussels infrequent to frequent. *Fish*: Herbivore-detritivores increasingly dominant. Occurrence of great river species common in impounded reaches.	River reaches where boats are always necessary to access sample points; occasionally necessary to pull boats through shallow areas.

[a] There is overlap between estimated ranges of drainage areas.

TABLE 1-1. Continued.

Bioassessment protocol class	Drainage area (range, km^2)[a]	Strahler order	Functional features (Ecological)	Narrative definition (Sampleability)
E. Great rivers	>25,000	> 8th	**Habitat:** Canopy opening extensive, even in forested streams, with stream surface largely unshaded. Channel dominated by unstable substrates. Left and right banks often independently affected by physical, hydrologic, and stressor conditions as a result of laminar flow along banks. A single habitat type may prevail for kilometers along a bank. Reaches frequently defined by large dams, which can limit the habitat heterogeneity and biotic diversity of a reach, especially true upstream of dams. System largely defined by FPOM. Allochthonous inputs of organic matter from upstream and lateral inputs are significant. Only autochthonous production is by phytoplankton. **Typical biotic assemblages:** *Algae*: Phytoplankton. Water retention time sufficient for assemblages to establish. Main channel unsuitable for macrophytes or periphyton due to turbidity, swiftness of current, and scarcity of stable substrates. Macrophytes potentially abundant particularly on river margins and in backwaters. *Benthic macroinvertebrates*: Dominated by collectors. Mussels potentially locally abundant but not ubiquitous. *Fish*: Regular occurrence of great river species. Planktivores, herbivore-detritivores common.	River reaches where boats are always necessary to access sample points. Habitat types are frequently large and thus may require the development of habitat-specific expectations for biotic assemblages. Consequently, complete assessment may require sampling and assessment of different habitats.

[a] There is overlap between estimated ranges of drainage areas.

Among the protocols discussed in this document, several were reviewed in detail before, during, and after the LR-BP research, and thus contributed directly to the development of the LR-BP (i.e., Ohio Environmental Protection Agency [Ohio EPA], US Geological Survey-National Water Quality Assessment [USGS-NAWQA] and the USEPA-Environmental Monitoring and Assessment Program [USEPA-EMAP]). Other programs and protocols discussed represent current research USEPA is conducting (i.e., USEPA-EMAP-Great Rivers Ecosystems [GRE]), and programs USEPA is currently collaborating with (Ohio River Valley Water Sanitation Commission [ORSANCO]).

The LR-BP represents an integrated approach to sampling in that the protocols for algae, benthic macroinvertebrates, and fish can be applied using the same sampling design. They are designed for rivers and, depending on the scale and scope of programmatic data needs, can be used for regional and site-specific studies. Also, while the protocols are not intended for application to great rivers (Table 1-1), adjustment of one or more components of the protocols will make them better suited for those kinds of systems. The LR-BP for physical habitat is not presented herein because refinements to the protocol are being field tested.

Much like the USEPA-EMAP and USGS-NAWQA protocols, the LR-BP for algae, benthic macroinvertebrates, and physical habitat are transect-based. This design has many desirable features for field studies; and as long as the first point is selected at random, remaining points based on that point can be considered random as well (Cochran 1977). The simplicity of this type of design makes it easy to execute without mistakes and results in significant time saving in the field. It also results in the drawn sample being spread more evenly over the population (Cochran 1977, Manly 2001). A common concern expressed about the transect approach to sampling is that the most productive habitat of a study reach may fall between transects and thus go unsampled. This will occur, but sampling what is perceived to be the most productive habitat is equivalent to selectively visiting the nicest house in a neighborhood and using it as a measure of the mean living conditions in that community. Another concern expressed is that at some sites, a standardized protocol may sample greater distances than required to achieve the data quality requirements (e.g., % of total species) set by the study. This is in all probability true, but if a standardized protocol is to be applied at all sites, it must adequately sample all, or a predetermined percentage, of the sites.

The combination of field-based comparative studies and collaborative field tests involving State agency biologists, Tribal members, and academic researchers was critical in ensuring the resulting protocols were consistent with the criteria established for the products. Findings from these studies, justification for follow-up research, and the performance of developed methods are discussed in this document where applicable.

Chapter 2.0 Elements of Biomonitoring

This chapter…

- reviews critical elements of a large river bioassessment program – split into design, methods, and interpretation elements
- introduces the major assessment elements: habitat, chemistry, and biology

2.1 Bioassessment Elements

Biological monitoring and assessment consists of evaluating sites using specified biological indicators, then repeating that evaluation consistently over time. In designing and implementing a large river biological monitoring program the goal is to develop a program that is comprehensive, accurate, cost-effective and meets stated objectives. Without clear strategies for data use, even good assessment techniques can be inefficient and unproductive. Similarly, making program decisions and setting objectives without understanding the quality of the underlying data is problematic. While bioassessment programs exist in all 50 States and several Tribes (USEPA 2002), the rigor and quality of these programs, and how they are used varies (Carter and Resh 2001, USEPA 2002). To facilitate a consistent understanding of the components that make a successful program, USEPA is developing a list of critical elements for successful bioassessment programs (Barbour and Yoder 2004). This chapter builds off that list and discusses those critical elements that must be applied to a successful large river assessment program.

The Methods and Data Comparability Board (MDCB) of the National Water Quality Monitoring Council (NWQMC) recommends that agencies ask basic *why-who-where-when-what-how* questions (Figure 2-1) when designing an effective program. Developing and documenting the basic bioassessment elements are critical for establishing and maintaining a high-quality, flexible program. This process also identifies the program constraints critical for understanding the limitations of data interpretation and management actions.

The design elements important for assessment can be broken into a few components that define a flexible and productive program: design, methods, and interpretation (Table 2-1). These elements are as important for large rivers as they are for any ecosystem assessment program.

2.1.1 Design Elements

Study design is the foundation of any monitoring program. Design elements include study design objectives, temporal and spatial coverage, classification, reference conditions, and criteria (Table 2-1). Design questions may be driven by regulatory requirements, program goals, and research questions. In any case, it is strongly recommended that analytical or statistical specialists cooperate with field crews and program managers during this phase. Too often, assessment programs are created without a clear sense of how the data are going to be used, only to find out that the design chosen was inappropriate to provide the answers or data quality needed to meet assessment program objectives. However, it is also possible to create data quality objectives (DQOs) that are unattainable or technically too difficult to implement.

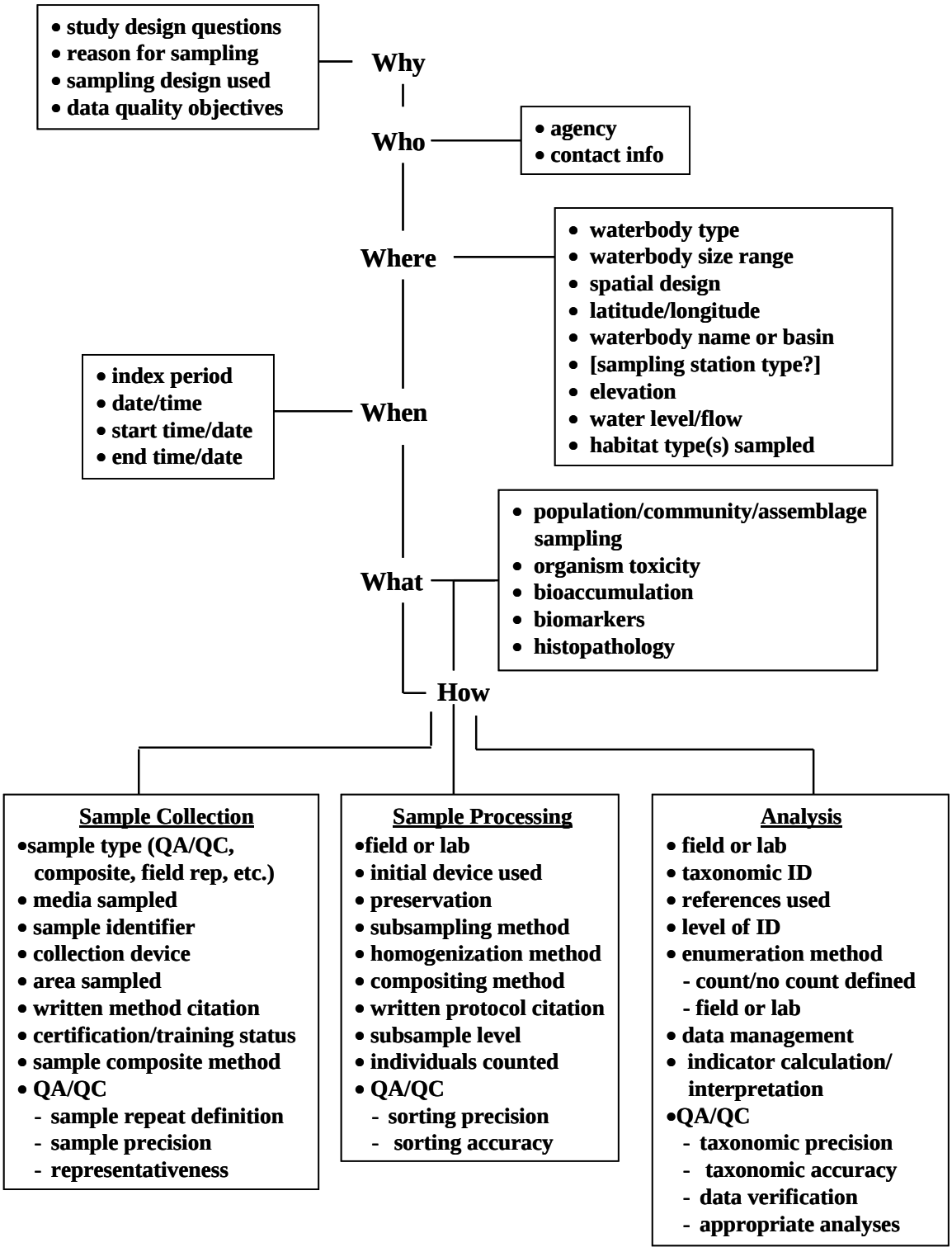


FIGURE 2-1. Data elements for biological assessment programs (modified from NWQMC 2006).

Although flexibility should always be incorporated, a good design will include documentation and presentation of programmatic limitations, as well as quantification of the level of uncertainty associated with any conclusion. Clearly defined questions are the first component of a good design.

TABLE 2-1. Important elements for a large river biological assessment program.

Design
• design objectives
• temporal coverage
• spatial coverage
• classification
• reference conditions
• reference criteria
Methods
• indicators
• sample collection
• sample processing
• data entry and storage
• QA/QC
Interpretation
• ecological attributes
• biological indices
• diagnostic capability
• performance evaluation

Design questions are usually derived from programmatic needs. Biological assessment data are used in a variety of programs (e.g., 305[b] and 303[d] reporting, source water assessments, NPDES permitting). Developing a monitoring program that meets multiple needs requires an understanding of the information required by each program, and thus, cooperation among program personnel. Although this involves effort, cooperation at this point can help avoid the inefficiency of having 2 or 3 sampling crews collecting the same or similar data on the same river for multiple programs. Including program requirements in question development is, therefore, also essential for a good design.

While developing the design questions it is also critical to develop DQOs to determine the quantity and quality of data needed (USEPA 2000b). DQOs are quantitative and qualitative statements that clarify objectives, define appropriate data, and specify tolerable levels of decision error. Each program will likely be able to define how the data will be used to answer their questions or meet their needs. Often, program requirements can be described in data quality terms (e.g., determine with 90% confidence whether a site is impaired, whether there is more than a 20% change in condition over time for a 2nd determination with 90% confidence, whether an outfall causes a decrease in biological condition). Each DQO has to be described to define the design elements needed to meet the assessment precision and accuracy required by the program.

For example, programs needing only to separate extremely disturbed from minimally disturbed sites will require less precision than programs designed to detect small departures in ecosystem condition (Figure 2-2). Greater precision may be required in large river work since the condition gradient, in many cases, is already restricted.

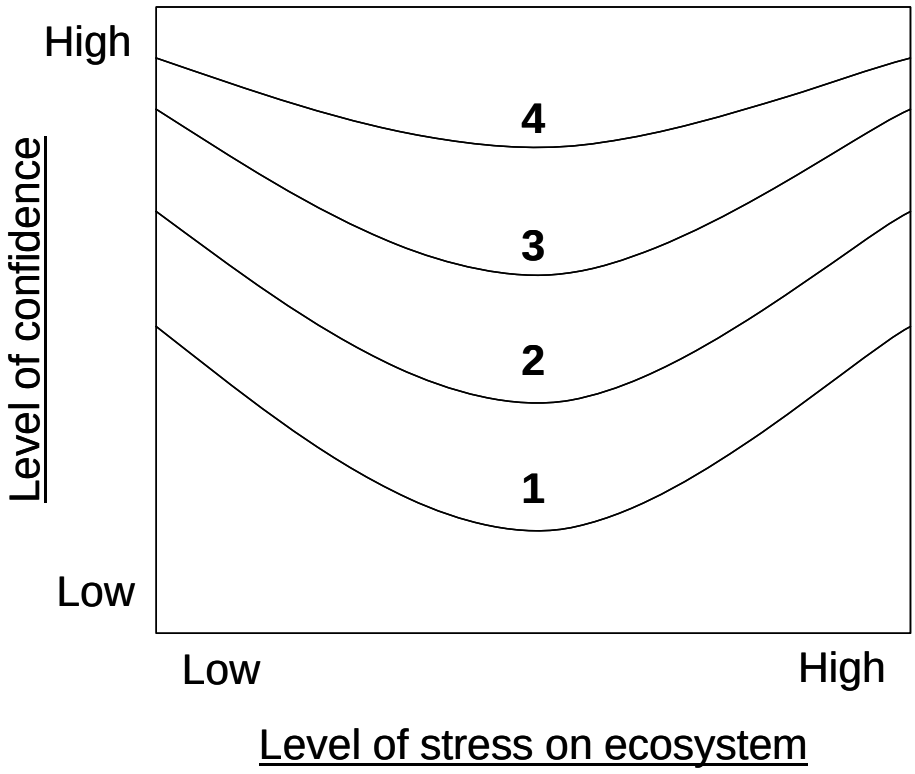


FIGURE 2-2. Conceptual illustration of confidence in detecting different level of stress on an ecosystem as a function of assessment precision (with 4 being most precise) (modified from Barbour and Yoder 2004). Intermediate levels of disturbance (stress) are typically more difficult to evaluate than are the high and low extremes.

Describing the quality of data necessary to meet project DQOs for the different assessment methods is critical (see Chapter 3). While this may seem more like a methodological element of assessment, it is important to include performance evaluation as a part of programs from the beginning. This includes being able to document and report the quality of each step, from data collection to site assessment. Performance elements (precision, accuracy/bias, representativeness, completeness, and sensitivity) must be included in the study design and incorporated into assessment program costs. Measurement quality objectives (MQOs) document method performance, as well as program technical staff, and are measurement goals needed to meet the programmatic DQOs. In general, MQOs do not specify the methods, but provide criteria for evaluating acceptability of data produced by a protocol or a program.

Precision, calculated on final assessments, can be used to identify errors and to determine the repeatability of site assessments. For example, assessment precision is generally evaluated using repeat sampling for some sites by the same team (to evaluate intra-team precision) or by different crews (to evaluate inter-team precision) (Barbour et al. 2006 [in press]). Precision also affects the ability of a method to detect an effect.

A biological assessment protocol is a series of methods, each of which produces information contributing to final site assessment and evaluation (Diamond et al. 1996, Barbour et al. 1999, Stribling et al. 2003). As such, each method has the potential of introducing error into final assessments (Figure 2-3). The relative importance and acceptability of different error sources and magnitudes are defined through use of data quality and measurement quality objectives (Taylor 1988, USEPA 2000b, Stribling et al. 2003). Through the use of MQOs, noise in a dataset can usually be distinguished from signal. MQOs can also help directly identify which specific components of the protocol are contributing to noise.

To ensure that DQOs and MQOs are met, it is necessary to develop a quality assurance project plan (QAPP) detailing the quality assurance and quality control (QA/QC) steps. Elements of QA/QC and suggestions for maintaining MQOs are described in Chapter 3 as well, but should include procedures for documenting the error associated with each components of the assessment process (Figure 2-3).

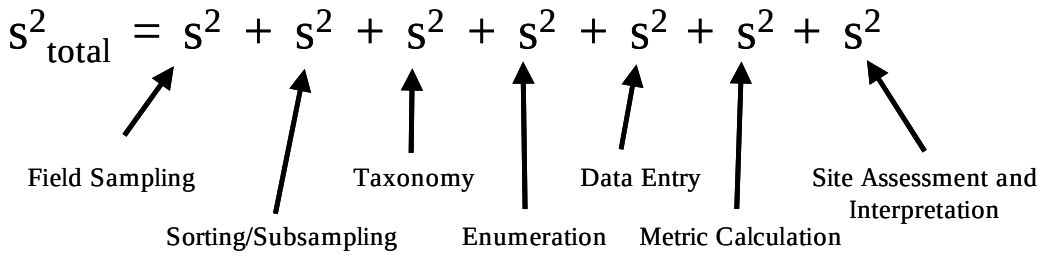


FIGURE 2-3. Total error or variability (S^2) associated with a biological assessment is a combined result of each component of the process (Barbour and Yoder 2004, modified from Taylor 1988). Example shown is from a benthic macroinvertebrate protocol.

Once design questions, DQOs, and MQOs, are defined, the remaining design elements (temporal and spatial coverage, classification, reference conditions, and reference criteria) can be addressed. Because taxa differ in the timing of their life cycles, the biological communities integrate environmental effects over time to varying degrees depending on specific environmental requirements or the natural history of the assemblage. A season-specific index period is used as a cost-effective way to decrease the natural variability in data on biological assemblages associated with seasonality by decreasing between-year variability and increasing sampling crew efficiency. Selection of an optimal sampling index period should take into consideration recruitment cycles (e.g., reproduction, emergence, growth, and migration for macroinvertebrates and fish or growing season for periphyton). Index periods ought to consider not only a particular season, but for some taxa also the time of day. If event-based sampling

outside the index period is an important design objective, then multiple index periods or temporally inclusive indices can be used (e.g., predictive models using Julian sampling day as a predictor).

River systems also vary spatially and reducing this variability through some kind of stratified sampling approach will improve the assessment precision. Biological assemblages vary with watershed size, so biological samples for any site can only be used to represent an area of similar physical dimensions and flow. Within rivers there is a great variety in size. Similarly, rivers will change if land-use alters its course. If specific land-use impacts are of design concern, then additional sites may be required to characterize this effect.

Of critical importance is determining what a representative and appropriate large river sampling site is. River systems are hierarchically arranged (Figure 2-4), with the reach often being the common sampling scale for biomonitoring programs (Frissell et al. 1986). Further discussion on sampling reach length is provided in Chapter 3.

The spatial density and placement of sites are also important for characterizing large rivers. One site at the bottom of a watershed provides limited information about average condition of the whole river system, although it is a reflection of the impact of that watershed on the receiving body of water. Unbiased estimates of the status and trends in watershed condition are most efficiently achieved using probability-based sampling designs, where the number of sites increases the confidence in the estimate of average condition. Rotating basin designs that cover an entire region over a set period of time (usually years) are an efficient way to apportion effort and reduce costs within a probabilistic design. However, many programs also require specific targeted sampling for particular program needs (e.g., NPDES permit compliance or specific stressor studies). Thus, multiple-objective programs will likely use a mixture of probabilistic and targeted sampling.

Classification of rivers within a large region may stratify coldwater vs warmwater rivers or blackwater vs clearwater rivers. Basins may be a basis for classifying rivers for fish sampling. Ecoregions are also a commonly used classification structure. Any natural biogeographic factor constraining the ecological community is worth exploring as a possible dimension for classification, but is retained in classification only if its use reduces (i.e., partitions) natural variability.

Reference condition approaches have traditionally been used to evaluate the biological condition of streams and rivers. The term "reference" is used in a variety of contexts. In this document, "reference" is used in its general context, to signify the benchmark against which biological condition is being assessed for any given sample. There are 4 types of reference condition (Stoddard et al., 2006). *Historic* conditions represent the biological condition that existed prior to human disturbance or the Reference Condition for Biological Integrity (RCBI); *minimally disturbed* condition is that found in large rivers minimally impaired or unimpacted by human disturbance; *best attainable* conditions represent that theoretical condition attainable under the application of all available best

Four Types of Reference Conditions (Stoddard et al. 2006)

1. Historic conditions
2. Minimally disturbed conditions
3. Best attainable conditions
4. Least disturbed conditions

management strategies; and *least disturbed* refers to the best available current conditions. Because large rivers integrate impacts from large areas and have many historical as well as present impacts, there are likely few segments of large rivers in the USA that are in historic or minimally disturbed conditions and most are substantially altered. Describing a reference condition for these systems is a challenge. In most programs, data for sites in least disturbed condition will have to be used as the benchmark condition for building assessment tools and the least disturbed condition is what will be used for "reference" in this discussion. Two common approaches for assessing biological condition are the site-specific or regional based reference condition approaches.

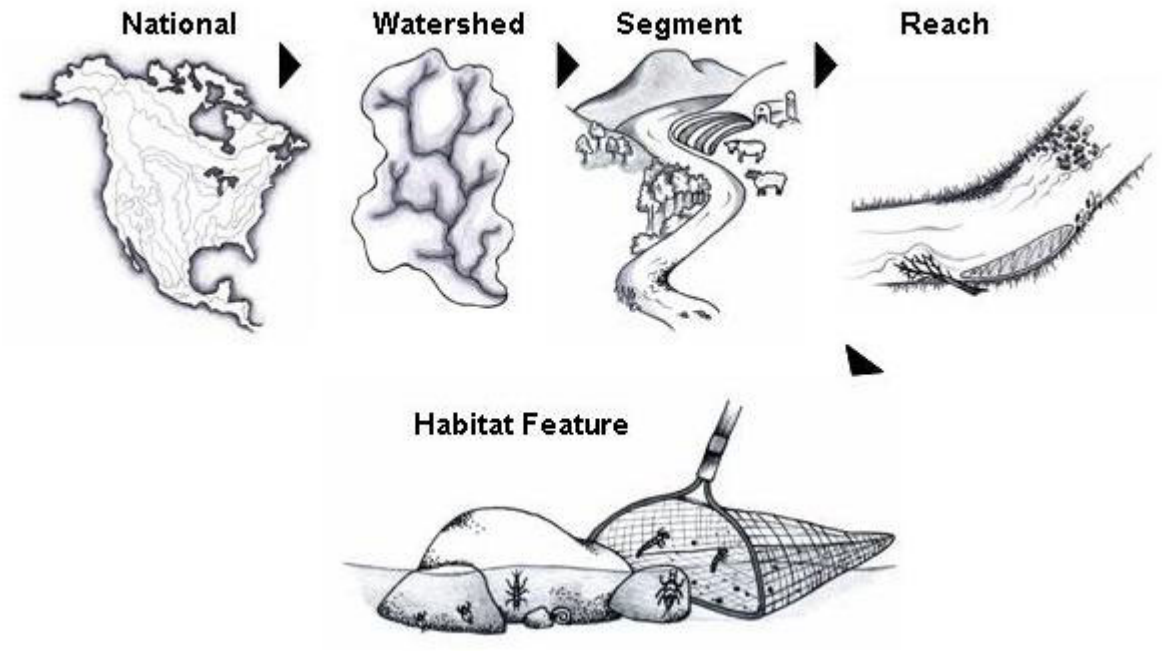


FIGURE 2-4. Environmental features sampled are nested in a spatial hierarchy.

In the case of the *site-specific reference approach,* the reference site is typically an upstream reach of comparable physical setting to the downstream site (e.g., below some impact) or a paired reach of similar physical setting to the one being assessed. The average biological condition of the downstream test reach is then compared to the site-specific reference. The advantages of upstream reference sites, if carefully selected, are that they are often of similar habitat condition, reducing the variability associated with habitat differences and they have a similar upstream water quality context to the downstream site (i.e., they experience the same set of upstream point and non-point sources). Disadvantages of the approach, however, include a limited capacity for extrapolation, logistical issues with mobile taxa, limited statistical power, pseudoreplication issues, and the comparatively high level of effort needed for assessing a state-wide set of test sites (Hughes et al. 1986, Barbour et al. 1995, Bailey et al. 1998, Reynoldson and Wright 2000).

The *regional reference site approach* defines a population of least impaired sites within a resource class. Both of the common biological indices, the multimetric index and the river

invertebrate prediction and classification system score (RIVPACS-type) (Wright et al. 2000), are developed using a population of least-impaired sites. The multimetric approach selects metrics that best discriminate between reference and study sites for a specific region. The RIVPACS-type approach derives a list of expected taxa for each test site based on their physical similarity to reference sites and the taxonomic composition of those reference sites. The list of expected taxa is, in essence, the average taxonomic composition of reference sites weighted toward those most physically similar to the test site.

Reference sites are usually defined using a set of *reference criteria* typically derived from data other than those indicators being calibrated. This allows avoidance of explicit circularity in the development of biological indicators. Criteria often include land cover, water chemistry, and habitat characteristics; and a site must meet all the criteria to be considered an appropriate reference site. As stated above, given the history of large rivers and their landscape position, historic and minimally impaired reference sites are no longer available for the majority of continental large rivers; and least disturbed conditions offer the most likely option for reference based approaches.

Reference criteria, well developed and documented, are used to evaluate the degree of human impact and to eliminate sites that have undergone excessive disturbance. Abiotic parameters are recommended as the principal criteria in defining the characteristics that become the basis for biological reference. If the same reference sites will be used to develop specific chemical water quality criteria (e.g., nutrient criteria), then these specific parameters should also be avoided as criteria. Factors reflecting anthropogenic stress (e.g., human population density, road density, land use/land cover, permitted outfalls, channelization, riparian condition, dams, etc.) should be used. Candidate reference sites can be identified within a randomly-drawn sample or selected from the entire population based on synoptic studies or prior knowledge. Remotely sensed data are best used as a first screening tool, followed by site reconnaissance and sampling, if needed. Field crews should also verify the suitability of a site as reference from the site visit, and a site should be eliminated if circumstances indicate a non-reference condition. The reference site selection process, including the criteria used, should be carefully documented. It is also important to recognize that some stressors may not be identifiable.

In the case of wadeable streams, it is often possible to identify entire watersheds of minimally disturbed or least disturbed conditions. Again, this is not possible with large rivers, since large portions of almost every large river catchment are disturbed to some extent. Reach-based spatial organization likely offers the best approach for defining reference conditions. In this approach, large rivers are split into segments or reaches using traditional geomorphic characteristics (such as those described by Frissell et al. 1986) or simple systematic criteria (e.g., 500-m reaches) and proximate stressors and land cover characterized for each reach or segment. The definition of proximate can be defined as simple linear distance (e.g., 40 km) or flow based distances (e.g., 0.5 or 1 day residence times above the segment), and reference criteria evaluated within that distance. The EMAP Great Rivers Ecosystems Research Program is using a reach based approach for defining reference sites (T. Angradi, personal communication). In this approach, Great Rivers were split into 500-m reaches. The proximate upstream (variable distances above the site depending on physical parameter considered) physical and chemical conditions were

evaluated and scored. Sites with the least impaired scores were used as candidate reference sites. In this approach, upstream tributary effects are scored based on their catchment characteristics.

Another potential approach is to create *theoretical reference conditions*. In this approach, expected conditions for assemblages are modeled. Historical data or data associated with least disturbed adjacent wadeable/non-wadeable systems are merged to define reach-specific expectations. These expectations for an invertebrate assemblage are based on the habitat characteristics for a site. These habitat characteristics can be existing or modeled, based on historical geomorphology, for example. Once an expected condition is defined, the existing condition can be compared with what is expected. The assessment is based on the difference. This approach has been used experimentally on the Missouri River (B. Poulton, personal communication). Expected macroinvertebrate richness and relative abundances were calculated for Missouri River reaches based on the richness and relative abundance of taxa from least disturbed regional water bodies with habitat types similar to those modeled for the Missouri. Where possible, functional equivalent taxa replacements were used. By combining these approaches, biologists were able to calculate metric values for a site and compare it to the theoretical reference.

Regardless of the approach, once the reference site population is established it is often used as the basis for listing criteria under the 303(d) requirements of the CWA. Commonly, a percentile of the reference population is used. Any non-zero percentile will, by definition, list a certain percent of reference sites as impaired. The percentile selected is often chosen to balance the error associated with calling a truly unimpaired site impaired with that of calling a truly impaired site unimpaired. Therefore, selecting the appropriate percentile depends on the condition of the population of reference sites. If very relaxed criteria were used, then a higher percentile would likely be more protective. If more stringent criteria were used, then lower percentiles would be sufficient.

2.1.2 Method Elements

Programmatic methods include indicator selection, sample collection and processing, data entry and storage, and QA/QC (Table 2-1). Indicators are comprised of physical and chemical habitat attributes and the biological assemblages selected for sampling. There are a variety of chemical parameters that can be measured in situ (e.g., using multiprobes or handheld single parameter meters) or in the laboratory on water samples collected in the field (APHA 1998). Similarly, methods exist for assessing physical habitat (see Chapter 4). Principal biological indicators include macroinvertebrates, fish, and periphytic algae; although some larger river programs have included vascular plants (Yin et al. 2000) and some use phytoplankton (Moulton et al. 2002). Zooplankton also has potential in large river programs, depending on water residence times.

Biological indicators provide a measure of biological condition, thereby addressing regulatory needs within the CWA related to the "…biological integrity…" of the nation's waters. They also integrate chemical and physical stressors over much larger spatial and temporal scales than can be assessed through direct measurement. Assemblages differ in how well they integrate stressors over these scales as a result of differences in their life history characteristics. For example, fish are longer-lived than algae, but are much more mobile. So algae may better integrate stressors in

one point in space over shorter time periods, whereas fish may integrate stressors over larger spatial scales. Sampling multiple assemblages provides a more comprehensive assessment and is generally preferable to single assemblage approaches (ITFM 1995).

Standardized *sample collection and processing* methods are necessary to establish the validity and reliability of biological data (Barbour et al. 1999). Evaluating the appropriate methods for a program includes consideration of target assemblages, river type, number of samples, reach length, and field methods. This document reviews a number of sampling approaches and makes recommendations for habitat, algae, macroinvertebrates, and fish methods. Processing considerations include proper preservation, labeling, transport, sorting, subsampling and taxonomic identification and are discussed in each methods chapter.

Individual species vary in their stressor tolerance and habitat preference, so taxonomic resolution is an important issue. Individual species can be thought of as individual units of information, like pixels in a digital photograph. The more pixels, the better a picture's resolution. But there is a trade-off in cost because species-level identification can take more time for certain taxa, and can also have much greater associated uncertainity. Lower resolution taxonomy (e.g., family level) provides less information, but may be sufficient depending on the question. For some questions, abundance, age-structure, or biomass information are also informative and may strengthen an assessment. Whatever indicators are chosen, whatever resolution selected, and whatever population level data are desired, documenting the selection rationale is necessary. Chapters 5 through 7 describe specific assemblages and assessment methods.

Data entry and storage are those methods or software programs used to enter and catalogue information. A variety of platforms exist for data management and storage, both of which are important to ensure long-term integrity of the data. This process must also consider how data will be extracted and manipulated. Given the potential benefits of web-based information transfer, data management planning should include dissemination of program information in web platforms as well as data entry storage and access. Broadly available, spatially integrated data (e.g., in a GIS) can improve stakeholder access and program visibility. This process should be carefully thought out and well-documented.

Essential components of a strong and defensible assessment program are documented standard operating procedures (SOP) for the collection and processing of all samples along with a detailed QAPP. Programs that lack these documented procedures yield data that are open to question. A number of critical issues need to be considered in collecting any sample and the SOP provides the detail and rationale of each (e.g., considerations for field sampling including habitat type to sample, gear type to use, number of samples to take, reach length to sample and field methods to use). The same is true for sample processing where considerations include proper preservation, labeling, transport and transfer, subsampling, and taxonomy. Developing an effective QAPP that addresses these elements is critical for assuring reliable data that meets the MQOs.

Quality assurance and control provisions are important for tracking and minimizing sources of error in monitoring programs. QA/QC procedures establish routines and documentation to ensure proper sampling, processing, data entry and data analysis methods are followed and that

systematic error is documented. Each of the assemblage chapters covers important QC procedures appropriate for each method and analysis step.

2.1.3 Interpretation Elements

Interpretation elements are the analytical methods used to analyze data and to assess both river condition and assessment program quality. These elements include determining and developing ecological metrics, biological indices (e.g., multimetric indices or RIVPACS-type scores) and thresholds, diagnostic capabilities, and performance evaluations.

Once assessment information has been evaluated, it is applied in some program context such as determining trends, preparing monitoring reports (e.g., 305[b]), or evaluating the impacts of certain discharges or catastrophic events on a water resource. In any context, a critical element of the assessment process is the ability to define biological condition adequately enough to detect when changes have occurred. This includes being able to define biological endpoints and thresholds of change. Thresholds will depend, in part, on the precision of the data; therefore, more precise methods will lead to more sensitive response variables. But these attributes will be dictated in large part by the DQOs outlined in the design.

Metrics are particular aspects of the structure and function of the biological assemblage that are of interest because they are judged to be ecologically-significant and respond to disturbance. Metrics include aspects of taxonomic composition, abundance, stressor tolerance, organism condition, and feeding type. These metrics are most often synthesized into *biological indices* that represent biological condition relative to reference condition. A common interpretation approach in the US is the multimetric index (such as the Index of Biological Integrity [IBI]) (Karr et al. 1986, Barbour et al. 1999) that combines several assemblage attributes into one, dimensionless index. Individual metrics are selected to reflect a variety of biological characteristics that respond to human influence in predictable and consistent ways. As such, they reflect a wide range of information regarding the structure and function of the assemblage. Useful metrics are ecologically-relevant and sensitive to stress. RIVPACS-type empirical models, more widely used in Europe, predict the assemblage of organisms expected for a site in the absence of stress, and are derived from reference sites. The ratio of observed (O) taxa at a site to those expected (E) gives the O/E ratio, the proportion of expected taxa actually observed, a straightforward measure of impairment. Methods for developing multimetric or RIVPACS-type models are available from a variety of sources (Hughes et al. 1998, Barbour et al. 1999, Karr and Chu 1999, Hawkins et al. 2000, Wright et al. 2000, Klemm et al. 2003 and see Chapter 8).

Thresholds are a measured level of biological condition above which the support of a designated use is indicated. Use attainment thresholds are often derived from the distribution of index scores in the population of sampled reference sites (e.g., the 25^{th} percentile of reference). Differences exist in the percentiles chosen and the distributions used (reference vs entire), but these thresholds determine which sites do not attain a designated use. Therefore, for defensibility, it is critical to carefully document threshold development and selection. Beyond attainment thresholds, other thresholds can be set to clearly distinguish higher or lower assessment categories. These thresholds are often useful for identifying sites of concern or

prioritizing conservation efforts. Multiple thresholds are also the basis for the tiered aquatic life use approach (Davies and Jackson 2006).

Another common application of assessment data is *diagnosis*. It is often not enough just to know that some change in biological condition has occurred. Frequently, knowledge about the likely source of that change is desired to stop or reverse the impact. Diagnosing causes of change requires integrating biological data with the physical and chemical data collected. It depends upon developing patterns and response signatures from a database that includes a range of stressors and biological responses. This capability is restricted to programs that have targeted their designs to incorporate these ranges of disturbance and response. In addition, a stressor identification protocol has been described in detail (USEPA 2000a, Suter et al. 2002, Cormier et al. 2002) and is discussed in the data analysis section of this document (Chapter 8).

A critical component of an assessment program structure is *performance evaluation* which provides information critical for gauging how well data collection meets programmatic needs. Performance evaluation includes everything from field collection audits and taxonomic checks to data entry verification and index calibration. It also should include regular training and external program review. If done correctly, a performance-based methods approach should provide a documented record of methodological quality and program performance. This topic is dealt with in detail in Chapter 3, and with each of the assemblage chapters (Chapters 5, 6, and 7).

In the following sections, some of the specific assessment elements are introduced, namely the biological, chemical and physical habitat elements, as well as reference conditions and data management.

2.2 Physical Habitat Quality

Physical habitat consists of the structural features of the riverine environment that influence the life history of the biota. Habitat and biological diversity are closely linked (Raven et al. 1998) and the loss or damage of habitat is one of the principal stressors to biota (Karr and Dudley 1981, Karr et al. 1986). There are a variety of habitat assessment approaches from highly quantitative methods designed to describe the geomorphic condition of streams and riparian zones as well as the biotic habitat condition (e.g., Kauffman and Robison 1997), to more qualitative methods using visually scored elements, principally designed to grade the biotic and adjacent riparian habitat alone (Barbour et al. 1999). When combined with land use/land cover data for adjacent and catchment areas, it is possible to draw an accurate picture of physical factors acting on a reach which helps with the initial stressor identification for impaired river sites. Documentation and assessment of large river physical habitat is covered in Chapter 4.

2.3 Chemistry

Biological data are not usually collected alone, and are often accompanied by a variety of physical and chemical measures. These data are key assessment elements, providing direct measures of water quality, many of which have associated standards. These data can be critical for helping characterize stressors and for interpreting biological assessment results.

A variety of chemical measures generally accompany biological sampling, and are used to characterize the chemistry of the water upon which the biota depend. They can be split into two general categories, field measures and laboratory measures.

Field measures are collected with hand-held instruments and, in general, include dissolved oxygen, conductivity, turbidity, and pH. Temperature, while not a chemical measure, is also often collected with these instruments. Some hand-held instruments have probes for measuring other constituents (e.g., chlorophyll, nitrate, etc.), but many are still under refinement.

Laboratory measures are analyzed from water samples collected in the field and transported to the laboratory. They can include common measures such as nutrients (e.g., total phosphorus and nitrogen) and simple cations and anions (e.g., sulfate and chloride). These analytes have established impacts and links to stressors, and their low analytical costs should permit their analysis as part of routine monitoring. Less common laboratory measures include heavy metals, pesticides, aromatic and aliphatic hydrocarbons and emerging contaminants such as pharmaceuticals and personal care products. The costs of these less common measures are often much higher. Although technological improvements will likely reduce these costs, the costs for their analysis in routine biomonitoring without a clear objective for their use, makes them a lower priority.

2.4 Biology

Biological assemblages are the central focus of biomonitoring programs, as they provide a direct measure of biological condition relative to biological integrity, a stated goal of the CWA. But the biota also integrate the effects of multiple stressors in space and time. These environmental sentinels provide a way of detecting stressors that may be so variable in time (e.g., pulses of metal effluent associated with storms) or space (e.g., bank erosion) that it is neither logistically nor economically feasible to monitor them directly. For example, episodic pollutants cause mortality that is reflected in changes in community structure long after the event. Similarly, sediment inputs associated with spatially variable erosion will have impacts far from the source, helping to integrate this variability into a distinct biological response.

A variety of taxonomic assemblages, have been used for biological monitoring. The three primary assemblages are algae, macroinvertebrates, and fish. Use of aquatic macrophytes for biological monitoring has shown some promise, and if the reader is interested, literature is available (Rogers and Owens 1995, Angradi 2006). In subsequent chapters we present specific advantages and disadvantages for algae, benthic macroinvertebrates, and fish. In this chapter, we simply introduce these three groups and the different ways in which they can be used.

Algae are primary producers with rapid reproductive rates and short life spans, which means they are indicators of short-term impact (Stevenson and Smol 2003). They are fairly sensitive to a variety of physical and chemical factors. As primary producers, many taxa are especially sensitive to nutrient pollution and will respond directly (Stevenson and Smol 2003). This has led to their use in the development of nutrient criteria. Similarly, these organisms will likely respond more directly than other organisms to certain contaminants (e.g., herbicides). Sampling is relatively easy for many of the common algal taxa. In wadeable streams, this has primarily

focused on periphyton or attached algae, especially diatoms (Stevenson and Smol 2003). In non-wadeable systems, the phytoplankton, or unattached free-floating taxa may also provide an appropriate algal assemblage for use in assessment. Algae can be characterized in terms of both individual taxonomic change or in terms of whole assemblage biomass (or chlorophyll) response (Stevenson and Smol 2003).

Benthic macroinvertebrates are invertebrates visible to the naked eye that live attached to substrates in very high abundances in most streams and rivers. They are the primary consumers in most systems and are an important link between primary resources and higher trophic levels, including many important recreational and commercial fish. Most macroinvertebrates are relatively sessile, which means they are excellent for use in evaluating site-specific impacts. They have a variety of life cycles, with short-lived and long-lived taxa, and thus provide a way of integrating impacts over a variety of time scales. These organisms are relatively easy to identify to the family level and many are easy to identify to genus. In addition, they are highly variable in terms of their tolerance to different stressors, providing important information for interpreting cumulative stressor impacts. Collection methods are relatively easy, straight-forward, and inexpensive. Wadeable stream methods have focused primarily on these benthic groups. However, large rivers may develop a substantial zooplankton assemblage which, though too small to be called macroinvertebrates, are also relatively easy to enumerate and may be useful indicators of water quality and physical stressors.

Fish are a diverse group of organisms that represent a variety of habitat uses. They are relatively longer lived organisms and include many mobile species, so they can potentially integrate effects over longer spatial and temporal scales. The environmental requirements and life histories of many fish species are well understood, meaning that the presence or absence of taxa can often be easily interpreted. Many fish species are consumed by humans and, therefore, they provide an assessment metric that is directly related to human health. In addition, many aquatic life uses are linked to fisheries, providing a direct measure of those uses. Fish are generally easy to collect and to identify to species. Most can be identified in the field and released, unharmed.

For all three assemblages, the goal of method selection should be to provide an approach that is as precise and responsive as necessary, given the constraints of time and effort. A method need not represent the entire assemblage of organisms, unless that is an explicit goal of the assessment. Rather, the focus should be on sampling those elements of the assemblage that give the most consistent and precise responses in meeting the program objectives.

2.5 Data Management

For any environmental data to be useful, they must be organized, accessible and secure. The flow of all types of data needs to be specified, beginning with pre-sampling logistical information and ending with fully QC'd pieces of data entered into a database. Thus, prior to implementation of monitoring programs, several administrative decisions should be made:

- Which person or agency will be responsible for data management and security?
- Will the database be made available on the internet?

- Will all data be housed at a single location on a single server? (that is, will there be a central repository for all data?)
- Who will ensure that sponsors and stakeholders have access to the database?
- Will all data be uploaded to USEPA's STORET?

Without addressing these and other questions and developing a data management system prior to beginning fieldwork, ultimate uses of the data in analyses can be chaotic and time-consuming. An efficient data management system will capture not only primary data, such as direct field observations and results of laboratory analyses, but also data such as location (place names and latitude-longitude), date and time; and ancillary such as (for biology) functional feeding groups, behavioral habit and stressor tolerance values. It is also important that data be easily accessed and exported to basic spreadsheet, statistical analysis or mapping software for analysis and interpretation.

USEPA is developing a biological data management system linked to STORET, which provides a centralized system for storage of biological data and associated analytical tools for data analysis. The field survey file component of STORET provides a means of storing, retrieving, and analyzing biosurvey data, and will process data on the distribution, abundance, and physical condition of aquatic organisms, as well as descriptions of their habitats. Data stored in STORET become part of a comprehensive database that can be used as a reference, to refine analysis techniques or to define ecological requirements for aquatic populations. Data collected using the RBPs (Barbour et al. 1999) can be readily managed with STORET field survey file using header information presented on the field data forms (http://www.epa.gov/owow/monitoring/rbp/) to identify sampling stations.

Habitat and physical characterization information may also be stored in the field survey file with organism abundance data. Parameters available in the field survey file can be used to store some of the environmental characteristics associated with the sample, including physical characteristics, water quality, and habitat assessment. Physical parameters include stream depth, velocity and substrate characteristics, as well as many others. STORET also allows storage of other pertinent station or sample information in the comments section.

Entering data into a computer system can provide substantial time savings. An additional advantage to computerization is analysis documentation, which is an important component for a QA/QC plan. An agency conducting rapid bioassessment programs can choose an existing system within their agency and/or use the STORET system developed as a national database system.

Data collected as part of state bioassessment programs are usually entered, stored, and analyzed in easily obtainable spreadsheet programs. This method of data management becomes cumbersome as the database grows in volume. An alternative to spreadsheet programs is a multiuser relational database management system (RDMS). Most relational database software is designed for the Windows operating system and offers menu driven interfaces and ranges of toolbars that provide quick access to many routine database tasks. Automated tools help users quickly create forms for data input and lookup, tables, reports and complex queries about the data. The USEPA is developing a multiuser RDMS that can transfer sampling data to STORET.

This relational database management system is called Ecological Data Application System (EDAS), and allows the user to input, compile and analyze complex ecological data to make assessments of ecosystem condition. EDAS includes tools to format sampling data so it may be loaded into STORET as a batch file. These batch files are formatted as flat ASCII text and can be loaded (transferred) electronically to STORET. This will eliminate the need to key sample data into STORET.

By using tables and queries as established in EDAS, a user can enter, manipulate and print data. The metrics used in most bioassessments can be calculated with simple queries that have already been created for the user. New queries may be created so additional metrics can be calculated at the click of the mouse each time data are updated or changed. If an operation on the data is too complex for one of the many default functions, then the function can be written in code (e.g., visual basic access) and stored in a module for use in any query. Repetitive steps can be handled with macros. As the user develops the database, other database elements such as forms and reports can be added.

Table design is the foundation of a relational database, such as EDAS (Figure 2-5), because they function as data containers. Tables are related through the use of a unique identifier or index. In the example database "StationId" links the tables "ChemSamps", "HabSamps" and "BenSamps" to the "Stations" table. The chemical parameters and habitat parameters table act as reference tables and contain descriptive data (e.g., measurement units and detection limits). This method of storing data is more efficient than spreadsheets, because it eliminates a lot of redundant data. Master taxa tables are created for the biological data to contain all relevant information about each taxon. This information does not have to be repeated each time a taxon is entered into the database.

Input or lookup forms (Figure 2-6) are screens that are designed to aid in entering or retrieving data. Forms are linked to tables so data go to the right cell in the right table. Because of the relationships among the tables, data can be updated across all the tables that are linked to the form. Reports can be generated in a variety of styles, and data can be exported to other databases or spreadsheet programs.

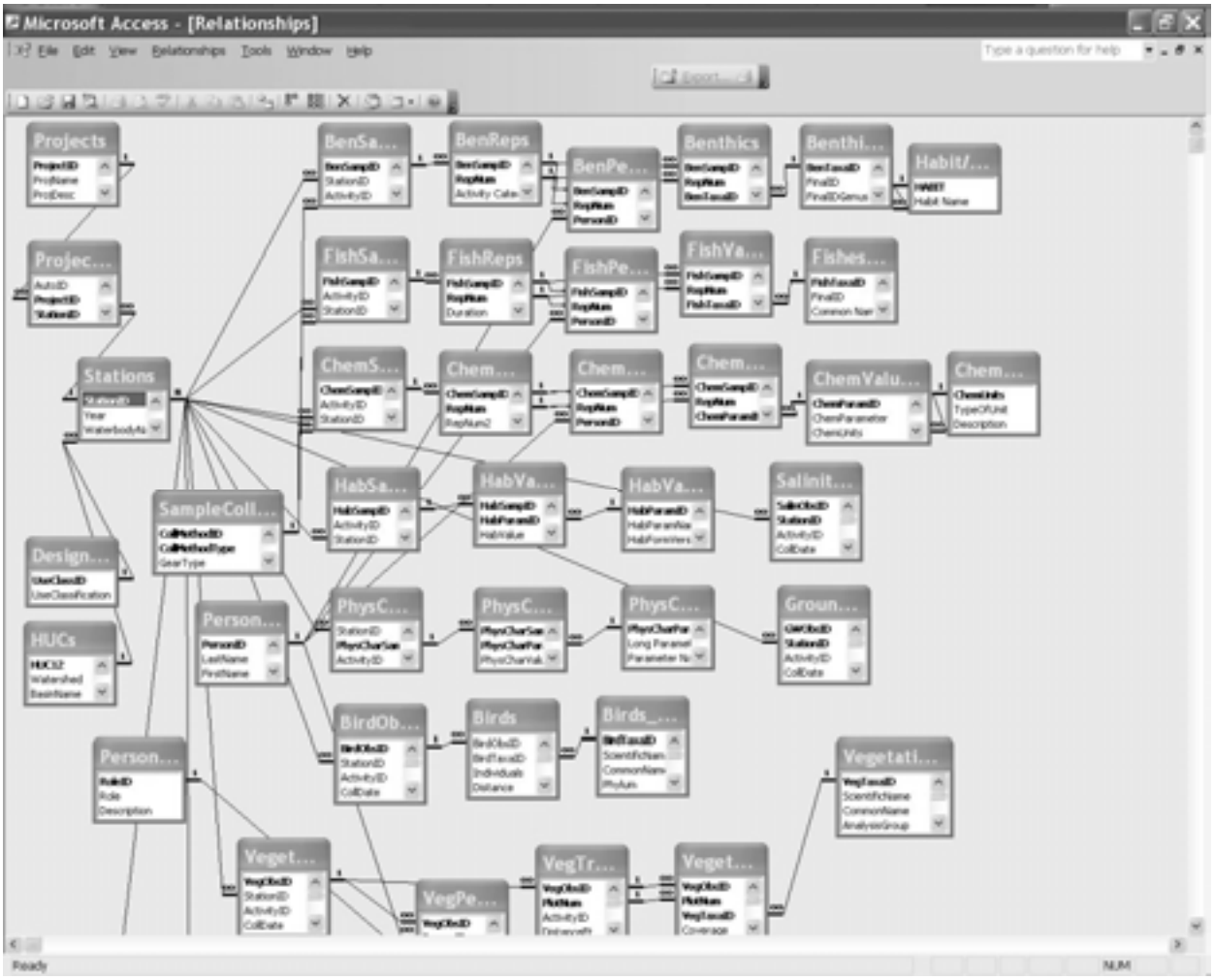


FIGURE 2-5. Example of the relationship of data tables in a typical relational database.

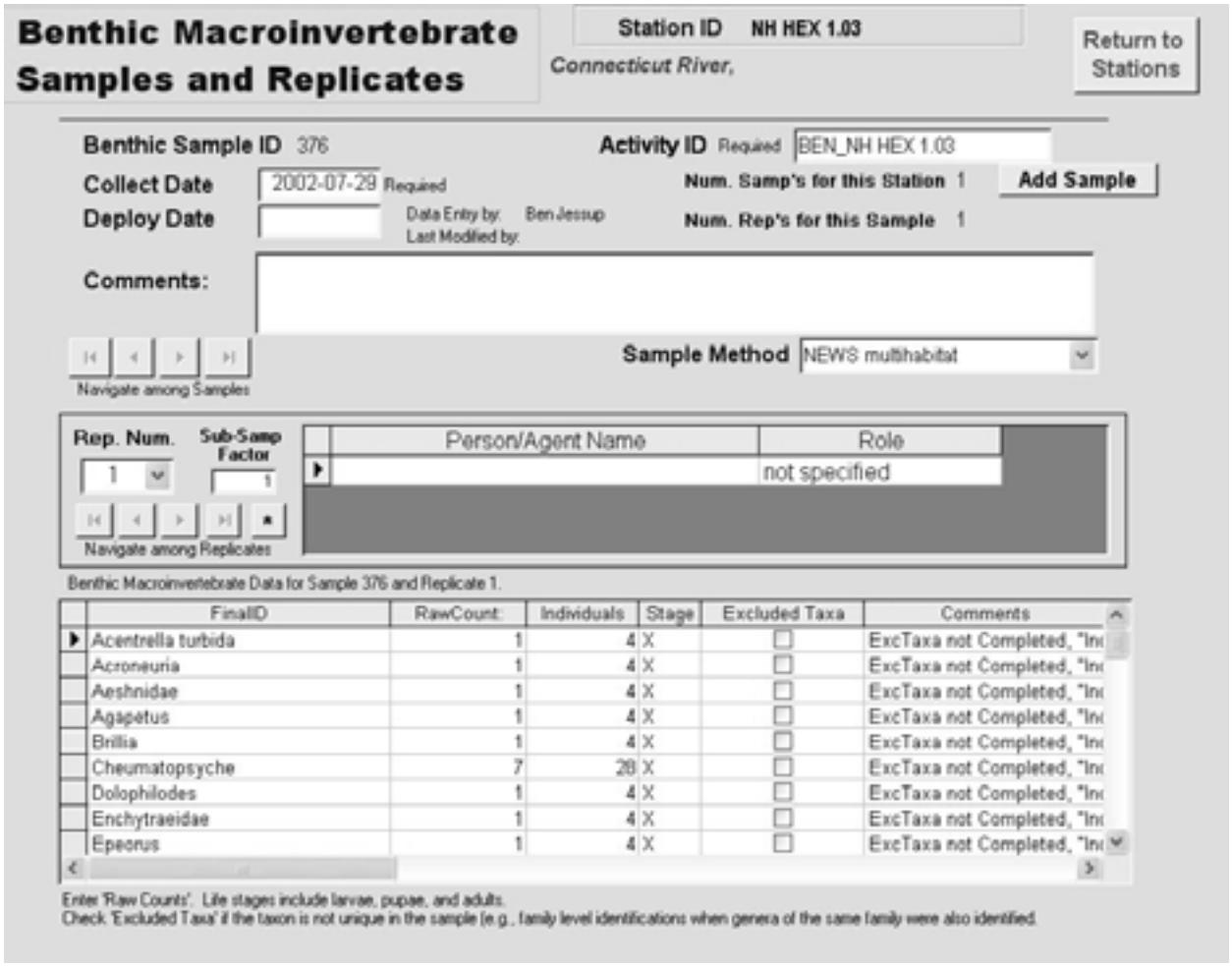


FIGURE 2-6. Example input or lookup form in a typical relational database.

Chapter 3.0 Study Design, Data Quality, and the Performance-Based Methods System

This chapter…

- reviews regional, site-specific, and gradient study designs
- describes methods for ensuring data quality objectives are met

Study design…

- consists of a variety of approaches depending on study questions
- is critical for an assessment program to meet its objectives and its data quality goals

3.1 Types of Study Designs

A variety of monitoring designs exist for biological assessment, all of which depend on the question(s) to be addressed. The major designs used for most assessment programs are commonly based on one of three objectives organized by spatial scale: regional assessments, site-specific assessments, and gradient studies.

In defining an objective, it is important to characterize the population of interest, whether it includes all reaches or segments (stretches between tributaries of a given size) in a region, all hydrologic unit codes (HUCs) of a certain size, all segments below publicly owned treatment works (POTW) sites, or the segment below one specific discharge. Once the objective and population are defined, it is easier to select a study design. Important aspects of statistically powerful designs include the concepts of random sampling and sample allocation. Stratified random sampling reduces variances, allowing a more precise measure of the variable of interest, and is therefore ideal for statistical rigor. But not all levels of a study need employ random selection. The important point is to randomize at the level of the question. For example, if we are interested in the average number of invertebrate taxa found in rivers of a particular region using a specified protocol, taking 100 random samples from one river reach would not be an appropriate allocation of effort because only that one river reach would be characterized. It would be more appropriate to take one sample from each of 100 randomly selected river reaches throughout the basin because the river is the level of the question. Suggestions of pseudoreplication are relevant and potentially valid criticisms only in the context of the questions being asked. To avoid trying to answer a particular question with inappropriate data, it is advisable to work with statisticians or analytical staff familiar with study design and data analysis. The purpose of Section 3.1 and its subsections are to provide an overview of the different sampling designs that could be used for large river systems.

3.1.1 Sampling Effort and Sampling Reach Length

It is challenging to balance required sampling effort with available resources, while simultaneously maintaining focus on programmatic goals and objectives. While increased sampling effort can be justified for increases in precision, there often are substantial increases in the cost of sampling (Reynolds et al. 2003, Hughes et al. 2002, Lyons 1992b, Cao et al. 2001). As Angermeier and Smogor (1995) point out, comparisons of estimates based on insufficient sampling effort can be confounded because real differences in assemblage structure may be indistinguishable from method error. In a bioassessment context, this can translate to a

decreased ability to distinguish among sites of varying condition (Patton et al. 2000). However, identifying the most appropriate sampling effort in assemblage-level studies is often ignored.

A number of issues have emerged that are worthy of discussion regarding sampling reach length for non-wadeable rivers. On these systems, sampling reach lengths are generally larger than those in wadeable systems - a result of the scaling-up to accommodate the magnitude of the resource. The approach used can result in relatively long (i.e., kilometers) or short sampling reaches (< 1 kilometer). Long reaches may mask small scale habitat conditions and impairments that may be of interest to resource managers. They may also weaken the ability of the data to detect linkages between local river conditions and the drivers of those conditions. In designs where a long sampling reach is warranted, however, several small sub-reaches, and thus multiple data points, can be used to estimate spatial variability. Such short reach lengths highlight small scale conditions which may, simultaneously, reduce their utility for estimating broader-scale characteristics. Both perspectives are justified at times, and selection of the appropriate reach length for a study should depend on the questions being addressed by the study.

The development of a scientifically-sound sampling design for large rivers must include some discussion of the sampling effort to be exerted at a given sampling location and along the river (Lyons 1992b, Angermeier and Smogor 1995, Paller 1995, Peterson and Rabeni 1995, Patton et al. 2000, Cao et al. 2001, Cao et al. 2002, Hughes et al. 2002, Dauwalter and Pert 2003, Reynolds et al. 2003, Maret and Ott 2004, Fayram et al. 2005, Flotemersch and Blocksom 2005, Hughes and Herlihy [accepted]). Any description of sampling effort includes specifying the spatial scale over which the sample(s) will be collected (channel length), the amount and types of habitats that will be sampled within that length, and the field sampling method to be used (Reynolds et al. 2003). Further, the field sampling method is typically described by detailing gear, specific habitat types, intensity, and often, an estimated number of person-hours per sample (or site). Estimates and inferences regarding assemblage attributes (e.g., number of taxa, metrics, and IBI scores) are sensitive to sampling effort (Angermeier and Karr 1986, Angermeier and Smogor 1995, Rosenzweig 1995, Patton et al. 2000, Cao et al. 2002, Reynolds et al., 2003, Flotemersch and Blocksom 2005, Hughes and Herlihy [accepted]) because riverine habitat is heterogeneous with non-uniform distribution of organisms among habitat types (Angermeier and Smogor 1995). The number of taxa collected at a given site will, thus, increase with sampling effort, and will also vary with biogeography, sampling method and efficiency, behavior and abundance of the assemblage being sampled, and patchiness of the targeted habitat components.

Ideally, the sampling effort applied is the minimum that will allow stated objectives to be addressed as required by a study (Angermeier and Smogor 1995, Patton et al. 2000). As an example of how the question can influence the required effort, estimates of species' relative abundances have been shown to require less sampling effort for a given accuracy than estimates of the absolute number of species (Angermeier and Smogor 1995). For a bioassessment program, potential cost savings realized through the use of efficient sampling protocols translate to opportunities to enhance other aspects of a study design or program (Patton et al. 2000). This section will focus on issues related to definition of the appropriate sample unit for large river bioassessments. In other words, what is the channel length that will be sampled?

3.1.1.1 What is a Reach?

In a hierarchical context (Figure 2-4), Frissell et al. (1986) defined the word "reach" as a length of stream between breaks in channel slope, local side-slopes, valley floor width, riparian vegetation, and bank material. They further added that the reach is sometimes the least physically discrete unit in the hierarchy, but an exceedingly useful scale for describing medium- and long-term effects of human activities on streams. We use the term "sampling reach" to describe the site from which samples are collected. In linear systems, such as rivers, it is quantified as some channel length.

Many factors relevant to sampling reach length decisions in wadeable streams (e. g., Patton et al. 2000, Lyons 2002) will influence those same decisions in larger, non-wadeable rivers. Paller (1995) suggested that streams with low species richness may require greater reach length-to-width ratios (l:w) to attain precise estimates of maximum species richness (MSR). However, large Oregon rivers with low fish species richness required less sampling effort to attain MSR relative to rivers with a higher species richness (Cao et al. 2001, Hughes et al. 2002). Paller (1995) also found that the relative importance of sampling depth may depend on the behavior of individual species (e.g., substrate or open-water orientation), or upon width-to-depth (w:d) ratios. Many large rivers have an abundance of habitats supporting fish species that are difficult to efficiently sample (e. g., those associated with deep, turbid, or swift-moving waters or off-channel habitats); they can be more frequent in some regions of the country than others. For these kinds of rivers and species, Angermeier and Smogor (1995) found that greater sampling effort is necessary to attain and adequately characterize fish assemblage structure.

3.1.1.2 Approaches for Sampling Reach Length Determination

In most applications, the channel length over which data are collected is the same for physical habitat measures and biota. Exceptions to this would be measures that characterize the larger watershed of the reach, and water grab and phytoplankton samples collected at a single point in a reach. The logistical advantages to using the same reach length for multiple indicator parameters collected over the extent of the reach are clear, because the same persons can collect different data at the same place and time. However, variable reach lengths may be justified, depending on the indicator for which the sample(s) are being taken. For example, because biota move down and upriver, an argument could be made that the channel length over which physical and chemical habitat data are collected should exceed that over which assemblage information is collected.

Different approaches have been used for determining the channel length used for bioassessment of large rivers, most involving consideration of several factors including the question being addressed by the study, the level of resolution (precision and accuracy) required to address the question, and the statistical approach that will be used to analyze any resulting data. Just as critical is ensuring that sampling reach length is balanced with available resources. The following discussion is intended not as an exhaustive review of the topic, but as an overview with examples.

The reach lengths for most studies were set based on judgment, past history, or the need to match some other aspect of sampling or management activities. However, recent research has been conducted on the selection of sample reach lengths by evaluating the response of biological parameters (e.g., species accumulation curves, assemblage metrics; IBI scores) as a function of geomorphology (e.g. channel widths, meander wavelengths, riffle-pool sequences). Most of these studies have used fish assemblages (Gammon 1976, Lyons 1992, Meador et al 1993, Penczak and Mann 1993, Angermeier and Smogor 1995, Paller 1995, Yoder and Smith 1999, Patton et al. 2000, Cao et al. 2001, Lyons et al. 2001, Hughes et al. 2002, Reynolds et al. 2003, Maret and Ott 2004, Flotemersch and Blocksom 2005, Hughes and Herlihy [accepted]), although a few have used benthic macroinvertebrates (Bartsch et al. 1998, Li 2001, Poulton et al 2003, Flotemersch et al. 2006). Whether MSR of the local or regional fish assemblage, form of the final indicators (metric or index scores), or geomorphic characteristics should drive reach length determinations should depend on by programmatic considerations and the overall questions being addressed.

Biological Approach

The rationale for using biological measures for determining reach length is that in bioassessment we are, by definition, assessing the condition of biota. Therefore, the sampling effort required to produce reliable indicator results (metrics, indices) seems to be a logical determinant of reach length. In most cases, this question is addressed by over-sampling at a series of sites that cover the gradient of conditions to be included in a study and then determining the reach length for which the required data quality has been achieved. Reach length is then determined based on when a specified indicator asymptote is reached (Lyons 1992b, Angermeier and Smogor 1995, Paller 1995, Patton et al. 2000, Cao et al. 2001, Lyons 2001, Hughes et al. 2002, Reynolds et al. 2003, Maret and Ott 2004), when some level of similarity has been attained (Cao et al. 2001, 2002), or variability of that measure has been reduced to a desired level (e.g., Flotemersch et al. 2006, Hughes and Herlihy [accepted]).

Design specifics have varied among these studies, resulting in differing conclusions. Hughes et al. (2002) sampled 100 wetted channel widths, and through data analysis, determined that 85 channel widths were needed to collect 95% of the species obtained in 75% of the reaches sampled; collection of all fish species in a reach was calculated to require 300 channel widths on average. Those findings resulted in a field sampling design specification of 100x wetted width (Peck et al. [in press]). Hughes and Herlihy (accepted) determined that 50 channel widths were needed to obtain IBI scores exceeding those obtained from 100 channel widths less than 10% of the time. In contrast, Flotemersch and Blocksom (2005) examined the effect of reach length on the variability of IBI metrics from samples covering up to 2 km, and determined that at shallow river sites 1 km total shoreline shocked was sufficient for limiting the change in metric scores to 20%. Additional recommendations were provided for deep river sites. These three studies began with different reference conditions (100 channel widths vs 1000 meters), different maximum distances (100 channel widths vs 2 km), and different values for acceptable variability (5, 10, and 20%), and thus produced different results.

Physical Approach

Fixed Length vs Multiples of the Wetted Width (MWW)

Another difference among study results is how the final reach length is framed. Some studies propose reach lengths as a function of multiples of the wetted width (MWW) of the channel (e.g., Lazorchak et al. 2000, Hughes et al. 2002, Reynolds et al. 2003, Maret and Ott 2004, Peck et al. [in press]) while others support the use of a fixed distance (Flotemersch and Blocksom 2005).

The MWW approach follows the logic that as a system gets bigger, the effort required to sample the habitat components of the system at an equivalent level should increase proportionally. In other words, a fixed length of 500 m on a river 100 m wide could potentially miss or under-represent habitat components (such as bar, glide, pool, inside bend, outside bend that recur at longer intervals). One argument against this logic is that differing amounts of sampling effort are being applied across sites, by definition. A difficulty encountered with this approach in wide or impounded rivers is long reach lengths (e.g., 5 km for a 100 m wide river if 50 channel widths are the protocol). It is possible that pre-impoundment wetted width could be used in these cases (although the information is often not readily available), or that impoundments could be sampled like lakes.

Others have set reach length as a fixed distance (Flotemersch and Blocksom 2005) rather then as MWW. Proponents of a fixed distance endorse the ease of application in the field and utility in planning field activities (Patton et al. 2000, Lyons 2001, Flotemersch and Blocksom 2005). Opponents argue that using a fixed distance results in unequal sampling effort relative to river size, and that studies of fixed lengths have had lower data quality objectives regarding reference condition, maximum level of effort, and acceptable levels of variability.

A second argument against fixed lengths is that where the reaches do not encompass a sufficient number of habitat units, the biological differences detected may be due to differences among the habitat units of the sites. This becomes a greater concern as river width increases. For example, the Ohio River Valley Water Sanitation Commission (ORSANCO) conducts biological sampling on the Ohio River using 500 m reaches (http://www.orsanco.org/watqual/aquatic/electro.asp). The problem of the reach not including all habitats of a meander is addressed by the development and use of habitat specific criteria for soft, hard, and mixed bottom types. But such criteria ignore the often substantial diel migrations of larger fish species.

Meander Cycles

An alternate approach to using the response of biological parameters for setting reach length is to set it independent of the biology using the geomorphology of the system. This approach has its origins in work conducted by Leopold et al. (1964) who proposed that in meandering streams, 20 times the bankfull channel width typically encompasses at least one complete meander wavelength of the system. Because fluvial characteristics are repetitive and cyclical (Dunne and Leopold 1978), this distance should theoretically include all major habitat types within a given geomorphic reach and, by default, be available to all resident biota of those habitats. Given this,

the logic behind using geomorphic meanders as a basis for setting reach length for bioassessment is clear.

However, in altered large rivers, the identity, extent, and boundaries of habitat units of a meander are often non-distinct, obscured by turbidity or impoundments, or removed by anthropogenic alteration of the channel (straightening, armoring, and dredging). These conditions can render identification of a meander an impractical option for setting reach length and highlight the value of the finding by Leopold et al. (1964) that one meander roughly equates to 20 times the wetted channel width.

Following this guidance, NAWQA uses 20x wetted width, and sets a minimum length of 500 m (to help ensure representativeness of biological data), and a maximum of 1000 m (to minimize crew fatigue) (Fitzpatrick et al. 1998). However, such inconsistent levels of effort could potentially lead to difficulties in interpretation.

Ultimately, it is the quantity and quality of information required that will dictate the level of effort that can and should be expended at each sampling location. Thus, application of the data quality objectives process, including quantification of desired indicator performance, and testing of the capacity of sampling design to meet those objectives (both site specific and area wide), should drive the appropriate reach length.

3.1.2 Regional or Area-wide Assessments

In this document, regional assessments are defined as those that assess water resource quality across a broad region for status and trends monitoring. These studies are typical of designs used to meet the 305(b) reporting requirements under the CWA, and often result in estimates of the proportion of waterbodies in a certain condition (i.e., good, fair, or poor; or attaining and not attaining).

Representativeness is a critical factor given that the objective is to estimate a parameter (e.g., mean condition) from a subsample of a larger population (e.g., all large river reaches in the region). An important note with large rivers is that it may be possible to sample the entire population in some regions. For example, in more arid regions, there may be a limited number of large river segments. If the segment can be sufficiently characterized with a reach-based sample, it is conceivable that the entire population of segments can be sampled, allowing calculation of the absolute mean and variance for the population. Most often, however, the population of segments or sample units will be large, making a census impossible. Some inference of the average condition and variance will have to be made using randomized selection of sampling reaches (Larsen 1997, Urquhart et al. 1998). To reduce bias in the final estimate (e.g., percent of river miles impaired), probability-based designs for site selection are appropriate.

The first step in this type of design is to organize continuous, linear systems like rivers into representative units. For large rivers, this could be river segments, a standard hierarchical unit defined by lengths of rivers between tributaries of a given size. The second step is creating an approach for sampling these segments randomly. This might mean creating a list of “sample

units" (or list frame), applying a code to each unit, and randomly sampling them based on specified rules. Sample units could also be selected using a grid placed over a region, selecting grid cells at random, and sampling large river segments within them. This approach can also be used hierarchically, so that large grids (tier 1) are randomly selected and then small grids (tier 2) within tier 1 grids are randomly selected for sampling (two-stage sampling). A benefit of this approach is that not all rivers would have to be digitized beforehand, which can be costly if these data do not exist. Only those segments within selected tier 1 grids would need to be digitized (Rathbun 1999). However, this is an unlikely problem for large rivers given the availability of existing digital information for rivers throughout the USA (e.g., USEPA's river reach file [RF3] coverage or USGS national hydrography dataset [NHD]). The EMAP program used a grid selection approach as part of its probabilistic design (Overton et al. 1991, Stevens 1997, Stevens and Olsen 1991, 1999).

Although there are a variety of probability-based designs, only the simple random, stratified random, and systematic random approaches are discussed here. In *simple random* designs (Rathbun 1999), the entire pool of segments is the population, and sampling sites are selected randomly (Figure 3-1). This is the most basic probabilistic design. One drawback of this approach is that natural variability among sample units will increase the sample size needed to attain a given precision. Results of pilot studies should be used to determine appropriate methods, the level of precision (repeatability) a method is capable of, and, thus, how many samples are necessary to detect a desired change.

This natural variability can be partially controlled during sampling design by partitioning the region into strata based on underlying, scientifically defensible, natural classes (e.g., ecoregions or stream orders) using a *stratified random* design. The sample units are then selected randomly from these strata (Figure 3-1). The strata should be selected to maximize the differences *among* strata and minimize the differences *within* strata. By partitioning the natural variance among segments within strata, this design can achieve the same precision using a smaller sample size than a simple random sampling design, thereby reducing costs (Rathbun 1999). Sampling allocation may be made proportional to the size of the strata (e.g., if 10% of the segments are coastal plain, then 10% of the total sample effort would be randomly selected coastal plain segments) or can be apportioned based on the within-stratum variance, if known. However, at the very least, two sites are needed within any stratum to generate an average or variance estimate. Using too many strata could lead to poor variance estimates of the river overall, and thus, stratification should only be used with caution.

Systematic random is an approach for sampling site selection where the starting point (i.e., the first site) is selected at random, and those following lie at regular intervals. For example, the initial sampling location might be a 500-m segment with the midpoint at River KM 100. That point would have been randomly selected from within the 25-km distance encompassing the wadeable/non-wadeable transitional zone. Then, a reach midpoint would be located every 50 km downstream to the confluence with a channel of the same size or larger (or to tidal zone, or to estuary). Each sampling reach produces a random sample. Results from this design are used for estimating overall condition of the river system (as a mean value), or examining cumulative downstream effects. Additional information on different types of monitoring designs can be

found on the EMAP website:
(http://www.epa.gov/nheerl/arm/designpages/design&analysis.htm).

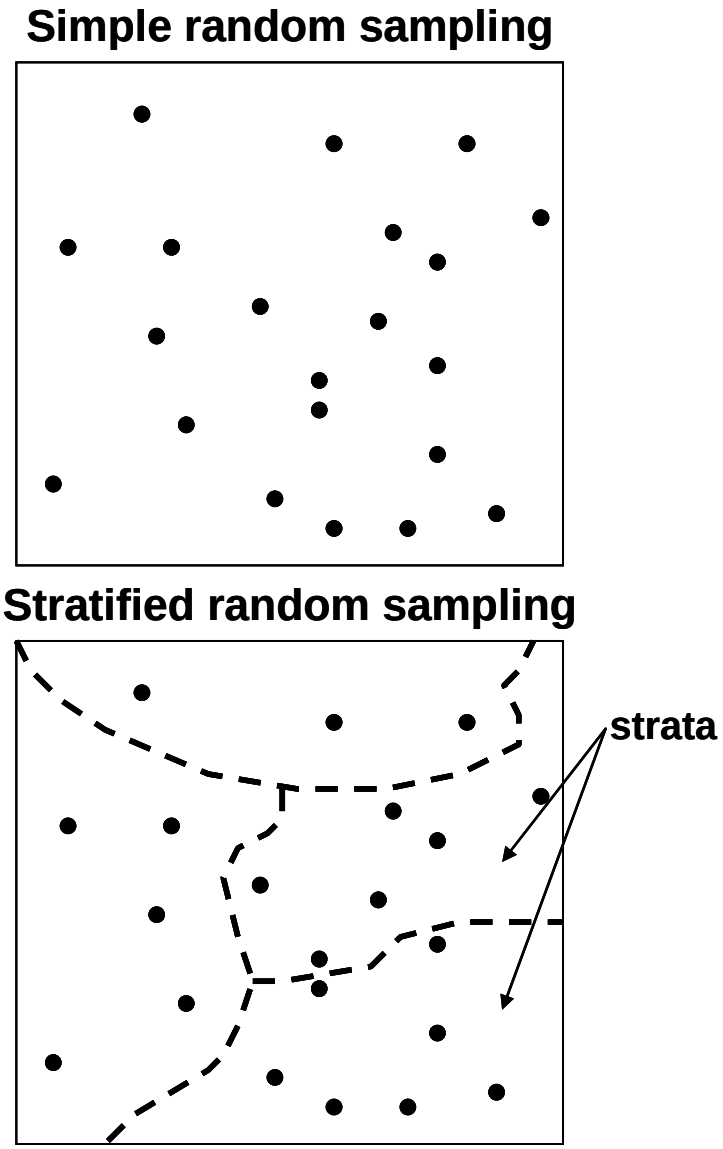


FIGURE 3-1. Examples of two-dimensional probabilistic sampling designs.

Quantifying trends in resource condition is often an important objective for regional assessments. Although there are different approaches for allocating sampling effort over time, only two are covered in this document: permanent station and serially alternating (Rathbun 1999). *Permanent station* approaches use a random sample of *n* sites that are all sampled during each time interval. This option provides the least spatial coverage but may provide the highest temporal resolution of trends, if temporal autocorrelation is weak. It is noteworthy that if resources allow sampling the entire population of large river segments, a permanent station temporal design is appropriate

then as well. *Serially alternating* designs (Table 3-1) or "rotating" designs partition the random sites within a stratum into sub-sets of sites that are sampled at regular intervals (e.g., every four years). This design was proposed for EMAP (Messer et al. 1991) and is the smaller scale probabilistic design used by the Maryland Biological Stream Survey (MBSS) for wadeable streams (Maryland Department of Natural Resources 1999).

TABLE 3-1. A serially alternating or rotating design for site sampling. In this example, all of the randomly selected sites are split into four sample sets. Sample sets would be serially sampled, such that each set is visited three times over 12 intervals (modified from Rathbun 1999).

		Sampling Interval (years, seasons, etc.)											
		1	2	3	4	5	6	7	8	9	10	11	12
Sample Set	1	X				X				X			
	2		X				X				X		
	3			X				X				X	
	4				X				X				X

3.1.3 Site-specific Assessments

Site-specific assessments focus on particular sites or small sets of sites, usually for the purpose of assessing the effects of a specific and known stressor source (e.g., effluent) or the effectiveness of a given intervention (e.g., restoration). Other site assessments may be performed for a unique question driven by a specific request (e.g., Is this segment of river comparable to the reference condition?). These objectives can be achieved through a variety of designs.

Traditionally, site-specific studies have been conducted using upstream vs downstream sampling, with a completely random selection of sampling locations some distance above and below the point of interest. One way of describing this is as a control-impact (CI) design. This sampling design is only able to compare the condition of the downstream reach to that of the upstream reach and use that as part of a weight-of-evidence argument for an impact. Drawing a conclusion that any effect is specifically due to the effluent is difficult because: 1) the effluent input is not replicated, and 2) since effluent pipes are not generally randomly placed, the local physical setting also likely influences the upstream and downstream conditions. As a result, it may be impossible to rule out other factors related to the upstream or downstream environment as responsible for observed differences. This effect can be reduced by comparing mean differences between the control and impacted sites to mean differences between comparable river segments without impacts. Samples through time can be used as replicates; but the impacted site would be pseudoreplicated, so there is only one true impact replicate (Hurlbert 1984). Still, some level of repeated sampling would improve weight-of-evidence arguments.

One option to reduce some of the limitations of CI analysis is to design a study to collect data prior to an impact and compare it to data collected after the impact begins. This design is referred to as a before-after (BA) design. The BA analysis requires a sufficient amount of before (pre-stressor or pre-effect) data so that the two sets of data can be analyzed as independent samples using two sample tests (t-test or analysis of variance) (Smith 2002). It is best to randomly assign the sampling dates to avoid systematic trend errors. As in the previous case, however, causal inference is problematic because observed trends may be due to climatic

differences or other natural events before and after the impact. Thus, this design does not have the controls that would account for natural, widespread changes. In addition, the impact may affect the variance structure rather than the mean, making detection difficult. Lastly, sufficient *before* data are often not available or bias exists because of *when* the sample was collected (sample timing), either of which would affect statistical power and inference. However, the BA approach could be used in building a weight-of-evidence assessment.

Incorporating a control site into the design of a BA approach provides some control of natural variability associated with time. In this design, data are also collected from the control site before and after the impact. It is best to randomly assign the sampling dates. The data are analyzed using a two-way ANOVA (BA and CI) with interaction (BA x CI), and the design is known as a before-after control-impact (BACI) design (Smith 2002). Such designs have been criticized because the sites are not randomly assigned and there is only one treatment area (Hurlbert 1984). One way around this statistical hurdle is to pair sampling at the control and impact areas and sample several times, resulting in a before-after control-impact paired design (BACIP) (Eberhardt 1976, Stewart-Oaten et al. 1986). The BACIP designs are treated much like a repeated-measures design with multiple times on one site instead of multiple treatment replicates. Each site-pair-time combination is treated as a unit. The ANOVA models in this analysis have BA, CI, sample time, and interaction (BA x CI) terms (Smith 2002). However, a simpler analysis of this design calculates differences for values collected at each site-pair-time unit and compares mean differences before and after the impact (Stewart-Oaten et al. 1986). This has also been called the paired BACI or BACI paired series approach (Smith 2002).

Variations on the BACI models have included increasing the number of randomly selected control sites (asymmetrical BACI design) (Underwood 1991, but see criticism from Stewart-Oaten and Bence 2001), including additional impact sites (Ellis and Schneider 1997), and using multivariate extensions of BACI (Faith et al. 1991, Kedwards et al. 1999).

3.1.4 Gradient Studies

The last class of designs discussed here are those that investigate the nature of the response to specific stressors. Rather than attempting to answer a yes-or-no question, these approaches investigate ecological response to gradients in stressor levels. The objective is to provide information to improve future management actions. An example would be to ask how biological condition changes in response to increasing urbanization density. Is the response linear or non-linear? Are there thresholds in the response? Such information can help land use planners manage future development differently. Another objective might be to define the response of a particular taxon to a known stressor. This information could be helpful in developing stressor tolerance values for taxa.

The main design approach in these studies is regression, where samples are collected along the entire gradient of the factor of interest (e.g., conductivity) and ecological response is measured. If pure hypothesis testing were the desired goal (e.g., do benthic IBIs respond to urbanization density?), then the levels of the independent variable should be controlled by the experimenter and all else left equal. This is not really possible for most assessment designs because there is rarely the opportunity to control land use intensity, but randomization schemes could be used to

reduce site selection biases. One factor that must be considered is that as samples are taken further downstream in a large river basin, the influence of small, degraded streams on downstream water resource quality is often masked due to the overwhelming differences in flows. However, if there is less interest in testing the hypothesis and more in defining or modeling the relationship, then the level of control on the independent variable is less important. It is easier to use existing gradients and to define the response with regression models.

Simple linear regression is used to define the response of one dependent variable (y-axis) to an independent variable (x-axis). Further, multiple linear regression is used to explore the response of a dependent variable to several independent variables (individual, transformed, or combinations of independent variables). With multiple linear regression, the relative effects of several potential explanatory variables can be examined simultaneously using a variety of approaches, setting a fixed multi-variable model, adding one predictor at a time, or starting with all of the independent variables and removing one at a time. In any case, the effectiveness of gradient designs depends on bracketing the gradient as well as possible. It is important to realize that certainty about responses is highest in the region where there are the most data (usually along the middle of a gradient), and lowest where data are least (usually at the extremes). This information must be extrapolated if there is interest in responses beyond the range of the gradient used to develop the models. Extrapolation is risky and any model should only be applied with great caution beyond the range of the independent variables used. With multiple regression using a number of transformed variables, this range is often difficult to identify.

3.2 Coordinating Sampling Design with Management Objectives

As discussed in Chapter 2, it is important to understand and present the specific questions, general goals, and potential uses for the assessment results; the DQOs that correspond to these goals (for the ultimate data user), and the quality of the measurement data that are necessary for the DQOs to be met (MQOs) (Figure 3-2). For biological monitoring and assessment practitioners, the following questions are common: *How healthy is the river? Is the river getting better? What is the condition of our watershed?* If there is general agreement that ecological indicators (in particular, multimetric indices of biological integrity [IBI, Karr et al. 1986]) and the ratio of observed to expected (O/E) taxonomic diversity (Wright 2000) provide the most appropriate information about overall water resource health or condition, then important decisions concerning the spatial placement of sampling sites and frequency of sampling.

Next, the study design process should allow specification of the spatial scale needed to address the objective: *Is the assessment intended to be for a single, particular river reach (e.g., 1 km, 10 km, the entire 2-km reach between two cities) or for all non-wadeable reaches within an entire watershed (at whatever scale the watershed might be defined)?* That is to say, is the objective to make defensible statements of condition for individual sites, for area-wide scales, or both?

Answering questions at area-wide scales requires aggregating multiple site-specific assessments to the scale of interest. However, if there is a probability-based component to the site-selection process (Stevens and Olsen 1991, Larsen 1997, Urquhart et al. 1998), then data can be used at

multiple spatial scales; from site-specific, to watershed-wide, to region-wide scales. Answers can be expressed in the following forms:

- The overall biological condition of River X at River Mile 27.14 is "fair" (IBI, 42 ± 9.4).
- The mean biological condition of non-wadeable river reaches in Watershed Y is "good" (IBI, $\bar{x}$ = 74, n = 12, 90% CI = ± 7).

These questions can be answered in a credible and defensible manner, only if data of sufficient quality and quantity are collected. Once the data user settles on the types of questions s/he is asking (or is being asked) and the kinds of answers that would be satisfactory (e.g., with known and acceptable confidence), then data of the required power and sensitivity should be specified in the DQOs.

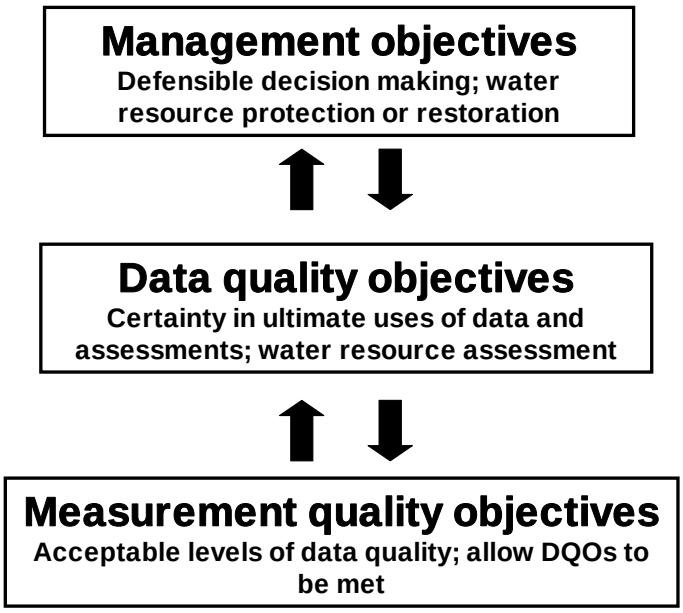


FIGURE 3-2. The relationship among management, data quality, and measurement quality objectives.

3.3 Data Quality Objectives

DQOs are statements of the level of uncertainty that a decision maker is willing to accept in decisions made on the basis of the measurement data (Smith et al. 1988, USEPA 2000b). An example DQO statement by a data user is:

This monitoring program should be able to detect a 20% change in mean biological index score (sensitivity), 80% of the time (power), with 95% confidence (certainty).

From this, or a similar statement, if there is a known or estimated precision value, a power analysis can be performed to help determine how many samples or sites are necessary to be able

to meet the stated DQOs (Osenberg et al. 1994, Urquhart et al. 1998). The greater the variance associated with an indicator, the larger will be the number of samples necessary to detect true change. Figure 3-3 presents the results of a power analysis, which show that 10 samples are necessary to be able to detect a 20% decrease in mean ($\bar{x}$) indicator value, with 95% confidence. How those 10 sites are arrayed throughout the landscape (or watershed) is dependent on the spatial scale of the question to be answered. For example, if one wants to have this level of data quality for three watersheds of different sizes, each watershed would need to be samples at 10 randomly-selected sites, regardless of its size. The key is to ensure that the locations are selected without bias.

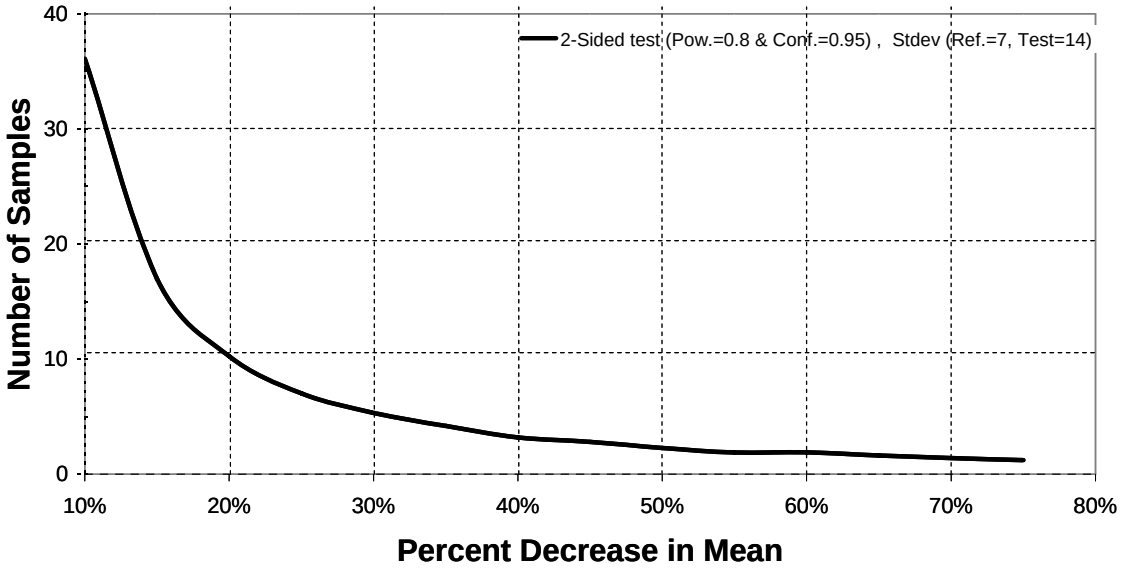


FIGURE 3-3. Results of power analysis showing the relationship between number of samples and the ability to detect differences (or changes) in mean index score (Stribling and Davie 2005). The index tested is multimetric and calibrated for Level 4 ecoregions in the Georgia Piedmont; benthic macroinvertebrate sampling methods are those of the Georgia Environmental Protection Division for wadeable streams.

If a certain level of data quality cannot be assured, then it is possible that the required DQOs cannot be met, with a resulting increased uncertainty (diminished defensibility) in addressing management objectives.

3.4 Measurement Quality Objectives and Performance Characteristics

Data quality is "the magnitude of error associated with a particular dataset" (Taylor, in Keith 1988). Overall error can be segregated into two types: random and systematic. Controlling error in datasets is necessary to ensure that reliable information is available to ecosystem managers and other decision makers. *Random error*, or variability, is error associated with natural variability; efforts to manage this kind of error are focused largely on sampling design such as by definition of temporal strata (e.g., seasonal index periods), stratification of sampling locations

(site classes), and randomized site selection (Figure 3-4). *Systematic*, or method, error results from how samples are taken and processed, and its control is largely through effective QA/QC. Although random and systematic error are often not completely independent (i.e., there is interaction between them in particular measurement systems), they do in some manner individually contribute to the overall variability of the final result. In fact, if some aspect of a sampling design is incorrect and gets implemented, data produced can exhibit substantial systematic error. However, it is possible to partition the potential error sources and use various control techniques to manage the error.

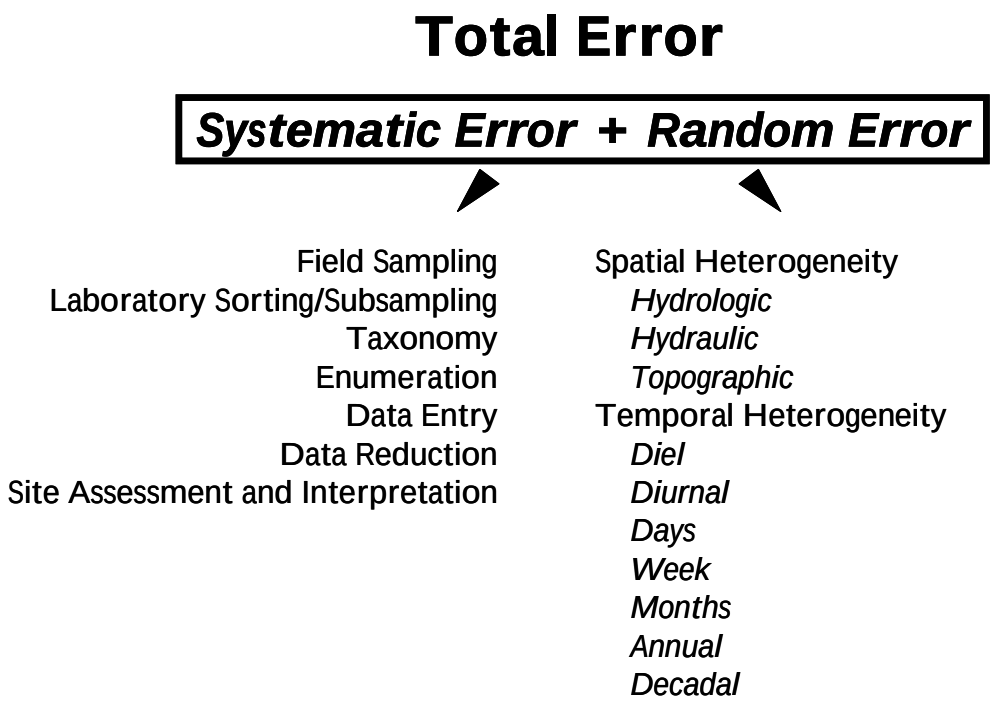


FIGURE 3-4. The overall variability of any measurement system results from both systematic error and random error. In biological assessment protocols, variability results from each step of the process and the spatial and temporal distribution of the samples.

An approach for ensuring that only data of known and acceptable quality are used is to establish and apply measurement quality objectives (MQOs). They can be established for any aspect of the biological assessment process, and MQOs may be quantitative or qualitative. Because a biological assessment protocol is a series of methods (Stribling et al. 2003), it is necessary to either describe the quality of data produced by each method or to assume sufficiency and acceptability. Different indicators require different activities to arrive at the endpoint. For example, the assessment process using the benthic macroinvertebrate assemblage is made up of at least seven methods or activities (see Chapter 6), and the quality of data and information produced by each can affect subsequent activities. Estimates of field sampling precision are directly affected by how the samples are processed (i.e., laboratory sorting, subsampling, and taxonomy). If laboratory activities are not performed at an acceptable level, any discussion of field precision may be meaningless. The magnitude of error that adversely affects a data user's

ultimate interpretation of an endpoint is likely unknown, or at least poorly understood. Routine documentation of data quality at each step of the bioassessment process improves defensibility of the end result. Acceptability of different rates and magnitudes of error is dependent on the needs of the data user. In the respective chapters on assemblage, components of the assessment process are segregated for purposes of defining performance characteristics.

PERFORMANCE CHARACTERISTICS DEFINITIONS

Precision – the nearness of two different measures of the same property (Taylor 1988, Taylor and Kuyatt 1994).

Accuracy – the nearness of a measurement to its true value, or analytical truth (Taylor 1988, Taylor and Kuyatt 1994, Clark and Whitfield 1994); the inverse of bias.

Bias – distance from a known value caused by systematically favoring some outcomes over others (Smith et al. 1988, Clark and Whitfield 1994); the inverse of accuracy.

Representativeness – that a value or entity depicts the property it is intended to depict.

Completeness – a measure of the number of valid data points relative to the planned number of data points (Smith et al. 1988).

Sensitivity – amount of change an indicator can detect relative to an independent variable (such as a disturbance gradient).

Although the importance of different performance characteristics should be determined by the ultimate data user, those data users should understand the potential error source interactions. The performance characteristics most commonly discussed are precision, accuracy, bias, representativeness, and completeness. Others which may be of importance and concern include selectivity and interferences, though they are often thought of as components of bias.

Individual performance characteristics are relevant to some components of the assessment process, but not to others because they may not be applicable. Further, some can be described quantitatively (QN) and others qualitatively (QL). Although there is differential rigor in how these aspects of data quality are communicated, and use of "na" may seem particularly trivial, it may be important. For example, it is important for non-specialists reviewing biological assessments to know that the concept of accuracy is not relevant to field sampling, while it is highly relevant to the final assessment of conditions. The analytical truth for benthic macroinvertebrate field sampling would be all organisms, in totality, present at a site. This value would be impossible to document, even with an enormous sampling effort. Table 3-2 presents formulas and explanations for quantitative performance characteristics. Documenting performance characteristics for a protocol or a program demonstrates the level of data quality that is achievable, and the quality of data associated with a program, project, or dataset.

TABLE 3-2. Formulas and explanations for quantitative performance characteristics.

Relative percent difference (RPD) – field sampling precision

This statistic represents the proportional difference between two measures and is calculated using the equation:

$$RPD = \left(\frac{|A - B|}{A + B} \times 2 \right) \times 100,$$

where *A* is the metric or index value of the first sample and *B* is the metric or index value of the second sample **(**Berger et al. 1996**)**.

Root mean square error (RMSE) – field sampling precision

A kind of generalized standard deviation, this precision statistic is a pooled standard error for a set of *k* group means, usually associated with a one-way ANOVA, and is calculated by:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{\sum df_{1...k}}},$$

where y_{ij} is the i^{th} individual observation in group *j*, *j* = 1…*k* (Zar 1999).

Coefficient of variability (CV) – field sampling precision

This statistic is a unitless measure of precision calculated from the RMSE by:

$$CV = \frac{RMSE}{\bar{\bar{Y}}} \times 100,$$

where $\bar{\bar{Y}}$ is the mean of the dependent variable (e.g., metric, index; Zar 1999).

Detectable difference (DD) – sensitivity of biological metrics, index, or O/E score

The detectable difference of the indicator defines the bracket around the observed mean (of metric, index, or O/E score) within which the true mean will be found with specified confidence, and thus, of the smallest difference between values that is significant. The implicit assumption here is that the frequency of repeat sampling is adequate to provide precision estimates representative of natural variability in the context of the method or protocol being used. Also, since the distribution is unknown, degrees of freedom (df) is set for an unlimited number of samples, or ∞ (Zar 1999). For a 90% detectable difference of a single observation (i.e., p = 0.10), the RMSE value is multiplied by 1.64 (from a standard t-table, e.g., Zar [1999]):

$$DD_{90} = RMSE \times 1.64$$

for 95% detectable difference (p = 0.05), the t-value multiplier is 1.96; and so on. With additional replicate samples, the detectable difference is divided by the square root of the number of replicates:

$$DD_{90}(2-tailed) = (RMSE \times 1.64)/\sqrt{n}$$

TABLE 3-2. (Continued)

Percent completeness (%C) – field sampling, taxonomy, site assessment and interpretation

Percent completeness is a measure of the number of valid samples that were obtained as a proportion of what was planned, and is calculated as:

$$\%C = \frac{v}{T} x100,$$

where *v* is the number of valid samples and *T* is the total number of planned samples. For percent taxonomic completeness, *v* is the number of specimens in a sample that were identified to the target taxonomic level and *T* is the total number of specimens in the sample.

Percent sorting efficiency (PSE) – sorting/subsampling bias

Percent sorting efficiency is calculated as:

$$PSE = \frac{A}{A+B} \times 100,$$

where *A* is the number of organisms found by the original sorter, and *B* is the number of missed organisms recovered (sort residue recoveries) by the QC laboratory sort checker.

Percent difference in enumeration (PDE) – taxonomic precision

Precision of sample counts is determined by calculating percent difference in enumeration by comparing results from two independent laboratories or taxonomists using the formula:

$$PDE = \frac{|n_1 - n_2|}{n_1 + n_2} \times 100,$$

Percent taxonomic disagreement (PTD) – taxonomic precision

Precision of taxonomic identifications is determined by calculating percent taxonomic disagreement by comparing genus-level taxonomic results from two independent taxonomists, using the formula:

$$PTD = \left[1 - \left(\frac{comp_{pos}}{N}\right)\right] \times 100,$$

where $comp_{pos}$ is the number of agreements and *N* is the total number of organisms in the larger of the two counts (Stribling et al. 2003).

TABLE 3-2. (Continued)

Discrimination efficiency (DE) – accuracy of site assessment and interpretation

The accuracy of the Index of Biological Integrity (IBI) and individual metrics is characterized as their capacity to correctly identify stressor conditions (physical, chemical, hydrologic, and land use/land cover) and is quantified as discrimination efficiency using the formula:

$$DE = {}^{a}\!/_{b} \times 100\,,$$

where *a* is the number of stressor sites identified as below some specified acceptance threshold, and *b* is the total number of stressor sites.

3.5 Performance-based Methods Systems

Performance-based methods systems (PBMS) require that acceptable data quality be defined relative to MQOs. Once MQOs are established, any protocol or program producing data meeting those acceptance criteria are acceptable for use. Using a PBMS enhances monitoring programs in that it:

- Provides the means to objectively screen data quality and quantify acceptable measurement error,
- Improves credibility and defensibility of biological assessments,
- Allows for communication of the data quality to secondary user(s), and
- Provides the necessary information for determining comparability among programs, protocols, methods, and data.

The PBMS (Figure 3-5) integrates decisions on the acceptability of data quality with their utility for management decisions (see the website for the Methods and Data Comparability Board of the National Water Quality Monitoring Council (NWQMC) (http://acwi.gov/methods/) for more information on PBMS). If performance characteristics are documented for one program or dataset, it looks similar to what should be routine QA/QC. If documented for two, determination of comparability between the two programs is relatively straightforward (Figure 3-6).

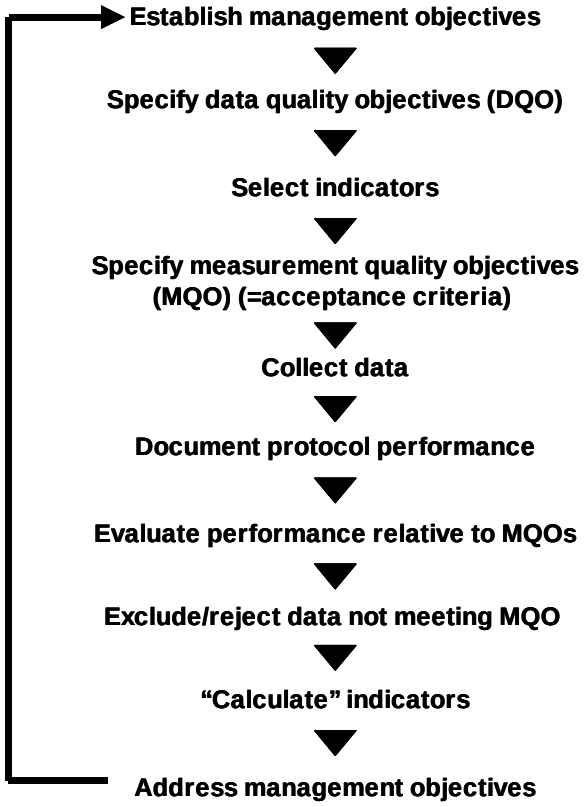


FIGURE 3-5. Use of MQOs and performance characteristics to ensure defensibility of management decisions (USEPA in preparation).

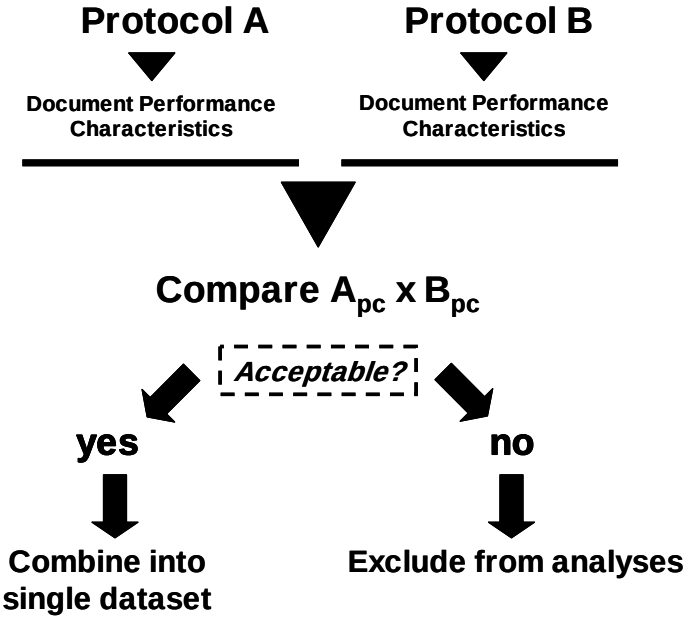


FIGURE 3-6. Framework for analyzing the comparability of multiple biological assessment protocols.

Chapter 4.0 Habitat Assessment and Physicochemical Parameters

with contributions from **JoAnna L. Lessard**[1]

This chapter…

- summarizes a variety of large river habitat assessment approaches

Habitat assessments…

- are important because of their established link to biological impairment
- are less developed than other assessment methods making definitive recommendations difficult
- should include consideration of bank/channel condition, instream habitat, and local and watershed scale disturbances

4.1 Introduction

This chapter provides an overview and summary of selected large river physicochemical and habitat assessment protocols representing a cross-section of field methods currently in use in the USA. *Habitat* refers to all aspects of the physical and chemical environment and the biotic interactions within an ecosystem. In the *Rapid Bioassessment Protocol for Wadeable Streams,* the definition of *habitat* was narrowed to those instream or riparian features that influence the structure and function of the aquatic community (Barbour et al. 1999). Physicochemical habitat condition provides the template for aquatic life and determines what can live in an aquatic system. Habitat diversity explains much of the variation in biological diversity in rivers (Gorman and Karr 1978, Vannote et al. 1980, Raven et al. 1998, Voelz and McArthur 2000). Habitat characteristics are important for classifying streams, identifying disturbance gradients and determining their effects, and are the basis for stream restoration efforts. Altered habitat structure is considered one of the major stressors to aquatic systems that leads to a loss of biological integrity (Karr et al. 1986). Evaluating habitat quality, therefore, is critical to any assessment of ecological condition and should be performed at the same time and location(s) as biological sampling.

This chapter discusses selected monitoring programs that evaluate physical habitat quality using a range of protocols, methods, and levels of effort. Several of the programs assess physicochemical habitat condition to document broad temporal and spatial patterns, collect baseline data, document influences of watershed disturbances, and evaluate general habitat quality, even as an independent indicator of ecosystem condition (e.g., Environmental Monitoring and Assessment Program [EMAP] and National Water-Quality Assessment Program [NAWQA]). These programs generally include more in-depth measurements and characterizations of a broad list of parameters. Other programs evaluate physical habitat quality primarily to describe potential drivers of biological condition, using biological patterns as the final measures of ecological condition, and for making management decisions (i.e., streams are designated as "impaired" based on biological condition). These programs generally measure a reduced number of parameters or use more qualitative and visual-based methods. Protocols developed and used to meet the programmatic objectives are designed so that habitat assessments can be completed at the same time biological samples are taken (e.g., Ohio Environmental Protection Agency's Qualitative Habitat Evaluation Index [QHEI; Rankin 1989], and Michigan

[1] Garcia and Associates, 2601 Mission Street, Suite 600, San Francisco, CA 94110

Department of Environmental Quality's Non-Wadeable Habitat Index [NWHI; Wilhelm et al. 2005]). Physical habitat data resulting from any of these programs can be utilized to evaluate sources of ecological stress, but statistical analysis and predictive models usually require more quantitative data. When beginning any river assessment program, the objectives of the program must be clear so that the data collected will be of the quality needed (i.e., appropriate precision) to meet those objectives. Future uses of the data should also be considered because post-hoc and metadata analyses cannot be made without data for important habitat features.

The assessments performed by most water resource agencies include a general description of the site, a physical characterization and water quality assessment, and a visual assessment of instream and riparian habitat quality (Barbour et al. 1999). These data, along with quantitative measurements of select physical parameters, provide an integrated picture of many of the factors influencing the biological condition of a system. These assessments, however, have mostly been developed for medium-to-high gradient wadeable streams (Wang et al. 1998, Barbour et al. 1999, Wilhelm et al. 2005). The methods, metrics, and criteria for the physical and chemical habitat assessment of large rivers are still in the early stages of development and use; and we expect the protocols to evolve over time. However, similar to wadeable streams, evaluating the habitat quality of large rivers can be accomplished by characterizing selected physicochemical parameters in conjunction with a systematic assessment of physical structure. Through this approach, key features can be measured, rated, or scored to provide a useful assessment of habitat quality.

Due to their extensive drainage basins, large rivers are often highly impacted by the cumulative effect of all upstream activities (e.g., land use and point-source discharges), which commonly lead to chemical and organic pollution. These larger systems also tend to have long histories of physical habitat alteration from dams, diversions, land use, and channelization. Large rivers, therefore, often present a truncated gradient of conditions skewed toward the impacted or degraded condition. These issues illustrate the need to adequately sample and characterize physicochemical habitat condition.

State and federal agencies have developed physical assessment protocols intended for use in rivers (Table 4-1). This chapter summarizes specified habitat methods for five programs, which were selected because they represent a cross-section of current large river methods. This chapter does not recommend a specific habitat assessment protocol primarily because habitat methods for large rivers are less developed than other large river assessment elements, and there is a lack of consensus on the most suitable approach. Ultimately, the selection of an approach will depend on whether the principal objective of users is to: 1) thoroughly characterize the physical habitat of large river reaches as the primary indicator of ecological conditions, or 2) in concert with biological surveys, to characterize only those physical habitat elements most likely contributing to the capacity of a river to support the survival and reproduction of the biota. The approach in this document, though it is still being refined, is presented as a compromise between these two options. The other approaches reviewed are presented as examples to allow flexibility in program design, depending on user needs. Parameters considered critical for assessment of habitat condition are discussed and described, including those common to most stream assessments as well as those thought to be more important for large river systems. EMAP is used as a case study example in the following text box.

Program Highlight

Environmental Monitoring and Assessment Program-Surface Waters (EMAP): Physical Habitat Characterization for Non-wadeable Rivers

This program focuses on evaluating ecological conditions on regional and national scales. The habitat assessment protocol describes procedures for collecting both quantitative and qualitative data about environmental measures or attributes of indicators of non-wadeable stream physical habitat and ecosystem condition. These procedures were developed based on standard or accepted methods, modified as necessary to accommodate EMAP sampling requirements; and the entire protocol was intended for use in field studies sponsored by EMAP. EMAP studies a proportional reach length of 100X the wetted width (western rivers) or 40X the wetted width (eastern rivers). Within each reach, EMAP samples or characterizes physicochemical habitat both longitudinally and at 11 equidistant cross-sectional transects. There are seven general physical habitat attributes used for EMAP non-wadeable river assessments: channel dimensions, channel gradient, channel substrate and type, habitat complexity and cover, riparian vegetation cover and structure, anthropogenic alterations, and channel-riparian interaction (Kaufman 1993, 2000). Expected values, however, change with stream size.

Components of Physical Habitat Assessment:

1. Thalweg Profile: At 10 equidistant places between each transect, record habitat type, presence of back-water or side-channel habitat, characterize substrate type; at 20 equidistant places between each transect, tally mid-channel wood snags and place in size classes, measure maximum thalweg depth.
2. Littoral/Riparian Cross Section: At each of the 11 transects measure/estimate from one chosen bank, gradient between transects, bearing between transects, wetted-width, mid-channel bar width, bankfull width and height, incision height, bank angle, riparian canopy cover in four directions from bank, shoreline substrate between water and 1 m up the bank.
3. Littoral Plots: At each transect, at the chosen bank, delineate 20 x 20-meter littoral plots that the water and transect line bisect.
   3a. In the wet half of these plots, determine littoral depth at five places, dominant and subdominant substrate size; tally large wood by size class and other fish concealment habitat.
   3b. In the dry half of these plots, determine estimate of areal cover by class and type, record and tally human disturbance types and their proximity to the channel, and estimate distance from bank, the diameter, height, and species of largest riparian tree.

Components of Water Chemistry Sampling:

1. Field Measurements: In an area of flowing water in the channel, take field probe measurements of specific conductance, dissolved oxygen, and water temperature.
2. Samples for Lab Analysis: At the same place as field measurements, obtain a 4-L cubitainer water sample and place on ice and in the dark. Also obtain two, 60-ml syringe samples of river water and then place on ice and in the dark. The cubitainer sample is analyzed for major ions, nutrients, iron, manganese, turbidity, and color. The syringe samples are analyzed for pH and Dissolved Inorganic Carbon. Water samples must be delivered overnight to analytical laboratory for analysis (syringe) or stabilization (cubitainer).

TABLE 4-1. Major large river program habitat approaches. Detailed program method comparisons are provided in the following sections.

Program	Protocol Summary	Citation
Primary objective: characterizing long-term spatial and temporal patterns in habitat condition as its own independent indicator of ecosystem condition.		
USEPA EMAP-Surface Waters *(Summarized in Program Highlight box)*	National and regional program for characterizing status and trends on ecological condition. Characterize seven general physical habitat attributes: channel dimensions, channel gradient, channel substrate size and type, habitat complexity and cover, riparian vegetation cover and structure, anthropogenic alterations, and channel-riparian interaction. Primarily quantitative measures along 11 transects in reaches 100X or 40X the wetted width. Estimated level of effort: 2 people, 2 day per reach.	Kaufmann 2000
USGS NAWQA	National program to characterize water quality condition and develop an understanding of factors influencing quality. Quantitative measures taken to characterize habitat at 4 hierarchical scales: basin, segment, reach, and microhabitat. At reach scale, primarily quantitative measures along 11 transects in reaches 20X stream width within 500-1000 m lengths. Estimated level of effort: unknown, probably 1 reach per day.	Fitzpatrick et al. 1998
Primary objective: evaluating habitat to understand biological condition.		
Large River Bioassessment Protocol (LR-BP)	Scaled down version of the EMAP program protocols. Characterize 6 of 7 EMAP attributes: channel dimensions, channel substrate size and type, habitat complexity and cover, riparian vegetation cover and structure, anthropogenic alterations, and channel-riparian interaction. Reach length set to correspond to biotic assemblages being sampled. Semi-quantitative measures from six transects. Still in development. Estimated level of effort: 1 person, a – 2 day per reach.	Blocksom and Flotemersch 2005, Flotemersch and Blocksom 2005
Non-Wadeable Stream Habitat Index (NWHI)	A multimetric index developed for characterizing habitat in Michigan non-wadeable streams and rivers. Features used in index include: riparian width, large woody debris, aquatic vegetation cover, sediment deposition, bank stability, substrate size, and off-channel habitat. Primarily quantitative measures along 11 transects scored as metrics within 2,000-m reaches. Estimated level of effort: 2 people, 1 day per reach.	Merritt et al. 2005, Wilhelm et al. 2005
Qualitative Habitat Evaluation Index (QHEI)	A multimetric index developed for characterizing habitat in Ohio streams. Composed of six variables: substrate, instream cover, channel morphology, riparian zone and bank erosion, pool/glide and riffle/run quality, and gradient. Primarily qualitative scoring of metrics over the entire 500 m length of study reach. Estimated level of effort: 1 hour per reach.	Rankin 1989

4.2 Site Location and Other Descriptive Information

Site location and other logistical or geographical details are often very important for site relocation and data interpretation and sometimes function as metadata for other analyses. All physical habitat data sheets should include identical header information sufficient to identify the station and location where the survey was conducted, date and time of the survey, and the name(s) of the investigator(s) responsible for the quality and integrity of the data. The river name and basin should identify the watershed and tributary sampled; the location of the station should be described in a narrative to help identify specific locations and access to the station for repeat visits. The river kilometer (RKM), if applicable, and latitude/longitude are examples of specific locational data for each station. Data sheets should include a section for notes on the weather conditions on the day of the survey and immediately preceding the survey. This information is important to interpret the effects of storms on the sampling effort. A photograph can be very helpful in identifying station location and documenting habitat conditions. Any observations or data not requested but deemed important by the field observer should be recorded. A hand-drawn map is also useful to illustrate other features such as major landmarks, vegetative zones, and buildings that might be used to aid in data interpretation. Record the origination type (such as glacial, montane, swamp, and bog) of the stream, if known. As the size of the river increases, a mixture of tributary origins is likely.

4.3 Sample Reach Characterization: Transects

Physical habitat has components that change both longitudinally (e.g., sinuosity, gradient) and laterally (e.g., cross-sectional shape and substrate). Each of the highlighted protocols sample both the longitudinal and cross-sectional characteristics of each sample reach (except the QHEI, which characterizes habitat within the reach as a single index score based on a visual assessment). Four of the protocols specify that 11 equidistant transects be delineated along the entire reach. Simonson et al. (1994) found that 11 transects in a sample reach was sufficient to achieve approximately 80% accuracy of estimated mean values of fish habitat characteristics. The EMAP, MDEQ, and NAWQA protocols all use 11 transects for physical habitat assessment. The LR-BP protocol pares the EMAP methods to six transects within a 500-m reach (see Blocksom and Flotemersch 2005).

4.4. Channel and Bank Characteristics

4.4.1 Water Depth

Depth profiles are used to characterize pools, river size, channel complexity, and proportions of habitat types (e.g., riffle, run, pool) (Kaufmann 2000). Water depth is generally measured longitudinally along the thalweg (i.e., flow path of the deepest water) or laterally across each transect. For large rivers, all protocols recommend depth measurements be made with a depth pole, sounding rod, or Sonar. The EMAP protocol specifies detailed thalweg maximum depth measurements be made at 20 or 10 equally spaced intervals (for 100X or 40X reach lengths, respectively) between each of the 11 transects during the downstream float of the reach. EMAP also specifies 5 equally spaced littoral zone depth measurements be made at each transect within 20 x 10-m littoral plots at each bank (~1 m from bank). The MDEQ and LR-BP protocols

specify a similar longitudinal approach with measurements made every 40 m or 25 m, respectively along the reach. The LR-BP also specifies that cross-sectional depth measurements be collected for discharge calculations at three points along the reach (0 m, 250 m, and 500 m). NAWQA specifies depth measurement at each transect.

4.4.2 Wetted and Bankfull Width

Width characteristics provide information on the channel type (e.g., constrained vs having a broad floodplain) and stream size, which sets important boundaries for biological interactions and riparian influences in rivers. Wetted width is particularly important for habitat assessments because it is used as the multiplier to set the sampling reach length and is used to calculate a width:depth ratio, which can be used as an index of channel shape (Merritt et al. 2005). Bankfull width indicates the boundary between normal high flows and flood stage, and helps characterize the dynamics of the channel and the extent of the floodplain. All protocols recommend using a laser rangefinder for channel width measurements. For transect-based methods, width measurements are made at each transect along the sample reach. For large rivers and streams, where channels are more easily delineated and not obscured by vegetation, channel width and floodplain extent can often be characterized and estimated using remote sensing techniques (with appropriate ground truthing and calibration; see Section 4.6). Where available, these resources should be utilized, especially for programs with assessment objectives across large geographic areas.

4.4.3 Sinuosity

Sinuosity (i.e., channel curviness) describes energy conditions, habitat diversity, and is also related to gradient (i.e., lower gradient rivers tend to have more sinuosity) (Fitzpatrick et al. 1998). The NAWQA protocol calculates sinuosity from the ratio of curvilinear length of the reach (estimated using a map wheel or GIS) to the segment valley length (straight line distance between upper and lower reach boundaries). The MDEQ protocol uses a mean width:depth ratio from the 11 transects as an indicator of channel shape (Merritt et al. 2005). The QHEI scores sinuosity by visual estimates in the channel morphology metric (Rankin 1989).

4.4.4 Gradient

Gradient is an indicator of the energy available for water and sediment movement through a reach, which controls the types of habitat that will be present in a river system. Therefore, it is an important determinant of distributions of stream organisms. Because of the difficulty of measuring this variable, scientists often estimate it using maps or the elevation differences between dams. The NAWQA protocol specifies measuring the upstream and downstream elevation (using a map, GIS or GPS unit), subtracting the two (upstream-downstream) and dividing this number by the segment valley length of the reach (Fitzpatrick et al. 1998). EMAP specifies measuring upstream and downstream elevation change with a clinometer or Abney level between each transect (Kaufmann 2000). The MDEQ and QHEI protocols estimate gradient from topo maps, digital elevation models (DEMs), or a GIS. It is important to note that for low-gradient and impounded rivers, the gradient is essentially zero when measured using

tools such as a clinometer or Abney level because they cannot detect very small changes in elevation.

4.4.5 Bank Characteristics

4.4.5.1 Bank Stability

Bank stability is an important indicator of both past and present disturbance and can often be used to predict future problems. Unstable banks add sediment to rivers and do not offer the structural or functional services that stable, vegetated banks provide to instream organisms. Bank stability is a function of bank angle, height, substrate type, and vegetative cover. The NAWQA protocol specifies using a clinometer to measure or estimate bank angle, a surveyor's rod and level to measure or estimate bank height, and visual estimates of substrate and vegetative cover at each transect, all of which are combined into one bank stability index score. Simpler protocols include estimating bank angle and height at each transect by comparing it to a categorical chart on the field sheets (EMAP) or visually estimating it (LR-BP) and scoring bank stability along the entire reach (MDEQ, QHEI).

4.4.5.2 Riparian/Floodplain Condition

Other than stability, lateral bank characteristics include the condition and characteristics of the riparian zone and floodplain. Floodplains are the lateral low-land areas adjacent to streams that include what is typically considered the "riparian area", but the true floodplain region often extends laterally far beyond what is normally evaluated during stream assessments. Floodplain zones are typically delineated by the temporal flooding pattern of the area (e.g., seasonal wet-weather floodplain, 10-year floodplain, 100-year floodplain), while riparian assessments often focus on areas delineated by some predetermined distance. Riparian condition measures include the vegetated width (i.e., buffer width), longitudinal continuity, and substrate and vegetation type. Floodplain characteristics beyond the "buffer", however, are also vital to the ecological function of rivers and are not typically a part of monitoring programs. Floodplains are very important for hydrological control, inorganic transport and storage, nutrient dynamics, and processing and transport of organic matter (Junk et al. 1989, Craft et al. 2002, Mouw et al. 2003, Poole et al. 2002). In addition, many organisms rely on floodplain inundation during important life history stages (Junk et al. 1989). Documentation of floodplain extent, vegetated cover, disturbance, and temporal patterns of inundation would contribute a great deal to habitat assessment programs aimed at understanding the drivers of ecological condition of large rivers. All of these features provide information on likely stressors that may be influencing instream organisms and also identify targets for riparian and floodplain restoration efforts.

All protocols discussed in this chapter visually estimate riparian parameters either at each transect within 10 x 20-m landward plots (i.e., assumes 20-m buffer is adequate) [EMAP, LR-BP, MDEQ], along 30-m lateral extensions of transects (i.e., assumes 30-m buffer is adequate) [NAWQA], or along the entire reach (riparian width is scored) [QHEI]. All protocols score or evaluate riparian areas higher or "in better condition" when they are wider, continuous, stable, and dominated by dense native vegetation. In many alluvial southern floodplain rivers, shifting sands, unstable substrates, unstable sliding banks, and unvegetated bare ground are all signs of a

naturally functioning riparian system. The biota inhabiting these sandy rivers are generally tolerant omnivores that thrive under such conditions. Overly stable banks, in this setting, would therefore be a sign of degradation, which illustrates the need to not only include both riparian and floodplain condition measures in assessments, but also to anchor the evaluations to a regional reference condition. All habitat metrics must be calibrated to local or regional reference sites before determining how to score habitat features that vary regionally. Programs may also include criteria for native vegetation in riparian and floodplain condition estimates, especially in areas where invasive species are a problem.

The Center for Environmental Research in Germany has studied the potential for using biological indicators for floodplain assessment in a report called "Development of a Robust Generally Applicable Indicator System for Ecological Changes in Floodplain Systems" (available online at www.ufz.de/index.php?en=1770). Floodplain hydrology models are also becoming more commonly used to predict patterns of inundation and water storage in floodplains. More work needs to be done on the relationship between biological condition of rivers and floodplain characteristics, especially for large rivers where floodplain connectivity is likely much more important. In developing an assessment protocol for large rivers, it is advisable to incorporate some sort of floodplain evaluation procedure to augment more formalized habitat assessments.

4.4.6 Channel Alterations (Unnatural Disturbance)

Characterizing direct channel modifications provides an important historical perspective on the anthropogenic disturbances to which a river system has been exposed. This information helps put many other habitat characteristics into perspective and will influence expectations for the biota. These activities are generally listed as comments in the site characterization and often influence the scores of other bank/channel parameters (e.g., channelization, riparian vegetation removal, logging) (MDEQ and LR-BP). The EMAP protocol records the presence/absence of 11 categories of disturbance at each transect. The NAWQA protocol includes noting any unnatural disturbances but specifically recommends noting water management activities or hydromodifications. The QHEI protocol includes human activities in the scoring of its riparian condition metric.

4.5 Instream Habitat

Instream habitat refers to the physical, chemical, and biological attributes within a stream channel that influence the structure and function of the aquatic community.

4.5.1 Physical Characteristics

As a subset of instream habitat, the term physical characteristics specifically refers to the types and distribution of physical habitat features present in a channel. Documentation of these features provides data for classifying streams, identifying disturbance gradients and determining their effects, and helping guide stream restoration efforts.

4.5.1.1 Geomorphic Channel Units (GCUs)

Geomorphic channel units (e.g., riffles, runs, pools, etc.) are fluvial geomorphic habitat types that describe scouring, channel shape, and overall habitat patterns in streams. The abundance and distribution of GCUs are noted, mapped, or tallied in each of the protocols. The QHEI protocol uses GCU abundances in its habitat diversity/quality score. NAWQA also uses them to establish the reach, unless the units exceed a maximum length of 1000 m.

4.5.1.2 Discharge

Flow regime affects biological condition and other instream factors (e.g., habitat structure, water quality) (Poff et al. 1997). Species distributions, abundances, and competitive interactions all rely upon natural flow regimes (Poff and Allan 1995, Reeves et al. 1995, Greenberg et al. 1996, Poff et al. 1997). In-situ discharge measurements can be difficult in large rivers, require cross-sectional depth and water velocity measurements (made with a current meter) (e.g., LR-BP calculates discharge at three places: top, bottom, and middle of reach), and only give an indication of the discharge on that day. It is preferable, if a gauging station exists nearby, to obtain these data and calculate mean annual discharge, 50% exceedence flow, and an estimate of flow variability (NAWQA, MDEQ). NAWQA does not require flow characterization at non-wadeable sites if no USGS gauging station exists for that stream. Many large rivers have gauges along them and these data can be used to simulate discharge at different sites along the stream using area-weighted adjustments. In addition, hydrologic models can be used to simulate hydrology for a site and can be calibrated using nearby gauge data. These simulated data can then be analyzed for standard measures of hydrologic behavior (e.g., mean flow, 7Q10, flow duration, flood frequency, etc.)

4.5.1.3 Substrate Size

Substrate is influenced by geology, climate, topography, and disturbance. Mountainous rivers are naturally characterized by fundamentally different substrate patterns than coastal or alluvial rivers. Substrate is an important habitat feature for benthic organisms because it influences habitat stability, interstitial habitat quality, refugia, and nesting habitat. Measurements of substrate types or sizes are, therefore, important components of physical habitat assessment. The MDEQ adopted EMAP's method of characterizing substrate types visually or by feel with a pole in the thalweg along the entire reach and in littoral plots at each transect. The QHEI scores substrate by the two most dominant types in the reach. The LR-BP uses substrate characterizations to guide macroinvertebrate sampling (Chapter 6, and Blocksom and Flotemersch 2005). NAWQA does not require substrate sampling at non-wadeable sites.

4.5.1.4 Embeddedness

Embeddedness is a measure of the percent of substrate (gravel-sized or larger) surface area covered by sand or finer particles. Embeddedness is another measure of substrate condition and is an important indicator of disturbance and potential stressors to benthic organisms. The MDEQ protocol specifies estimating the percentage of the wetted width covered in silt at each transect.

The QHEI similarly estimates coverage of silt and scores the reach by the percentage covered. NAWQA does not require embeddedness estimates.

4.5.1.5 Large Woody Debris (LWD) (i.e., Snags, Root Wads, Down Trees, etc.)

Large woody debris (LWD) offers habitat for attachment, feeding, and cover for stream organisms. Wood can be the only stable substrate in naturally sandy waters or those with high siltation problems. Instream wood habitat is related to the production of both macroinvertebrates and fish (Angermeier and Karr 1984, Benke et al. 1985, Lisle 1986, Dollof and Warren 2003). The EMAP, NAWQA, LR-BP, and MDEQ protocols specify tallying LWD habitat pieces within the channel during the downstream float of the thalweg and transect sampling. The EMAP also places LWD into length and diameter classes to estimate surface area. The QHEI includes LWD presence in the instream cover metric.

4.5.1.6 Aquatic Vegetative Cover

After wood, instream vegetation is the next most important organic source of stable substrate for attachment, feeding, and cover. In large rivers, where depths, clarity, and flows often do not allow for vegetated growth near the center, there can be extensive littoral areas which provide essential marginal habitat for stream organisms. Similar to embeddedness, vegetative cover is primarily estimated in terms of areal coverage. The EMAP, LR-BP, and MDEQ estimate vegetative cover in littoral plots at each transect. The NAWQA records presence/absence of cover of any type (mineral or vegetative) at 22 points along the shoreline. The QHEI specifies a visual reach-wide estimate of all fish coverage, including vegetation.

4.5.1.7 Riparian Cover

This parameter refers to the amount of the stream channel influenced by the shade of riparian vegetation, mainly trees and shrubs, and is an important feature for organisms in that it moderates water temperatures and provides habitat. Riparian vegetation is also an important food source and the level of riparian cover will influence the abundance of organisms that rely on allochthonous resources. NAWQA takes two measurements of riparian cover at each transect, open canopy angle and riparian canopy closure. Open canopy angle is measured (with a clinometer or compass) from the center of the channel to the tallest object on each bank. These angles are subtracted from 180 to give the open angle. Proportional estimates can be calculated by dividing the open angle by 180 (Fitzpatrick et al. 1998). Riparian canopy closure is measured with a spherical densiometer at each transect (Fitzpatrick et al. 1998). EMAP also uses a spherical densiometer and measures canopy cover at each transect in four places: left bank, right bank, upstream, and downstream. The LR-BP, MDEQ, and QHEI protocols visually estimate and score riparian cover, but the MDEQ and LR-BP protocols make estimates at each transect.

4.5.1.8 Off-channel Habitat

Off-channel habitats (e.g., backwater areas) are often important spawning and nursery habitat as well as refugia during high flow (Merritt et al. 2005). All protocols note, map, or tally off-channel habitat within the sample reach. The MDEQ protocol has a metric that is scored by the

number of off-channel habitats that exist in the reach. The QHEI scores the presence of these areas in the instream cover metric. The presence of off-channel habitats is more likely to be documented by study designs using sampling reach lengths based on repeating geomorphic units or multiples of the wetted width (e.g., EMAP). Remote sensing techniques can add information on the location, number, and types of off-channel habitats if the resolution is high enough.

4.5.1.9 Temperature

Although not specifically discussed in any of the protocols except EMAP, water temperature measurements are commonly made during habitat assessments. Biological communities inhabiting coldwater rivers are markedly different from those in warmwater rivers, and many states have established temperature criteria for each type. Temperature should also be measured if dissolved oxygen is being measured because these two parameters are related. Temperature can be measured with a thermometer, temperature meter, or temperature field logger. Logger measurements characterize daily temperature fluctuations, annual mean temperatures, and seasonal extremes; whereas point measures on the day of sampling are primarily useful for locating thermal pollution or calibrating dissolved oxygen readings. They are less helpful for interpreting organism data.

4.5.2 Chemical Characteristics

Water quality is an essential component of habitat quality and must be assessed along with physical habitat condition to make sense of biological trends and to aid stressor identification. Chemical characteristics include all dissolved constituents which influence pH, conductivity, trophic status, and toxicity. EMAP outlines protocols for water sample collection for detailed laboratory analyses and also in-situ sampling of certain water quality parameters. The LR-BP collects river water samples as outlined in Kaufmann (2000). Detailed water quality analyses suggested by the EMAP program are acid neutralizing capacity (pH), nutrient enrichment or dissolved inorganic carbon (for trophic condition), chemical stressors (nutrients, cations, anions, iron, and manganese), and classification of the water chemistry type (Herlihy and Hendricks 2000, Lazorchak et al. 2000). Water quality studies conducted in the NAWQA program are extensive and focus on assessing physical and chemical characteristics of stream water, including suspended sediment, dissolved solids, major ions and metals, nutrients, organic carbon, and dissolved pesticides, and on relating these characteristics to hydrologic conditions, sources, and transport (see Shelton 1994 for more details and Tables 4-2 and 4-3). This section only describes the sampling included in most protocols.

4.5.2.1 Dissolved Oxygen and Conductivity

Dissolved oxygen is used (with temperature) to determine, in part, suitability of the habitat for biota. Specific conductance (i.e., conductivity) is a measure of the capacity of the water to move an electrical current and is related to ionic strength (many ions can be stressors). In-situ measurements of dissolved oxygen and conductivity are easily obtained with a field meter or data logger containing the appropriate probes. Care must be taken, however, to calibrate the meter and check the probes and membranes regularly, ideally before each field day. Point measures of these factors provide limited information for management or biological analyses, but

they can help indicate where a problem may lie in order to guide more intense sampling efforts or study. Diel dissolved oxygen data can be collected by deploying dissolved oxygen probes with logging capacity and used to examine oxygen behavior over 24-hour periods. This can identify oxygen sags, which typically occur in early morning before most field crews sample. Low dissolved oxygen, even for short periods of time, can be stressful for many taxa.

TABLE 4-2. Analytical strategy for basic fixed sites in NAWQA (Shelton 1994).

Field measurements
Dissolved oxygen, pH, and Alkalinity
Specific conductance (consider hourly)
Temperature (hourly for 1 year)
Laboratory analyses
Total Suspended Sediment
Major constituents: Dissolved solids, Major ions and metals (Calcium, Chloride, Fluoride, Iron, Magnesium, Manganese, Potassium, Silica, Sodium, Sulfate),
Nutrients: Nitrogen (Total,Total dissolved, Ammonia, Nitrite, Nitrate), Phosphorus (Total, Total dissolved, Ortho), Organic carbon (Suspended, Dissolved)

TABLE 4-3. Analytical strategy for intensive fixed sites not required by the basic fixed site analyses in NAWQA (Shelton 1994).

Field measurements
Specific conductance (hourly or daily for 1 year)
Laboratory analyses
Dissolved Pesticides:
Amides, Carbamates, Chloropheoxy herbicides, Dinitroanalins, Organochlorines, Organophosphates, Pyrethroids, Triazine herbicides, Uracils, Ureas
Miscellaneous
Actifluorfen, Dicamba,1-Napthol, Bentazon, 2,6-Diethylanaline, Norflurazon, Bromoxynil, Dinoseb, Picloram, Chloramben, DNOC, Propargite, Clopyralid, Esfenvalerate

4.5.2.2 Nutrients

Nutrient concentrations (nitrate, nitrite, ammonium, total phosphorus, ortho-phosphorus, micronutrients, etc.) are important indicators of human disturbance and trophic status of rivers. EMAP procedures specify sampling 4 L of bulk water that is kept cold and shipped overnight to an analytical laboratory. Shelton (1994) provides detailed descriptions of various samplers, sample techniques, storage, and QA/QC directions.

4.6 Remote Sensing Applications for Habitat Assessment

Remote sensing refers to data on the spectral qualities of objects gathered by sensors located some distance from those objects. For habitat assessments, remotely sensed data usually are gathered from one of three sources: interpretation of satellite images, aerial photographs, or infrared photographs. The collected images, which are used to build GIS databases of watershed land use and land cover data layers, can also be used to measure many instream habitat parameters. Remote sensing technology is increasingly being used by scientists to collect data and analyze environmental parameters at much smaller scales. The characteristics of the images (i.e., resolution, spatial coverage, temporal relevance, and spectral range) determine the utility of the image for gathering habitat and watershed data (Faux et al. 1998, Legleiter et al. 2004, Boivin et al. 2005). Remotely sensed information can be particularly useful for large river habitat assessment because the size of such systems is more conducive to broad spatial analyses. Habitat features that have shown potential for this type of analysis are: channel width, stream shape and sinuosity, sedimentation and sediment grain size, riparian and catchment vegetation patterns (type and coverage), watershed land use and land cover, riparian corridor width and extent, type and extent of off-channel habitats (e.g., floodplain, wetlands, and side channels), aquatic vegetation type and coverage, water temperature and other watershed disturbances (Faux et al. 1998, Mertes et al. 1993, Mertes 2002, Poole et al. 2002, Whited et al. 2003, Charbonneau et al. 2004, Lymburner et al. 2004, Legleiter et al. 2004, Boivin et al. 2005). Remote sensing should be used to augment field measurements, or in some circumstances even replace field measuring. For example, Forward Looking Infrared (FLIR) has been used as an efficient and inexpensive tool for monitoring stream temperatures at the watershed scale and even to individual habitats in western streams (Norton et. al 1996, Faux et al. 1998). It also holds promise for use in detailed sub-meter accuracy channel morphology on a watershed scale.

Laser imaging detection and ranging (LIDAR) is another remote sensing technology that provides precise and accurate topographic resolution. LIDAR is being used by fluvial geomorphologists and will likely become a valuable and efficient tool for accurate characterization of river floodplain and channel geomorphology, including depth and width profiles. It can also be used to accurately measure riparian characteristics, including tree height, biomass, density, and leaf area. Clearly this technology holds great promise for large river habitat assessment.

There are some constraints to obtaining these data for programs without a GIS expert on staff or access to remote sensing images. Over time, however, these barriers will be reduced and inclusion of such data where available is encouraged. Tools currently available for characterizing stream and watershed characteristics are: Analytical Tools Interface for Landscape Assessments (ATtiLA), National Land Cover Database (NLCD) and aerial photographs. Regions and programs with well developed GIS capabilities should plan to use these data for habitat assessment purposes as much as possible. Remote sensing is an important additional tool that will allow efficient, safe, and inexpensive characterization of many important habitat features over large spatial scales. As the technology of these tools improves, their applications to large river assessment programs will undoubtedly increase.

4.7 Unnatural Disturbances

Historically, pollution and hydrological modifications were the dominant disturbances to fresh waters. These problems continue today and also include extensive transformations of the landscape including mining, forest harvest, agriculture, urbanization, industry, and recreation, which have resulted in a wide variety of environmental impacts (Richter et al. 1997, Bryce et al. 1999). Over the last 30 years, legislation and new technologies have led to progress in treating point sources. Notably less successful are efforts to address diffuse pollutants and non-point sources, which have become the dominant inputs to river ecosystems and are extremely difficult to manage (Smith et al. 1997). The transition from undisturbed to human-dominated landscapes has altered ecosystems on a global scale and made the quantification of land use/land cover a necessary component to any study of ecosystem condition (Meyer and Turner 1994, Vitousek et al. 1997, Carpenter et al. 1998). The primary human induced changes to large rivers fall into three categories: land use alterations, direct hydrological changes, and channel modification. Within each of these categories are several human activities that have been linked to stream degradation. Assessments of large rivers, therefore, should include at least a cursory survey of the disturbance history of the waterbody so that changes in habitat leading to stressors of ecological condition can be linked to their sources. The stressor identification process and the development of stressor-source relationship models are necessary first steps in developing restoration plans.

4.7.1 Land Use Alterations

Agriculture and urban development have long been linked to physical habitat degradation of streams (Richards et al. 1996, Roth et al. 1996, Wang et. al. 1997, Allan 2004). There are hundreds of studies that document statistical relationships between land use and measures of stream condition (Allan 2004). The extent of land use transformations nationwide is substantial. For example, agriculture is the dominant land use in many large river watersheds in the USA – the area of six major hydrologic units (the Lower Mississippi, Upper Mississippi, Southern Plains, Ohio, Missouri, and Colorado) are more than 40% agriculture (Allan 2004). Due to the recognition of the importance of watershed land cover to ecological condition, many programs include watershed analyses to evaluate causes of stream habitat degradation. Wilhelm et al. (2005) developed the non-wadeable stream habitat index (NWHI) based on scores of seven habitat variables (riparian width, LWD, aquatic vegetative cover, embeddedness, bank stability, thalweg substrate, and off-channel habitat). These variables were selected for the NWHI because of the strong relationship with catchment and riparian disturbance gradients. As discussed above, immediate access to land cover data may be limited for some programs, but the need to include such data in river assessment must be recognized and set as a programmatic goal.

4.7.2 Hydrological Modification

Confounding the effects of land use is the extent of direct hydrological modification to stream and river ecosystems. When irrigated agriculture and hydropower are common, dams and diversions convert rivers to eutrophic impoundments and alter hydrologic behavior. Such alterations may create ideal pond habitats for alien invasive species, form impassable barriers to migration, reduce channel complexity, or eliminate some aquatic environments altogether.

Altered stream flows are well known to be associated with poor channel habitats, erosion, bank instability, and lower base flows (Poff et al. 1997). Species distributions, abundances, and competitive interactions all rely upon natural flow regimes (Greenburg et al. 1996, Poff and Allan 1995, Poff et al. 1997, Reeves et al. 1995). Ecological perspectives on human disturbance and biological responses require consideration of how human actions directly and indirectly affect stream and river channels and flow (volume, duration, fluctuations, and timing). Dynesius and Nilsson (1994), Graf (2001), and Reisner (1986) offer excellent summaries of how human water management practices have fundamentally altered rivers. The location, dam characteristics, and impoundment features, where available, should be included as a data layer for GIS development and also taken into account during habitat assessments.

4.7.3 Channel Modification

Many human disturbances alter channels indirectly by causing excessive sedimentation that fills pools, increases bank steepness, and reduces habitat complexity. In addition to those indirect impacts on channels, humans directly alter channels through dredging, wetland and floodplain draining/filling, channel straightening, and even active channel filling and development. These activities are used for navigation, flood control, and near-stream and shoreline development, all of which are typically accompanied by additional habitat stressors during their implementation. While information on the time-since-alteration, extent of alterations, frequency of impact (e.g., annual dredging), etc. may require some research, they are ancillary data that may be very helpful in understanding habitat patterns and setting restoration objectives.

Chapter 5.0 Algae

with contributions from **Lei Zheng**[1]

This chapter…

- presents methods for sampling periphyton and phytoplankton
- recommends a periphyton-based approach

Algae are…

- a basal food resource for much of the riverine food web
- important biological indicators
- the most responsive indicators for nutrients

5.1 Introduction

Algae are a highly diverse group of photosynthetic organisms with unicellular reproductive structures. They have important functions in aquatic habitats as producers of organic matter and play a vital role in inorganic nutrient retention, transfer and cycling (Stevenson 1996). Large bodies of freshwater, such as large rivers, are usually dominated by diatoms, which are generally referred to as microalgae. The degree to which components of the algal assemblage are used in bioassessment and monitoring programs across the country varies. Diatoms, for example, are widely used as indicators, whereas cyanobacteria (commonly referred to as blue-green algae) and green algae are only occasionally used. This is in part because of differences in taxonomic development, availability of tolerance values, and availability of protocols. The routine use of algae as indicators is also more limited often due to a lack of expertise within monitoring entities. Another factor limiting use is the substantial spatial and temporal variability in species composition even without changes in water quality (Wetzel 2001). The use of cyanobacteria has recently increased because of a need to monitor the occurrence and extent of harmful algae blooms.

It has long been recognized that pollution can change the structure and function of the natural algal assemblage, especially diatoms (Patrick et al. 1954, Patrick 1977), and thus have substantial utility for biological assessments. A number of algal metrics and indices (a majority of which are diatom metrics) have been developed and used to indicate various environmental changes. Most of them belong to one of three categories of methods. The first category is the saprobic system and its derivatives in which diatom assemblages are characterized by their tolerance to organic pollution (Kolkwitz and Marsson 1908, Liebmann 1962, Sladecek 1973). The second category is based on the classification of diatoms according to their sensitivity to all types of pollution (Fjerdingstad 1950, 1965, Coste 1974). Fjerdingstad (1950, 1965) classified diatom species according to their ability to withstand varying amounts of pollution and then described communities in terms of dominant and associated species. The third category of methods is based on the diversity of diatom assemblages. These methods include plotting the number of species against the number of individuals per species (Patrick 1968) and calculating diversity indices (review by Archibald 1972).

This chapter provides brief reviews of several different protocols for sampling periphyton and phytoplankton in a variety of ways (Hill and Herlihy 2000, Stevenson and Bahls 1999, Moulton et al. 2002). The LR-BP for periphyton presented here is an amalgam of methods used by these

[1] Tetra Tech, Inc., 400 Red Brook Blvd., Suite 200, Owings Mills, MD 21117

programs. If field sampling methods other than those recommended are more suitable to your program, they should be thoroughly tested to ensure that they return data of sufficient quality and provide the capacity to address their intended and stated purposes.

5.1.1 Benthic Algae Overview

The benthic algal assemblage in streams and rivers is increasingly being used as an indicator of environmental condition (USEPA 2002). Sampling is generally active through scraping rocks, sticks, or other substrata, or passive by use of artificial substrata. In streams where flow and substratum characteristics create efficient interactions between water and the benthic algal assemblage, benthic algae reflect recent water chemistry (Lowe and Pan 1996). However, in large rivers, suitable attachment surfaces may only occur along banks. In some cases, little suitable substrates may be present for sampling, which may limit the utility of benthic algae as indicators of water chemistry in some rivers. This is particularly relevant in impounded systems where light and flow rates are reduced.

Periphyton assemblage composition is strongly influenced by land-water interactions, and also by river size and the level of human disturbance. In relatively undisturbed rivers, primary productivity is directly correlated with stream order because the surface area of substrata available for periphyton production is increasing and light penetration is adequate. With the increase of ecosystem disturbance (e.g., deforestation and agriculture), periphyton production declines with increasing river size and turbidity (Naiman 1983). The appropriate sampling depth for periphyton in rivers, therefore, will depend heavily on turbidity. It should be noted, however, that periphyton photosynthesis can occur at relatively low light intensities (e.g., 5-25 µmol m^{-2} s^{-1}) (Wetzel 2001).

Bioassessment programs use algal surveys for two primary purposes: 1) to quantify biomass and 2) to characterize species composition. Benthic algal biomass can be generally characterized by different measures, including cell density/biovolume, chlorophyll *a*, ash-free dry mass (AFDM) and dry mass measurement. Qualitative field observation of algal status also helps to identify environmental impairment in rivers. When combined with chemistry information and other biological metrics, qualitative site ranking of the algal assemblage can help decision making. The Kentucky Division of Water (DOW) (2002) uses a 1 (lowest quality) to 5 (highest quality) scoring system and a number of criteria to describe the algal assemblage. The criteria include phytoplankton density, presence/absence of floating algal mats, diversity of several divisions (e.g., chrysophytes, chlorophytes, cyanobacteria, rhodophytes) and the thickness and color of filamentous algae.

5.1.2 Phytoplankton Overview

Phytoplankton is that portion of the plankton composed of algae and cyano bacteria. In general, phytoplankton diversity and biomass are much greater in high order rivers than in low order streams, although their productivity is also often limited by light, as is true for periphyton. The sampling depth of phytoplankton is also regulated by flow, turbidity, and light. In deep, well-mixed large rivers or shallow rivers (i.e., 2-3 m in depth), one phytoplankton sample collected at the depth of 0.5 to 1 m may be adequate. Usually, it is desirable to sample the main channel of

the rivers and avoid inlets, backwater, and sloughs areas. If it is determined that phytoplankton distribution is variable or patchy in a very heterogeneous river channel, compositing samples from multiple locations in a reach is recommended. The planktonic assemblages in general (i.e., phytoplankton and zooplankton) are potentially useful indicators of environmental condition because they are important to the trophic structure of larger rivers, and they are likely sensitive to a number of anthropogenic disturbances, including flow regulation, habitat alteration, invasive species, and contamination by nutrients, metals, and herbicides (Angradi 2006).

Important issues to consider prior to launching a program using periphyton or phytoplankton as a biological indicator include:

- sampling period,
- quantitative and/or qualitative samples,
- collection method to use,
- substrata to sample,
- target indicator to use,
- whether to composite samples,
- sample locations, and
- level of taxonomic identification.

Additional issues to consider associated with phytoplankton include:

- hydrologic seasonality,
- distance from impoundments,
- presence of flushable backwaters, and
- water residence time.

5.2 Discussion on Algal Methods

The protocols in this section have largely been designed for specific applications. However, most can be adapted to meet the differing needs of researchers and resource managers, depending on specific objectives for individual programs and projects. A few questions should be addressed before selecting a field protocol, including, will the focus of the sampling be quantitative or qualitative? If the focus is quantitative, how many parameters will be measured? Is the targeted habitat a single habitat type or multihabitat? Other aspects of the protocol to consider include reach lengths, sampling points and transects, and algal count methods.

Biomass is often the primary concern when extensive algal growth and associated nutrient enrichment are present. For this type of assessment, quantitative sampling to characterize algal biomass is the best approach. However, algal species composition, especially for diatoms, is a useful tool for metrics and indicator development, and can be characterized as relative abundance of individual taxa in a sample. Table 5-1 summarizes the advantages and disadvantages of various algal measures.

Different field sampling methods for freshwater algae can yield similar results (taxonomic composition and relative abundance) providing considerable flexibility in selection of field

techniques. This is likely due to the general ubiquitous distribution of algae in water bodies. As a result, field efficiency can be increased by allowing for the coordinated collection of multiple assemblages at the same collection points of a single design. For example, to facilitate the collection of periphyton sampling from a study reach without significant increases in field time, periphyton samples are regularly collected using the collection techniques discussed in Sections 5.3 and 5.4, but using the field design developed for the LR-BP for benthic macroinvertebrates (Chapter 6).

TABLE 5-1. Advantages and disadvantages of selected algal methods.

Measures	Purpose	Advantages	Disadvantages
Rapid periphyton survey	Quantifying macroalgae and periphyton cover and thickness in a stream reach.	Provides relative biomass of dominant macroalgae and periphyton without laboratory processing and counting.	Requires 3-10 transects for algal cover. Increases field time.
Chlorophyll *a*	Frequently used for indirectly estimating algal biomass.	Measures only the algal portion of the biomass.	Has a relatively short holding time (24 hours) before filtering. Samples must be kept on ice, in a freezer, or in liquid nitrogen in the field, and in the dark prior to laboratory freezer storage and later analysis.
AFDM	Direct measure of algal biomass.	Adds little additional field time. Easy to analyze in the laboratory.	Can include debris and other organic material in the sample. The proportion of algae, bacteria and debris can significantly change the AFDM/dry mass ratio in a sample.
Dry Mass	Direct measure of algal biomass.	Adds little additional field time. Easy to analyze in the laboratory.	Silt can account for a substantial proportion of dry mass in some samples. The proportion of algae, bacteria and debris can significantly change the AFDM/dry mass ratio in a sample.
Cell Density / Biovolume	Estimates the total number algal cells in a sample area.	Provides the most accurate and reliable estimates of total algal standing crop.	Costs more and requires longer processing time.

5.3 Field Sampling Methods

Although there have been efforts to develop broadly-consistent sampling protocols, some differences remain. Basic sampling approaches for periphyton and phytoplankton are provided by the USEPA RBP (Stevenson and Bahls 1999 in Barbour et al. 1999), USEPA EMAP (Hill and Herlihy 2000), USGS-NAWQA program (Moulton et al. 2002), and the USEPA EMAP-GRE (Table 5-2).

TABLE 5-2. Major large river periphyton and phytoplankton sampling methods.

Program	Protocol Summary	Citation
USEPA RBP (periphyton)	Representative samples taken from natural materials (organic and inorganic) and from artificial substrata, and are scraped, drawn, or washed into sample containers; all microhabitat types sampled, or all surfaces from artificial substrata scraped. As appropriate, composite sample preserved or frozen (analyzed for taxonomic composition, biomass, condition index).	Stevenson and Bahls 1999 (from Barbour et al. 1999)
USEPA EMAP-Surface Waters Non-wadeable Streams and Rivers (periphyton)	Individual sample units are taken at eleven transects over a 40 or 100X sampling reach length, on each bank. Use stiff-bristled brush to dislodge periphyton from defined area of rock or wood, wash into sample container as composite sample. Preserve or ice, as necessary. Syringe used to draw sample from soft sediment (analyzed for species composition, relative density, chlorophyll *a*, biomass, enzymatic activity).	Hill and Herlihy 2000
USGS NAWQA Program (periphyton)	Qualitative and quantitative samples taken from epilithic, epidendric, epiphytic, epipelic, and epipsammic habitats over a 500 to 1000 m sampling reach. Use, as appropriate, tools to scrape from rock, wood, or other plant material, and some suction device or spoon to draw soft sediment. For quantitative samples, 25 representative subsamples with controlled effort, and composited into one sample jar. Preserve on ice as necessary (analyzed for species composition, relative density, chlorophyll *a*, and biomass.	Moulton et al. 2002, Porter et al. 1993
USGS NAWQA Program (phytoplankton)	A subsurface grab or depth/width-integrating sampler is used to collect a quantitative whole-water sample. A 1-L sample is sufficient for productive, nutrient-enriched; larger volumes up to about 5 L may be necessary for unproductive, low-nutrient rivers. Subsample volumes may range from 50 mL to more than 500 mL. Subsamples are prepared for chlorophyll a, particulate organic carbon, and biomass.	Moulton et al. 2002, Porter et al. 1993
USEPA EMAP GRE (phytoplankton)	A quantitative phytoplankton sample is collected as a ~2-L composite and preserved with formalin. Samples are analyzed for assemblage structure, body size distribution, and trophic structure. Separate water samples are collected for chlorophyll *a* analysis.	Angradi 2006

Periphyton data were found to be more consistent across different field designs (Charles Lane, US Environmental Protection Agency, personal communications) than are benthic macroinvertebrate data or fish data (Blocksom and Flotemersch 2005, Flotemersch and Blocksom 2005). Collection points for periphyton are, thus, relatively flexible and can be placed according to the needs of the less-flexible designs for benthic macroinvertebrates and fish, thus increasing field efficiency. The field sampling design of the LR-BP for periphyton, as presented in this document, is configured for field compatibility with that of benthic macroinvertebrates (12 sampling zones) and fish LR-BPs (Chapters 6 and 7).

5.4 The Large River Bioassessment Protocol (LR-BP) for Periphyton

Each sampling site consists of a 500-m reach. The GPS coordinates correspond to the downstream end of the sampling reach. At each site, there are a total of six transects. Transect A is located at the downstream end of the reach (0 m) with the remaining five transects at 100 m, 200 m, 300 m, 400 m, and 500 m from the downstream end (Figure 5-1). At each transect, a 10-m sample zone (5 m on each side of transect) on each bank defines the area that will be searched for a substratum suitable for collecting a periphyton sample. The zone extends from the edge of the water to the midpoint of the river, or to a depth of 1m.

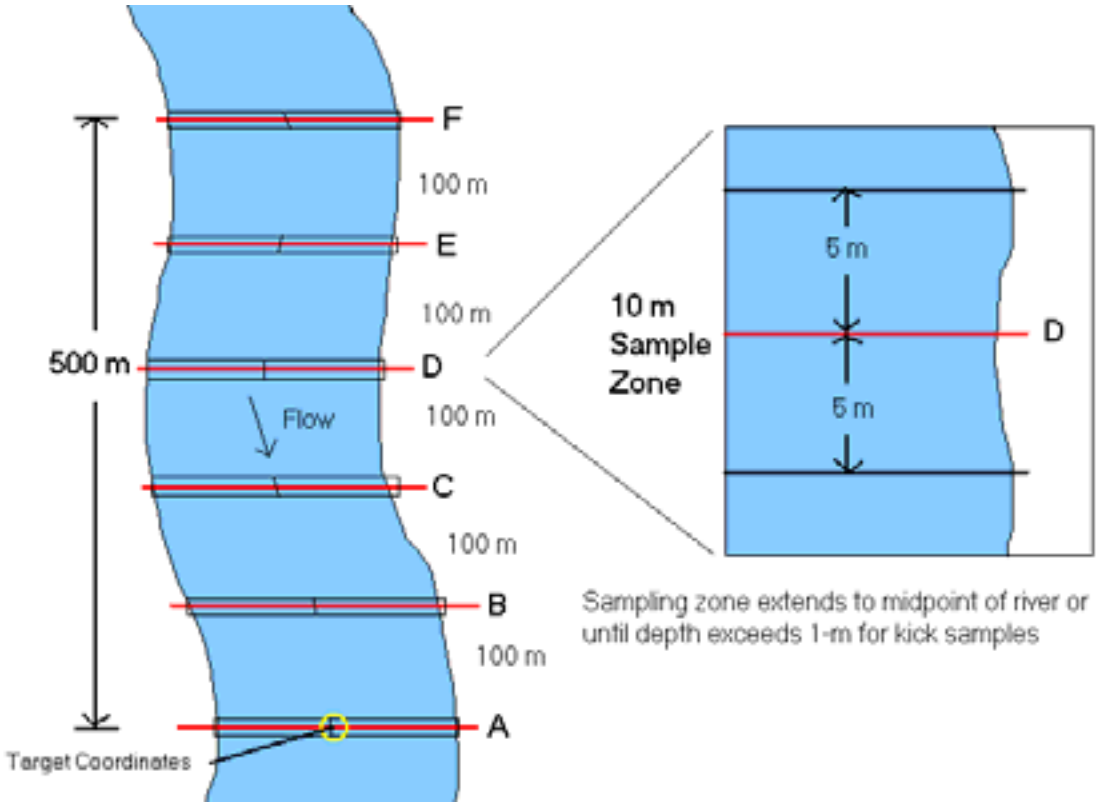


FIGURE 5-1 Example of the six transects and 12 sample zones for collection of periphyton in large rivers using the LR-BP design.

5.4.1 Substrata Selection

At each of 12 sampling zones established, a suitable substratum is selected for collection of a single sample. The substratum selected for sampling should be collected from a location where light penetration reaches the bottom such that it can support algal development. Hill and Herlihy (2000) suggested that the sample be collected from a depth no deeper than can be reached by submerging your arm to mid-bicep depth. If water at a site is >1m deep at the water's edge or the bank is steep, the substratum may be sampled by reaching out of the boat. If a suitable substratum cannot be located or safely sampled, the transect can be bypassed and the exception noted on the field forms.

Often, the selected substratum may arguably be the richest habitat, but this is not the guiding factor in the selection of a suitable substratum. The substratum selected should be one with a high likelihood of producing a quality sample, that is, one that strikes a balance between being representative of the sample station and being suitable for processing in the laboratory. Samples containing excessive sediment are less desirable because they generally take much longer to process, require less than ideal levels of dilution, and can result in poor measures of chlorophyll. The best samples are those from surfaces with a well-developed algal assemblage (e.g., biofilm, algal mat) and a minimum of non-productive sediment. As a guide, epilithic (rock), epidendric (wood) and epiphytic (plant) substrata are preferred (in that order), but other substrata can be sampled, including non-natural surfaces. An example would be when submerged rocks have been covered by a layer of sediment while a suspended piece of woody debris has not. The sample of substratum that is selected is usually small enough (<15 cm diameter) and can be easily removed from the river. At some sampling stations, this may not be the best substratum, but rather the most suitable for the protocol. If the majority of substrata present at a sampling station are so large as to prohibit removal from the river, a longer section of PVC pipe can be used (as described by Hill and Herlihy 2000) that has been fitted with a gasket to seal around the delimited area. The sample can then be removed with a long barrel syringe.

5.4.2 Sample Collection

The following sample collection procedure is a modified version of that outlined by Hill and Herlihy (2000). Once the substratum has been identified, the periphyton sample is collected by removing attached algae from a defined area. Several options exist for delineating the area. Two of the more common apparatuses used for this are a short section of PVC pipe (Hill and Herlihy 2000), and the barrel of a syringe fitted with an O-ring (Porter et al. [1993], and citations therein). The O-ring on the syringe provides for a better seal on the substratum. Two other approaches are the plastic frame of a 35 mm or medium format slide and a rubber mat with an opening. The slide frame is preferred by some because it is more flexible and form-fitting than a section of PVC pipe or the barrel of a syringe. The rubber mat is likewise flexible with the added feature of covering the area outside of that delineated and when rinsed, reduces the potential for sample contamination.

Program Highlight

USGS National Water Quality Assessment (NAWQA) Algal Assessment Protocols for Non-wadeable Streams and Rivers

The USGS NAWQA program has developed a suite of protocols for the collection of algae from non-wadeable streams and rivers (Moulton et al. 2002). These include protocols for the active collection of qualitative and quantitative periphyton samples that use artificial substrata and phytoplankton samples. The sampling reach length and location used for the collection of algal samples are determined on the basis of a combination of repeating geomorphic channel units (Meador et al. 1993). However, given the realities often faced on large river systems, minimum and maximum reach lengths of 500 and 1000 m, respectively, have been deemed acceptable (Meador et al. 1993).

In general, periphyton samples are collected from the surfaces of natural substrata in relation to the presence of microhabitats in the sampling reach by scraping, brushing, siphoning, or other methods appropriate to each microhabitat (Porter et al. 1993). Periphyton is sampled in erosional habitats by removing the designated substratum from the stream, dislodging the attached material from a predetermined area on the upper surface of the substratum with a stiff-bristled brush, and then washing the material into a sample bottle. In depositional habitats, a predetermined area of soft sediment is collected using a syringe or a spoon and transferred to the sample bottle. Sampling is conducted at locations chosen to represent combinations of natural and anthropogenic factors (Porter et al. 1993).

Qualitative Multihabitat Sampling Method

For this protocol, periphyton samples from all instream microhabitat types present in the sampling reach are composited (Porter et al. 1993).

Quantitative Targeted-Habitat Sampling Method

The goal of quantitative periphyton sample collection is to measure relative abundance and density of taxonomically representative periphyton within: (1) a richest-targeted habitat (RTH) which supports the taxonomically richest assemblage of organisms within a sampling reach and (2) a depositional-targeted habitat (DTH) where organisms are likely to be exposed to sediment-borne contaminants for extended periods of time. Another quantitative method is the use of artificial substrata, which don't necessarily target any specific habitat. In both RTH and DTH, the protocols specify sampling from five representative substrata at five locations within the designated reach. This results in a final composite sample (for both the RTH and DTH) that is composed of, at most, 25 subsamples each (if five substrata are available).

Richest-Targeted Habitat (RTH)

Typical RTH areas include riffles in shallow, coarse-grained, high-gradient systems, or woody snag habitats in sandy-bottomed systems. At each of the five locations, samples are taken from five representative substrata (25 total samples). In order of preference, samples are taken from epilithic, epidendric, and epiphytic substrata. A simple sampling device is used to quantify the size of the sampled area (Porter et al. 1993). The device consists of a 60-cc syringe barrel fitted with a rubber O-ring on one end. The end with the rubber o-ring is placed flat on the substratum surface so that a seal is formed. A brush is then placed through the syringe barrel and used to dislodge the attached periphyton from the surface of the substratum. The sample area is then washed with a squirt bottle and the dislodged periphyton is rinsed into the sample collection container. If the substratum surface is irregular so that the rubber o-ring cannot form a seal, the periphyton can be brushed from the entire substratum and the entire substratum is then fitted with aluminum foil. The substratum is discarded and the foil is returned to the laboratory so that the surface area of the substratum can be determined. If bedrock is to be sampled, then a PVC pipe sampler is used. The periphyton from all 25 subsamples are composited into one sample jar.

An example of a DTH area is an organically-rich depositional area such as a pool. If epilithic or epidendric substrata are available in the DTH area, then periphyton should be collected in the same manner as they are collected from the RTH areas. However, if these substrata are not present, then epipelic or epipsammic microhabitats should be sampled. In order to sample epipsammic or epipelic habitats, the top half of a disposable 47-mm plastic petri dish is gently pushed into the streambed sediment. Then, a small sheet of Plexiglas or a spatula is slipped under the petri dish top so that the sediment is trapped inside. The contents are then rinsed into a sample jar. Because the volume of the petri dish top can be measured, then the sample can be quantified. Five sediment samples are taken for the entire reach. All DTH samples (sediment and any other available substratum samples) are composited into one sample jar.

Artificial Substratum Sampling Method

When natural substrata cannot be sampled because of inaccessibility of the microhabitats, cost of sample collection or safety issues, artificial substrata can be used in sampling reaches. These limitations occur in large rivers and should be considered when designing a sampling program for this type of system. Samples obtained from artificial substrata typically have reduced heterogeneity compared to those obtained from natural substrata but can be used to compare water quality among rivers with disparate periphyton microhabitats. However, data from artificial substrata cannot be compared with data from natural substrata. If artificial substrata are used for one or more stream reaches in a basin, it is recommended that they be used at all sites so that meaningful water quality interpretations can be made. The advantages and limitations of artificial substrata are discussed in Porter et al. (1993).

Quantitative Phytoplankton Protocol

Phytoplankton are more reflective of conditions in the open water column, whereas periphyton represent conditions at the sediment/substratum-water interface. Quantitative phytoplankton samples are obtained by collecting representative whole-water samples. A sample volume of 1 L is sufficient for samples collected from productive, nutrient-enriched rivers as indicated by water color, but a larger sample volume is required for samples collected from unproductive, low nutrient rivers as indicated by water transparency. Phytoplankton samples, taken in conjunction with water chemistry sampling, are taken with a depth-integrating sampler. Alternatively, quantitative phytoplankton samples can be collected with a water sampling bottle or with a pump. If chlorophyll is not to be measured, the entire sample is preserved with buffered formalin. For chlorophyll measurements, an unpreserved subsample is withdrawn from the phytoplankton sample, and the aliquot is filtered onto a glass fiber filter. The filtered subsample volume should be sufficient to ensure that adequate algal biomass is retained on the filter. Filters are then wrapped in aluminum foil and immediately stored on dry ice (Porter et al. 1993).

Place the substratum in a plastic funnel which drains into a 500-ml plastic bottle with volume graduation. Use the area delimiter to define an area on the upper surface of the substratum. Dislodge attached periphyton from the substratum within the delimiter into the funnel by brushing with a stiff-bristled toothbrush for 30 seconds. Take care to ensure that the upper surface of the substratum is the surface that is being scrubbed, and that the entire surface within the delimiter is scrubbed. It may be necessary to mark the area contained in the delimiter with a tool, remove the delimiter and proceed with processing the surface.

In systems where the material within the area of the delimiter takes the form of a thick mat rather than a biofilm, the delimiter is placed on the substratum particle as previously described, but before the surface is scrubbed, a micro-spatula or spoon may be used to scoop out the material. This activity may or may not need to be followed by brushing. A minimal volume of river water from a bottle is used to wash the dislodged periphyton from the funnel into the 500-ml bottle. This process is repeated on left and right banks of all six transects and samples are composited into the same 500-ml bottle.

After samples have been collected from all 12 sample zones, the 500-ml bottle is thoroughly mixed regardless of substratum type. The sample container should be placed on ice until preservation. The total estimated volume of the composite sample is recorded. This volume will be used for subsequent preparation of samples for laboratory processing. The total area sampled for the reach will be 12X the area of the delineator minus those sample zones, if any, that were not sampled. Place samples on dry ice. Within 24 hours, place 48 mL of the samples into a separate container and preserve with 2 ml of 10% buffered formalin.

5.5 Laboratory Processing

As many as five different types of samples could be shipped to laboratories for analyses: Chlorophyll (Chl) *a*, Biomass (AFDM and dry mass), Acid/Alkaline Phosphatase Activity (APA), algal mass nutrient contents, and algal ID/enumeration samples. The Chl *a* and AFDM analyses estimate algal total biomass by area in the sampling reach. Standard methods have been developed for these measurements (APHA 1998). The APA and periphyton C:N:P ratio are optional measurements of algal nutrient limitation for determining which nutrient is limiting an aquatic system (Hill and Herlihy 2000). They are not as frequently used as algal biomass parameters. Algal ID/enumeration samples are used for algal taxonomy and algal cell density/biomass measurements.

5.5.1 Chlorophyll a and AFDM Analyses

Chl *a* analysis methods have been developed for both phytoplankton and periphyton (APHA 1998). After samples are filtered through a glass fiber filter, they are extracted using 90% aqueous acetone (APHA 1998) and stored for 24 hours at 4 °C in the dark. Three techniques for measuring Chl *a* in solution are the spectrophotometric, the fluorometric, and the high-performance liquid chromatography (HPLC) techniques. Hydrochloric acid (HCl) is used to correct phaeophytin *a* concentration. Dry mass and AFDM can be determined by weighing dry algal mass and weighing ash after incinerating organic material at 500 °C for 1 hour (APHA 1998).

5.5.2 Taxonomy and Enumeration: Soft-bodied Algae

The methods summarized here are a combination of the protocols provided by Barbour et al. (1999) and Moulton et al. (2002). Direct laboratory analysis results in density and abundance values for both soft-bodied algae (non-diatom) and diatoms. However, in many cases, data from diatoms only are sufficient for indicators of river condition.

Specification of laboratory procedures (counting, subsampling, and taxonomy) is extremely important for developing datasets for algae-based indicators. Subsamples of soft algae are used to determine density or biovolume of major taxa. After appropriate dilution, a Palmer-Maloney counting cell at 400X magnification can be used for both identification and enumeration. A common procedure for counting soft algae is to count 300 natural units (i.e., each individual filament, colony, or isolated cell). This procedure prevents a colonial or filamentous cluster from dominating a count and allows the assemblage structure to be assessed. All algae should be identified to the lowest possible taxonomic level, recording its name and density. Many of the non-diatom taxa, however, are not easily identified to the species level without culturing of the taxa. For most purposes of biological assessment, genus level identification is usually adequate. References for soft algae taxonomy are *Freshwater Algae of North America* (Wehr and Sheath 2003), and *Algae of the Western Great Lakes Area* (Prescott 1962). Although cell density of each taxon is recorded during the soft algae count, it is also recommended to convert cell density to biovolume (Hillebrand et al. 1999) or bio-surface area (APHA 1998) to account for the actual biomass of algae. Biovolumes and biosurface areas of all common taxa (relative abundance >5%) in any sample are determined by measuring at least 15 cells of each taxon present. Density and live:dead ratios of diatoms are recorded in the Palmer-Maloney soft algae count. Identification of diatoms is not necessary in the soft algae count.

Another algal counting method is a simple wet mount procedure used to identify small algal cells under high magnification (1000X). Sedgewick-Rafter counting chambers (large modified slides with 1-ml wells) are also often used for counting large filamentous algae and larger cells under lower magnification (100X).

5.5.3 Taxonomy and Enumeration: Diatoms

Diatom subsamples can be digested with hydrogen peroxide or nitric acid to remove organic matter (Patrick and Reimer 1967). Permanent slides are prepared using Naphrax, a high refractive index mounting medium, following APHA (1998). Approximately 300 diatom cells (600 valves) are counted at random or fixed transects and identified to the lowest possible taxonomic level (usually species or subspecies).

Both soft algae and diatom counts require highly trained taxonomists to perform consistent identification. The four-volume series on the diatoms of Middle Europe (Krammer and Lange-Bertalot 1986, 1988, 1991a, 1991b) is the primary diatom reference used by many taxonomists. This series updates diatom structure with detailed micrographs of each taxon. In *The Diatoms of the US*, Patrick and Reimer (1967, 1975) include the major algal taxa found in the USA. The recent treatments by Round et al. (1990) provide details on diatom ultrastructure. Terminology necessary to work through the keys and descriptions of genera is presented in the glossary.

Other taxonomic references include the diatom naming conventions adopted by the Academy of Natural Sciences of Philadelphia (Morales and Potapova 2000).

5.6 Data Entry

Taxonomic nomenclature and counts are frequently entered into the data management system directly from handwritten bench or field sheets. Depending upon the system used, there may be an autocomplete function that helps prevent misspellings. There are two methods for assuring accuracy in data entry. One is the double entry of all data by two separate individuals, and then the performance of a direct match between databases. Where there are differences, it is determined which database is in error, and corrections are made. The second approach is to perform a 100% comparison of all data entered to handwritten data sheets. Comparisons should be performed by someone other than the primary data enterer. When errors are found, they are hand-edited for documentation, and corrections are made electronically. The rates of data entry errors are recorded, and then, in the overall database are segregated by data type (e.g., fish, benthic macroinvertebrates, periphyton, header information, latitude longitude, physical habitat and water chemistry).

5.7 Data Reduction (Metric Calculation)

The current literature identifies various diatom metrics that indicate responses to different environmental stressors (Van Dam et al. 1994). Metrics summarized by Stevenson and Bahls (1999) are used by several States (Bahls 1992, Kentucky DOW 2002). Of the metrics discussed in the RBP manual, nine represent metrics of biotic condition while six others are diagnostic metrics. Hill et al. (2000, 2003) and Kentucky DOW (2002) developed periphyton indices of biotic condition using multiple metric approaches. Van Dam (1994) and Kelly (1998) also provided valuable information on diatom autecological indices that have been widely used in Europe. Table 5-3 lists the most commonly used metrics.

5.7.1 Diversity Metrics

Two periphyton metrics are measurements of taxa richness (i.e., total taxa and Shannon diversity) and are estimated from the count of taxa found in a target number of cells (e.g., 300 cells). Diversity metrics are less persuasive in indicating nutrient enrichment because of the confounding effect of algal diversity increasing with increased nutrient levels in oligotrophic systems as a consequence of adding homogenizing species.

5.7.2 The Pollution Tolerance Index (PTI) of Diatoms

An example of a water quality assessment method based on the pollution tolerance of diatom assemblages is the pollution tolerance index (PTI), which is used by the Kentucky Department of Environmental Protection (Kentucky DOW 2002). The PTI is similar to that used by Lange-Bertalot (1979) and resembles the Hilsenhoff biotic index (HBI) for macroinvertebrates (Hilsenhoff 1987). There are three categories of diatoms according to documented pollution tolerance, with the most tolerant taxa assigned a value of 1 and the most sensitive taxa assigned a value of 3. For the PTI, the categories are expanded to four with the resulting values ranging

from 1 to 4. Similarly, percent sensitive and percent tolerant taxa can be derived from this method. The formula used to calculate PTI is:

$$PTI = \sum \frac{n_i t_i}{N},$$

where n_i is the number of cells counted for species i, t_i is the tolerance value of species i (1-4), and N is the total number of cells counted. Tolerance values have been generated from several sources, including Patrick and Reimer (1966, 1975), Lowe (1974), Patrick (1977), Descy (1979), Lange-Bertalot (1979), Kelly (1988), Sabater et al. (1988), and Bahls (1992).

TABLE 5-3. Diatom and non-diatom metrics summarized from various sources.

All Algae Metrics	Diatom Metrics
Taxa richness of non-diatoms or all algae	Total number of diatom taxa (TNDT)
Species dominance	Shannon diversity index
% cyanobacteria	Percent community similarity (PSc)
Number of Divisions represented by all taxa	Pollution tolerance index
Chlorophyll *a*	Percent sensitive diatoms
Ash-free dry mass (AFDM)	Percent live diatoms
Phosphatase activity	Van Dam's diagnostic metrics
Autotrophic index	Simple autecological indices
	Percent Epithemiaceae
	Percent motile diatoms
	Percent *Achnanthidium minutissimum*

5.7.3 Percent Community Similarity (PSc)

Percent community similarity (PSc) by Whittaker (1952) is an example of a water-quality assessment method based on the diversity of diatom assemblages. The PSc was chosen for use in diatom bioassessments because it shows assemblage similarities based on relative abundances and gives more weight to dominant taxa than to rare taxa. The PSc should only be used when comparing a study site to a control site, or when conducting multivariate cluster analysis. If the emphasis is comparing a study site to a regional reference condition (i.e., a composite of sites), the PSc should not be used. The PSc values range from 0% (no similarity) to 100% (identical).

The formula for calculating PSc is:

$$PS_c = 100 - 0.5\sum_{i=1}^{s} |a_i - b_i|,$$

where a_i is the percentage of species i in sample A, and b_i is the percentage of species i in sample B.

5.7.4 The Autotrophic Index

Because periphyton is found on or in close proximity to the substratum, dry mass (DM) and ash-free dry mass (AFDM) values are used as assessment tools. The AFDM is an estimate of total organic material accumulated on the substratum. This organic material includes all living organisms (e.g., algae, fungi, bacteria, and macroinvertebrates) as well as non-living detritus. The DM values are used in conjunction with chlorophyll *a* as a means of determining the trophic status of rivers through the use of the autotrophic index (AI). The formula used to calculate AI is:

$$AI = DM\ (mg/m^2)/\text{Chlorophyll } a\ (mg/m^2).$$

High AI values (i.e., >200) indicate that the assemblage is dominated by heterotrophic organisms and can indicate poor water quality (Weber 1973, Weitzel 1979, Matthews et al. 1980). This index should be used with discretion because non-living organic detritus can artificially inflate the AFDM value. One option is to modify the AI to include AFDM and invert:

$$AI = \text{Chlorophyll } a\ (mg/m^2)/ \text{AFDM}\ (mg/m^2)$$

In this form, the index is positively related to the autotrophic proportion of the assemblage instead of the heterotrophic proportion. Also, since chlorophyll *a* / AFDM values normally are about 0.1%, the modified index would have better statistical properties than the original index.

5.7.5 Diagnostic Diatom Metrics

Diatom species have different sensitivities to different types of pollution (e.g., nutrients, metals, pH, salinity). Thus, stressor-specific metrics may help to diagnose environmental pollution in aquatic systems. A number of diatom metrics have been developed to assess environmental impairment (Table 5-2). For example, % *Eunotia* species has been used to assess acidic condition (e.g., in association with acid mine drainage), and % Epithemiaceae taxa has been used to indicate nitrogen limitation (Kentucky DOW 2002). The diatom indices in Van Dam et al. (1994) are among the most complete diatom autecological references for diagnosing various environmental conditions. These indices are:

- *Trophic state index*. Eutrophic and hypereutrophic diatoms indicate elevated concentrations of nutrients that are important for diatom growth: nitrogen, phosphorus, inorganic carbon and silica. Diatom species are assigned nutrient tolerance values ranging from 1 – 6. As nutrient concentrations increase, the mean tolerance value of diatoms present increases from 1 to 6, and the proportion of eutrophenic diatoms (indicator values from 5 to 6) will increase. Therefore, the index or % eutrophenic taxa will also increase.

- *Index of nitrogen uptake metabolism*. Indicator values of nitrogen uptake from autotrophic to heterotrophic taxa range from 1 to 4. When nitrogen concentrations increase, the percentage of obligate nitrogen autotrophs will decrease, but obligate nitrogen heterotrophs will increase. Therefore, the index value will increase with organic enrichment.
- *Saprobity index*. The index characterizes waters with light to heavy loads of organic matter and with low or no oxygen. The index value will increase from 1 to 5 as organic loads (e.g., from agricultural and wastewater discharges) increase.
- *pH index*. Diatoms are extremely sensitive to pH. The index value ranges from 1 to 5, inferring acidic to alkaline conditions. The index indicates pH value from below 5.5 to above 8.5.
- *Oxygen demand index*. Oxygen demand is also classified for many diatoms, ranging from 1 to 5, indicating very low (i.e., <10% saturation) to very high (i.e., 100% saturation) dissolved oxygen. It is also an indication of organic degradation.
- *Salinity index*. A diatom-based salinity index that was formulated from statistical relationships between salinity data and diatom assemblages. The index value ranges from 1 to 4, indicating salinity from <0.2 to 9.0%.

5.8 Site Assessment and Interpretation

Although the use of diatoms for assessing stream condition is well established, their use for determining biological impairment (such as for purposes of water quality standards programs) is not as widespread and has not been extensively used in large river systems. Taken collectively, diatoms span a very wide autecological spectrum (e.g., ranging from ultraoligotrophic to hypertrophic conditions); but within this broad range, individual species display relatively consistent environmental tolerances (or autecological values), even across wide geographic distributions. These characteristics make diatom assemblages potentially useful indicators of biological impairment. The diatom diagnostic metrics discussed in Section 5.5.5 quantifies algal status and environmental characteristics in a sampling station for particular stressors. Other algal metrics (e.g., algal biomass, AFDM, Chl *a*), while not stressor-specific, can also indicate human disturbance. However, for resource management, a simple multimetric biotic index is much more practical for decision making and for implementing protection goals.

The reference approaches based on an IBI or other integrated variables (e.g., observed/expected ([O/E] ratio) could be used for periphyton assessment. Hill et al. (2000, 2003) developed a periphyton IBI that included eight metrics (i.e., species richness; species dominance; relative abundances of acidobiotic, eutraphentic, and motile diatoms; standing crops from Chl *a* and biomass; and APA). The Kentucky DOW (2002) also developed a diatom bioassessment index (DBI) comprising six metrics (i.e., diatom richness, Shannon index, PTI, Siltation index, *Fragilaria* richness, and *Cymbella* richness). A well-designed index should respond to both specific and multiple environmental stressors in a predictable manner and should not be affected by river size or watershed area. To determine an impairment threshold, a reference approach can be used. Hill et al. (2003) adopted 75^{th}, 25^{th}, and 5^{th} percentile scores of the reference site distribution to set thresholds for excellent, good, fair or poor conditions, respectively, and thus set protective goals for algal status. Another approach that has been widely used in macroinvertebrate assessment is the O/E approach. Mid-Atlantic streams were assessed using

diatoms by comparing the O/E approach to other metrics and multimetric approaches (R. J. Stevenson, personal communication). The results indicated that the approach to quantifying loss of native diatom taxa with increasing nutrients was not successful, but diatom autecologies are likely correlated with nutrient concentrations.

5.9 Performance Characteristics for Biological Assessments Using Algae

5.9.1 Field Sampling

Quantitative (QN) performance characteristics for field sampling are *precision* and *completeness* (Table 5-4). Repeat samples for purposes of calculating precision of field sampling are obtained by sampling from two adjacent 500-m reaches (Figure 5-2). For algae, samples from the adjacent reaches (also called quality control [QC] or duplicate samples) must be laboratory-processed prior to data being available for precision calculations. These precision values are statements of the consistency with which the sampling protocols:

- characterized the algal biota of the river and
- were applied by the field team,

and thus, reflect a combination of natural variability, laboratory error, and systematic error (see Chapter 3).

The number of reaches for which repeat samples are taken varies, but a rule-of-thumb is 10% of the total number of sampling reaches constituting a sampling effort (whether a programmatic routine or an individual project), and they would be randomly selected from that list. Values for calculation of precision are dry mass (DM), ash free dry mass (AFDM), chlorophyll *a*, biovolume/density and some measure of species composition. In effect, whatever the indicator values that are to be used for site assessment, are also used to calculate relative percent difference (RPD), root-mean square error (RMSE) and coefficient of variability (CV) (Table 3-2). Acceptance criteria for each of these would be established based on programmatic capabilities demonstrated via pilot studies, or through analysis of existing datasets produced using the same protocols. These criteria are not data quality thresholds beyond which data points should be considered for discarding. Rather, they are flags for potential problems (errors) in sample collection or processing, are used to help determine the sources of the problems, and can be used to help develop recommendations for corrective actions.

Percent completeness (Tables 3-2, 5-4) is calculated to allow communication of the number of valid samples (however validity is judged) that were collected as a proportion of those that were originally planned. This value serves as one summary of overall data quality for a sampling effort, and it demonstrates confidence in the final results.

TABLE 5-4. Error partitioning framework for biological assessments and biological assessment protocols for algae. There may be additional activities or performance characteristics, and they may be quantitative (QN), qualitative (QL), or not applicable (na).

Component Method or Activity	Performance Characteristics				
	Precision	Accuracy	Bias	Representativeness	Completeness
1. Field sampling	QN	na	QL	QL	QN
2. Laboratory subsampling	QN	na	QL	QL	na
3. Taxonomy	QN	QL	QL	na	QN
4. Data entry	na	QN	na	na	QN
5. Data reduction (e.g., metric calculation)	na	na	QN	na	na
6. Site assessment and interpretation	QN	QN	QL	QL	QN

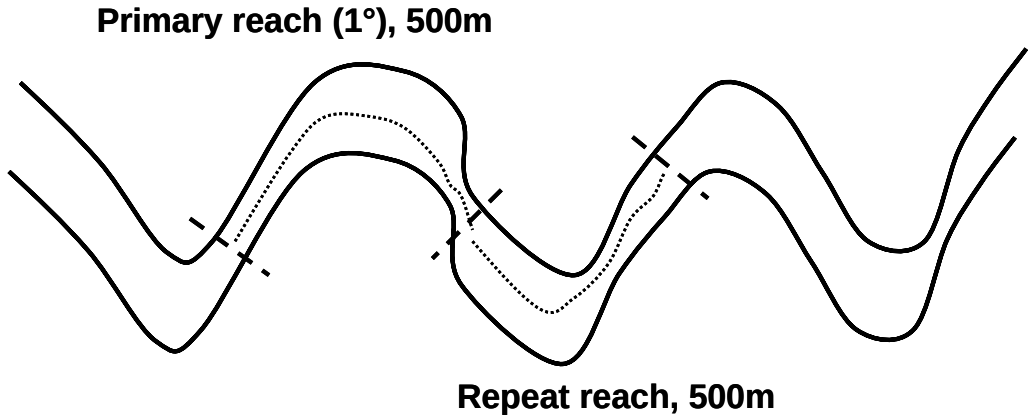


Figure 5-2. Adjacent reaches (primary and secondary) on a fluvial channel.

Qualitative (QL) performance characteristics for field sampling are *bias* and *representativeness* (Table 5-4). Attempts to minimize the bias associated with the LR-BP for algae include the fact that sample portions are taken are from areas with hard surfaces (rock, wood) that are distributed among 12 sample zones (Figure 5-1) and composited; sampling is not restricted to small, limited areas. The LR-BP field sampling method is intended to depict the algal assemblage present in the shore-zone area (out to a 1-m depth) that the large river has the conditions to support.

Accuracy is considered "not applicable" to field sampling (Table 5-4) because efforts to define analytical truth would necessitate a sampling effort excessive beyond any practicality. That is,

the analytical truth would be all algae or algal taxa that exist in the river (shore zone to 1-m depth). There is no sampling approach that would collect all individual algal cells or filaments.

5.9.2 Laboratory Sorting/Subsampling

Precision is a QN characteristic of performance for laboratory subsampling of algae (Table 5-4). Subsampling of algae (specifically, for diatoms and phytoplankton) occurs as pipettes are used to draw liquid from a sample to prepare slide mounts. Comparison of the results from multiple slides prepared from the same sample provides information on the precision of subsampling, which is calculated using RPD and CV (Table 3-2) with measures of species composition as the input variables. Precision is an indication of how well the sample is mixed; it is not necessary to do this for every sample. Serial subsampling and precision estimates should be done on approximately 10% of all samples collected as part of a project and on two timeframes. First, they should be done and the results documented and reported, to demonstrate what the laboratory is capable of in application of the subsampling method. Second, they should be done periodically to demonstrate that the program routinely continues to meet that level of precision. Representativeness of the sorting/subsampling process is addressed as part of the standard operating procedure (SOP). Considered as "not applicable", estimates of *accuracy* are not necessary for characterizing laboratory sorting performance.

5.9.3 Taxonomy

Precision and *completeness* are QN performance characteristics used for taxonomy (Table 5-4). Precision of taxonomic identifications is calculated using percent taxonomic disagreement (PTD) and percent difference in enumeration (PDE) (Table 3-2), both of which rely on the raw data (list of taxa and number of individuals) from whole-sample re-identifications. The primary taxonomy is completed by the project taxonomist (T1); re-identifications are performed by a secondary or QC taxonomist (T2) as blind samples. The number of identifications in agreement between the two sets of results, as an inverse proportion of the total number of individuals in the sample ((1-[number of agreements])/N), is precision of the taxonomic identifications. The percent difference in sample counts by each of the taxonomists is percent difference in enumeration (PDE). These two values are evaluated individually and can be used to indicate the overall quality of the taxonomic data; and if there is a problem, they can help identify what is causing the problem. The number of samples for which this analysis is performed will vary, but 10% of the total sample lot (project, program, or year, or other) is an acceptable rule-of-thumb. Exceptions are that large programs (>~500 samples) may not need to do >50 samples; small programs (<~30 samples) will likely still need to do at least 3 samples. In actuality, it will be program-specific and the number of samples re-identified will be influenced by multiple factors, such as how many taxonomists are doing the primary identification (there may be an interest in having 10% of the samples from each taxonomist re-identified), and how confident the ultimate data user is with the results. Mean PTD and PDE across all re-identified samples is an estimate of taxonomic precision (consistency) for a dataset or a program. Percent taxonomic completeness (PTC; [Table 3-2]) quantifies the proportion of individuals in a sample that are identified to the specified target taxonomic level (lowest practical taxonomic level, species, genus, family or other, including mixed levels).

Accuracy and *bias* are QL performance characteristics for taxonomy (Table 5-4). Accuracy requires specification of an analytical truth. For taxonomy, it is a) the museum-based type specimen (holotype, or other form of type specimen), b) specimens verified by recognized expert in that particular taxon or c) unique morphological characteristics specified in dichotomous identification keys. Determination of accuracy is considered "not applicable (na)" for production taxonomy (most often used in routine monitoring programs) because that kind of taxonomy is focused on characterizing the sample; taxonomic accuracy, almost by definition, would be focused on individual specimens. Bias in taxonomy can result from use of obsolete nomenclature and keys, imperfect understanding of morphological characteristics, inadequate optical equipment, and poor training. Neither of these performance characteristics is considered necessary for production taxonomy, in that they are largely covered by the estimates of precision and completeness. For example, although it is possible that two taxonomists would put an incorrect name on an organism, it is considered low probability that they would put the same incorrect name on that organism.

5.9.4 Data Entry

Efforts to understand the quality of data entry activities may seem trivial. However, the impact of errors can be substantial, and, if undiscovered and uncorrected, can become amplified through the assessment process. This QN performance characteristic (*accuracy*) simply quantifies the number of correctly-entered data values as a proportion of the total number of data values entered. The process involves having a QC person, distinct from the staff doing the primary data entry, check all data values (100%) against the original handwritten datasheets. With the datasheets as the analytical truth, the rate of errors is the accuracy of the data entry (Table 5-4). As errors are found, they are corrected electronically. For their wadeable streams program, Mississippi Department of Environmental Quality (MDEQ) found that the two data types with the highest error rates were the datasheet header information (e.g., stream name, latitude/longitude, date of site visit, and name of field staff) and streambed particle size data (MDEQ 2006). This allowed corrective actions to be focused where needed. All other performance characteristics are considered not applicable.

5.9.5 Data Reduction (Metric Calculation)

For most biological assessment programs, raw data are the list of taxa found at a site (in a sample) and the number of individuals recorded for each taxon. Preparation of those data for analysis requires conversion to metrics or other terms; metric calculation is a form of data reduction. When electronic spreadsheets or other data manipulation techniques are used, queries are often built to perform both complex and simple calculations. If queries are not performing as intended, or links to the raw data are incorrect, errors in metric values can occur. Precision of data reduction is a QN performance characteristic (Table 5-4) that helps ensure database/ computer calculation routines are performing as intended. A subset of metric values is hand-calculated using only the taxonomic and enumeration data, which are then compared to those that result from the computer queries. A recommended approach involves calculating one metric for multiple samples (e.g., systematic, every third sample), as well as all metrics for at least one sample. If differences are found, each value should be checked for errors in the calculation process (hand calculator vs computer algorithm), and corrections made.

5.9.6 Site Assessment and Interpretation

QN performance characteristics for site assessment and interpretation are *precision*, *accuracy*, and *completeness* (Table 5-4). Site assessment precision is based on the narrative assessments from the associated index scores (e.g., good-fair-poor) from reach duplicates. It quantifies the percentage of duplicate samples that receive the same narrative assessments as the original. These comparisons are done for a randomly selected 10% of the total sample lot. Table 5-5 shows that, for this dataset, 79% of the replicates returned assessments of the same category (23 out of 29); 17% were 1 category different (5 of 29); and 3% were 2 categories different (1 of 29). Accuracy is the proportion of samples for which the biological index correctly identifies sites as impaired; the calculation is discrimination efficiency (DE) (Table 3-3). DE is a value that is developed during the index development and calibration process. Percent completeness (%C) is the proportion of sites (of the total planned) for which valid final assessments were obtained; a site assessment is considered valid when data of sufficient quality and quantity are available for that assessment.

QL performance characteristics for site assessment and interpretation are bias and representativeness (Table 5-4). The final assessment of a site can be biased if a small number of reference or stressor sites are used during the calibration process; low numbers of stressor sites can potentially result in high discrimination efficiencies that are spurious. If interpretation of assessment results fails to take into consideration abnormal or extreme hydrologic or climatic events, or other non-natural catastrophic and localized events, results could be considered non-representative of ambient conditions.

TABLE 5-5. Assessment results shown for sample pairs taken from 29 sites, each pair representing two adjacent reaches (back to back). Assessment categories are 1-good, 2-fair, 3-poor and 4-very poor.

Site	Replicate 1		Replicate 2		Categorical Difference
	Narrative	Assessment Category	Narrative	Assessment Category	
A	Poor	3	Poor	3	0
B	Poor	3	Poor	3	0
C	Good	1	Good	1	0
D	Poor	3	Very Poor	4	1
E	Fair	2	Fair	2	0
F	Poor	3	Fair	2	1
G	Poor	3	Poor	3	0
H	Very Poor	4	Very Poor	4	0
I	Very Poor	4	Very Poor	4	0
J	Poor	3	Poor	3	0
K	Poor	3	Poor	3	0
L	Very Poor	4	Very Poor	4	0
M	Very Poor	4	Very Poor	4	0
N	Poor	3	Fair	2	1
O	Poor	3	Poor	3	0
P	Poor	3	Poor	3	0
Q	Poor	3	Very Poor	4	1
R	Poor	3	Poor	3	0
S	Fair	2	Very Poor	4	2
T	Fair	2	Fair	2	0
U	Good	1	Good	1	0
V	Poor	3	Fair	2	1
W	Fair	2	Fair	2	0
X	Poor	3	Poor	3	0
Y	Poor	3	Poor	3	0
Z	Very Poor	4	Very Poor	4	0
AA	Poor	3	Poor	3	0
BB	Fair	2	Fair	2	0
CC	Poor	1	Poor	1	0

Chapter 6.0 Benthic Macroinvertebrates

By Brent R. Johnson[1], James B. Stribling, Joseph E. Flotemersch and Michael J. Paul

This chapter…

- reviews existing large river macroinvertebrate sampling methods
- recommends a bank-oriented multi-habitat approach

Macroinvertebrates are…

- important components of large river food webs
- proven indicators of biological condition
- responsive to a wide range of stressors

6.1 Introduction

Benthic macroinvertebrates include aquatic insects, crustaceans, annelids, mollusks, nematodes, planarians, bryozoans, cnidarians (*Hydra*), and nemerteans. They inhabit sediments or live on bottom substrates of aquatic ecosystems. At least some representatives of this assemblage can be found in virtually every freshwater environment on Earth. Macroinvertebrates, specifically, are invertebrates retained by a mesh size of 500 µm (Hauer and Resh 1996). While early developmental stages may pass through a mesh of this size, 500-595 µm is generally considered suitable for biomonitoring purposes (e.g., Klemm et al. 1990, Barbour et al. 1999, Lazorchak et al. 2000). Smaller mesh sizes are required for ecological studies that focus on life histories and secondary production, and those that include meiofauna. Macroinvertebrates play a critical role in the transfer of energy from basal resources (e.g., algae, detritus and associated microbes) to vertebrate consumers in aquatic food webs, and they serve as the primary food resource for many commercially and economically important fish species.

Benthic macroinvertebrate are the most common faunal assemblage used in bioassessments of wadeable streams and rivers (e.g., Rosenberg and Resh 1993, Barbour et al. 1999, USEPA 2002, Carter and Resh 2001). After careful sampling using standardized field collection methods, laboratory species identification and enumeration, evaluation of structural and functional attributes of the assemblage are used to evaluate biological condition. The following factors have contributed to their becoming so widely used in biomonitoring programs (modified from Barbour et al. 1999):

- Macroinvertebrates are ubiquitous and abundant in most streams and rivers, including headwater streams where fish may be absent.
- Macroinvertebrates are relatively sedentary in the aquatic environment so they are good indicators of local condition.
- Many taxa are long-lived (1 year or more) and, thus, integrate short-term disturbances and reflect long-term site condition.

[1] US Environmental Protection Agency, National Exposure Research Laboratory, 26 W. Martin Luther King Blvd., MS 642, Cincinnati, OH 45268

- Macroinvertebrates are diverse in their habitat requirements, feeding modes and tolerance to pollutants and other stressors (e.g., low dissolved oxygen, temperature changes and sedimentation). They, therefore, provide valuable information about ecosystem health and source(s) of impairment.
- In most cases, sampling macroinvertebrate assemblages is relatively easy, requiring few people and inexpensive gear.

Despite their widespread use in streams, benthic macroinvertebrates have rarely been incorporated into formal bioassessments of large rivers. There is a general belief that macroinvertebrate assemblages become less diverse and more tolerant in large rivers (i.e., that the replacement of sensitive stoneflies and other "coldwater" taxa is a common occurrence). The unstable fine sediments typical of many large river bottoms generally support fewer taxa than smaller streams and rivers that have larger substrate sizes (Allan 1995). Due to the long history of benthic sampling in smaller streams, most of the common quantitative and qualitative methods for sampling macroinvertebrate assemblages require easy access to substrates.

Macroinvertebrate sampling in large rivers presents programs with several difficulties common to all assemblage surveys relating to spatial scale and sampling logistics:

- The diversity of habitat types in large rivers (e.g., back channels, inlets, floodplain wetlands) makes it difficult to obtain a standardized and representative sample.
- Balancing the appropriate reach length with time and cost constraints for macroinvertebrate assessment is more difficult as repeating habitat units are spaced farther apart and meander wavelength increases.
- Identifying reference conditions for large rivers is difficult due to the large areas of intensive human land use.
- Identifying specific stressors or causes of impairment, as required by the CWA §303(d), is more difficult in large rivers because of the cumulative impact of multiple stressors that result from disturbances within large drainage areas.
- Large river macroinvertebrate sampling is more costly and hazardous than on wadeable streams because it typically requires use of a boat on navigable waterways that are often subject to commercial traffic.

Despite these obstacles, many researchers have sampled large river macroinvertebrate assemblages for inventory and monitoring purposes or for targeted sampling around point sources of pollution. More recently, efforts have increased to standardize large and great river macroinvertebrate assessment programs (Lazorchak et al. 2000, Merritt et al. 2005, Angradi 2006). There is a lack of assessment information that characterizes the condition of large rivers and the need for these bioassessment programs has risen with this recognition. Table 6-1 provides a brief summary of five of these large river bioassessment programs. Michigan DEQ's macroinvertebrate bioassessment program is also highlighted in this chapter.

TABLE 6-1. A comparison of large rivers program macroinvertebrate sampling approaches.

Program	Protocol Summary	Citation
USEPA EMAP-Surface Waters	An acceptable sampling point is identified in an area away from the river margin and less than or equal to 1 m depth. Two kick net samples are taken at each of 11 transects and composited. Samples are placed in a bucket, detritus is removed without removing the macroinvertebrates. Samples are placed in plastic jars and filled with 95% ethanol to preserve the sample.	Lazorchak et al. 2000
USGS NAWQA Program	The types of instream habitats are recorded and semi-quantitative samples are taken to determine relative abundance when it is possible. Semi-quantitative samples are taken from the richest targeted habitat (RTH). Typically, this is riffle habitat or woody snags. A 0.25 m^2 area is sampled using a slack sampler (500-µm mesh) in riffles. Two snags are sampled by disturbing snags upstream as a sampler for woody snag sampling. Area of the snags sampled is estimated for that habitat. Qualitative samples: Proportional multi-habitat samples are taken along the study reach. Samples are taken with a D-frame kick net and visual collections and some grab collections are made. Water depth and substrate type are recorded. Large debris is removed along with large crayfish, hellgrammites and mussels. The sample is placed in a standardized bottle with a 10% buffered formalin solution.	Moulton et al. 2002
Ohio Environmental Protection Agency (OEPA)	Quantitative methods: A modified Hester-Dendy (H-D) multiple-plate artificial substrate sampler, with eight plates and 12 spacers, is placed in the river and tied to a concrete construction block. In rivers more than four feet deep, a floater is attached to keep it within four feet of the surface. Whenever possible, the samplers are placed in runs. A sample consists of three multiple-plate samplers. Samples are retrieved by cutting them from the block and placing them in one-quart plastic containers while still under water. Formalin is added to make a 10% solution. Qualitative samples are collected at the same time for organisms in the natural substrate. Qualitative methods: Each station is sampled at least once between June 15 and September 30. If possible, a riffle, run, pool, and margin are sampled at each site. Organisms are collected using a triangle ring frame 30-µm mesh dip net and field picked with forceps for at least 30 minutes until no new taxa can be identified. The organisms are preserved in 70% ethanol. In both methods, a station description sheet is filled out and the length of time spent sampling is recorded.	Ohio EPA 1989
Kentucky Division of Water (KDOW)	The 20-jab method is used augmented by dredge samples, a wood sample, and rock picking along a 300-meter reach of the river. The sample is placed in a 600-µm mesh washing bucket where the macroinvertebrates are removed and placed in 70% ethanol. When possible, 15 large rocks and 6 m of wood are picked and washed.	Kentucky DOW 2002
Michigan Department of Environmental Quality (MIDEQ)	The individual habitat types are counted. Habitats must be within the littoral area and large enough to collect a 15-second sample. A 15-second sample is taken for every habitat type with a D-frame net, with a mesh size of 500 µm. The net is emptied into a bucket or pan filled with water. Detritus is removed before placing the sample in a 500-µm sieve to remove excess water. The sample is placed in 95% ethanol.	Merritt et al. 2005

PROGRAM HIGHLIGHT

Qualitative Biological and Habitat Survey Protocols for Michigan's Non-Wadeable Rivers Submitted to the Michigan Department of Environmental Quality (Michigan DEQ) (Merritt et al. 2005)

The Michigan DEQ is responsible for water quality monitoring in the state. As part of their Strategic Environmental Quality Monitoring Program, they have conducted or are conducting biological and habitat surveys across the state to assess more than 80% of their stream and river miles. The specific goals of their program are to:

1. determine whether waters of the state are attaining standards for aquatic life,
2. assess the biological condition of the waters of the state,
3. determine the extent to which sedimentation in surface waters is impacting indigenous aquatic life,
4. determine whether the biological condition of surface waters is changing with time,
5. assess the effectiveness of best management practices (BMPs) and other restoration efforts in protecting and restoring biological integrity and physical habitat,
6. evaluate the overall effectiveness of DEQ programs in protecting the biological integrity of surface waters,
7. identify waters that are high quality or not meeting standards, and
8. identify the waters of the state that are impacted by nuisance aquatic plants, algae, and bacterial slimes.

The Michigan DEQ has an existing rapid assessment protocol for wadeable streams, but it is not applicable for their non-wadeable rivers. They contracted with Michigan State University scientists to develop a non-wadeable method for assessing macroinvertebrate and habitat condition.

Michigan DEQ Macroinvertebrate Sampling Methods

The Michigan DEQ macroinvertebrate method was developed using data from 45 locations on 13 non-wadeable rivers from across the state. The approach requires sampling between June and September during stable discharge and is designed to take approximately 0.5 days for a two-person crew. The sampling unit is a 2000-m reach split into 11 equally spaced transects. Along each transect, two littoral (20-m long X 10-m wide) plots are established. One plot, chosen by a coin flip, is sampled at each transect. If large woody debris (LWD) is present along eight of the 11 transects, then only LWD is sampled. If not, then all available habitats are sampled in each plot (fine particulate organic matter (FPOM), sand, gravel, cobble, LWD, and macrophytes). Each available habitat is sampled for 15 seconds using a D-frame dip net with 500-μm mesh. If flow is insufficient, nets are swept through the habitats. For cobble, a cobble of at least 15-cm in width is placed in a bucket and brushed with a toilet brush. Similarly, LWD is brushed either above the kick net or the kick net is swept through the water. The net is swept through macrophytes for 15 seconds to dislodge organisms. Each sample is placed in a white enamel pan with water and the nets are cleaned. The pan material is sieved (500 μm) to remove excess water and placed into a bucket with 95% ethanol. Individual transect samples are composited into one bucket. A plankton splitter is used to divide the composite sample into quarters. All the individuals in the quarter sample are counted and identified to family level. The macroinvertebrate data are used to calculate 13 individual metrics combined into an overall multimetric score for each site. The individual metrics are Plecoptera richness, EPT richness, Diptera richness, percent dominance, percent Diptera, total richness, functional feeding group diversity, and the ratio of (#scrapers + #collector-filterers)/(#collector-gatherers + #shredders). Individual metrics are scored differently depending on whether the multihabitat or LWD sampling methods are used, and different metrics are weighed differently based on how much among-site variability they explained. Final scores are broken into four classes: 0-15 (poor), 16-30 (fair), 31-45 (good) and 46-60 (excellent). For detailed descriptions of the metric development, please contact Michigan DEQ.

This chapter provides a review of several different active and passive methods for benthic macroinvertebrates in large rivers. It also gives recommendations for a protocol (Flotemersch and Blocksom 2004, Flotemersch et al. 2006) borne from some of these methods. If field sampling methods other than those recommended here are more suitable for a particular program, they should be thoroughly tested to ensure that they return data of sufficient quality and provide the capacity to address their intended and stated purposes.

6.2 Field Sampling Methods

Numerous studies have demonstrated that dramatic differences can exist among large river benthic sampling methods (Anderson and Mason 1968, Rabeni and Gibbs 1978, Slack et al. 1986, Diamond et al. 1994, Humphries et al. 1998, Leland and Fend 1998, Hoffman 2003, Poulton et al. 2003, Blocksom and Flotemersch 2005). Benthic grab/dredge samples or the use of artificial substrates have historically been the most common collection methods for large river macroinvertebrates and they remain common choices for many researchers. More recently, however, active sampling methods, such as kick net or D-net sampling along the shoreline and scraping large woody debris (LWD), have become more common in an effort to assess a river reach and to sample the most productive (per unit area) habitats for macroinvertebrates. Flow regime and substrate stability are major factors influencing distribution of large river macroinvertebrates. The location of benthic sampling within the channel can greatly influence results (e.g., high-velocity main channel vs low-velocity shoreline areas; fine sediments vs vegetation or larger mineral substrates). Most sampling methods, however, are only appropriate for, or artificially represent, one substrate type or area. A combination of methods and sample locations may prove best for assessment, but the choice of these methods should depend upon specific management questions and available resources. Numerous authors have provided comprehensive reviews of benthic macroinvertebrate sampling methods (Rosenberg and Resh 1982, Flannagan and Rosenberg 1982, Klemm et al. 1990, Merritt et al. 1996). The following sections provide a brief review of sampling methods as they relate to large river sampling.

6.2.1 Passive Methods

Passive methods include artificial substrate samplers defined by Klemm et al. (1990) as "devices made of natural or artificial materials of various composition and configuration that are placed in the water for a predetermined period of exposure and depth for colonization." Artificial substrate samplers can be used to obtain qualitative and quantitative macroinvertebrate samples and they have been recommended for use in deep or turbid waters and in areas with muddy, sandy, or otherwise unstable bottoms (Taylor and Kovats 1995). Exposure periods are typically four to six weeks to allow for colonization of biofilm and subsequent macroinvertebrate fauna and samplers are usually deployed at 1- to 3-m depths. Deployment depth is chosen so that receding or rising waters during the exposure period will not leave samplers dry or too deep to retrieve and so the samplers will be in the photic zone. Typically, 4 or 5 Hester-Dendy's (H-D's) or 3 rock baskets are placed per sampling reach and the data are composited from all samplers retrieved. Placing multiple samples per reach and compositing data also helps buffer the effects of loss or vandalism. Upon retrieval, samplers are slowly lifted to the water surface. If possible, a net is placed downstream or around the sampler to collect any organisms that fall off or leave the samplers during removal. The samplers are placed in a bucket and the substrates are scraped or brushed into the bucket. The bucket contents are then sieved and preserved for laboratory processing. Alternatively, some choose to return the complete sampler to the laboratory for processing. Some advantages and disadvantages to using artificial substrate samplers are summarized in Table 6-2.

6.2.1.1 Rock Basket Samplers

Rock baskets are passive samplers that typically consist of plastic or wire baskets (e.g., square or cylindrical barbeque grilling baskets) filled with native rock or gravel. Baskets are typically tied to a rope that is fastened on the shore and then dropped into the river. Standard-sized quarry rocks can be used in baskets to help standardize surface areas and facilitate density calculations.

Rock basket samplers can have the advantage of providing a natural substrate with irregular surfaces and interstitial spaces that mimic those of the natural environment. However, rock baskets have the disadvantage of being slightly less standardized and quantitative than H-D type samplers. Rock baskets (similar to Figure 6-1) have been successfully used in Ohio (Anderson and Mason 1968, Mason et al. 1973), Maine (Rabeni and Gibbs 1978), Pennsylvania (Hoffman 2003) and along the Missouri River (Poulton et al. 2003). Rock-filled trays are similar to baskets and have been used to sample smaller streams (e.g., Townsend and Hildrew 1976, Clements 1991), but they are not as effective in large rivers due to their instability in fast currents.

TABLE 6-2. Advantages and disadvantages of artificial substrate samplers.

ADVANTAGES/DISADVANTAGES
Numerous researchers have described artificial substrate samplers and their relative advantages and disadvantages (Rosenberg and Resh 1982, Flannagan and Rosenberg 1982, Klemm et al. 1990, Merritt et al. 1996). Some of these are given below.
Advantages
1) Allow quantitative collection of benthic macroinvertebrates from sites that cannot be effectively sampled using other conventional benthic sampling methods.
2) Can be used effectively in shallow or deep water, making them useful for sampling throughout the large river mosaic.
3) Easy to use and usually require less time and effort in the field than active methods. The ease of deployment and retrieval helps reduce sampling variability associated with the operator.
4) Generally accumulate very little debris during incubations making sample processing more efficient.
5) Can be especially effective in reflecting water quality as a result of the standardized habitat they provide.
Disadvantages
1) Require two trips to the sample site (for deployment and retrieval) that can add time, cost and other logistical constraints.
2) Measure colonization potential rather than the resident assemblage.
3) Loss of individuals when retrieving the sampler can bias results.
4) Can effectively indicate water quality, but not sediment or other habitat quality.
5) Exact placement of individual sampler units can skew results (e.g., high vs low velocity).
6) Damage or loss of artificial substrates can occur due to vandalism, high flows, shifting channels or they may be left dry during drought conditions.

FIGURE 6-1. Rock-filled wire basket used as introduced substrate.

6.2.1.2 Multiplate Samplers

The most common type of artificial substrate samplers are variations of the H-D multiplate sampler (Hester and Dendy 1962). Many monitoring programs use these samplers for assessment of both point and non-point sources of pollution in large rivers. Configurations may vary greatly in size, shape, and number of plates used, but all consist of round or square plates (typically made of Masonite board or porcelain) with spacers placed in between and bolted together to form stacks (Figure 6-2). Spacing between plates is typically varied to provide different refuge sizes and flow regimes within the stacks. Stacks are tied together and attached horizontally to a brick or cinder block and placed on the river bed (Figure 6-2). Alternatively, stacks may be positioned vertically by screwing the bolts into the anchor blocks. These samplers have been successfully used on many large rivers, notably as part of standard programs in Florida, Wisconsin, and Ohio.

6.2.1.3 Other Passive Methods

Although rock baskets and H-Ds are by far the most common artificial substrates used in benthic studies, a number of other passive samplers may be used. Beak trays are round metal trays with expanded mesh inserts for colonization (Beak et al. 1973). Upon retrieval, a lid is lowered by rope to cover the tray and the sampler is lifted from the water. Beak trays can be effective in collecting macroinvertebrates from unstable or sandy substrates, but they have been shown to collect fewer taxa and individuals than multiplate and rock basket samplers (Slack et al. 1986). Flannagan and Rosenberg (1982) described several other types of samplers of various size, shape, and composition that have been placed on the substrate or suspended in the water column for sampling benthic macroinvertebrates. These include mesh bags, boards, tiles, bricks, plastic sheets or ropes (vegetation mimics), and buried pots, baskets or trays filled with organic or inorganic materials. However, many of these devices are inadequate due to the depth, elevated turbidity, and high flows of many large rivers. Drift nets are another passive method that can be used to sample large river macroinvertebrates if flow is adequate (Lazorchak et al. 2000), but

studies have shown drift net data are highly variable compared to other methods if not deployed properly (Blocksom and Flotemersch 2005). Poor performance of drift nets can be attributed to low velocities, length of deployment periods, and deployment season. Macroinvertebrate drift densities peak at night (Resh and Rosenberg 1984), so evening deployment of drift nets would be required to maximize their effectiveness.

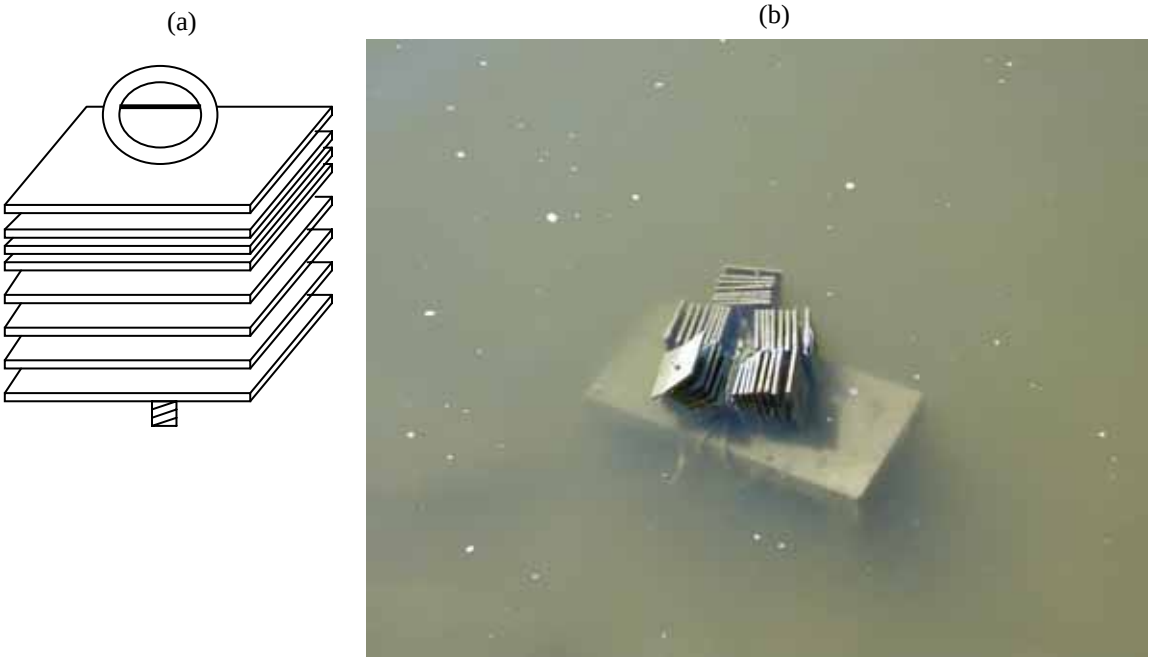


FIGURE 6-2. a) Modified Hester-Dendy multiplate artificial substrate sampler; b) Exposed Hester-Dendy sampler attached to cinder block anchor.

6.2.2 Active Methods

Active methods for sampling macroinvertebrates include a wide variety of sampling approaches that can be grouped into two categories: deep water and shallow water. Active methods are quantitative, semi-quantitative or qualitative and can be used alone or in combination. All active methods have the advantage of only requiring one trip to the sample site, thereby reducing travel cost and effort over passive methods. In addition, these methods focus on measuring or characterizing the existing macroinvertebrate assemblage at a site rather than colonization potential. Disadvantages include a generally high degree of sample variability and high sample debris accumulation that increases sample processing time.

6.2.2.1 Deep Water: Main Channel Sampling

Deep habitats of large rivers can be sampled from a boat using various dredge or bottom grab sampling devices described by Klemm et al. (1990) (e.g., Peterson, Ponar, Ekman, van Veen samples). These samplers are specifically designed for sampling less-stable substrates (e.g.,

sand, silt) usually found in depositional areas. Grab samplers are lowered to the bottom and penetrate the sediments under their own weight. Jaws of the samplers are forced shut by weights, levers, springs or cables to retrieve samples from a known surface area. Although these samplers are most commonly used in deep water, some can be adapted to shallow waters by rigging samplers on poles or by physically pushing samplers into the substrate. Bottom-grab samplers are available in several different designs, each with their own subtle advantages and disadvantages for specific habitats or substrate types (see Klemm et al. 1990 for a review) (Table 6-3).

TABLE 6-3. Advantages and disadvantages of bottom grab samplers.

ADVANTAGES/DISADVANTAGES
Advantages:
1) Requires only one site visit for sample collection, thus reducing overall cost and effort.
2) Results in a sample of the macroinvertebrate assemblage at the site.
3) Effective in sampling deepwater habitats not reachable by most conventional methods.
4) Effective for sampling organisms that burrow in soft sediments and are often the most abundant in large rivers (e.g., oligochaetes and burrowing mayflies).
5) Requires little training and can collect standardized, quantitative benthic samples.
Disadvantages:
1) Usually operated "blind," due to elevated turbidity common on large rivers, with little or no knowledge of specific substrate type that is being sampled (i.e., silt, sand or gravel).
2) Ineffective at sampling rocky or hard substrates.
3) Organisms often lost in "washout" as devices are lifted onto the boat and removed from water.
4) "Jaws" of many samplers can be easily blocked by debris.
5) Some dredges are heavy and cumbersome, occasionally requiring a mechanical winch.
6) Using these methods, reducing sampling variability by stratification is difficult due to the patchy distribution of organisms in sand and silt substrates.
7) Proper operation of many dredge samplers prevents them from being used in habitats with significant flow rates.

Deep waters of large river main channels can also be sampled by SCUBA divers. A diver-operated dome sampler contains a battery-operated pump that moves materials dislodged by a diver into a Nitex mesh sample bag (Gale and Thompson 1975). This quantitative method can be used to successfully sample a variety of deepwater habitats, including coarse substrates. Divers can also operate other devices for sampling benthos, including suction samplers, grab samplers, and corers; and can be used for placement and retrieval of artificial substrates (Gale and Thompson 1974, Klemm et al. 1990). A major advantage of using SCUBA divers is that the divers can see the habitats, making proportional or habitat-specific sampling of river bottoms

more feasible. However, cost, logistical and safety constraints usually render this method impractical for widespread and routine application.

Although more frequently applied in lakes (Muli and Mavuti 2001) and oceans, benthic trawls have also been used to sample the macrobenthos of deep large river main channels. Wright et al. (2000) used benthic trawls to survey the macroinvertebrate fauna of the Thames River. Similarly, benthic trawls have been used in estuarine sections of the Lower St. Johns River in Florida (Mason 1998) and in the Columbia River estuary (Jones et al. 1990). For additional information on trawl selectivity and efficiency, consult Stokesbury et al. (1999).

6.2.2.2 Shallow Water: Shoreline Sampling

Approaches for large river shoreline sampling are similar to well-developed methods for wadeable streams (Ohio EPA 1989, Barbour et al. 1999, Klemm et al. 2000, Flotemersch et al. 2001, Moulton et al. 2002, Merritt et al. 2005). They are often used in large rivers to help avoid logistical constraints encountered in deepwater sampling from a boat in the main channel (see Table 6.4 for a description of advantages and disadvantages). These methods often involve wading in shallow near-shore areas of larger rivers. Even though the wadeable shore zone only accounts for a small proportion of the entire river channel, it may be the most productive and diverse zone for benthic macroinvertebrates (Wetzel 2001). The shallows along main-channel margins have the greatest light penetration for benthic algae and aquatic macrophytes. Allochthonous organic matter also accumulates in the shallows as a result of direct riparian inputs and from backeddies and currents that deposit LWD and FPOM along the shore. The shoreline substrates of many large rivers tend to be dominated by LWD and other stable substrates, such as cobbles and boulders. As a result of their relatively high habitat complexity and productivity, large river shorelines are similar to the highly productive littoral zones of lentic ecosystems. This is particularly true of large, deep rivers where flow is heavily regulated.

Most sampling approaches used for wadeable streams can be used in the littoral areas of large rivers. Active sampling methods along the shoreline include a variety of qualitative, semi-quantitative, and quantitative techniques. When sampling larger substrate types that can be easily handled (e.g., rocks, woody debris/snags, macrophytes), macroinvertebrates may be removed by scrubbing the substrate with a soft brush or picking them individually with forceps. Conventional dip net-based methods include kicks, dips, jabs, or sweeps in one or more habitat types. D-frame or rectangular kick nets are commonly used at the wadeable margins and are most effective when flow is adequate to carry dislodged organisms into the net. Surber and Hess samplers (which quantitatively sample fixed areas) can also be used, but require greater flow velocity than do dip net methods. Although kick nets are most commonly used; grab samplers, corers, and suction samplers can also be used to sample fine sediments along the shoreline. Table 6-4 list some general advantages and disadvantages of active shoreline benthic sampling.

6.2.2.3 Snag Sampling

Sampling woody debris or “snags” (usually >10 cm in diameter) is another method that can be used either in the deep waters of the main channel, from a boat, or in shallow shoreline areas.

These substrates are natural and stable and have been recognized as some of the most productive macroinvertebrate habitats of large rivers, particularly in rivers dominated by unstable sandy bottoms (e.g., Benke et al. 1985, Benke 2001, Merritt et al. 2005). Snags are most frequently sampled by placing a dip net on the downstream side and gently scrubbing the snag surface with a soft brush, allowing the current to carry dislodged material into the net. Although a regular dip net is often used, Angradi (2006) describes a specialized "snag net" that resembles a D-frame net except that the frame is constructed so that the net fits over half the circumference of the snag. Snag bags have also been used to collect macroinvertebrates from woody debris (Growns et al. 1999). Snags have an advantage over artificial substrates because, in addition to providing stable habitats, they are natural substrates and the decomposing wood and associated biofilms serve as a food resource for macroinvertebrates. However, irregular size and shape often make it difficult to standardize the area sampled. The length of time the snag has been in the water, or the period of colonization, is also typically unknown. Yet it may be possible to use conditioned snag habitats for preliminary bioassessment, or "bioreconnaissance," efforts on large rivers. Snag sampling is currently being incorporated into both large river and great river macroinvertebrate sampling protocols of the USEPA (Angradi 2006, Johnson et al. 2004) and the Michigan DEQ (Merritt et al. 2005).

TABLE 6-4. Advantages and disadvantages of shoreline benthic sampling.

ADVANTAGES/DISADVANTAGES
Advantages:
1) Requires only one site visit for sample collection, thus reducing overall cost and effort.
2) Assesses the macroinvertebrate assemblage found in the study reach.
3) Doesn't require a boat, therefore reducing cost and hazards associated with boat operation, if shoreline sample zone is wadeable and easily accessible.
4) Shallow shoreline habitats are often readily observable, making it possible to target specific habitats or to sample habitats proportionately.
5) Dip-net methods can be used to sample a variety of both stable (e.g. rocks, woody debris, macrophytes, cobble) and unstable (e.g., sand, silt, muck) habitats, enhancing sample representativeness.
Disadvantages:
1) Samples can be variable due to diversity of habitat types and the patchy distribution of organisms, potentially requiring more replicate samples to reduce this variability.
2) Sorting macroinvertebrates from the debris of shoreline samples increases sample processing time and costs.
3) Difficult or impossible where there are steep drop-offs or sheer cliffs at rivers edge.

6.3 The Large River Bioassessment Protocol (LR-BP) for Benthic Macroinvertebrate Sampling

The LR-BP method is a hybrid of USEPA-EMAP (Lazorchak et al. 2000), USEPA-RBP (Barbour et al. 1999) and USGS-NAWQA (Moulton et al. 2002) sampling methods. The LR-BP uses transect sampling and can be applied in a systematic, unbiased manner for bioassessment. The LR-BP is a combination of semi-quantitative multi-habitat sampling methods applied in a systematic randomized fashion that has been studied for its performance characteristics and variability (Flotemersch et al. 2006) and was designed to be standardized, quantitative and user friendly. It incorporates proportional multi-habitat sampling and, therefore, should accurately reflect site condition. This method was shown to be responsive to a gradient of disturbance and can be used on a variety of large rivers (Flotemersch and Blocksom 2006).

The LR-BP specifies a reach length of 500 m because it: 1) has been shown to provide representative samples (Blocksom and Flotemersch 2006 [submitted]); 2) is manageable for investigators due to the entire reach usually being observable from a single point; and 3) works well for large river fish bioassessment when both banks are electrofished and, thus, provides comparable sampling reaches for both assemblages (1000 m total shoreline) (Flotemersch and Blocksom 2005). The target sample location (e.g., established by GPS coordinates for a probabilistic design) indicates the downstream end of the reach where sampling begins. At each site, there are a total of six transects. Transect A is located at the downstream end of the reach with the remaining five transects at 100 m, 200 m, 300 m, 400 m and 500 m (Figure 6-3). At each transect, a 10-m sample zone (5 m on each side of transect) on each bank defines where macroinvertebrates will be collected. The zone extends from the edge of water to the mid-point of the river or until depth exceeds 1 m (Figure 6-3), but sampling is largely bank-oriented except in shallow rivers. Six sweeps, each 0.5 m in length, are collected within the zone using a D-frame net (500-µm mesh). Each sweep covers 0.15 m^2 of substrate (i.e., net width of 0.3 m and a 0.5 m length of pass); therefore, six sweeps will cover an area of 0.9 m^2. The six sweeps are proportionately allocated based on available habitat within the 10-m sample zone (e.g., snags, macrophytes, cobble). This method negates the need for separate collection nets in the field and helps standardize the area sampled. If water at a site is more than 1 m deep at the waters edge, the six sweeps should be collected from a boat if possible. Each transect has two zones (one on each bank) and samples from the entire reach are composited into a single sample. This results in each sample containing debris and organisms from 12 separate zones (total of ~12 m^2) that represent the 500-m reach.

6.4 Field Preservation

In most macroinvertebrate sampling protocols, multiple steps are involved in processing samples in the field. Sample material is composited for the entire site, and then placed into a sieve bucket to drain excess water and allow washing of fine sediments. The number of samples comprising the composite sample will depend on the sampling method used at the site. Large objects (e.g., rocks, woody debris) are inspected, attached invertebrates are picked from them, and the objects are returned to the river. Each piece of substrate is then gently washed or scrubbed to remove attached organisms. Substrate pieces are removed from the bucket or sieve after cleaning.

After sieving, samples are typically transferred to a suitable container and preserved with ethanol (70% final concentration) or a 10% buffered formalin solution. Buffered formalin may be a better preservative for large river benthic samples as they typically contain a greater number of soft-bodied oligochaetes and leeches that are inadequately preserved by alcohol. Many investigators choose to first fix the sample in formalin and later transfer the sample to ethanol prior to laboratory processing (Klemm et al. 1990). In addition to externally labeling the sample container at the site, it is advisable to use an internal label. Additional details on field processing of macroinvertebrate samples are provided by Klemm et al. (1990).

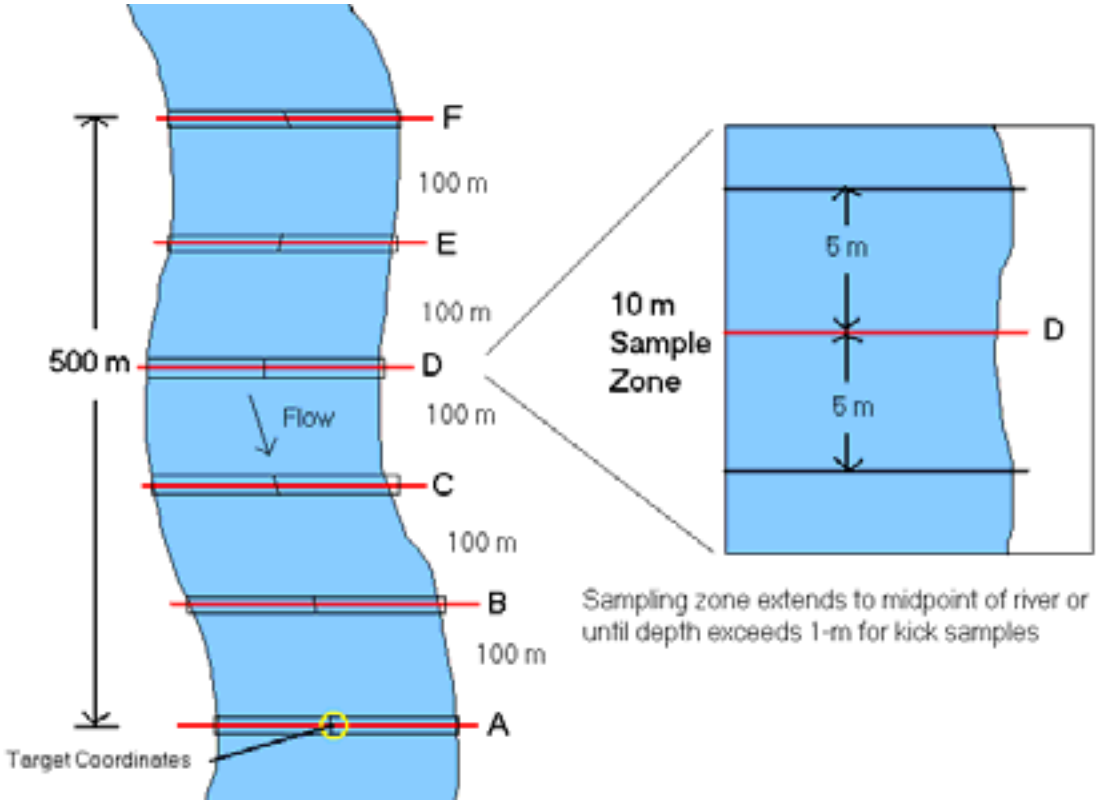


FIGURE 6-3. Example of the six transects and 6 sample zones for collection of benthic macroinvertebrates in large rivers using the LR-BP design.

6.5 Laboratory Processing

There are three components to laboratory processing of benthic macroinvertebrate samples: sorting/subsampling, taxonomic identifications and counts (i.e., enumeration). Several questions should be addressed prior to initiating laboratory processing.

- Will samples be sorted in their entirety, or will they be subsampled?
- If samples are to be subsampled, will the process be based on fixed volume or fixed count?
- If fixed count, what is the target (e.g., 100, 200, 300, 500 organisms)?

- Is there a target taxonomic level (e.g., genus), the lowest practical taxonomic level, or does it vary by group?
- What, if any, rules are there for counting?

6.5.1 Sorting and Subsampling

Although it is widely recognized that subsampling helps to manage the level of effort associated with bioassessment laboratory work (Carter and Resh 2001), the practice has been the subject of much debate (Courtemanch 1996, Barbour and Gerritsen 1996, Vinson and Hawkins 1996). If a fixed count method is used, power analyses can determine the most appropriate number of targeted organisms (Ferraro et al. 1989, Barbour and Gerritsen 1996). Fixed organism counts vary greatly among monitoring agencies (Carter and Resh 2001), with 100, 200, 300 and 500 counts being most often used (Plafkin et al. 1989, Barbour et al. 1999, Cao and Hawkins 2005). As part of the LR-BP development process, Flotemersch and Blocksom (2005) provided an assessment of the effect subsample size had on metric performance from large river benthic samples. They concluded that a 500-organism count was best, based on examination of the relative increase in richness metric values (< 2%) between successive 100-organism counts. However, a 300-organism count was deemed sufficient for most study needs. Others have recommended higher fixed counts, including a minimum of 600 in wadeable streams (Cao and Hawkins 2005).

If organisms are missed during the sorting process, bias is introduced in the resulting data. Thus, the primary goal of sorting is to completely separate organisms from organic and inorganic material (e.g., detritus, sediment) in the sample. A secondary goal of sorting is to provide the taxonomist with a sample for which the majority of specimens are identifiable. Although it is not the decision of the sorter whether an organism is identifiable, straightforward rules can be applied that minimize specimen loss (Table 6-5). If a sorter is uncertain about whether an organism is countable, the specimen should be placed in the vial and not added to the rough count total.

TABLE 6-5. Example list of counting "rules": what not to count.

Organisms that should *not* be counted include:
a) Non-benthic organisms, such as free-swimming gyrinid adults or surface-dwelling veliids (Insecta:Heteroptera)
b) Empty mollusk shells (Mollusca:Bivalvia)
c) Non-headed worm fragments (Oligochaeta)
d) Terrestrial insects (incidentals)
e) Copepoda
f) Exuviae (molted "skins")

The sorting/subsampling process is based on randomly selecting portions of the sample detritus spread over a gridded Caton screen (Caton 1991, Barbour et al. 1999; Figure 6-4a, b). Prior to beginning the sorting/subsampling process, it is important that the sample be mixed thoroughly

and distributed evenly across the sorting tray to reduce the effect of organism clumping that may have occurred in the sample container. The grids are removed from the screen, placed in a sorting tray, and all organisms removed; the process is completed until the rough count by the sorter exceeds the target subsample size. This process should produce at least three containers per sample (all of which should be clearly labeled):

- Subsample to be given to taxonomist,
- Sort residue, to be checked for missed specimens, and
- Unsorted sample remains to be used for additional sorting, if necessary.

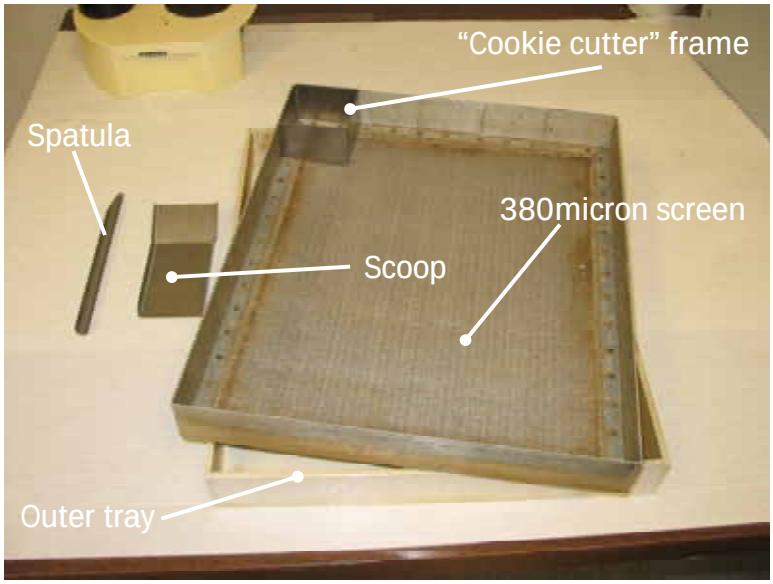


FIGURE 6-4a. Gridded screen (Caton 1991) used to facilitate subsampling.

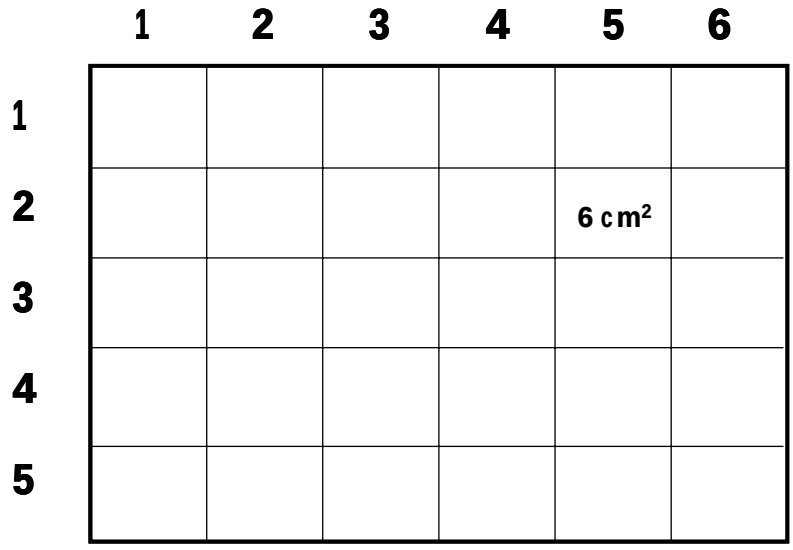


FIGURE 6-4b. Schematic diagram of the Caton gridded subsampling screen, consisting of 30 6-cm^2 grids.

6.5.2 Taxonomy and Enumeration

The next step of the laboratory process is identifying the organisms within the subsample. A major question associated with taxonomy is the hierarchical target levels required of the taxonomist, including order, family, genus, species or the lowest practical taxonomic level (LPTL). While family level is used effectively in some monitoring programs (Carter and Resh 2001), the taxonomic level primarily used in most routine monitoring programs is genus. However, even with genus as the target, many programs often treat selected groups, such as midges (Chironomidae) and worms (Oligochaeta), differently due to the need for slide-mounting. Slide mounting specimens in these two groups is usually necessary to attain genus level nomenclature, and sometimes even tribal. Because taxonomy is a major potential source of error in monitoring data sets (Stribling et al. 2003), it is critical to define taxonomic expectations and to treat all samples consistently, both by a single taxonomist and among multiple taxonomists. This, in part, requires specifying both hierarchical targets and counting rules.

An example list of taxonomic target levels is shown in Table 6-6. These target levels define the level of effort that should be applied to each specimen. If it is not possible to attain these levels for certain specimens due to, for example, the presence of early instars, damage, or poor slide mounts, the taxonomist provides a more coarse-level identification.

When a taxonomist receives samples for identification, depending upon the rigor of the sorting process (see Section 6.3.1), the samples may contain specimens that either cannot be identified, or should not be included in the sample (Table 6-6). The final screen of sample integrity is the responsibility of the taxonomist, who determines which specimens should remain unrecorded (for any of the reasons stated above). Beyond this, the principal responsibility of the taxonomist is to record and report the taxa in the sample and the number of individuals of each taxon.

Programs should use the most current and accepted keys and nomenclature. *An Introduction to the Aquatic Insects of North America* (Merritt and Cummins 1996) is useful for identifying the majority of aquatic insects in North America to genus level. By their very nature, most taxonomic keys are obsolete soon after publication; however, research taxonomists do not discontinue research once keys are available. Thus, it is often necessary to have access to and be familiar with ongoing research in different taxonomic groups. Other keys are also necessary for non-insect benthic macroinvertebrates that will be encountered, such as Oligochaeta, Mollusca, Acari, Crustacea, Platyhelminthes and others. Klemm et al. (1990) and Merritt and Cummins (1996) provide an exhaustive list of taxonomic literature for all major groups of freshwater benthic macroinvertebrates. Although it is not current for all taxa, the integrated taxonomic information system (ITIS; http://www.itis.usda.gov/) has served as a clearinghouse for accepted nomenclature, including validity, authorship and spelling.

TABLE 6-6. Example of taxonomic hierarchical targets used in benthic macroinvertebrate identifications.

TAXON		TARGET
PHYLUM ANNELIDA		
Class Branchiobdellida		Genus
Class Hirudinea		Genus
Class Oligochaeta		Genus
Class Polychaeta		Family
PHYLUM ARTHROPODA		
Class Arachnoidea		
	Acari	Genus
Class Insecta		
	Coleoptera	Genus
	Diptera	*Identify all to genus except in the following cases:*
	Chironomidae	Genus (tribe or subfamily, if specified)
	Dolichopodidae	Family
	Phoridae	Family
	Scathophagidae	Family
	Syrphidae	Family
	Ephemeroptera	Genus
	Heteroptera	Genus
	Lepidoptera	Genus
	Megaloptera	Genus
	Odonata	Genus
	Plecoptera	Genus
	Trichoptera	Genus
Class Malacostraca		Genus
	Amphipoda	Genus
	Decapoda	Genus
	Isopoda	Genus
	Mysidacea	Genus
Class Ostracoda		Genus
PHYLUM COELENTERATA		
PHYLUM MOLLUSCA		
Class Bivalvia		Genus
Class Gastropoda		*Identify all to genus except in the following cases:*
	Family Hydrobiidae	Family
PHYLUM NEMERTEA		Genus

6.6 Data Entry

Taxonomic nomenclature and counts are usually entered into the data management system directly from handwritten bench or field sheets. Depending upon the system used, there may be an autocomplete function that helps prevent misspellings. There are two methods for assuring

accuracy in data entry. One is the double entry of all data by two separate individuals, and then performing a direct match between databases. Where there are differences, it is determined which database is in error, and corrections are made. The second approach is to perform a 100% comparison of all data entered to handwritten data sheets. Comparisons should be performed by someone other than the primary data enterer. When errors are found, they are hand-edited for documentation, and corrections are made electronically. The rates of data entry errors are recorded and segregated by data type (e.g., fish, benthic macroinvertebrates, periphyton, header information, latitude and longitude, physical habitat, and water chemistry).

6.7 Data Reduction (Metric Calculation)

This section focuses on activities that convert raw data (taxa lists and counts) into numeric terms (metrics) to be used for subsequent analyses, (e.g., metric calculation). For example, Blocksom and Flotemersch (2005) tested 42 metrics relative to different sampling methods, mesh sizes, and habitat types (Table 6-7). Twenty-seven of the 41 metrics (66%) are taxonomically based. Those remaining require tolerance value and functional feeding group designations to calculate the metrics.

To ensure that database queries are correct and result in the intended metric values, a subset of values should be recalculated by hand. One metric is calculated for all samples, all metrics are calculated for one sample. When recalculated values differ from those values in the matrix, the 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.

6.8 Final Index and Site Assessment

Approximately 56 state or tribal agencies currently use macroinvertebrates in biomonitoring or bioassessment programs in the USA (USEPA 2002). Of these, more than 40 have developed an index of some type (multimetric or multivariate predictive) for use in site assessment. These indices are developed using reference sites. The final assessment for a site is usually determined based on a site score relative to the distribution of reference site scores. Approaches for scoring the reference distribution vary and depending on several factors (Barbour et al. 1999). The process for developing these indices is described in detail in Chapter 8.

TABLE 6-7. Benthic macroinvertebrate metrics evaluated by Blocksom and Flotemersch (2005) for responsiveness to measured disturbance gradients in large rivers.

Metric (by category)	Metric Description
Richness and diversity	
Number of taxa	The count of unique taxa in the sample. A standard level of identification (family, genus, species) must be defined for each taxonomic group
Number of Ephemeroptera, Plecoptera, Trichoptera (EPT) taxa	Number of taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)
Number of Ephemeroptera taxa	Number of mayfly taxa
Number of Plecoptera taxa	Number of stonefly taxa
Number of Trichoptera taxa	Number of caddisfly taxa
Number of Ephemeroptera, Trichoptera, and Odonata (ETO) taxa	Number of taxa in the insect orders Ephemeroptera (mayflies), Trichoptera (caddisflies), and Odonata (dragonflies and damselflies)
Number of Odonata taxa	Number of dragonfly and damselfly taxa
Number of Chironomidae taxa	Number of midge taxa
Number of Hemiptera taxa	Number of "true" bug taxa
Number of Coleoptera taxa	Number of beetle taxa
Number of Mollusca + Crustacea taxa	Number of mollusk (snails and clams) and crustacean (e.g., amphipods, copepods, decapods taxa
Shannon diversity	An index of richness and composition calculated as: $\Sigma -((n/N)*\text{Log}(n/N))/\text{Log}(2)$; where n is the number of individuals in a taxon and N is the number of individuals in the sample, summed for all taxa in the sample. The index is commonly standardized on log of 2 (as shown here) or the natural log (log e)
Composition and evenness	
Non-insects (%)	Non-insect individuals in the sample as a percentage of all individuals
Oligochaetes and leeches (%)	Percentage of worm and leech individuals
EPT individuals (%)	Percentage of mayfly, stonefly, and caddisfly individuals
Taxa in EPT (%)	Mayfly, stonefly, and caddisfly taxa in the sample as a percentage of all taxa
Ephemeroptera individuals (%)	Percentage of mayfly individuals
Plecoptera individuals (%)	Percentage of stonefly individuals
Trichoptera individuals (%)	Percentage of caddisfly individuals
Chironomidae individuals (%)	Percentage of midge individuals
Taxa in Chironomidae (%)	Percentage of midge taxa
Hemiptera individuals (%)	Percentage of "true" bug individuals
Odonata individuals (%)	Percentage of dragonfly and damselfly individuals
Coleoptera individuals (%)	Percentage of beetle individuals
Elmidae individuals (%)	Percentage of riffle beetle individuals
Number of individuals per taxon	The average number of individuals per unique taxon
Dominant taxon (%)	Individuals in the most numerous unique taxon as a percentage of all individuals
Dominant five taxa (%)	Individuals in the five most numerous unique taxa as a percentage of all individuals

TABLE 6-7. Continued.

Metric (by category)	Metric Description
Pollution tolerance	In all of the pollution tolerance metrics, degrees of pollution tolerance must be defined per taxon. This may be done categorically (e.g., sensitive, facultative, tolerant) or on a more continuous scale, as in the Hilsenhoff scale from 0 to 10. In addition, the pollution to which the organisms are responding may be general habitat and water quality stresses or specific (e.g., metals, sediments).
Number of intolerant taxa	Count of unique taxa that are sensitive to stresses (e.g., Hilsenhoff values 0 – 3)
Taxa as intolerant (%)	Sensitive taxa in the sample as a percentage of all taxa
Intolerant individuals (%)	Sensitive individuals in the sample as a percentage of all individuals
Number of tolerant taxa	Count of unique taxa that are tolerant of stresses (e.g., Hilsenhoff values 7 – 10)
Taxa as tolerant (%)	Tolerant taxa in the sample as a percentage of all taxa
Tolerant individuals (%)	Tolerant individuals in the sample as a percentage of all individuals
Hilsenhoff Biotic Index	The average individual pollution tolerance value for the sample. Calculated as: HBI = Σ (n)*(tolerance value)/N; where n is the number of individuals in a taxon and N is the number of individuals in the sample that have known tolerance values; summed for all taxa in the sample. Modifications of the published index (Hilsenhoff 1987) may include assignment of tolerance values to previously unrated organisms or of groups of organisms at genus, family, or order taxonomic levels.
Functional feeding groups	
Number of collector-filterer taxa	Number of unique taxa that feed on particles filtered from the water column
Collector-filterer individuals (%)	Filtering individuals in the sample as a percentage of all individuals
Number of collector-gatherer taxa	Number of unique taxa that feed on particles encountered among the substrates and detritus
Collector-gatherer individuals (%)	Gathering individuals in the sample as a percentage of all individuals
Number of predator taxa	Number of unique taxa that feed on living animal organisms
Predator individuals (%)	Predatory individuals in the sample as a percentage of all individuals
Number of scraper taxa	Number of unique taxa that feed on algae and bacteria that are attached to the surfaces of hard substrates
Scraper individuals (%)	Scraping individuals in the sample as a percentage of all individuals

6.9 Performance Characteristics for Biological Assessments Using Benthic Macroinvertebrates

6.9.1 Field Sampling

Quantitative (QN) performance characteristics for field sampling are *precision* and *completeness* (Table 6-8). Repeat samples for purposes of calculating precision of field sampling are obtained by sampling two adjacent reaches, shown as 500 m in this example (Figure 6-5) and for which there are not dramatic differences in condition. This can be done by the same field team for intra-team precision, or by different teams for inter-team precision. For benthic macroinvertebrates, samples from the adjacent reaches (also called quality control [QC] or duplicate samples) must be laboratory-processed prior to data being available for precision calculations. Assuming acceptable laboratory error, these precision values are statements of the consistency with which the sampling protocols 1) characterized the biology of the river and 2) were applied by the field team, and thus, reflect a combination of natural variability and systematic error (see Chapter 3).

TABLE 6-8. Error partitioning framework for biological assessments and biological assessment protocols for benthic macroinvertebrates. There may be additional activities and performance characteristics, and they may be quantitative (QN), qualitative (QL) or not applicable (na).

	Performance Characteristics				
Component Method or Activity	**Precision**	**Accuracy**	**Bias**	**Representativeness**	**Completeness**
1. Field sampling	QN	na	QL	QL	QN
2. Laboratory sorting/subsampling	QN	na	QN	QL	QN
3. Taxonomy	QN	QL	QL	na	QN
4. Data entry	na	QN	na	na	QN
5. Data reduction (e. g., metric calculation)	na	QN	QN	na	na
6. Site assessment and interpretation	QN	QN	QL	QL	QN

The number of reaches for which repeat samples are taken varies, but a rule-of-thumb is 10% randomly selected from the total number of sampling reaches constituting a sampling effort (whether yearly, programmatic routine, or individual project). Metric and index values are used to calculate relative percent difference (RPD), root-mean square error (RMSE), and coefficient of variability (CV) (Table 3-2). Acceptance criteria for each of these would be established based on programmatic capabilities demonstrated via pilot studies, or through analysis of existing datasets produced using the same protocols. These criteria are not data quality thresholds beyond which data points should be considered for discarding. Rather, they are flags for potential

problems (errors) in sample collection or processing. They are used to help determine the source(s) of the problems and to help develop recommendations for corrective actions. (K. Blocksom U.S. Environmental Protection Agency, personal communication) characterized performance measures for the benthic macroinvertebrate LR-BP (Table 6-9) (field sampling precision and metric sensitivity) when sample reaches are categorized according to mean thalweg depth.

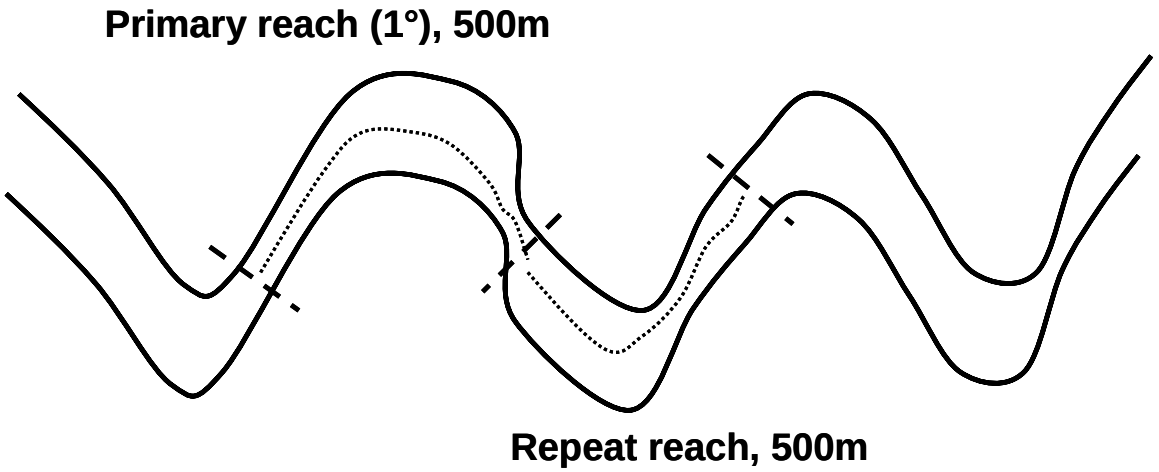


FIGURE 6-5. Adjacent reaches (primary and repeat) on a river channel.

TABLE 6-9. Precision and sensitivity of field sampling using the LR-BP for benthic macroinvertebrates (K. Blocksom, US Environmental Protection Agency, personal communication).

Metric	Mean*		Field Variance		Field CV (%)		Variance (field +lab)		DD (field+lab)†	
	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow
Total Taxa	43.7	56.4	17.3	6.4	9.5	4.5	56.4	51.7	14.7	14.1
EPOT Taxa	7.6	16.6	1.1	0.1	13.6	2.0	0.2	0.2	0.9	0.8
% Tolerant Indiv.	50.7	32.5	10.4	25.2	6.4	15.4	47.9	80.6	13.6	17.6
% Chironomidae	49.0	33.0	73.6	25.7	17.5	15.4	158.2	88.1	24.6	18.4
% Dominant Taxon	34.0	19.8	62.3	18.5	23.2	21.8	137.0	72.7	22.9	16.7

*"Deep" and "Shallow" refer to different depth categories of sampling reaches

†Based on $\alpha = 0.05$; n=1

Percent completeness (Tables 3-2, 6-8) is calculated to communicate the number of valid samples collected as a proportion of those that were originally planned. This value serves as one summary of overall data quality for a sampling effort and it demonstrates confidence in the final results.

Qualitative (QL) performance characteristics for field sampling are *bias* and *representativeness* (Table 6-8). Attempts to minimize the bias associated with the LR-BP for benthic macroinvertebrates include two components of the field method. First, it is not limited to one or a few habitat types (it is multihabitat and samples stable undercut banks, macrophyte beds, root wads/snags, gravel/sand/cobble). Second, allocation of the sampling effort is distributed throughout the entire 500-m sampling reach by use of six evenly-spaced transects, preventing the entire sample from being taken in a shortened portion of the reach. The LR-BP field sampling method is intended to depict the benthic macroinvertebrate assemblage physical habitat in the large river shore-zone (out to a depth of 1m).

Accuracy is considered "not applicable" to field sampling (Table 6-8), because efforts to define analytical truth would necessitate a sampling effort excessive beyond any practicality. That is, the analytical truth would be all benthic macroinvertebrates that exist in the river (shore zone to 1-m depth). There is no sampling approach that will collect all individual benthic macroinvertebrate organisms.

6.9.2 Laboratory Sorting/Subsampling

Precision, *bias*, and, in part, *completeness* are QN characteristics of performance for laboratory sorting and subsampling (Table 6-8). Precision of laboratory sorting is calculated by use of RPD with metrics and indices as the input variables (Table 3-2). If, for example, the targeted subsample size is 300 organisms, and that size subsample is drawn twice from a sorting tray without re-mixing or re-spreading, metrics can be calculated from the two separate subsamples. RPD would be an indication of how well the sample was mixed and spread in the tray; the "serial subsampling" and RPD calculations should be done on two timeframes. First, these calculations should be done, and the results documented and reported to demonstrate what the laboratory (or individual sorter) is capable of in application of the subsampling method. Second, they should be done periodically to demonstrate that the program routinely continues to meet that level of precision. Bias of the sorting process is evaluated by checking for specimens that may have been overlooked or otherwise missed by the primary sorter; checking of sort residue is performed by an independent sort checker. The number of specimens found by the checker as a proportion of the total number of originally found specimens is the percent sorting efficiency (PSE) (Table 3-2), and quantifies sorting bias. This exercise is performed on a randomly-selected subset of sort residues (generally 10% of total sample lot), the selection of which is stratified by individual sorters, by projects, or by programs. As a rule-of-thumb, an MQO could be "less than 10% of all samples checked will have a PSE ≤90%". Representativeness of the sorting/subsampling process is addressed as part of the standard operating procedure (SOP) that requires random selection of grid squares (Figure 6-4) with complete sorting, until the target number is reached within the final grid. Percent completeness for subsampling is calculated as the proportion of samples with the target subsample size (±20%) in the rough sort. Considered as "not applicable", estimates of *accuracy* are not necessary for characterizing sorting performance.

6.9.3 Taxonomy

Precision and *completeness* are QN performance characteristics that are used for taxonomy (Table 6-8). Precision of taxonomic identifications is calculated using percent taxonomic disagreement (PTD) and percent difference in enumeration (PDE) (Table 3-2), both of which rely on the raw data (list of taxa and number of individuals) from whole-sample re-identifications. The primary taxonomy is completed by the project taxonomist (T1); the re-identifications are performed by a secondary, or QC taxonomist (T2) as blind samples. The number of identifications in agreement between the two sets of results, as an inverse proportion of the total number of individuals, is precision of the taxonomic identifications, or "percent taxonomic disagreement (PTD)". The percent difference in sample counts by each of the taxonomists (not the sorters) is "percent difference in enumeration (PDE)". These two values are evaluated individually, and can be used to indicate the overall quality of the taxonomic data. They can also be used to help identify the source of a problem. The number of samples for which this analysis is performed will vary, but 10% of the total sample lot (project, program, year, or other) is an acceptable rule-of-thumb. Exceptions are that large programs (>~500 samples) may not need to do >50 samples; small programs (<~30 samples) will likely still need to do at least 3 samples. In actuality, the number of re-identified samples be program-specific and will be influenced by multiple factors, such as, how many taxonomists are doing the primary identification (there may be an interest in having 10% of the samples from each taxonomist re-identified), and how confident the ultimate data user is with the results. Mean PTD and PDE across all re-identified samples are estimates of taxonomic precision (consistency) for a dataset or a program. Percent taxonomic completeness (PTC; [Table 3-2]) quantifies the proportion of individuals in a sample that are identified to the specified target taxonomic level (lowest practical taxonomic level, species, genus, family, or other, including mixed levels). Results can be interpreted in a number of ways: the individuals in a sample are damaged or early instar, many are damaged with diagnostic characters missing (such as, gills, legs, antennae, etc.) or the taxonomist is inexperienced or unfamiliar with the particular taxon.

Accuracy and *bias* are QL performance characteristics for taxonomy (Table 6-8). Accuracy requires specification of an analytical truth. For taxonomy, it is 1) the museum-based type specimen (holotype, or other form of type specimen), 2) specimen(s) verified by recognized expert(s) in that particular taxon or 3) unique morphological characteristics specified in dichotomous identification keys. Determination of accuracy is considered "not applicable" for production taxonomy (most often used in routine monitoring programs) because that kind of taxonomy is focused on characterizing the sample; taxonomic accuracy, by definition, would be focused on individual specimens. Bias in taxonomy results from use of obsolete nomenclature and keys, imperfect understanding of morphological characteristics, inadequate optical equipment, and poor training. Neither of these performance characteristics is considered necessary for production taxonomy, in that they are largely covered by the estimates of precision and completeness. For example, although it is possible that two taxonomists would put an incorrect name on an organism, it is considered low probability that they would put the same incorrect name on that organism.

6.9.4 Data Entry

Efforts to understand the quality of data entry activity may seem trivial. However, the impact of errors can be substantial, and, if undiscovered and uncorrected, can become amplified through the assessment process. This QN performance characteristic quantifies the number of correctly-entered data values as a proportion of the total number of data values entered. The process involves having a QC person, distinct from the staff doing the primary data entry, check all data values (100%) against the original handwritten datasheets. With the datasheets as the analytical truth, the rate of errors is the *accuracy* of the data entry (Table 6-8). As errors are found, they are corrected electronically and the corrected value recorded. For their wadeable streams program, Mississippi DEQ found that the two data types with the highest error rates were the datasheet header information (e.g., stream name, latitude/longitude, date of site visit, names of field staff) and streambed particle size counts (Mississippi DEQ 2003). This allowed corrective actions to be focused where needed. All other performance characteristics are considered not applicable.

6.9.5 Data Reduction (Metric Calculation)

For most biological assessment programs, raw data are the list of taxa found at a site (in a sample) and the number of individuals recorded for each taxon. Preparation of those data for analysis requires conversion to metrics or other terms; metric calculation is a form of data reduction. When electronic spreadsheets or other data manipulation techniques are used, queries are often built to perform both complex and simple calculations. If queries are not performing as intended, or links to the raw data are incorrect, errors in metric values can occur. *Accuracy* of data reduction is a QN performance characteristic (Table 6-8) that helps ensure database/ computer calculation routines are performing as intended. A subset of metric values is hand-calculated using only the taxonomic and enumeration data, which are then compared to those that result from the computer queries. A recommended approach involves calculating one metric for multiple samples (e.g., systematic, every third sample), as well as all metrics for at least one sample. If differences are found, each value should be checked for errors in the calculation process (hand calculator vs computer algorithm), and corrections made.

6.9.6 Site Assessment and Interpretation

QN performance characteristics for site assessment and interpretation are *precision*, *accuracy*, and *completeness* (Table 6-8). Site assessment precision is based on the narrative assessments from the associated index scores (good, fair, poor) from reach duplicates and quantifies the percentage of duplicate samples that are receiving the same narrative assessments. These comparisons are done for a randomly-selected 10% of the total sample lot. Table 6-10 shows that, for this dataset, 79% of the replicates returned assessments of the same category (23 out of 29); 17% were 1 category different (5 of 29); and 3% were 2 categories different (1 of 29). Accuracy is the proportion of samples for which the biological index correctly identifies sites as impaired; the calculation is discrimination efficiency (DE) (Table 3-2). DE is a value that is developed during the index development and calibration process. Percent completeness (%C) is the proportion of sites (of the total planned) for which valid final assessments were obtained.

QL performance characteristics for site assessment and interpretation are bias and representativeness (Table 6-8). The final assessment of a site can be biased if a small number of reference or stressor sites are used during the calibration process. Low numbers of stressor sites can potentially result in high discrimination efficiencies that are spurious. If interpretation of assessment results fails to take into consideration abnormal or extreme hydrologic or climatic events, or other non-natural catastrophic and localized events, results could be considered non-representative of ambient conditions.

TABLE 6-10. Assessment results shown for sample pairs taken from 29 sites, each pair representing two adjacent reaches (back to back). Assessment categories are 1-good, 2-fair, 3-poor and 4-very poor.

Site	Replicate 1		Replicate 2		Categorical Difference
	Narrative	Assessment Category	Narrative	Assessment Category	
A	Poor	3	Poor	3	0
B	Poor	3	Poor	3	0
C	Good	1	Good	1	0
D	Poor	3	Very Poor	4	1
E	Fair	2	Fair	2	0
F	Poor	3	Fair	2	1
G	Poor	3	Poor	3	0
H	Very Poor	4	Very Poor	4	0
I	Very Poor	4	Very Poor	4	0
J	Poor	3	Poor	3	0
K	Poor	3	Poor	3	0
L	Very Poor	4	Very Poor	4	0
M	Very Poor	4	Very Poor	4	0
N	Poor	3	Fair	2	1
O	Poor	3	Poor	3	0
P	Poor	3	Poor	3	0
Q	Poor	3	Very Poor	4	1
R	Poor	3	Poor	3	0
S	Fair	2	Very Poor	4	2
T	Fair	2	Fair	2	0
U	Good	1	Good	1	0
V	Poor	3	Fair	2	1
W	Fair	2	Fair	2	0
X	Poor	3	Poor	3	0
Y	Poor	3	Poor	3	0
Z	Very Poor	4	Very Poor	4	0
AA	Poor	3	Poor	3	0
BB	Fair	2	Fair	2	0
CC	Poor	1	Poor	1	0

Chapter 7.0 Fish

with contributions from **Blaine D. Snyder**[1]

This chapter…

- reviews existing methods for large river fish sampling
- recommends a margin-oriented boat electroshocking sampling approach

Fish are:

- important consumers in large river food webs
- established indicators in biological assessment programs
- valuable connections to cultural, recreation, and economic interests

7.1 Introduction

Fish assemblages are commonly used as indicators of ecological condition because they represent an important component of the aquatic community, and are of heightened interest to the public (Hocutt 1981, Barbour et al. 1999, Simon 1999, McCormick and Peck 2000, Lazorchak et al. 2000, USEPA 2002). Many States designate aquatic life use-support narratives based on fish assemblage characteristics. Narrative expressions such as "maintaining coldwater fisheries", "fishable", or "fish propagation" are prevalent in State standards. Fish are good indicators of ecological condition because they are relatively long-lived, mobile, feed at every trophic level (e.g., herbivores, omnivores, and predators), and can be relatively easy to identify to species (Plafkin et al. 1989). There are both advantages (many as described above) and disadvantages to using fish in bioassessment programs that should be considered when developing a large river biological assessment program (Table 7-1).

Fish bioassessments use structural and functional attributes of the ichthyofaunal assemblage to evaluate biological condition. This involves careful sampling using standardized field collection techniques, species identification and enumeration, and analyses using measured biological attributes (e.g., density, biomass, etc.) or metric calculations (e.g., feeding types, pollution tolerance measures, taxonomic affiliations, etc.). Data produced by an appropriate fish sampling protocol can be used to assess use attainment, develop biological criteria, prioritize sites for further evaluation, provide a reproducible impact assessment, and evaluate status and trends of the fish assemblage.

Karr (1981) developed a fish assemblage assessment approach known as the index of biotic integrity (IBI), which is commonly used in biological assessment and monitoring programs. The IBI incorporates the zoogeographic, ecosystem, and population aspects of the fish assemblage into a single, ecologically based index. Calculating and interpreting the IBI for a particular area involves a sequence of activities including: fish collection, data tabulation, regional metric selection, and calibration of metrics to expected values. A detailed description of this assessment approach is presented in Karr et al. (1986). Regional IBI modifications and applications are described in Leonard and Orth (1986), Moyle et al. (1986), Hughes and Gammon (1987), Wade and Stalcup (1987), Miller et al. (1988), Steedman (1988), Simon (1991), Lyons (1992a), Simon and Lyons (1995), Lyons et al. (1996), Simon (1999), Lyons et al. (2001) and Emery et al. (2003).

[1] Tetra Tech, Inc., 400 Red Brook Blvd., Suite 200, Owings Mills, MD 21117

TABLE 7-1. Advantages and disadvantages to using fish as bioindicators.

ADVANTAGES/DISADVANTAGES
Advantages:
• Fish are good indicators of long-term effects and broad habitat conditions because fish are relatively long-lived (3-10+ years). • Fish can be sampled year round; seasonal changes in distributional patterns must be considered. • Fish assemblages generally include a range of species that represent a variety of trophic levels (omnivores, herbivores, detritivores, insectivores, planktivores, and piscivores). • Fish are relatively easy to collect and identify to the species level by trained fishery professionals. • Most specimens can be identified in the field and released unharmed, requiring minimal laboratory follow-up. • Environmental requirements, life histories, and distributions of many fish species are well known. • Contaminants often induce identifiable morphological deformities that can be used as indicators of condition (Sanders et al. 1999, Smith et al. 2002). • Fish have high social and cultural value (e.g., sport, subsistence, and commercial fisheries). • Fish are at the top of the aquatic food web and are consumed by humans, making them important for assessing ecological and human health risk. • Aquatic life uses are typically characterized in terms of fisheries (coldwater, coolwater, warmwater, sport, forage). Monitoring fish provides direct evaluation of "fishability" and "fish propagation".
Disadvantages:
• Because of the seasonal mobility of some species, they may be less indicative of localized disturbances. • The initial cost of sampling gear is often considerable (investment cost of gear is offset by minimal lab cost). • Safety concerns are increased due to use of 500-1000 volts (when using boat electrofishing gear) and the potential hazards associated with night electrofishing (Graham 1986). • May require that agencies collaborate to facilitate sampling (possible advantage).

Many studies have shown strong associations (i.e., correlations) between fish IBI results, physical and chemical habitat condition, and human activities that alter stream and river habitat (e.g., dams, agriculture, urban development, etc.) (Karr et al. 1985, Berkman et al. 1986, Leonard and Orth 1986, Ohio EPA 1987, Steedman 1988, Karr 1991, Yoder and Rankin 1995, Ohio EPA 1999). Most of the studies using fish IBIs have been conducted in wadeable streams systems, and the application of IBIs for large river assessment is relatively limited (Hughes and Gammon 1987, Ohio EPA 1987, Oberdorff and Hughes 1992, Hugueny et al. 1996, Ganasan and Hughes 1998, Simon 1999, Gammon and Simon 2000, Lyons et al. 2001, Araujo et al. 2003, Emery et al, 2003, Mebane et al. 2003, Stoddard et al. 2005).

In this chapter, we provide an overview of several different programs (Table 7-2) that have developed and successfully applied different fish sampling protocols to biological assessments in large rivers, including USEPA/EMAP (McCormick and Hughes 2000, Angradi 2006), USGS/NAWQA (Moulton et al. 2002), and ORSANCO (Emery et al. 2003). Although specific definite protocols of sampling and assessment using fish are not proposed, different approaches and techniques are covered and techniques for documenting method performance are suggested. Whatever methods are selected for your program, they should be thoroughly tested to document the quality of data they can produce. It is up to you, as the data user, to ensure these data will meet project objectives.

TABLE 7-2. A comparison of large river program fish sampling approaches.

Program	Protocol Summary	Citation
Environmental Monitoring and Assessment Program (EMAP-Non-Wadeable)	Focus on all but most rare fish. Data collected on richness, guild structure, abundance, size, and anomalies. Sample reach varies: 40X to 100X wetted width, fish sampled along one bank with 14-16 ft electrofishing boat; 50X wetted width, fish sampled along alternating shores by raft. Identified, counted, total length measured, anomalies recorded, and vouchering if needed.	McCormick and Hughes 2000, Hughes and Herlihy accepted, Lazorchak et al. 2000
Environmental Monitoring and Assessment Program: Great River Ecosystems (EMAP-GRE)	Focus on characterizing all but most rare fish in littoral habitat of great rivers. Data collected on species composition, size, and condition. Sample reach is 500 m long along one bank. Electrofishing zone extends out from shore 30 m or to a depth of 6 m, whichever is closer. Fishing time must be 30 minutes or longer at 3000 watts (adjusted as necessary to reduce injury to fish). Use a 5.5 m welded hull, aluminum electrofishing jon boat with 90 hp engine for travel and 25 hp engine for sampling. Identified, counted, total length measured, weighed, anomalies recorded, and vouchering if needed.	Emery et al. 2006
United States Geological Survey-National Water Quality Assessment (NAWQA)	Focus on sampling a representative portion of the fish assemblage. Reach lengths are 20X wetted width (500 m min to 1000 m max). Electrofishing and seining are used. Make two passes along one bank with electrofishing boat. Three seine collections composited from wadeable shoreline areas after electrofishing. Each electrofishing pass and seine composite samples all processed separately. Identified, counted, total length measured, weighed, anomalies recorded, and vouchering if needed.	Moulton et al. 2002
Ohio River Valley Water Sanitation Commission (ORSANCO)	Focus on condition of fish assemblages along Ohio River. Expected index values developed for different habitats. Sample from July to October along 500 m shoreline zones. Night electrofishing used with a boat unit. Identified, total length measured, weighed, anomalies recorded, and habitats noted.	Emery et al. 2003, and *See program highlight box*
Large River Bioassessment Protocol (LR-BP)	Focus on developing an unbiased and representative sample of fish assemblage within logistical and budgetary constraints. Uses a 2-3 person crew – one boat operator and 1-2 dippers. Targets main channel border habitats. Basic design sample either 500 m paired bank or 1000 m single bank. Reach length may be increased for study-specific needs. Sites <4m, mean thalweg depth are electrofished at day. Sites >4m are preferably electrofished at night. Identified, counted, total length measured, weighed, anomalies recorded, and vouchering if needed.	Flotemersch and Blocksom 2005

7.2 Methods

Several questions related to program development and method selection should be considered prior to beginning a fish bioassessment program:

- What fish sampling permits are required by the State?
- Which habitats should be sampled?
- What should the reach length be for each site?
- What is the appropriate time of day to sample?
- What method and sampling gear should be used?
- Should multiple samples be collected (for population estimates) or only one sample (for richness, relative abundance, and other metric calculations)?
- Should samples be composited or kept separate?
- What is the most appropriate spatial sample design?
- How will specimens be identified?
- What are the indicators that will be used?

It is strongly advised that consensus-driven responses to these and similar questions be prepared and signed-off on by key staff who will be involved in the collection, analysis, and use of the resulting data. This should occur prior to the data collection process and will greatly increase the level-of-success achieved by the project.

Most fish collection procedures use a multi-habitat sampling approach, sampling habitats in relative proportion to their local availability as determined during site reconnaissance. Sample reach lengths vary among studies, but generally attempts to encompass most if not all prevailing habitats. The exception is when habitat features are so large that a reach length encompassing all habitats is unrealistic. In such cases, the development of habitat specific criteria should be considered. When placing the reach, it is important that it not be located to purposefully avoid manmade obstacles such as bridges, rip-rap, road crossings, or channelization as the potential influence of these features are relevant to assessing overall river condition (Lazorchak et al. 2000). However, study-specific needs may necessitate that such features be avoided.

When compared to wadeable streams, accessibility issues on rivers are two-sided in that they can be reduced or increased depending on the nature of the site. This may be more true of studies assessing fish assemblages than other assemblages because of the substantially more bulky equipment required. If the site is remote yet the river is navigable, access by boat may be easier than by foot. However, in free-flowing low-gradient systems, navigation along the channel may be impeded by shallows, log jams, or other obstacles. In such cases where a pre-selected reach cannot, in a safe and reasonably efficient manner, be accessed, it should be left out and not forced; if this situation occurs, a replacement reach should be randomly selected from a list of alternate reaches. However, caution should be exercised to not declare sites unsampleable out of convenience. This could bias results by skewing sites sampled towards those that are potentially more impaired because of their accessibility or reason for accessibility (e.g., presence of a boat ramp because the site is impounded).

A habitat assessment is typically performed and physical/chemical parameters are measured concurrently or just prior to fish sampling, to document and characterize available habitat within the sample reach (Chapter 4). It is extremely important for experienced fisheries scientists to be involved in the adaptation and application of field protocols and the taxonomic identification of fishes. Since most protocols specify field identification and release of captured fish, fish bioassessment data quality and comparability are assured through the use of qualified fisheries professionals, consistent methods, and correctly applied quality control activities.

While electrofishing is most commonly used, it is only one of several fish sampling procedures that may be useful as part of a bioassessment program. Fish sampling methods can be broadly categorized as either passive or active. Passive sampling methods include those that use hoop (Figure 7-1), fyke, trap, and gill nets, or hook-based methods such as trotlines (Hubert 1996). Many of these methods have high species selectivity in that they are most effective for specific species, guilds, or size classes of fish and thus may only effectively sample a segment of the assemblage. However, because of their efficiency at collecting the targeted organisms, they are frequently used by resource managers to attain species-specific information. Active sampling methods include electrofishing, seining, and trawling (Hayes et al. 1996). Table 7-3 summarizes some advantages and disadvantages of common non-electrofishing sampling techniques.

While all fish sampling methods are generally considered selective to some degree, electrofishing has proven to be the most comprehensive and effective single method for collecting fishes (Figure 7-2) from streams and rivers (Vincent 1971, Gammon 1973 and 1976, Novotny and Priegel 1974, Ohio EPA 1987, Davis et al. 1996, Barbour et al. 1999, Simon and Sanders 1999). There are situations, however, where approaches other than electrofishing (e.g., seining, trawling) may be preferred or necessary. As an example, electrofishing is limited in some river reaches with endangered fish (Barbour et al. 1999) or mussels. This concern seems warranted since studies have shown that spinal injuries and associated hemorrhages occur in over 50% of fish examined internally subsequent to being electroshocked (Snyder 2003). Another example is when collections are required in river systems that lack the solutes necessary (e.g., conductivity <10 µS/cm) to effectively pass the electrical current through the water due to regional geological characteristics. This is less likely, however, in large rivers that accumulate solutes from diverse landscape types. Excessively high conductivities (e.g., 1,000-3,000 µS/cm; Hill and Willis 1994), such as in brackish waters and at sites where non-natural inputs artificially raise the conductivity, reduce the effectiveness of electrofishing as well (Reynolds 1996).

7.2.1 Electrofishing

Although numerous agencies electrofish, the equipment used, the electrofishing configuration, and the field design applied may vary greatly. Variables that often differ include the sampling design (e.g., habitat sampled), whether the electrofishing is conducted during the day or night (Section 7.2.1.1), the mesh of the dip net, number of netters, the power of the electrofishing unit and generator, and the size of the boat. An assortment of electrofishing equipment may be necessary to cover the range of habitats in large rivers. It is not uncommon for field teams targeting rivers to need equipment that ranges from tote-barges to large electrofishing boats.

FIGURE 7-1. Use of a hoop net as a passive fish sampling method.

TABLE 7-3. Advantages and disadvantages of non-electrofishing sampling approaches including passive (e.g., hoop, fyke, and gill nets, and trotlines) and active (e.g., seines, trawls) sampling gears.

ADVANTAGES/DISADVANTAGES
Advantages: • potentially a low cost alternative to electrofishing (design dependent) • no electrical components • require little specialized training • can yield precise data on the components of the fish assemblage the gear targets • effectiveness not impaired by conductivity or turbidity
Disadvantages: • selectivity of the gear • require multiple trips to a site (although seines only require one trip) • spatial coverage of a site may be limited • ineffective or difficult to deploy in swift water areas (e.g., runs or rapids) • passive methods only sample fish that are moving • potentially high mortality and bycatch (study dependent) • effort required to reduce disease transmission across sites • repair and maintenance of gears • large trawls require large boats

FIGURE 7-2. Net retrieval of fish stunned by boom-shockers, an active method.

The performance of equipment and field personnel can greatly affect the results of an electrofishing effort, potentially lending to bias. The type of boat and electrofishing equipment, equipment settings, electrode arrays, field conditions, and skill of the crew will all influence the catch. Among required skills is that the boat driver be able to navigate the boat in a manner that assures the safety of themselves and the crew while assuring the collection of a high quality sample. The boat operator is also usually required to monitor electrofisher performance to assure uniform application of the electrofishing field across the sampling site.

Once a sample has been collected, accurate fish identifications in the field are essential. This is an important component for quality control of bioassessment programs and requires extensive training and study, knowledge of regional distributions, and proper allowance of time in the field to do a thorough job. Regardless of skill level, some specimens will have to be returned to the laboratory for identification/verification. It is strongly recommended that field crews be adequately trained. Additionally, the crew lead should possess broad electrofishing experience attained under the leadership of a qualified professional.
Environmental factors can also influence electrofishing performance. These factors include time of day, wind, excessive amounts of flotsam or macrophytes, bottom substrate, water depth, cover, conductivity, temperature, water clarity, and any additional deviations from normal water conditions (e.g., flow rate, water level, dissolved oxygen, etc.) that might result in the collection of anything other than a representative sample. All of these conditions should be evaluated, recorded, and considered prior to initiating the collection of what should be a representative sample of the fish assemblage. Additionally, recent research that compared electrofishing designs in large rivers of the eastern and central USA (Flotemersch and Blocksom 2005) showed that the degree to which a river has been impounded plays a critical role in electrofishing performance. In short, daytime electrofishing was less productive per unit of effort at sites

where the depth exceeded 4 m, as is frequently the case at impounded sites. Hence, different electrofishing designs (e.g., day vs night; see Section 7.2.1.1) and metrics may be required to adequately describe different types of systems or parts of systems.

7.2.1.1 Day and Night Electrofishing

At riverine locations where the diel movements of fish significantly influence the efficiency of electrofishing efforts, varying the time electrofishing is conducted may aid in most efficiently meeting study objectives. Research comparing the catches between day and night electrofishing sessions has shown that catches can be significantly different (Sanders 1991, Andrus 2000, Dumont and Dennis 1997, Simon and Sanders 1999). For example, Sanders (1991) found that day sampling collected 9 species not collected during night sampling, while night sampling collected 17 unique species and 2 hybrids previously unreported from the study area. Overall, night catches contained significantly more species, higher numbers and weights of fish, and were compositionally more evenly distributed than day catches (Sanders 1991, Simon and Sanders 1999); all qualities advantageous to the bioassessment of a site. Andrus (2000) reported that in alcoves and main channel reaches of the Willamette River, Oregon, night electrofishing yielded more taxa and a greater abundance of fish than daytime electrofishing. Increase in catch was attributed to fish migrating into the alcoves at night and fish being more vulnerable to electrofishing at night because of positioning in the water column. Nonetheless, the data requirements of the study should be consulted prior to deciding whether day or night electrofishing is most appropriate. It would be improper to night electrofish for a study targeting a species almost exclusively collected during the day.

While preferable for some scientific collection purposes, some concerns have been raised about the safety and logistical problems of night electrofishing (e.g., navigation in the dark, fatique) (Graham 1986). Specific problems that have been cited include difficulty identifying shallow water, unexpectedly entering fast and shallow water, a limited ability to see downstream hazards such as log jams, and difficulty in setting reach lengths with electronic rangefinders (Andrus 2000). To address such problems, Andrus (2000) recommended visiting the site during the day prior to sampling so the crew can become familiar with the site and complete tasks that may be difficult in the dark.

7.2.1.2 Sample Reach

Considerable research has been conducted on the determination of sufficient sample reach lengths for large rivers (e.g., Gammon 1976, Penczak and Mann 1993, Yoder and Smith 1999, Cao et al. 2001, Lyons et al. 2001, Hughes et al. 2002, Maret and Ott 2004, Flotemersch and Blocksom 2005, Hughes and Herlihy, accepted). As seen in Table 7-4, reach lengths found suitable for rivers vary in length and form (i.e., fixed distance vs multiples of the wetted width). Some of these differences can be attributed to the geographic area of the work, system type (e.g., high gradient vs low gradient rivers), and the evaluation parameter(s) used to determine sample sufficiency.

Most electrofishing designs call for shocking a continuous length of shoreline. However, other options exist. For example, Hickman and McDonough (1996) discuss the development and use

of an electrofishing design on Tennessee River valley reservoirs that shocks 15 independent 300-m shoreline zones. Dominant habitat features in each electrofishing run and at each gill-net set are recorded to determine habitat influences on metric results. The electrofishing catch is supplemented with 10 overnight experimental gill net sets. To mitigate the effects of one sample on the next, a 50-m shoreline section between each electrofishing run is not sampled. For additional reading on this design, consult Jennings et al. (1995) and McDonough and Hickman (1999).

TABLE 7-4. A comparison of different reach lengths found suitable for bioassessment of rivers.

	Reach length	Geographic Area	Evaluation Parameter
Fixed Distance			
Gammon 1976	500-2000 m	Wabash River, Indiana	Assemblage parameters
Meador et al. 1993	500-1000 m	United States	Representative sample
Penczak and Mann 1993	500-1000 m	Pilica River, Poland	Species richness
Yoder and Smith 1999	500/1000 m	Ohio large/great rivers	Representative sample
Lyons et al. 2001	1600 m	Wisconsin rivers	Species richness
Flotemersch and Blocksom 2005	500-1000 m (both banks)	Mid-Western rivers	Assemblage parameters
Angradi 2006	500 m	Great Rivers	Representative sample
Emery et al. 2003	500 m	Ohio River	Representative sample
Multiples of the Wetted Width (MWW)			
Hughes et al. 2002	85 MWW	Oregon raftable rivers	Species richness
Maret and Ott 2004	30-40 MWW	Idaho rivers	IBI scores
Hughes and Herlihy, 2006	50 MWW	Oregon raftable rivers	IBI scores

When selecting a reach length, or conducting research for setting reach length, it is important to consider several factors including the question being addressed by the study, the level of resolution (precision and accuracy) required to address the question, and the statistical approach that will be used to analyze any resulting data. Ideally, the sampling effort applied is the minimum required that will allow stated study objectives to be addressed (Angermeier and Smogor 1995, Patton et al. 2000). Just as critical is ensuring that reach length is balanced with available resources, logistical constraints, and safety issues. A detailed discussion covering issues related to setting reach length for bioassessment of riverine biotic assemblages is provided in Chapter 3 of this document.

7.2.1.3 Generalized Electrofishing Protocols

Boat electrofishing techniques are often very similar across protocols (e.g., Ohio EPA 1987, Reynolds 1996, McCormick and Hughes 2000). Yoder and Smith (1999) provide insight into the intricacies involved in the successful navigation of an electrofishing boat within a sample reach. Table 7-5 lists activities that should be performed prior to leaving the launch site.

TABLE 7-5. Preparation activities onshore at launch site.

- Check generator oil and fill tank with gas (wipe up any spillage).
- Attach and inspect anode arrays.
- Attach the cathode.
- If the target site is in close proximity to the launching point, the anode booms should be positioned for electrofishing. Otherwise, travel with the booms in their stowed position.
- Complete all necessary electrical connections between the generator, the variable voltage pulsator box, and the anode booms.
- Review and confirm that all gear is in the boat.
- Assure crew members have donned personal floatation devices.
- If the target reach is in close proximity to the launch site, the crew can prepare for electrofishing activities. Otherwise, assure all gear is properly stowed for travel to the site.

Establishing the Reach

Upon arrival at the site, the first task is to delineate the targeted reach to be sampled (see Table 7-4 for common reach length examples).

- Examine the immediate area for influences of major tributaries and bridge/road crossings (i.e., the site should be sufficiently upstream to decrease influences on overall habitat quality). If an influence disruptive to the integrity of the sample exists that would fall within the estimated limits of the sample zone, the decision may be made to slide the reach either up or downstream. If the sample zone is relocated, an effort should be made to retain the original site identifier (i.e., latitude and longitude) in the reach.
- The exact location (i.e., latitude and longitude) of the downstream limit of the designated reach should then be recorded on the appropriate field data sheet(s). If a GPS unit is used to provide location information, the accuracy or design confidence of the unit should be noted.
- Designate (i.e., flag) the downstream extent of the reach on both the left and right banks (or on one particular bank if choosing a single bank sample). Several methods can be used for locating the upstream end of the reach, but the two most common techniques use a laser rangefinder or the "distance traveled" feature of a GPS unit or alternate equipment (e.g., depth finder) with GPS capabilities.

Preparing for Electrofishing of the Reach

With the sample reach established, the crew can prepare for electrofishing.

- Discuss the layout of the reach, any hazards or obstacles observed when the reach was established, and discuss how each will be addressed.
- Prepare the live well by filling it to a suitable level with fresh river water and assure aeration devices are functioning. A large live well (>300 L) should be used to ensure adequate holding capacity for all fish collected in a long reach (e.g., 500 m). A strong and reliable aerator should be used to maintain oxygen levels. If a large number of

fish are captured, it may be necessary to periodically change the water in the live well. Usually this is done after the electrofishing run has been completed, just prior to processing the fish, or continually during processing.

- Check all electrical connections (including on/off switches) and assure anodes (electrofishing booms) and cathode array (if equipped) are in position and secure.
- Crosscheck with crew that all safety gear (e.g., personal floatation devices, watertight rubber linesman's gloves, rubber footwear, hearing protection, and communication gear) is functional.
- Verify that all electrical switches are off, that all non-target organisms (e.g., cattle, waterfowl, and humans) are clear of the water, and that boat surfaces are dry.
- Test and record the conductivity of the water in the area of the sample reach. This information is needed to determine if the conductivity is within the performance specifications of the equipment, to determine preliminary settings for the variable voltage pulsator box, and to track changes at the site through time.
- In an area outside the target reach, start the generator and test for the proper functioning of all equipment (particularly on/off switches). Adjust the variable voltage pulsator box setting for effective electrofishing. Settings for electrofishing will vary greatly across sites. No single setting will work in all places, but a standardized approach for arriving at a setting can be achieved. If no one on the crew has experience in the area being sampled, it is advisable to find a local professional with experience and consult with them prior to heading for the field.
- Experienced or properly trained crew members will be able to determine the effectiveness of pulsator box settings and verify that fish within the electrofisher's field are rolled and relaxed but not rigid. Record pulsator box settings on the field sheets, reset shock-time timer to record total seconds the pulsator is engaged and fishing (often referred to as button time), and record start time.

Electrofishing the Designated Reach

- Collection via electrofishing can begin on either bank at the upstream end of the sample reach. The boat is piloted to proceed in a downstream direction along the main-channel shoreline habitat of each bank at a speed near or slightly exceeding the river velocity (Ohio EPA 1987, Reynolds 1996, Yoder and Smith 1999, McCormick and Hughes 2000). Proceeding at this speed serves two purposes. First, stunned fish will be moving at or near the speed of the boat and the netter(s) is (are) provided the best opportunity to collect stunned fish in front of and on the sides of the boat. Second, keeping the boat in motion serves to help standardize the effort among crews and across sites.
- The effective shoreline electrofishing area will vary within and among sites, but generally follows the shoreline in waters from 0.25 m to 3 m deep.
- As electrofishing proceeds through the reach, minor fluctuations in the observed amperage may be observed. If the amperage deviates significantly, the electrofishing settings (usually percentage applied) should be adjusted to maintain consistent, effective, and humane electrofishing. Significant adjustments to the settings should also be noted on field forms.

- Stunned and collected fish should be placed directly in the livewell as soon as possible. Fish should not be held in the net in the electrical field as this will increase the likelihood of mortality among stunned fish. Netters should collect all stunned fish and avoid being size selective (e.g., netting only large specimens). Try to net all fish seen, but in productive systems this may not be possible. If benthic fish are being missed, an option may be to pivot the boat occasionally or hold the net behind the anode and along the bottom so more are collected. Care should be taken to thoroughly maneuver the electrofisher around objects such as snags, downed trees, piers, boulders, and other potential fish cover until each object yields no more fish.
- During the electrofishing run, rare, sensitive, or excessively large fish may be encountered. If these fish appear overly stressed (as indicated by loss of righting response), the decision may be made to pause the electrofishing effort, process the fish, and release them behind the boat to ensure that they are not recaptured. If this occurs, be sure to pause the clock recording total time electrofished and restart it when fishing is resumed. Button time will not be affected.
- In shallow reaches of some rivers, there may be sections that are non-navigable. In such areas, a small boat (12-14′) with a crew of 2 may be used. With this configuration, using oars or push-poles, electrofishing can continue in depths as shallow as 12 cm. In areas where the water depth is shallower, a method used by the Ohio EPA is to have one or more crew members exit the boat and position it at the top of the reach, while the netter takes position downstream. When the okay is given, the operator engages the electrofisher as the netter proceeds upstream netting fish. When the netter reaches the boat, the crew (as a team) repositions the boat further downstream and repeats the activity. When the full extent of the non-navigable section of river has been electrofished, the full crew re-boards the boat, electrofishes the downstream extent of the riffle, and then proceeds downstream as described above.
- Upon arriving at the downstream extent of the reach, the variable voltage pulsator and the generator should be turned off and both the button time and total time should be recorded. Shocking a 500-m reach on one bank generally should take between 20-30 minutes (depending upon flow and fish abundance). All fish should be field processed immediately. Fish that appear overly stressed, or are known to become stressed with increased holding times, should be processed first. If additional passes are part of the study plan, it may be advisable to retain fish in holding nets to eliminate the possibility of repeat capture during additional passes or on the opposite bank. If fish are returned directly to the river, they should be released in an area that ensures that they are not recaptured.
- If multiple passes or both banks are sampled, fish from the first bank (or pass) should be processed before proceeding with subsequent effort. After electrofishing and processing has been completed, fish can be released with the exception of voucher specimens that need to be identified in the laboratory or those retained for other purposes (e.g., fish tissue sampling, histopathological analysis).

Ancillary Data Collected to Characterize the Electrofishing Event

A number of variables are commonly recorded to characterize prevailing conditions while electrofishing, some of which have already been mentioned (e.g., button time, total time, conductivity). Beyond their utility for data analysis, many are highly useful to crews revisiting the site. Many of these variables should be considered as critical data elements to be included in any electrofishing activity. These items include, but are not limited to:

- Location of target site (e.g., latitude and longitude, position of sampling reach to latitude and longitude, and landmarks),
- Time of day electrofishing occurred,
- Habitat variables that may already be part of a larger habitat assessment (e.g., maximum and mean widths, dominant habitats, secchi depth, depth characterization),
- Factors that affect sampling efficiency (e.g., field team's ability to see and net stunned fish, whether polarized sunglasses were worn, prevailing flow conditions, river stage, water clarity, and water color), and
- Items useful to crews conducting future sampling events (e.g., directions, access points, difficulty of access, land owner contact information, potential safety concerns).

7.2.1.4 Electrofisher Configuration and Design (Electrode Arrays)

The size, surface areas, and shape of the electrodes are the most important element of an electrofishing system (Novotny 1990). In combination with the water conductivity, the array configurations determine the system's electrical resistance and the distribution of field intensity that determines the unconfined size and shape of the effective field for a specified voltage output (Snyder 2003). For reasons of sampling efficiency and reduced injury to both fish and incidentally shocked humans and other animals, direct current (dc) is most frequently used to power contemporary electrofishing boats (Reynolds 1996, Snyder 2003).

A contemporary dc powered electrofishing configuration consists of anode and cathode arrays. The anode array usually extends in front of the boat suspended from booms. Anode designs vary greatly among electroshocking boats. Common designs include use of aircraft cable or flexible conduit suspended between two booms, suspension of one or two metal spheres in the water (i.e., Coffelt Sphere), and suspension of a cluster (i.e., umbrella) of cable droppers from one or two booms.

Regarding the cathode side of the circuit, in its simplest form, the boat hull is used as the cathode although this is generally advised against (Reynolds 1996). A commonly employed alternative is using a cathode array attached to the front, side, or on both the front and side of the boat. In most cases, the array is constructed of metal cable suspended into the water from the boat hull. Use of a cathode array (rather than using the boat hull as the cathode) can effectively increase the efficiency of the electrofishing unit by concentrating the effective electroshocking field to the viewable area of the netters. The smaller electrical field requires less power to produce and is generally more stable and uniform (design dependent). Combined, these factors can extend

equipment life reducing the load on the electrical equipment, reduce the injury rate to fish, and reduce the risk presented to incidentally shocked humans and other animals.

Electrode arrays tend to be easily serviced due to their location and can be reconfigured if necessary to meet site conditions. In general, longer arrays are preferred in deeper waters and shorter arrays may be preferred in habitats where longer arrays may become snagged. To assure efficiency, it is recommended that the surface area of the cathode array be a minimum of twice that of the anode array with some recommending a surface area 10 to 20 times that of the anode array (Bob Hughes, Oregon State University, personal communications). Keeping the surface of the arrays clean and free from build-up is vital to maintaining the performance of the configuration.

For a comprehensive discussion on the basic principles of electrofishing, refer to Reynolds et al. 1996, and the support section available online at www.Smith-Root.com and the list of additional reference and training materials included therein. For additional information on implications of electrode configuration, refer to Beaumont et al. 2006. For a summary of how anode and cathode configuration can influence the extent of the harmful effects of electroshocking on fish, refer to Snyder (2003).

7.2.1.5 Electrofishing Field Team Safety and Organization

Adequate education, training, and experience of all members of the fish collection team are critical for assuring the safety of all personnel and the quality of the data (Barbour et al. 1999). At least one biologist with training and experience in electrofishing techniques and fish taxonomy must be involved in each sampling event. All field team members must be trained in boating safety and electrofishing safety precautions and unit operation procedures identified by the electrofishing unit manufacturer. Any crew member that will be driving the field vehicle, with or without a boat in tow, should attend a safe driving course. It is also recommended that at least 2 fish collection team members are certified in CPR (cardiopulmonary resuscitation). If electrofishing will take place in white water rivers, white water safety courses such as those developed by Rescue 3 International (www.rescue3.com) are also highly recommended.

Proper maintenance of all equipment is an important component of safety in the field. For a boat-based electrofishing crew, this includes maintenance and repair of the boat, motor, and trailer, plus regular inspection of all components of the electrofishing configuration. It is also recommended that the electrofishing boat be annually inspected by a professional electrician for shorts, voltage differences, and general wear of electrical components.

When electrofishing, each team member must be insulated from the water and the electrodes even when in a boat and not wading in the river; therefore, insulative footwear (e.g., knee boots, chest waders) and rubber gloves (linesman's gloves) are required. Likewise, dip net handles must be constructed of insulating materials (e.g., wood, fiberglass). The electrofishing boat should be equipped with functional safety switches (usually standard equipment from electrofisher manufacturers) so that each member of the crew has the ability to interrupt the flow of electricity when needed. Field team members must not reach into the water unless the electrodes have been removed from the water or the electrofisher has been disengaged.

Furthermore, every effort should be made to keep the electrofishing gloves dry. If they become excessively wet, electrofishing should stop and the gloves should be dried. Additionally, efforts should be made to minimize water on the boat deck(s).

The priority of the boat operator is overseeing the safe operation of the boat and general safety of the crew. The netters are likewise charged with assuring safety of the crew. Key responsibilities include providing the operator with information about obstacles in the water while the boat is in motion, and assuring conditions in the holding tank are suitable for specimen survival. Two-person crews are generally used at shallow sites where smaller, lighter electrofishing boats are needed to successfully navigate the river. At sites where the depth of the river permits the use of a larger electrofishing boat, an optional second netter is often added to the crew. Note that the addition of a second netter will increase capture efficiency and potentially influence the collection. Thus, crew configuration should be documented on field sheets and should be part of the permanent record for the resulting data. Table 7-6 provides a checklist if items and gear needed for boat electrofishing. For additional reading on the safety and logistics of ecological sampling on large rivers, see Flotemersch et al. (2001).

7.2.2 Seining

Seining in streams and rivers is generally conducted with a beach seine consisting of uniform mesh, two wings, and a bunt section that holds the catch (Hayes et al. 1996). Scientists with the US Army Corps of Engineers Waterways Experiment Station in Vicksburg, Mississippi regularly use beach seines to evaluate changes in species composition in response to riverine habitat changes (J. Killgore, US Army Corps of Engineers, personal communications). Their research shows that seining often collects equivalent numbers of species as electrofishing, if not more, and provides data that meets their specific study objectives. However, species collected are usually limited to small bodied species. Consequently, the approach may not be appropriate for study objectives that require adequate sampling of large fish that comprise an important component of the fish assemblage in rivers. Their research also shows that effectiveness of seine hauls declines with increasing river size. Seining is also hindered at river locations where physical habitat is complex (e.g., boulders, large amounts of woody debris) or miry (i.e., soft and watery) substrates hinder foot travel.

7.2.3 Trawling

Trawling in inland rivers has recently received increased interest. Trawls are funnel-shaped nets that are towed along the bottom (bottom trawls) or in the water column (midwater trawls). As the net is towed through the water, fish enter the net, become exhausted, and drift to the cod end (rear) of the net until retrieved (Hayes et al. 1996). Variations in net configuration determine what is retained and survivorship (Herzog et al. 2005). Trawling is a commonly used method for sampling oceanic and estuarine habitats (Hayes et al. 1996) and reservoirs (Matsushita and Shida 2001), but has only been used to a limited extent in rivers (Pitlo 1992, Gutreuter et al. 1995, Dettmers et al. 2001, Wildhaber et al. 2003, Stewart and Barko 2005). Herzog et al. (2005) describes the successful application of the Missouri trawl for sampling benthic species in moderate to large size rivers. Stewart and Barko (2005) discuss the use of the same trawl configuration for collecting darter species undersampled by seining. Given these results, it

seems likely that use of trawling will increase in studies targeting benthic species that may be undersampled by other methods (e.g., inventory, monitoring of threatened and endangered species). Similarly, the method's selectivity for benthic species limits its use as a stand-alone method for bioassessment purposes.

TABLE 7-6. Field equipment supply checklist for fish sampling via electrofishing.

- scientific collection permit(s)
- boat, motor, and trailer
- boat electrofisher and associated equipment (generator, variable voltage pulsator, anode poles, cathode, gasoline)
- dip nets
- insulated waterproof gloves (linesman gloves)
- insulated footwear
- polarized sunglasses (day-time electrofishing only)
- lights/flashlights (for night sampling)
- livewells with functioning aerators and water circulation
- jars for voucher/reference specimens
- waterproof jar labels
- 10% buffered formalin (formaldehyde solution)
- measuring board[a]
- balance (gram scale)[b]
- fish sampling field data sheet
- taxonomic references (fish keys)
- laser range finder
- topographic maps
- copies of field protocols
- pencils, clipboard
- first aid kit
- US Coast Guard required safety equipment (personal floatation devices, fire extinguisher, etc.)
- cell phone
- global positioning system (GPS) unit
- tool box

[a] Needed only if program/study requires length frequency information
[b] Needed only if total biomass and/or the index of well-being are included in the assessment

7.3 The Large River Bioassessment Protocol (LR-BP) for Fish

The fish LR-BP is based on results of a study conducted on several Mid-Western rivers using an electrofishing design that permitted examination of the effects of designs and distances on fish assemblage metrics (Flotemersch and Blocksom 2005). While the results of the study likely apply to many rivers outside the study area, consultation of other more regionally specific literature is advised (Table 7-4).

The study concluded that depth plays a critical role in the response of fish assemblages to electrofishing and the resulting metric values. For example, at sites with a mean thalweg depth < 4 m, a daytime main-channel border design that includes electrofishing 1000 m along a single bank or 500 m on paired banks was sufficient to characterize sites for bioassessment purposes. At sites with a mean thalweg depth > 4 m, results were more variable. Therefore, at such sites,

the LR-BP protocol suggests that a switch from daytime to nighttime electrofishing be considered. If night electrofishing is not feasible, the LR-BP suggests increasing the electrofishing distance at these sites to a 1000-m paired-banks design or a 2000-m single-bank design. In addition, metrics based on fish species prone to diel movements should be interpreted with caution.

The fish LR-BP is quantitative and designed to support bioassessment and monitoring activities of states, regions, tribes and other agencies. It is designed to collect samples that are as unbiased and representative as possible within the logistical realities of fieldwork and constraints of time and budget, and are indicative of the ecological condition of a site when compared to sites of known condition. This sampling approach is not appropriate for qualitative studies that strive to maximize the number of species as a measure of local (alpha) diversity, although data collected using the fish LR-BP could be used to supplement qualitative investigations.

7.4 Sample Processing in the Field

The accurate identification of each fish collected is essential, and species-level identification is required (including hybrids in some cases). Field identifications are acceptable; however, voucher specimens may be retained for laboratory verification, particularly if there is any doubt about the correct identity of the specimen. Because the collection methods used are not consistently effective for young-of-the-year fish, and because their inclusion may seasonally skew bioassessment results, fish less than 20 mm total length are not identified or included in standard samples. During the identification process, be as precise as the data quality objectives require. Common variables that are recorded during the identification process include total count, length, weight, and the presence of external anomalies. Measurement of length may take the form of an actual measurement or placing specimens in size classes.

While processing fish, an assessment of the condition of the fish is often conducted. A widely used and reliable approach for documenting external anomalies as indicators of fish assemblage condition is to record DELT anomalies (deformities, erosions, lesions, and tumors) (Sanders et al. 1999). This is especially true for sites degraded by multiple and cumulative stressors. Documentation of such anomalies is an effective way to communicate information about degraded water quality to resource managers, the regulatory community, and to the general public. Guidelines for more extensive assessment of external and internal anomalies can be found in Goede and Barton 1990, Adams and Ryon 1994, Adams et al. 1993, 1996, Schmitt et al. 1999, and Smith et al. 2002.

7.5 Quality Control in the Field

Quality control must be a continuous process in fish bioassessment and should include all program aspects, from field collection and preservation to habitat assessment, sample processing, and data recording. Field validation should be conducted at selected sites and involves the collection of a duplicate sample taken from an adjacent reach upstream of the initial sampling site. The adjacent reach should be similar to the initial site with respect to habitat and stressors. To mitigate the effects of intersegmental fish movement, a section of shoreline (e.g., 50m) between successive electrofishing reaches is not sampled. Sampling QC should be performed

on a routine basis to document sampling error (field sampling precision) associated with a dataset and program; as a rule-of-thumb this can result from sampling adjacent reaches from a randomly selected subset of reaches..

Field identifications should be conducted by qualified, trained fish taxonomists who are familiar with local and regional ichthyofauna. Questionable records are prevented by: 1) requiring the presence of at least one experienced/trained fish taxonomist on every field effort, and 2) preserving selected specimens and those that cannot by readily identified in the field for laboratory verification or examination by a second qualified fish taxonomist. An approach for documenting taxonomic precision is suggested in Section 7.7.3. If being retained, specimens must be properly preserved and labeled. When required, chain-of-custody forms must be initiated following sample preservation, and must include the same information as the sample container labels.

All field equipment must be in good operating condition, and a plan for routine inspection, maintenance, and calibration must be developed to ensure consistency and quality of field data. Field data must be complete and legible, and should be entered on standardized field data forms and/or digital recorders. While in the field, the field team should possess sufficient copies of standardized field data forms and chains-of-custody for all anticipated sampling sites, as well as copies of all applicable standard operating procedures (SOPs) (see also Chapter 2).

7.6 Fish-based Index of Biotic Integrity

Approximately 22 different fish-based indices of biotic integrity have been developed for the assessment of streams and rivers in various regions and of differing types (Simon and Lyons 1995). Among these indices (which vary in terms of the number and complement of metrics), Table 7-7 summarizes seven examples that focus on the assessment of large rivers of the USA. Additional examples from outside the USA include Oberdorff and Hughes (1992), Hugueny et al. (1996), Ganasan and Hughes (1998), and Araujo et al. (2003). For a review on the use of environmental guilds for assessment of the ecological condition of rivers, consult Welcomme et al. (2006). This paper includes a list of ecological guilds, their typical behavior, reaction to changes in hydrograph, and typical species and can be used as a guide for the development of guild classification at the level of individual basins.

In reviewing the table, it is important to keep in mind that rivers vary in physical nature, as do the fauna and flora they support. Consequently, the metrics necessary to assess river condition, as well as metric response, may vary. For example, low levels of stressors (e.g., nutrients and thermal loading) may initially increase metric scores, and lower them at higher stressor levels. This is often observed in cold oligotrophic rivers, but not in warm water rivers. Also in such rivers, we see increases in centrarchids, catostomids, and cyprinids, simply because they are better adapted to such conditions than salmonids, cottids, and petromyzontids. Likewise, biomass increases with nutrient and thermal enrichment of cold oligotrophic systems.

TABLE 7-7. Fish metrics selected for inclusion in biological indexes developed for large rivers.

Metric	Response to General Stressors*	OH EPA Iwb[1]	OH EPA[2] FIBI	ORSANCO ORF (In)[3]	WI FIBI[4]	Wabash River IBI[6]	PN-IBI[5]	OR[7]
Species richness and composition								
# native spp.	decrease		X	X	X	X		X
# sunfish spp.	decrease		X	X		X		
# sucker spp.	decrease							X
# sucker spp.(round-bodied)	decrease		X	X	X	X		
% round bodied suckers	decrease		X		X			
% intolerant individuals	decrease						X	
# intolerant spp.	decrease		X	X	X	X		X
% tolerant individuals	increase		X	X		X	X	
# of great river spp.	decrease			X				
% great river individuals	decrease					X		
% simple lithophils	decrease		X	X	X	X		
% non-native individuals	increase			X				X
# non-native spp.	increase						X	
% riverine spp.	decrease				X			
# minnow spp.	decrease							X
# riverine spp.	decrease				X			
% common carp	increase						X	X
% coldwater individuals	decrease						X	
# native coldwater spp.	decrease						X	
# salmonid age-classes (whitefish omitted)	decrease						X	
% catchable salmonids	decrease							X
# sculpin age-classes	decrease						X	
% sculpin individuals	decrease						X	
# sculpin spp.	decrease							X
shannon h (numbers)	decrease	X						
shannon h (biomass)	decrease	X						
Trophic composition								
% omnivores	increase		X			X		X
% invertivores	decrease		X	X	X			
% top-piscivores	decrease		X	X				
% detritivores	increase			X				
% insectivorous	decrease					X		X
% macrovorous	decrease					X		
Fish abundance and condition								
# DELT anomalies	increase		X	X	X			
% DELT anomalies	increase					X	X	X
total biomass of catch	decrease	X			X			X
catch per unit effort (no./distance)	decrease					X		X
catch per unit effort (no/time)	decrease	X	X	X			X	

1 Ohio EPA 1987b: Ohio EPA's Index of Well Being (Iwb)
2 Ohio EPA 1987b: Ohio EPA's Fish Index of Biological Integrity (FIBI) for boatable rivers
3 Emery et al. 2003: ORSANCO's Ohio River Fish Index [ORF(In)]
4 Lyons et al. 2001: Wisconsin's Fish Index of Biological Integrity (WI FIBI) for large warm-water rivers
5 Mebane et al. 2003: Pacific Northwest Rivers Index of Biotic Integrity (PN-IBI)
6 Gammon and Simon 2000: Wabash River Index of Biotic Integrity (PN-IBI)
7 Hughes and Gammon 1987: Wilamette River IBI, Oregon (OR)
* Low levels of stressors (e.g., nutrients, thermal loadings) may initially increase metric scores, and lower them at higher stressor levels.

Program Highlight

Ohio River Valley Water Sanitation Commission (ORSANCO)
Fish Population Monitoring
Protocols for Non-wadeable Rivers

The Ohio River Valley Water Sanitation Commission (ORSANCO; the Commission) is an interstate agency charged with abating existing pollution in the Ohio River basin and preventing future degradation of its waters. ORSANCO conducts water quality monitoring and assessments on behalf of the Ohio River mainstem states (Illinois, Indiana, Kentucky, Ohio, Pennsylvania, and West Virginia). The Bimonthly Manual Sampling Program entails the collection of water column grab samples for water quality analysis. Fish assemblages are assessed using ORSANCOS's Ohio River fish index (ORFIn) for evaluating fish assemblage data.

ORSANCO Fish Assemblage Monitoring

ORSANCO developed an index to assess the condition of fish assemblages along 1,580 km of the Ohio River. Representative fish samples were collected from over 700 reaches, including 318 "least-impacted" sites, via standardized nighttime boat-electrofishing. A total of 55 candidate metrics were evaluated (based on attributes of fish assemblage structure and function) to derive a multimetric index of river health for the Ohio River. Metric evaluations considered the variability of these metrics spatially (by river kilometer) and temporally, and their responsiveness to human disturbances (e.g., effluents, turbidity, and embedded substrates). The resulting Ohio River fish index (ORFIn) comprises 13 metrics (Table 7-6) selected because they responded predictably to measures of human disturbance or reflected desirable features of the Ohio River. Two metrics were retained (the number of intolerant species and the number of sucker species [family Catostomidae]) from Karr's original index of biotic integrity. Six metrics were modified from indices developed for the upper Ohio River (the number of native species; number of great-river species; number of centrarchid species; the number of DELT abnormalities; percent individuals that are simple lithophils; and percent individuals that are tolerant species). They included three trophic metrics (the percent of individuals that are detritivores, invertivores, or piscivores), one metric of catch per unit effort, and one metric based on the percent of individuals as nonindigenous fish species. The ORFIn was responsive (i.e., significant negative correlations) to anthropogenic disturbances on substrate and water quality and was significantly lower in the first 500 m below point source discharges than at least-impacted sites nearby. Incorporation of the ORFIn into Ohio River assessments represents an improvement over other physicochemical protocols.

ORSANCO typically conducts fish assemblage studies every year from July through October. Fish samples are taken via electrofishing boat along 500-m shoreline zones at randomly selected sites. Each 500-m zone is marked with fluorescent orange paint or a surveyor's flag. Dissolved oxygen, conductivity, temperature, pH, secchi depth, river stage, and general weather are recorded before sampling begins. Each sample reach is electrofished by boat at night. The fish are netted, weighed, measured, species recorded, any unusual abnormalities are noted, habitats within the zone are recorded, and GPS coordinates are taken at the upstream, midpoint, and downstream section of the zone. These data are then used to calculate the ORFIn score for each site. Each site is classified into one of these habitat classes based on substrata composition. The ORFIn score is then compared with a habitat specific biocriteria value and the proportion of sites falling below the threshold is estimated as the proportion of the pool that is impaired.

7.7 Performance Characteristics for Biological Assessments Using Fish

7.7.1 Field Sampling

Quantitative (QN) performance characteristics for field sampling are *precision* and *completeness* (Table 7-8). Repeat samples for purposes of calculating precision of field sampling are obtained by sampling two adjacent reaches (i.e., adjacent 1000-m single-bank reaches or adjacent 500-m paired-bank reaches [Figure 7-3] or other (see Section 3.1.1). Fish samples from the adjacent reaches (also called quality control [QC] or duplicate samples) must be processed prior to data being available for precision calculations. These precision values are statements of the consistency with which the sampling protocols:

- characterized the biology of the river and
- were applied by the field team,

and thus, reflect a combination of natural variability and systematic error (see Chapter 3).

TABLE 7-8. Error partitioning framework for biological assessments and biological assessment protocols for fish. There may be additional activities and performance characteristics, and they may be quantitative (QN), qualitative (QL), or not applicable (na).

Component Method or Activity	Performance Characteristics				
	Precision	Accuracy	Bias	Representativeness	Completeness
1. Field sampling	QN	na	QL	QL	QN
2. Laboratory sorting/subsampling	na	na	na	na	na
3. Taxonomy	QN	QL	QL	na	QN
4. Data entry	na	QN	na	na	QN
5. Data reduction (e.g., metric calculation)	Na	QN	na	na	na
6. Site assessment and interpretation	QN	QN	QL	QL	QN

The number of reaches for which repeat samples are taken varies, but a rule-of-thumb is a randomly selected 10% of the total number of sampling reaches constituting a sampling effort (whether a yearly, programmatic routine, or an individual project). Metric and index values are used to calculate relative percent difference (RPD), root-mean square error (RMSE), and coefficient of variability (CV) (Table 3-2). Acceptance criteria for each of these would be established based on programmatic capabilities demonstrated via pilot studies, or through analysis of existing datasets produced using the same protocols. These criteria are not data quality thresholds beyond which data points should be considered for discarding. Rather, they are flags for potential problems (errors) in sample collection or processing, are used to help

determine the source(s) of the problems, and can be used to help develop recommendations for corrective actions.

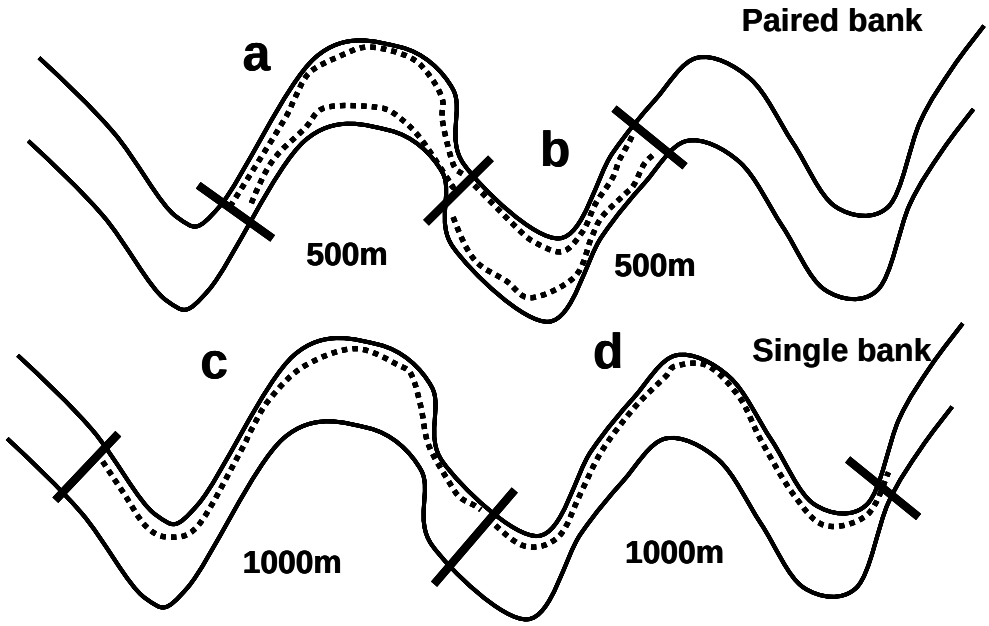


FIGURE 7-3. Two different scenarios for obtaining repeat reaches for large river fish bioassessments. Paired 500-m banks shown by a + b, and 1000-m single-bank approach by c + d (dotted line is where sampling is performed).

Percent completeness (Table 3-2) is calculated to allow communication of the number of valid samples (however validity is judged) that were collected as a proportion of those that were originally planned. This value serves as one summary of overall data quality for a sampling effort and it demonstrates confidence in the final results.

Qualitative (QL) performance characteristics for field sampling are *bias* and *representativeness* (Table 7-8). Attempts to minimize the bias associated with the LR-BP for fish for example, include two components of the field method. First, it is not limited to one or a few habitat types, (it is multihabitat and samples all shore-zone habitats within the reach. Second, allocation of the sampling effort is distributed throughout the entire sampling reach by use of a continuous electrofishing pass, preventing the entire sample from being taken in a shortened portion of the reach. The LR-BP field sampling method is intended to depict the fish assemblage that the physical habitat in the large river shore-zone has the capacity to support.

Accuracy is considered "not applicable" to field sampling (Table 7-8) because efforts to define analytical truth would necessitate a sampling effort excessive beyond any practicality. That is, the analytical truth would be all fish that exist in the river (shorezone electrofishing reach); there is no sampling approach capable of capturing every fish.

7.7.2 Laboratory Sorting/Subsampling

All laboratory-oriented performance characteristics are considered "not applicable" since most fish bioassessment methods assume field processing or sorting of fishes.

7.7.3 Taxonomy

Precision and *completeness* are QN performance characteristics that are used for taxonomy (Table 7-8). Precision of taxonomic identifications is calculated using percent taxonomic disagreement (PTD) and percent difference in enumeration (PDE) (Table 3-2), both of which rely on the raw data (list of taxa and number of individuals) from whole-sample re-identifications. The primary taxonomy is completed by the project taxonomist (T1); the re-identifications are performed by a secondary, or QC, taxonomist (T2) as blind samples. Since large river fish samples are typically processed in the field, this re-identification process would need to be conducted in the field. The "secondary" taxonomist could be a member of the electrofishing crew, or a second taxonomist could be brought on site on occasion, solely for the purpose of these performance checks. The number of identifications in agreement between the two sets of results, as an inverse proportion of the total number of individuals in the sample ((1-[number of agreements])/N), is precision of the taxonomic identifications. For example, the percent difference in sample counts by each of the taxonomists is "percent difference in enumeration (PDE)". PTD and PDE are evaluated individually, and can be used to indicate the overall quality of the taxonomic data, and if there is a problem, to help identify what is causing the problem. The number of samples for which this analysis is performed will vary, but 10% of the total sample lot (project, program, year, or other) is an acceptable rule-of-thumb. Exceptions are that large programs (>~500 samples) may not need to do >50 samples; small programs (<~30 samples) will likely still need to do at least 3 samples. In actuality, it will be program-specific and the number of samples re-identified will be influenced by multiple factors, such as how many taxonomists are doing the primary identification (there may be an interest in having 10% of the samples from each taxonomist re-identified) and how confident the ultimate data user is with the results. Mean PTD and PDE across all re-identified samples is an estimate of taxonomic precision (consistency) for a dataset or a program. Percent taxonomic completeness (PTC; [Table 3-2]) quantifies the proportion of individuals in a sample that are identified to the specified target taxonomic level (lowest practical taxonomic level, species, genus, family, or other, including mixed levels). Results can be interpreted in a number of ways: the individuals in a sample are damaged juvenile, or hybrid (increasing the difficulty of identification), many are damaged with diagnostic characters missing (such as coloration, fins, etc.), or the taxonomist is inexperienced or unfamiliar with the particular taxon.

Accuracy and *bias* are QL performance characteristics for taxonomy (Table 7-8). Accuracy requires specification of an analytical truth. For taxonomy, the analytical truth includes: 1) the museum-based type specimen (holotype, or other form of type specimen), 2) specimen(s) verified by a recognized expert in that particular taxon, or 3) unique morphological characteristics specified in dichotomous identification keys. Determination of accuracy is considered "not applicable" for this kind of taxonomy (most often used in routine monitoring programs) because it is focused on characterizing the sample; taxonomic accuracy, by definition, would be focused on individual specimens. Bias in taxonomy results from use of obsolete

nomenclature and keys, imperfect understanding of morphological characteristics, inadequate optical equipment, and poor training. Neither of these performance characteristics is considered necessary for field fish taxonomy in that they are largely covered by the estimates of precision and completeness. For example, although it is possible that two taxonomists would put an incorrect name on an organism, it is considered a low probability that they would put the same incorrect name on that organism.

7.7.4 Data Entry

Efforts to understand the quality of data entry activity may seem trivial. However, the impact of errors can be substantial, and, if undiscovered and uncorrected, can become amplified through the assessment process. This performance characteristic quantifies the number of correctly-entered data values as a proportion of the total number of data values entered. The process involves having a QC person, distinct from the staff doing the primary data entry, check all data values (100%) against the original handwritten datasheets. With the datasheets as the analytical truth, the rate of errors is the *accuracy* of the data entry (Table 7-8). As errors are found, they are corrected electronically. For their wadeable streams program, Mississippi DEQ found that the two data types with the highest error rates were the datasheet header information (e.g., stream name, latitude/longitude, date of site visit, names of field staff) and streambed particle size data (Mississippi DEQ 2003). This allowed corrective actions to be focused where needed. All other performance characteristics are considered not applicable.

7.7.5 Data Reduction (e.g., Metric Calculation)

For most biological assessment programs, raw data are the list of taxa found at a site (in a sample) and the number of individuals recorded for each taxon. Preparation of those data for analysis requires conversion to metrics (Table 7-7) or other terms; metric calculation is a form of data reduction. When electronic spreadsheets or other data manipulation techniques are used, queries are often built to perform both complex and simple calculations. If queries are not performing as intended, or links to the raw data are incorrect, errors in metric values can occur. *Accuracy* of data reduction is a QN performance characteristic (Table 7-8) that helps ensure database/ computer calculation routines are performing as intended. A subset of metric values is hand-calculated using only the taxonomic and enumeration data, which are then compared to those that result from the computer queries. A recommended approach involves calculating one metric for multiple samples (e.g., systematic, every third sample), as well as all metrics for at least one sample. If differences are found, each value should be checked for errors in the calculation process (hand calculator vs computer algorithm), and corrections made.

7.7.6 Site Assessment and Interpretation

QN performance characteristics for site assessment and interpretation are *precision*, *accuracy*, and *completeness* (Table 7-8). Site assessment precision is based on the narrative assessments from the associated index scores (e.g., good-fair-poor) of the reach duplicates. It quantifies the percentage of duplicate samples that receive the same narrative assessments. These comparisons are done for a randomly-selected 10% of the total sample lot. Table 7-9 shows that, for this example dataset, 79% of the replicates returned assessments of the same category (23 out of 29);

17% were 1 category different (5 of 29); and 3% were 2 categories different (1 of 29). Accuracy is the proportion of samples for which the biological index correctly identifies sites as impaired; the calculation is discrimination efficiency (DE) (Table 3-2). DE is a value that is developed during the index development and calibration process. Percent completeness (%C) is the proportion of sites (of the total planned) for which valid final assessments were obtained.

QL performance characteristics for site assessment and interpretation are *bias* and *representativeness* (Table 7-8). The final assessment of a site can be biased if a small number of reference or stressor sites are used during the calibration process. Low numbers of stressor sites can potentially result in high discrimination efficiencies that are spurious. If interpretation of assessment results fails to consider abnormal or extreme hydrologic or climatic events, or other non-natural catastrophic and localized events, results could be considered non-representative of ambient conditions.

TABLE 7-9. Assessment results shown for sample pairs taken from 29 sites, each pair representing two adjacent reaches (back to back). Assessment categories are 1-good, 2-fair, 3-poor, and 4-very poor.

Site	Replicate 1		Replicate 2		Categorical Difference
	Narrative	Assessment Category	Narrative	Assessment Category	
A	Poor	3	Poor	3	0
B	Poor	3	Poor	3	0
C	Good	1	Good	1	0
D	Poor	3	Very Poor	4	1
E	Fair	2	Fair	2	0
F	Poor	3	Fair	2	1
G	Poor	3	Poor	3	0
H	Very Poor	4	Very Poor	4	0
I	Very Poor	4	Very Poor	4	0
J	Poor	3	Poor	3	0
K	Poor	3	Poor	3	0
L	Very Poor	4	Very Poor	4	0
M	Very Poor	4	Very Poor	4	0
N	Poor	3	Fair	2	1
O	Poor	3	Poor	3	0
P	Poor	3	Poor	3	0
Q	Poor	3	Very Poor	4	1
R	Poor	3	Poor	3	0
S	Fair	2	Very Poor	4	2
T	Fair	2	Fair	2	0
U	Good	1	Good	1	0
V	Poor	3	Fair	2	1
W	Fair	2	Fair	2	0

TABLE 7-9. Continued.

Site	Replicate 1		Replicate 2		Categorical Difference
	Narrative	Assessment Category	Narrative	Assessment Category	
X	Poor	3	Poor	3	0
Y	Poor	3	Poor	3	0
Z	Very Poor	4	Very Poor	4	0
AA	Poor	3	Poor	3	0
BB	Fair	2	Fair	2	0
CC	Poor	1	Poor	1	0

Chapter 8.0 Data Analysis

This chapter…

- describes how to create IBI and RIVPACS models
- describes how to analyze regional, site-specific, watershed and gradient study design data
- describes different reporting approaches

Data Analysis…

- is critical in translating monitoring data into information for management action

8.1 Introduction

Assessment data are collected not only to help define the status of large rivers, but also to guide management decisions. Data must be translated into a format from which management decisions regarding water resources can be made. The strategies outlined in Chapters 2 and 3 on the elements of assessment and study design, provide users with approaches to clearly define questions and objectives to create an appropriate study design. These chapters also provide approaches for developing a thorough quality assurance plan (QAP) that will allow the study to meet and quantify measurement and data quality objectives (MQOs and DQOs). The tools introduced in those chapters lead to this section which outlines some common approaches for analyzing assessment data and presenting it in a way that is most useful for decision-making.

There are a variety of materials available detailing analyses of assessment data (Reckhow and Warren-Hicks 1997, Barbour et al. 1999). This section discusses the main approaches, but interested readers should also consult the existing literature (see Barbour et al. 1999). First, this chapter discusses two of the major biological analysis strategies used in assessment: the multimetric approach and the predictive modeling approach. Then, the chapter presents analysis approaches to be used under the major study designs introduced in Chapter 3 (watershed, site-specific assessments and gradient studies) and approaches that can be used for relating assessment data to stressors and stressor sources. Last is a brief discussion of the different approaches that can be used for reporting results.

8.2 Biological Analysis Strategies

Water quality data can be used as stand-alone data and analyzed as individual variables. However, it is common to combine physical habitat or biological data into habitat or biological indexes that synthesize multivariate data into one variable or score (e.g., index of biological integrity (IBI) - Karr et al. 1986, Hughes et al. 1998, Barbour et al. 1999, Karr and Chu 1999; predictive models - Moss et al. 1987, Novak and Bode 1992, Hawkins et al. 2000, Wright 2000; qualitative habitat evaluation index-Rankin 1989, non-wadeable stream habitat index - Wilhelm et al. 2005). It is important to note that indices are developed for specific methods. Data derived from different methods would have to be evaluated for comparability before being applied to an existing index or a different index developed.

8.2.1 Multimetric Indexes

Multimetric indexes of biotic condition for fish and benthic macroinvertebrates have been developed for many regions of North America and Europe and are generally accepted for biological assessment of aquatic resource quality. Some examples include IBI (Index of Biotic Integrity for fish; Karr et al. 1986), RBP (Rapid Bioassessment Protocol, Plafkin et al. 1989), ICI (Invertebrate Condition Index, Ohio EPA 1987), B-IBI (benthic IBI; Kerans and Karr 1994), SCI (Stream Condition Index; Barbour et al. 1996), and others (see Chapter 7 for large river fish index examples). A multimetric index is a simple sum or average of several standardized metrics. For index development, metrics are attributes of the biota that respond to anthropogenic stressors in consistent ways and are thus, useful indicators of stress (Barbour et al. 1999). Developing a multimetric index consists of three overall steps:

1. Classifying natural biological assemblages into relatively homogeneous groups, so that the species composition can be reliably predicted by geographic location or site characteristics;
2. Identifying metrics that respond to anthropogenic stressors; and
3. Aggregating standardized, non-redundant metrics that represent aspects of diversity, composition, sensitivity, and function into an index.

Data analysis for index development consists of characterizing reference conditions that will form the basis for assessment of degradation and calibration of the index to a gradient of human influence. This is a well-documented procedure (Davis and Simon 1995, Gibson et al. 1996, Barbour et al. 1999) and is described below. Reference site selection was described in Chapter 2.

8.2.1.1 Classification of Biological Resources

Index development requires a waterbody classification framework to partition natural variability. Classification frameworks can be geographic (e.g., ecoregions [Omernik 1995]), they may be based on continuous variables (e.g., catchment area, elevation), or they may be a combination. The framework should rely on characteristics that are intrinsic and independent of human influence (e.g., climate, topography, vegetation, soils, geology, elevation, waterbody type and size) and that account for differences in the composition of relatively undisturbed reference sites (Barbour et al. 1999, Hawkins et al. 2000b). Classification is best accomplished with reference sites that represent the range of natural conditions of the region (Chapter 2). Candidate reference sites that are based on least degraded physical habitat and water chemistry can also be used as the basis for river classification. Using quantitative criteria for reference site selection helps provide a consistent classification framework.

A result of the classification step is a set of rules that directs the partition of sites into biologically-meaningful natural classes or groups. These rules may be simple. For example, if elevation is above 2000 m, then a site belongs to the “mountain” class. Conversely, the rules may be complex, requiring multivariate discriminant equations to determine site class. Classes are initially determined by the biota of reference sites. However, because biological information is reserved for assessment, the rules generally do not use biological information.

The two basic approaches to developing rules are:

- Examining prior rules (e.g., ecoregions) with biological information. If a prior rule is found or modified that adequately explains biological variability, it is used for further index development.
- Developing posterior rules (e.g., using ordination) from a biological classification.

There is no clear distinction between prior and posterior rule development. Prior hypotheses could be applied (e.g., elevation and catchment area) to a biological classification to determine the rules for class boundaries. Rules may be fixed (the elevation example above) or probabilistic (a discriminant function). The key to classification is practicality within the region or state in which it will be applied; local conditions determine the classes.

The most common prior rules examined are geographic region, elevation or gradient, and measure of waterbody size (catchment area, stream order, surface area). As a guide for developing rules, landscape types or ecoregions are a very good start and account for much variability (e.g., Yoder and Rankin 1995, Barbour et al. 1996, Feminella 2000, Gerritsen et al. 2000a, Jessup and Gerritsen 2000). But in some landscapes, more continuous variables have done a better job accounting for variability (e.g., montane regions, Hawkins and Vinson 2000, McCormick et al. 2000, Pan et al. 2000, Van Sickle and Hughes 2000, Waite et al. 2000).

The general approach for confirming or testing prior classification rules is to examine alternative sets of prior classification rules to determine which rule yields the simplest classification of the reference sites in the data set and accounts for a substantial fraction of variability in biological composition among sites. The techniques for examining and testing the alternatives include: 1) ordination and examining prior classifications in ordination space (e.g., using unique labels for prior classes), 2) comparing prior classification to cluster analysis results using similarity analysis (Van Sickle 1997), and 3) multivariate analysis of variance (MANOVA) on prior groups.

Posterior rule classification involves using the biological data collected from reference sites to classify sites into groups based on similarity in taxonomic composition. This can be done using ordination or cluster analysis, and is the classification approach used principally in RIVPACS analysis (Hawkins et al. 2000a, Wright 2000). This approach does not use prior rules or adherence to any existing framework.

8.2.1.2 Selection and Evaluation of Metrics and Formation of a Multimetric Index

Metrics allow the investigator to use meaningful indicator attributes in assessing the status of assemblages and communities in response to perturbation. The definition of a metric is a characteristic of the biota that changes in some predictable way with increased human influence (Barbour et al. 1999). For a metric to be useful, it must have the following attributes: 1) ecological relevance to the biological assemblage under study and to the specified program objectives, and 2) sensitivity to stressors and a response that can be discriminated from natural variation. The purpose of using multiple metrics to assess biological condition is to aggregate

the information available from multiple structural and functional elements of aquatic communities into one score.

All metrics that have ecological relevance to the assemblage under study and that respond to the targeted stressors are potential metrics for testing. From this "universe" of metrics, some will be eliminated because of insufficient data or because the range of values does not sufficiently discriminate between natural variability and anthropogenic effects. In this step, investigators identify the candidate metrics that are most informative and, therefore, warrant further analysis.

Investigators should select the measures that are relevant to the ecology of rivers within a region to ensure that various aspects of the structure and function of the aquatic assemblage are addressed. Representative metrics should be selected from each of four primary categories: 1) richness measures for diversity or variety of the assemblage; 2) composition measures for identity and dominance; 3) tolerance measures that represent sensitivity to perturbation; and 4) trophic or habit measures for information on feeding strategies and guilds. Other metric categories (especially useful in fish multimetrics) include life history and reproductive strategies. Common metrics are shown in Table 8-1. Karr and Chu (1999) suggest that measures of individual health be used to supplement other metrics.

TABLE 8-1. Some potential metrics for periphyton, benthic macroinvertebrates, and fish that could be considered for rivers. Redundancy can be evaluated during the calibration phase to eliminate overlapping metrics.

	Richness Measures	Composition Measures	Tolerance Measures	Trophic/Habit Measures
Periphyton	• Total no. of taxa • No. of common nondiatom taxa • No. of diatom taxa	• % community similarity • % live diatoms • Diatom (Shannon) diversity index	• % tolerant diatoms • % sensitive taxa • % aberrant diatoms • % acidobiontic • % alkalibiontic • % halobiontic	• % motile taxa • Chlorophyll a • % saprobiontic • % eutrophic
Benthic macroinvertebrates	• No. Total taxa • No. EPT taxa • No. Ephemeroptera taxa • No. Plecoptera taxa • No. Trichoptera taxa	• % EPT • % Ephemeroptera • % Chironomidae	• No. Intolerant Taxa • % Tolerant Organisms • Hilsenhoff Biotic Index (HBI) • % Dominant Taxon	• No. Clinger taxa • % Clingers • % Filterers • % Scrapers
Fish	• Total no. of native fish species • No. of darter species • No. of sunfish species • No. of sucker species	• % pioneering species • Number of fish per unit of sampling effort corrected for drainage area	• No. of intolerant species • % of individuals as tolerant species • % of individuals as hybrids • % of individuals with disease, tumors, fin damage, and skeletal anomalies	• % omnivores • % insectivores • % top carnivores

It is generally not advisable to use metrics that are inherently unstable or variable due to their quantitative definition (e.g., ratio of scrapers to filterers, or ratio of EPT to Chironomidae in RBP 2; Plafkin et al. 1989). For example, ratios of two independent variables (x/y) should never be used as metrics because they range from zero (if x = 0) to undefined (if y = 0). Instead, use proportions of a total (x / (x + y)), which range from 0 to 1. Components of metric review include:

- Eliminating metrics that have too many zero values in the population of sites to calculate the metric at a large enough proportion of sites. Many zero values or a predominance of either very low values or very high values (close to 100%) indicate that the metrics may not have sufficient range to discriminate impairment. For example, the number of Plecoptera taxa (stoneflies) in even unstressed reference sites is often low, in the range of three to five genera. Although stoneflies are highly sensitive, there is not enough range (i.e., three taxa) to detect intermediate levels of impairment. This is why stoneflies are commonly grouped with mayflies and caddisflies to form the metric EPT taxa, which usually has sufficient scope (10-15 genera or more) to be a sensitive indicator.
- Using descriptive statistics (central tendency, range, distribution, outliers) to characterize metrics within the population of reference sites of each site class.
- Eliminating metrics where variability in the reference site population of a class is so large, they cannot discriminate among sites of different condition. The potential for each measure is based on containing enough information within a specific range of variability to discriminate among site classes and biological condition (reference vs degraded). Highly variable metrics (in unstressed sites) are poor indicators because their precision is low. This can also be characterized using signal/noise ratios across all sites (Kauffman et al. 1999).

It is important to understand the effects of various stressors on the behavior of specific metrics. If metric response is counter-intuitive or poorly understood on conceptual grounds, it is better to avoid using them.

The ability of a biological metric to discriminate between "known" non-stressed conditions and "known" stressed conditions (defined by physical and chemical characteristics) is crucial in the selection of core metrics for future assessments. Two general approaches to identifying responsive metrics to stressors are: 1) looking for categorical responses and 2) looking for response to gradients of stressors. The categorical approach is more common and analytically simpler, but does not provide potential diagnostic information. Examining response to gradients can only be done if measurements of the stressors exist in the data set.

Categorical Response: Examining categorical metric responses is based on comparing metric distributions in reference and degraded sites. The simplest comparison, and in many ways the most effective, is to examine box and whisker plots of the metric values in two groups of sites: reference sites and known "stressed" sites (defined by physical and chemical criteria, much like reference sites) (Figure 8-1). Box plots show several attributes of the distribution graphically: median, upper and lower quartiles, tails, outliers and/or minimum and maximum. Box plots of

two distributions (reference and stressed sites) show exactly how much the distributions differ or overlap with each other. Formal hypothesis tests are therefore not necessary; in fact, they are generally not meaningful because the question is not whether the reference and stressed sites differ (the subject of a hypothesis test), but whether a given metric can distinguish between them (e.g., Salsburg 1985, 1986, Yoccuz 1991, Suter 1996). Metrics having the strongest discriminatory power provide the most confidence in assessing biological condition of unknown sites.

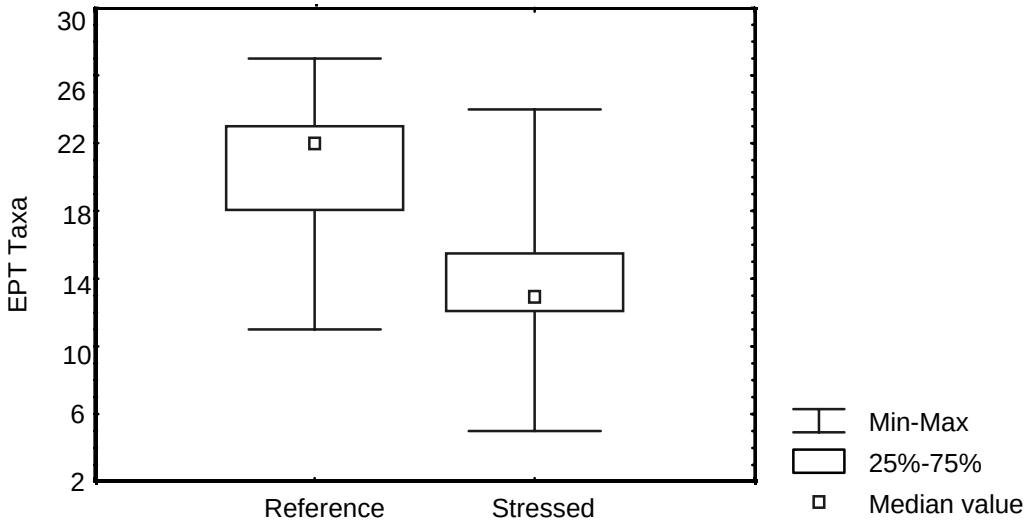


FIGURE 8-1. A box and whisker plot comparing the distribution of the number of EPT taxa, a common macroinvertebrate metric, in reference and stressed sites.

Discrimination efficiency (DE) is also used to evaluate metrics. The DE is the proportion of stressed sites that would be deemed different from reference if below a given threshold. For example, in Figure 8-1, if we choose the 25th percentile of EPT taxa (18 taxa) as a threshold, then all sites with fewer than 18 EPT taxa would be "different from reference". The discrimination efficiency for EPT is then the fraction of stressed sites with EPT taxa <18; 80% in this case.

Gradient Response: If quantitative measures exist for stressors or sources of stressors in the data set, then it is possible to examine the response of candidate metrics to those gradients using scatterplots. Measured stressors could include habitat, water chemical measures, water column contaminants, and sediment contaminants. Measured sources include land use or known discharges. Since many stressor measurements are correlated (e.g., pH, conductivity, sulfate), it is often advantageous to define stressor axes with principal components analysis (PCA) of chemical and habitat measures (e.g., Norton et al. 2000, Gerritsen et al. 2002) or some combined disturbance gradient (Fore 2004). In using the gradient approach, those metrics exhibiting the strongest response to stressors are usually selected as candidates. Other multivariate approaches can be used to identify responsive metrics, including canonical correspondence and canonical

correlation analysis of metrics and environmental variables to explore the relationship of metrics to certain stressors (Griffiths et al. 2001, 2002, 2003).

The final step in the process is combining candidate metrics into a multimetric index. The index should include metrics representing richness, composition, tolerance, and trophic aspects of the assemblage, contain minimally redundant metrics, and be able to discriminate reference from impaired sites with low variability. Metrics are standardized to a common scale so that all are weighted equally, and alternative combinations of metrics are examined for discrimination.

Responsive metrics are evaluated for redundancy. A metric that is highly correlated with another metric may not contribute new information. Redundancy among candidate metrics is determined from correlation analysis. A correlation matrix (Pearson) is calculated for all remaining candidate metrics. High correlation coefficients ($r > 0.7$) indicate strong linear relationships. A high correlation coefficient alone is not sufficient to eliminate one of a pair of correlated metrics (Karr 1991). Although there is no absolute threshold, $r > 0.7$ is generally used to indicate "forbidden combinations", and all pairs with $r > 0.7$ are examined with scatterplots to determine if there are nonlinearities in the relationship. If the scatterplot shows a curvilinear relationship, then both metrics may be retained because each one contributes information in a different part of the range.

The purpose of an index is to provide a means of integrating information from the various measures of biological attributes (or metrics). Metrics vary in their scale—they are integers, percentages, or dimensionless numbers. Prior to developing an integrated index for assessing biological condition, it is necessary to standardize core metrics via transformation to unitless scores. The standardization assumes that each metric has the same value and importance (i.e., they are weighted the same), and that a 50% change in one metric is of equal value to assessment as a 50% change in another.

Where possible, scoring criteria for each metric are based on the distribution of values from the population of sites, which include reference rivers. For example, the 95^{th} percentile of the data distribution is commonly used (Figure 8-2) to eliminate extreme outliers (e.g., Hughes et al. 1998, Gerritsen et al. 2000b). From this upper percentile, the range of the metric values can be standardized as a percentage of the 95^{th} percentile value, or other percentile (e.g., trisected or quadrisected), to provide a range of scores. Those values that are closest to the 95^{th} percentile would receive higher scores, and those having a greater deviation from this percentile would have lower scores. For those metrics whose values increase in response to perturbation, the 5^{th} percentile is used to remove outliers and to form a basis for scoring.

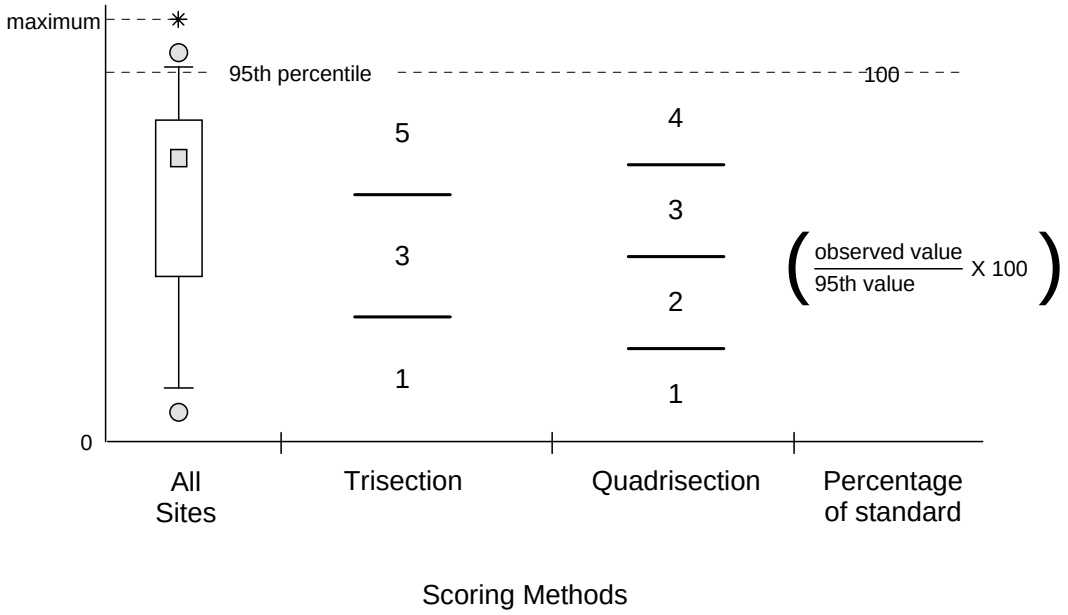


FIGURE 8-2. A comparison of different methods used for standardizing metric scores. The trisection method split the score distribution into 3 categories and the quadrisection, into 4. The last approach creates a continuous range of scores from 0 to 100, and the standardization formula depends on the response of the metric to disturbance.

Alternative methods for scoring metrics, as illustrated in Figure 8-2, are currently in use in various parts of the USA for multimetric indexes. A "trisection" of the scoring range has been well-documented (Karr et al. 1986, Ohio EPA 1987, Barbour et al. 1996, Fore et al. 1996). A "quadrisection" of the range has also been found to be useful for benthic assemblages (DeShon 1995). More recent studies are finding that a standardization of all metrics as percentages (0-100) of the 95th percentile value yields the most sensitive index, because information of the component metrics is retained (Minns et al. 1994, Ganasan and Hughes 1998, Hughes et al. 1998). Index development from statewide databases for Idaho (Jessup and Gerritsen 2000), Wyoming (Jessup et al. 2002), and West Virginia (Gerritsen et al. 2000b) are supportive of this third alternative for scoring metrics. The 95th percentile scoring method is as follows:

Scoring metrics that decrease with stress. The 95th percentile of metric values in all samples is assigned a unitless "best" or "standard" score of 100. Values between the minimum ("worst," usually 0) and the 95th percentile values are scored proportionally from 0 to 100 according to Equation 1:

$$score = \left(\frac{X - X_{min}}{X_{95} - X_{min}} \right) \times 100 \qquad \text{Equation 1}$$

where,

x = the calculated metric value

x_{95} = the 95th percentile of this metric's values in all samples

x_{min} = the minimum possible value, usually 0.

Scoring metrics that increase with stress. The 5th percentile of metric values in all samples is assigned a unitless best, or standard, score of 100. Values between the maximum (worst) value in the range and the 5th percentile value (standard, or best value) are scored proportionally from 0 to 100 according to Equation 2:

$$score = \left(\frac{X_{max} - X}{X_{max} - X_5} \right) \times 100 \qquad \text{Equation 2}$$

where,

x = the calculated metric value

x_5 = the 5th percentile of this metric's values in all samples

x_{max} = the maximum observed or possible value; e.g., 10 for HBI or 100% for percentage metrics.

In some States, trisected, quadrisected, or continuous scoring is based on percentiles of the reference distribution and not the entire range (Ohio EPA 1987, Stribling et al. 1998). After identifying redundant metric pairs, possible alternative indexes that exclude one of each redundant pair are built by averaging individual metric scores across different combinations or summing metric scores. Alternative configurations are examined for discrimination efficiency. The optimal index has no redundant pairs of metrics, has a high discrimination, and a mix of metrics from the richness, composition, tolerance, and trophic categories.

8.2.2 Predictive Models

Multimetric indicators are very broadly used across the US, and thus a more in-depth understanding exists on the use of these analytical approaches. However, there is growing interest in simultaneous use of predictive models in conjunction with multimetric approaches. To support this trend, this section provides detailed information related to development of predictive models, much moreso than for multimetric indicators in previous sections.

The River InVertebrate Prediction And Classification System (RIVPACS), developed as one bioassessment model for Britain, and AUStralian RIVer Assessment System (AUSRIVAS) are methods of bioassessment that predict an expected invertebrate assemblage in a river based on physical and chemical features of the river reach and surrounding landscape (Wright et al. 1984, Furse et al. 1984, Moss et al. 1987, Marchant et al. 1995, Wright 1995, Davies 2000, Simpson and Norris 2000, Wright 2000). These models compare the observed assemblage of macroinvertebrates at a test site to that expected in the absence of human disturbance (Observed:Expected; O/E) and assess biological condition based on a significant departure from 1.0 (where Observed = Expected). The observed assemblage is that found using standard sampling methods, whereas the expected assemblage is built using a model based on reference sites from across the sampling region. The approach is based on the concept that any site, in the absence of stressors, would likely have those taxa commonly found from physically similar reference sites. So, in essence, a site-specific reference condition is constructed for each test site based on the most probable assemblage of invertebrates expected at that test site in the absence of human disturbance. Conceptually, the expected taxa list is a weighted average of taxa frequencies found in reference sites. The weights represent the probability that a site falls in a particular group of reference sites based on physical similarity. Taxa from reference sites that

are physically very similar to a test site are weighted most. The approach has been applied successfully in the UK, Australia, and in several states in the USA (Wright et al. 1993, Hawkins et al. 2000, Paul et al. 2002).

This type of analysis proceeds in three main steps (Figure 8-3) described in detail below: 1) a cluster analysis of reference sites based on taxonomic composition to classify reference assemblage groups; 2) a discriminant analysis to develop linear models using physical variables to estimate the probability with which a test site belongs to each of the reference assemblage groups created in step 1; and 3) the prediction of the taxonomic composition of test sites based on group membership probabilities (step 2) and the frequency of taxa occurrence in each reference group.

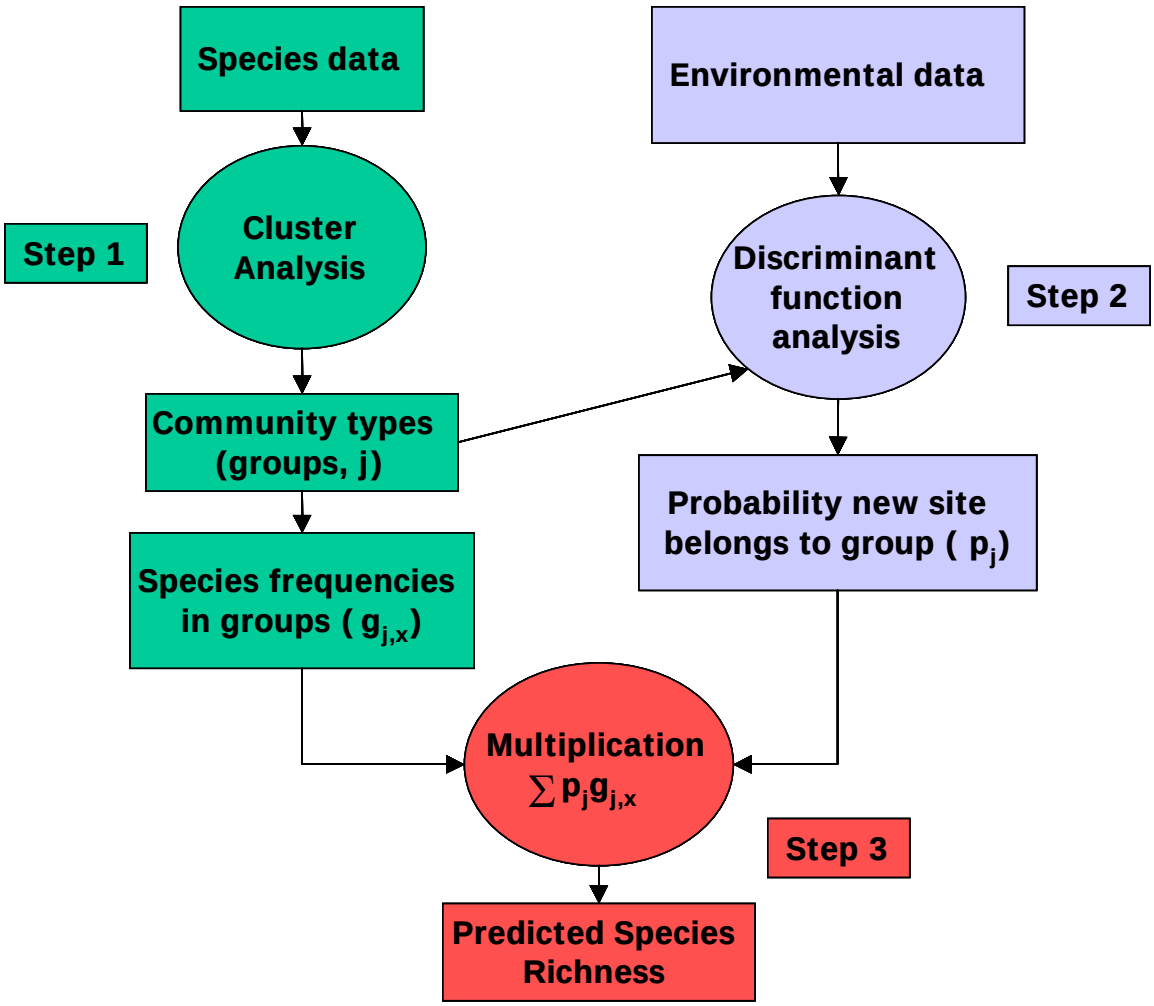


Figure 8-3. Schematic showing the three main steps involved in building RIVPACS-type bioassessment models.

8.2.2.1 Data Preparation for Predictive Models

RIVPACS models are built from variables considered relatively invariant to human disturbance (Wright et al. 1984, Hawkins et al. 2000, Wright 2000). Using established biogeographic factors that are minimally affected by human activity, it is possible to predict the expected assemblage for altered rivers. If alterable variables were used (e.g., nutrient concentrations, conductivity, forest cover), it would be difficult to discriminate the natural gradient from that caused by human activity; and confident prediction of an expected assemblage in the absence of human disturbance for a test site using this approach would be impossible. Commonly used variables for building RIVPACS models are shown in Table 8-2.

TABLE 8-2. Predictor variables commonly used for building multivariate predictive models.

Predictor Variables Used		Reference
<u>RIVPACS in United Kingdom</u>		Wright 2000
Mean depth	Slope	
Mean width	Discharge category	
Mean substratum	Mean air temperature	
Alkalinity	Annual air temperature range	
Altitude	Latitude	
Distance from source	Longitude	
<u>AUSRIVAS in Australia</u>		Simpson and Norris 2000
Longitude	Macrophyte taxa	
Latitude	Flow pattern	
Alkalinity	Macrophyte cover	
Altitude	Shading	
Distance from source	Bedrock	
Catchment area	Stream width	
Conductivity	Riffle depth	
Stream slope	Percent pebble	
Riparian width	Edge/bank vegetation	
Percent cobble	Vegetation category	
Percent boulder	Annual air temperature range	
Stream order	Percent gravel	
Discharge	Percent silt	
Percent sand	Percent clay	
<u>Models from California</u>		Hawkins et al.2000
Conductivity	Stream length	
Longitude	Mean width	
Catchment area	Sampling date	
Altitude	Slope	
Mean depth	Azimuth	
Latitude		

After a comprehensive dataset has been established, including ample reference sites across the range of natural environmental gradients sampled, the data must be prepared for analysis. As part of a preliminary analysis, all of the physical and chemical variables should be investigated graphically (e.g., frequency plot, normal quantile-quantile plot) to look for obvious lack of normality. Variables should be transformed as necessary. Common transformations include $\log_{10}$ for chemical concentrations and arcsine square-root transformations for percentile data (which are calculated using the ratio form of the percentile – 0 to 1.0) (Zar 1999). Transformed variables should also be inspected graphically. If necessary, tests for normality (e.g., Shapiro and Wilk's test) and equal variance (Bartlett's test) can be used to check these assumptions. Again, some departure from normality and equal variance is generally acceptable, especially since no hypotheses are being tested; but predictor models are being built using these techniques.

Many multivariate predictive models use an external validation (Hawkins et al. 2000). This consists of testing the final models with an independent set of data. Remember that models are built with reference sites only, so one approach is to set aside randomly selected reference sites (approximately 20%) before constructing the models. These are labeled as validation reference sites and are used to validate the models.

Other considerations in data preparation include some assessment of sampling the temporal and spatial variability in final scores. Estimating scores through time at a set of reference sites allows investigation of temporal stability of scores. Where this has been assessed, RIVPACS scores have exhibited marked stability (C.P. Hawkins, personal communication). Similarly, estimating scores in a number of replicate reaches within a set of reference sites examines the spatial stability (essentially the sampling error) associated with the models (see error estimation below). If different teams independently conduct each replicate, it is also possible to assess inter-team sources of sampling error. These all help identify the true error associated with model estimates (Clarke 2000).

8.2.2.2 Cluster Analysis

Once the environmental data are prepared for discriminant function analysis, the biological data should be prepared for cluster analysis. The cluster analysis is essentially the classification step in RIVPACS type modeling and is run only using reference sites and only using taxa that exist in reference sites. Software programs differ as to how data are prepared for analysis. Generally a site (rows) by taxon (columns) matrix is constructed with binomial data (0 or 1) entered into each cell to indicate the presence or absence of each taxon at each site. Cluster analysis can also be run using abundance data (commonly using Bray-Curtis similarity), which are commonly transformed using log (abundance), relative abundance, or fourth-root abundance. A cumulative taxa list is used, representing the entire list of taxa collected across the study and a record entered for each taxon at each site. At this point, two important factors need to be considered: taxon resolution and the exclusion of rare and/or common taxa.

Taxonomic resolution must be consistent among samples. This does not mean that all organisms must be identified to the same taxonomic level, but that a group (e.g., Diptera) is identified the same way among all samples. Thus, Diptera may be identified to family and Ephemeroptera to genus. In many real-world samples, fragments, juveniles, early instars, and pupae are not

identifiable to the target taxonomic level. These individuals are either not included in the data analysis, or they may be identified at the next higher taxonomic level. During data analysis, it is impossible to tell if records are different species or unidentifiable (e.g., damaged, too immature, etc.) individuals of the same species. There are two ways to use these records: 1) keep the species records or 2) collapse all of the species records to a higher level (Figure 8-4). Whatever choice is made, resolution decisions have to be applied consistently. In general, rules that keep the most data are preferable, but too much lumping can mask the unique elements that distinguish sites. Imagine models built from insect records at the order level only – there are only 13 unique aquatic or semi-aquatic orders of insects to use and the sites would look very similar. On the other hand, if species resolution is used, individuals that could not be identified to species (due to cost, specimen quality, or taxonomic expertise) would be lost. There is a trade-off between comparability of taxonomy among sites and maintaining as much information as possible. Taxonomic resolution rules (species, family, operational taxonomic unit, etc.) need be applied consistently across all sites – reference and test sites. So even though the cluster analysis step of RIVPACS uses reference data – the same taxonomic rules have to be applied to all sites.

The treatment of rare and common taxa in this step of the predictive model process is important as well. In general, rare taxa (occurring at less than 5% of reference sites) are often excluded because they contribute too much unique information for only a few sites and lead to under-clustering (over-splitting) (Hawkins et al. 2000). Likewise, common taxa (occurring at more than 95% of reference sites) are often excluded at this point because they can obscure unique differences among sites and lead to over-clustering (Hawkins et al. 2000). These taxa are not eliminated from the whole process, only from the cluster analysis. They are used later in the construction of expected communities for each site. Once the data have been prepared, with rare and common taxa removed and the validation set of reference sites set aside, a cluster analysis can be performed.

In this approach, the goal of cluster analysis is to produce as many groups as possible to simulate the continuous and dynamic assemblage structure that exists across any region and to minimize the number of unique small groups that would be too hard to predict accurately without overfitting the discriminant function models. Organisms exist along continuous environmental gradients with optima under certain conditions. Of course, there are a multitude of different environmental gradients and many different taxa. Therefore, modeling the distribution of all of those taxa and all of those continuous gradients would not be a trivial exercise. The cluster analysis step is used to dissect the distributions of taxa into as many small groups of co-occurring taxa as possible, much like how one learns to approximate curves by breaking them into small pieces using integral calculus. The ultimate result is a series of unique site clusters with similar taxonomic composition.

Original List

Taxon	Records	Taxon	Records
Family:		Family:	
Baetidae	19	Scirtidae	7
Genera:		Genera:	
Baetis	113	*Elodes*	1
Callibaetis	49	*Prionocyphon*	1
Centroptilum	18	*Scirtes*	1
Cloeon	10		
Heterocloeon	1		

Revised List

Taxon	Records	Taxon	Records
~~Family:~~		Family:	
~~Baetidae~~	~~19~~	Scirtidae	10
Genera:		~~Genera:~~	
Baetis	113	~~*Elodes*~~	~~0~~
Callibaetis	49	~~*Prionocyphon*~~	~~0~~
Centroptilum	18	~~*Scirtes*~~	~~0~~
Cloeon	10		
Heterocloeon	1		

FIGURE 8-4. A table demonstrating decisions made for lumping taxa upwards or discarding higher taxa records. In the case of the Baetidae, lumping all of the genera to the family level would obscure all of the unique information stored in those five genera, represented by the 191 reference site observations. Clearly removing the 19 records keeps the most information intact. In contrast, while three individual Scirtidae genera were identified, the vast majority of individuals could only be identified to family. Throwing out the seven records in favor of keeping the three genera records would lose the seven reference sites that had Scirtidae present. Clearly, the three genera records should be lumped to family unless there is 100% certainty the seven identified to family represent different genera.

Cluster analysis actually refers to a suite of different methods that group sites together based on their similarity with regards to many elements. Different cluster analysis approaches have been used in building bioassessment models. Approaches are split into agglomerative (lumping) or divisive (splitting) approaches. Agglomerative cluster analyses start with all of the sites separated and the sites with the greatest similarity are joined to form new groups. This is the most common type used in predictive modeling. Those groups and the remaining individual sites are then compared, and the next most similar elements are joined together – either two other sites or a third site is joined to the first group. The cluster analysis proceeds until all of the sites are grouped together into one large group. As agglomerative cluster analysis proceeds, however, there is less similarity among the elements being joined. By the end, the final group containing

all of the sites has the lowest overall average similarity. The common representation of the process of clustering is the dendrogram – which is a graphical way of viewing the clustering of sites. The axes usually contain some indication of the amount of unique information contained at each level of clustering.

There are a variety of agglomerative approaches, differing in the similarity or dissimilarity indices used and in the rules that are used to link similar elements together during clustering. The two most commonly used similarity indices are the Bray-Curtis and Jaccard indices. The most commonly used linkage methods are the flexible-beta method (with beta commonly set at -0.25), the unweighted pair-group method with arithmetic mean (UPGMA), and Ward's method. In practice, it may be best to explore a variety of approaches (or varying beta) and select the one that gives the best overall clustering.

Divisive cluster analysis is the second approach and works in the opposite direction from agglomerative clustering. Divisive analysis starts with all of the sites grouped together and they are split into the two most dissimilar groups. These two groups are then each split into two dissimilar groups (to yield 4 groups), and so on until either some pre-selected final number of groups is reached or all of the sites are split apart (Gauch 1982).

The most common divisive technique, the two-way indicator species analysis (TWINSPAN), is based on a correspondence analysis of site similarities (Hill 1979). Correspondence analysis is an ordination method that defines an axis along which the sites are ordered in terms of their taxonomic similarity. The mid-point of that axis is located and the sites are split along it into equal halves. Then, two new canonical analyses are run on the two new groups and those groups are split in half, and the process repeats accordingly.

Once a good cluster analysis is achieved, the selection of the optimum number of clusters is made. Obviously, the final cluster (one group) will not work. Likewise, using every individual site will not work. There is a point between these two extremes that represents the optimum number of clusters (Figure 8-5). This step also relies on professional judgment. A good rule-of-thumb is to draw a line in the middle of the cluster axis (e.g., 50% information remaining or other axis value indicating 50% of variance explained) and investigate how many clusters this resolves (Figure 8-5). A cluster consists of all the sites below the stem that is intersected by the line drawn. In many cases, there will be cluster nodes very near this line. So the line can be moved up and down until an optimum set of clusters is selected. The goal is to have as many clusters as possible to resolve the continuous distribution well, while at the same time avoiding very small clusters (<5 sites). Small clusters should either be joined to the next most similar cluster if possible or simply removed and placed with the test dataset. Once the final decision on the number of groups is made, the groups are numbered and each site within a group is given a group code. Again, the ideal approach may be to select three or four final clustering strategies and test each one through the rest of the analyses to see which produces the most precise and responsive models.

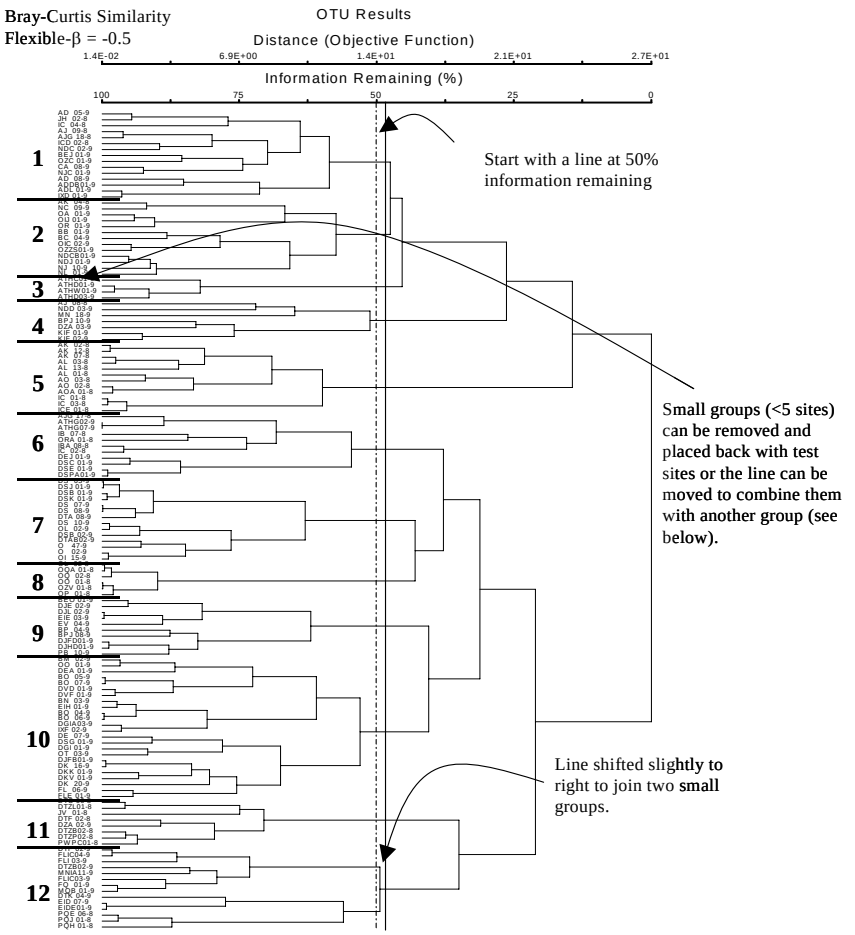


FIGURE 8-5. A final dendrogram used with a genera only dataset. This example starts with a line drawn at 50% information remaining to delimit groups and then the line is moved slightly to join two smaller groups, resulting in the final 12 groupings. Different software will produce different axes, but generally you start where half the variance is explained. The 14 group models would have worked as well. It would be best to test both groupings.

It is not uncommon to use an independent ordination of the sites using the same presence/ absence matrix as a check on the final cluster groupings (Wright et al. 1984). To do this, you use an appropriate ordination technique (e.g., non-metric multi-dimensional scaling or detrended correspondence analysis [DCA]) and give sites within each cluster group unique symbols. Visual assessment of the ordination can then be used to assess whether the groups are also unique in the new ordination space.

8.2.2.3 Discriminant Function Analysis

The goal of discriminant function analysis in predictive modeling is to generate a probability that a site belongs to each of the reference cluster groups generated by the cluster analysis. This probability is generated using environmental predictor variables available for each site. Discriminant function analysis (DFA) itself is a technique used when investigators have an

existing grouping structure and want to develop a model to predict the group membership of a new observation (Legendre and Legendre 1998). In some applications, we only want to know into which one group to assign a site. But in the RIVPACS approach, the desire is to generate the probability with which a new site belongs to each of the cluster groups. When a non-reference site has physical characteristics that resemble a mixture of a few different reference groups (e.g., along an ecotone), the expectation is a mixture of the most common taxa found in each of those different groups. The degree of mixture is generated using probabilities derived from discriminant function analysis. An important distinction should be made here. In this context, DFA is being used to build predictive models not to test hypotheses, so many of the statistical constraints are not applicable.

Discriminant function analysis is a mixture of MANOVA and multiple linear regression (MLR) (Statsoft 1994). Like classic ANOVA, MANOVA is a group means comparisons test that can determine if two or more groups are different with respect to many dependent variables simultaneously (Zar 1999). Its importance in discriminant function analysis is to decide if the groups identified with cluster analysis are indeed different with respect to a set of physical predictor variables. If they are not significantly different with respect to the variables, then those variables will not be much use in discriminating among the different groups.

Much like MLR, discriminant function analysis creates a set of equations that are used to predict to which group a site belongs. Unlike MLR, discriminant function analysis uses a canonical ordination approach, most like canonical correlation analysis, to construct linear equations (called discriminant functions) that are the combination of predictor variables that best discriminate among the groups. The number of discriminant functions (also called roots) is always one less than the number of groups or equal to the number of predictor variables, whichever is least. The first discriminant function explains the greatest variation among the different group means (it discriminates the best), the second function explains the second most, and so on. The coefficients in front of each predictor variable, when standardized, indicate which variable is most strongly contributing to the discrimination.

From these functions, a distance is calculated between each site and each group average. The Mehalanobis distance is often used in multivariate space. A site is assigned to the group centroid to which it is closest. But, more importantly, the probability that a site belongs to each group (which is what is needed for predictive models) is derived from the Mehalanobis distance to each group centroid. The closer a site is to a certain group centroid, the more it resembles the environmental characteristics of that group, and the higher the probability it belongs there. Sometimes, however, a site is so anomalous that the Mehalanobis distances to the centroids are very large. Most RIVPACS methods calculate a minimum distance that a site must be to any one centroid to be considered "within the experience of the model," and these are based on a chi-square value (the 99^{th} percentile chi-square value, degrees of freedom = number of groups - 1) (Moss et al. 1987). If none of the groups are within the critical chi-square distance to the site, the site is not assessed since a confident prediction of the probabilities cannot be made.

Most software programs will do all of the discriminant function analysis and most have stepwise options (forward, backward, or mixed) which allow users to choose criteria for selecting or removing variables until some final criterion is met. Entry and removal usually are determined

by F-to-enter and F-to-remove criteria, as in multiple regression. Similar to MLR, these F-values indicate the statistical significance of each variable to the overall discrimination; in essence, the significance with which an additional variable makes a unique contribution to the prediction of groups. Variables will be added in the order of their significance and will be added as long as they meet the criterion. As in MLR, the final model produced in a stepwise discriminant analysis may not be the global optimum. If possible, it is best to test different combinations of starting variables and see which model works best.

A novel approach for selecting an appropriate discriminant model has been developed using all-subsets modeling (Van Sickle et al. 2006). In this approach, all possible combinations of predictor variables are used and run through the calculation of O/E scores for calibration and validation reference data. The best predictor combinations are those that produce models that are the most precise (lowest standard deviation or root-mean square error of O/E scores in reference sites) while avoiding over-fitting (similar values for validation data). These models are available in the R open-source statistical programming language and offer an alternative to the traditional stepwise approach described here. One advantage of this approach is that it considers the universe of possible models, minimizing the risk of selecting locally optimal models. It also places a large value on avoiding overfit models, which is one of the more important risks when constructing these (or any) models.

Among the many statistics often generated from DFA, Wilk's λ is a common statistic used to indicate how well a model discriminates among groups (Pillai's trace, and Lawley Hotelling's trace are other similar statistics) (Zar 1999). Values range from 1.0 (no discrimination) to 0.0 (full discrimination). Wilks λ can be used to help select among the most discriminating models. The all-subsets modeling routines also use Wilks λ to evaluate and select the most discriminating models.

The ultimate test of model performance, in most cases, has been how well they predict the assemblage structure (i.e., how close the number of expected taxa matches the observed) of the reference sites for both the model building and validation datasets, while minimizing the risk of over-fit models. Highly discriminatory models are the goal, but over-fitting problems are also a threat. The all-subsets modeling routines include methods for evaluating the risk of over-fit models (Van Sickle et al. 2006). The value of independent set of reference validation data, however, cannot be overstated. Running the final model through the validation data will also provide an indication of model fit.

The classification of elements into distinct groups is a traditional focus of discriminant function analysis. Predictive modeling, is more interested in the group membership probabilities rather than exact group classification. However, the classification efficiency can also be investigated to look at general model fitting. In DFA, a classification matrix is a matrix of actual group membership vs predicted group membership, and is an *a posteriori* analysis, since it is looking at how well it predicts group membership of sites actually used to build the models. Therefore, it is not truly independent. In classic discriminant function analysis, the group classification functions derived from the discriminant functions are run for each site. A site is then assigned to whatever group classification score is highest. These are compared to the actual group to which each site belongs. In RIVPACS modeling, group classification efficiencies around 50% or less

are not uncommon, especially for small groups. This applies to the validation set as well, which is a more appropriate independent test of the classification efficiency. The all-subsets models actually compare DF classification efficiencies after leave-one-out cross-validation and resubstitution routines to evaluate appropriate model size and model fit.

The final step from the discriminant function analysis is the calculation of group membership probabilities, which is the final product of interest from this step. These membership probabilities were discussed above but need to be explained in detail. As described, the actual goal of the discriminant function analysis is to generate the probability with which each site belongs to each reference group. The cluster analysis was used to break the continuous distribution of communities into discrete pieces and the discriminant function analysis uses the physical characteristics of those groups, in a sense, to place a site back along that continuous gradient. Ideally, each test site would look physically just like one reference group. But what about those sites that fall somewhere among the physical characteristics of a number of groups? As mentioned earlier, those sites would have an equal probability of being in any one of the groups. Those probabilities are generated from the Mehalanobis distances. The Mehalanobis distance is a multivariate distance measure. It is the distance from any one site to the centroids of each of the different groups in multivariate space. The probability a site belongs in each group is derived from those distances – the closer a site is to one centroid, the higher the probability it belongs to that group. Many programs will calculate these probabilities using a variety of methods. In the original RIVPACS formulation, the probabilities were calculated using the formula:

$$p_j = q_j / \sum_{j=1}^{k} q_j , \qquad \text{Equation 3}$$

where p_j is the probability a site belongs to group j (of k different groups). The value q_j is a weighted distance measure and is defined as:

$$q_j = n_j \times e^{\left(\frac{-d_j^2}{2}\right)}, \qquad \text{Equation 4}$$

where n_j is the number of sites in group j, and d_j^2 is the square distance (e.g., Euclidean, Mehalanobis) between the site score and each group mean discriminant function score (Moss et al. 1987). These probabilities are the important outcome of the discriminant function analysis. They are combined with taxa frequencies in each group to predict the final taxonomic composition of a site. This will be explained in the next section.

Some sites are so far from any group centroid that an accurate determination of the probabilities cannot be made. The critical distance for a site to be accurately determined is calculated using the 99th percentile chi-square distribution value based on degrees of freedom equal to the number of groups. Since the Mehalanobis distances follow a chi-square distribution, any site that does not contain a distance less than the critical distance cannot be adequately assigned a probability and is considered "outside the experience of the model". These sites are often set aside and must wait until more reference sites with similar physical characteristics are assessed and the model is updated. If the sites are taken through the prediction analysis, any conclusions using O/E scores

generated from these sites need to be tempered by the fact that they are physically distinct from the reference groups used to construct the models.

8.2.2.4 Prediction of Taxa Composition

The final step in model building is to predict the number of expected taxa for a site. Before this step takes place, rare and common taxa removed before cluster analysis are added back into the database. These taxa, while rare over all sites, may be frequently found in one group and would be an important prediction for that group. Once these are reincorporated, the prediction analysis proceeds.

As mentioned before, the predicted taxa list for a site is not based solely on the taxa composition of the one reference group to which a site is most similar. If that were the case, one could simply find the group to which the site had the highest probability of belonging and compare the observed assemblage to the average assemblage composition of that one group. If each test site looked exactly like only one reference group, this would be fine. But sites are often physically similar to several groups, because the groupings frequently reflect subtle differences among reference sites (e.g., low gradient vs high gradient reaches within one basin). The sensible thing is to predict a mixture of taxa based on: 1) which group a site is most similar to and 2) which taxa are most frequently found in those groups. Therefore, essentially, a weighted average expected assemblage composition is calculated. This is done by using the probability a site belongs to each reference group as the weight and then multiplying this by the frequency of taxa in each reference group (Moss et al. 1987).

In order to do this, the frequency of each taxon in each reference group has to be estimated. This is done by calculating the frequency with which each taxon is found in each group (Table 8-3); $g_{j,x}$ = proportion of reference sites in group j containing taxon x. This value is calculated for each taxon in the master taxa list (over all sites). In the end, each taxon has a frequency with which it occurs in each reference group. Many taxa from the master list are not found in every group; therefore, they will have a probability of zero where they are absent; others are ubiquitous and have a value near 1.0 for every reference group.

Now that the probability of membership of any site in each reference group (p_j) has been generated from the discriminant function analysis and the frequency of every taxon x in each reference group ($g_{j,x}$), the probability of capturing (P_c) each taxon x at any site can be calculated (Table 8-4):

$$P_{c,x} = \sum_{j=1}^{k} p_j \times g_{j,x} \text{ , for } k \text{ reference groups.} \qquad \text{Equation 5}$$

TABLE 8-3. The first component of the prediction phase is to estimate average assemblage composition of reference groups. For each taxon, the fraction of reference sites containing each taxon is calculated. This is an estimate of the frequency ($g_{j,x}$) with which each taxon (x) is found in each group (j). A sample for a few taxa is shown here. Not all the reference sites could fit in the table. But for the first taxon (*Ablabesmyia)* in group 1, 11 of the 15 sites had that taxon; therefore, its frequency at that site is 11/15 = 0.73. Only 1 of the 15 sites in group 1 contained *Acroneuria*, therefore its frequency in group 1 is 1/15 = 0.07. This proceeds for all taxa (even the rare ones added back in) and for all 12 groups. Note that some taxa are fairly common across the groups (*Baetis*) whereas others are frequent in only a few groups (*Acroneuria*).

		Frequencies				
Group	1	0.73	0.07	0.33	0.73	0.27
	2	0.77	0.00	0.38	0.15	0.23
	3	1.00	0.00	0.00	0.50	0.25
	4	0.29	0.14	0.14	0.57	0.00
	5	0.75	0.58	0.00	0.75	0.17
	6	0.09	0.00	0.00	1.00	0.09
	7	0.93	0.00	0.36	1.00	0.00
	8	1.00	0.00	0.50	1.00	0.00
	9	0.00	0.00	0.00	1.00	0.00
	10	0.17	0.00	0.21	1.00	0.00
	11	0.00	0.00	0.00	0.88	0.00
	12	0.21	0.00	0.29	0.93	0.00
Site	Group	*Ablabesmyia*	*Acroneuria*	*Anopheles*	*Baetis*	*Basiaeschna*
AD 05-92	1	0	0	0	0	0
AD 08-92	1	1	0	0	1	0
ADDB01-92	1	0	0	0	1	0
ADL 01-92	1	0	0	0	0	0
AJ 09-87	1	1	0	1	1	1
AJG 18-87	1	1	1	0	1	1
BEJ 01-96	1	1	0	1	1	0
CA 08-98	1	1	0	0	1	0
IC 04-87	1	1	0	0	0	0
ICD 02-87	1	1	0	1	1	1
IXD 01-92	1	1	0	1	1	0
JH 02-84	1	0	0	0	1	0
NDC 02-95	1	1	0	0	1	0
NJC 01-95	1	1	0	0	0	1
OZC 01-96	1	1	0	1	1	0
AK 04-86	2	0	0	0	0	0
BB 01-96	2	0	0	0	0	0
BC 04-96	2	1	0	0	0	0
NC 09-95	2	0	0	0	0	0
NDCB01-95	2	1	0	0	0	1

TABLE 8-4. Having calculated the taxon frequencies ($g_{j,x}$, above) and the group probabilities (p_j, from the discriminant function analysis), the product of these values is used to calculate the probability of capturing each taxon at a site (P_c). For example, for *Ablabesmyia* at site AD 05-92, the probability that a site is in each group (p_j) is multiplied times the frequency of finding *Ablabesmyia* in each reference group (g_j). The sum of those products = 0.713, which is the probability of capturing *Ablabesmyia* at this site. The same calculation is made for all taxa.

Site AD 05-92

	Frequencies ($g_{j,x}$)			Probability of Group Membership (p_j)	($g_{j,x}$)(p_j)		
Group	*Ablabesmyia*	*Acroneuria*	*Baetis*		*Ablabesmyia*	*Acroneuria*	*Baetis*
1	0.73	0.07	0.73	0.657	0.479	0.046	0.480
2	0.77	0.00	0.15	0.012	0.009	0.000	0.002
3	1.00	0.00	0.50	0.136	0.136	0.000	0.068
4	0.29	0.14	0.57	0.015	0.004	0.002	0.009
5	0.75	0.58	0.75	0.096	0.072	0.056	0.072
6	0.09	0.00	1.00	0.081	0.007	0.000	0.081
7	0.93	0.00	1.00	0.000	0.000	0.000	0.000
8	1.00	0.00	1.00	0.002	0.002	0.000	0.002
9	0.00	0.00	1.00	0.001	0.000	0.000	0.001
10	0.17	0.00	1.00	0.001	0.000	0.000	0.001
11	0.00	0.00	0.88	0.000	0.000	0.000	0.000
12	0.21	0.00	0.93	0.000	0.000	0.000	0.000
				Probability of Capture (P_c) =	0.713	0.102	0.717

$P_c = \Sigma(g_{j,x})(p_j)$

Note that each probability of capturing a taxon is a continuous probability and not a discrete number. It is derived from the probability of group membership and the distribution of taxa frequencies. The expected number of taxa (E), then, is the sum of the capture probabilities of all the taxa at a site:

$$E = \sum_{x=1}^{i} P_{c_x} . \qquad \text{Equation 6}$$

This value is compared to the sum of the expected taxa (from the same master taxa list) actually observed (O) at the site. It is important to note that the number of observed taxa is the sum of only those expected taxa that are actually observed. The final ratio of these values (O/E), is the proportion of expected taxa actually observed at the site and is the indicator of biological condition. At relatively undegraded sites, one would expect to capture all the taxa frequently found in reference sites of comparable physical characteristics from the same region and O/E = 1.0. The lower the O/E ratio is, the fewer expected taxa actually captured.

Because the expected number of taxa is generated from a continuous frequency distribution over many reference sites within a group, capture probabilities can range from 0 to >1.0. It is possible

to have a test site with O/E >1.0, where there are more taxa captured than expected. This reflects a site where one observes many taxa with even partial probabilities of capture (e.g. 0.4), so that the sum of observed taxa (integers) is greater than the sum of expected (fractions <1.0). The average reference site O/E score, however, ought to be equal to 1.0 and this is used as a check on the adequacy of the model. If the mean reference O/E is significantly different from 1.0, then there is a problem with the model and it would need to be checked.

To this point, all taxa have been considered, regardless of their probability of capture at each site, which introduced some variability in comparing fractional expected data to integer observed data. Several RIVPACS-type model applications constrain the expected taxa list to only those taxa with a capture probability >0.5 (e.g., AUSRIVAS, Simpson and Norris 2000). This limits the list of taxa considered (both observed and expected) to only the most commonly expected. It is important to remember that the number of observed taxa is the sum of only those expected taxa that are actually observed. So if one only uses taxa with P_c >0.5 to estimate the expected number of taxa, one would only count actual observances of that same restricted taxa list, not all of the observed taxa.

The primary test of final model adequacy is running an independent set of validation reference sites through the model and calculating O/E scores (Hawkins et al. 2000, Simpson and Norris 2000). Therefore, a test of model robustness is that the O/E of the validation dataset is not significantly different from the O/E of the dataset used to construct the models, and neither of these means should be significantly different from 1.0 (Figure 8-6).

A second test of model adequacy is model precision. The objective is to create models with as low a variation about the mean reference score (1.0) as possible (i.e., to precisely predict the observed taxa). One rule of thumb is a standard deviation (or root mean square error) of mean reference O/E score of 0.15 to 0.20 or less (Figure 8-6). The lower the standard deviation is, the more precise the model and the greater the potential discrimination of degraded rivers. This also means that more degraded classes can be resolved.

A third way of assessing the model adequacy is to compare the model with a null model (Van Sickle et al. 2005). The null model for these predictive models is to simply compare the number of observed taxa at each reference site to the sum of the average frequency of taxa across all the reference sites without any clustering. In the null model, therefore, the number of expected taxa (E) is the same for every site. In essence, the null model ignores the cluster analysis and discriminant analysis and assesses how much extra precision one adds by going through those calculations. This is fairly straightforward to do, can be done for any capture probability threshold, and is a good check on the modeling effort.

One final important test of the models is whether or not the O/E scores respond to disturbance. There are a number of ways to evaluate disturbance response (Figure 8-7). Gradient approaches evaluate the response of O/E scores to a pre-determined disturbance gradient (e.g. water chemistry, land use, habitat, etc.). Another approach would be to rank degradation classes and test whether there are significant differences in mean O/E score between the reference class and the degradation classes, using either ANOVA or some other means comparison test.

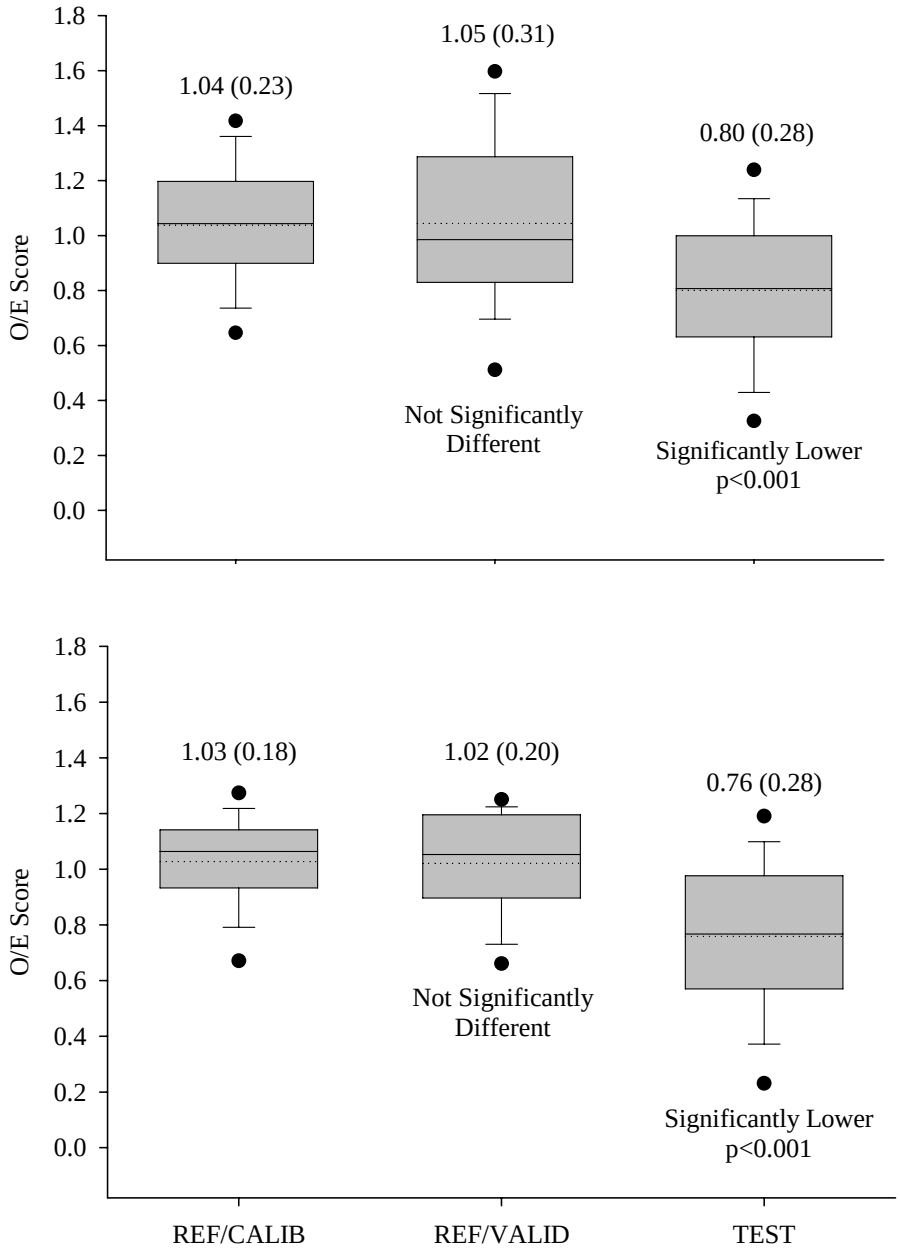


FIGURE 8-6. This figure shows the O/E score distributions for reference calibration, validation, and non-reference test site data. The dataset used genera only data and results for both the P_c>0.01 (top) and P_c>0.5 (bottom) taxa are shown. Mean scores are shown along with the standard deviations in parentheses. Reference validation (REF/VALID) scores were not significantly different from reference calibration (REF/CALIB) data scores, but non-reference scores (TEST) were significantly lower than both reference datasets (ANOVA, Tukey's HSD comparison). Whiskers indicate the 10th and 90th percentiles, and the boxes indicate the 25th and 75th percentiles. The solid line is the median and the dotted line the mean O/E scores. Filled circles indicate the 5th and 95th percentile values.

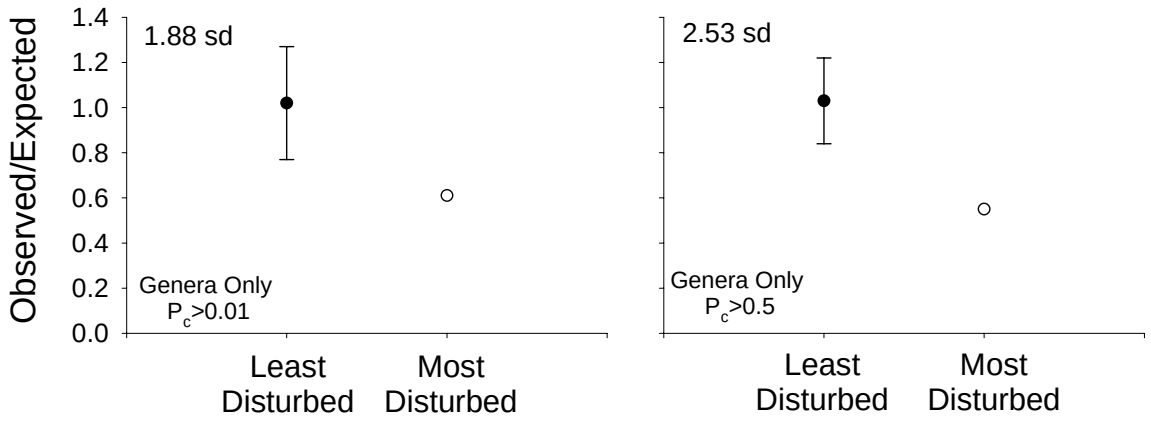


FIGURE 8-7. Comparisons of the discrimination between least and most disturbed sites using P_c>0.01 and P_c>0.5 taxa. One standard deviation around the mean least disturbed O/E scores is shown and compared to the mean O/E scores for the most disturbed sites. The number of least disturbed standard deviations between the least and most disturbed mean scores is also shown. The higher this value, the greater the discrimination.

8.2.3 Estimating Measurement Error

Variability has many possible sources. In dealing with bioassessment variability, the goal is to: 1) minimize variability due to uncontrolled measurement error and 2) characterize and partition the natural variability. When sampling rivers measurments (e.g., taxa richness) are often made at single points in space and time (riffle, 10-cm depth, 10 AM on 2 July). If the same measurement is made at a different place (pool, 1-m depth) or time (30 January), the measured value will be different. These two natural components of variability (space and time in this example) are called sample variability or sampling error (Fore et al. 1994). A third component of variability, called measurement error, refers to the ability to accurately measure the quantity of interest. Measurement error can be affected by sampling gear, instrumentation, errors in proper adherence to field and laboratory protocols, the choice of methods used in making determinations, and small-scale spatial variability at the sample site. The three basic rules of efficient sampling and measurement are:

1. Sample so as to minimize measurement error.
2. Characterize the components of variability that influence the central questions and reporting units.
3. Control other sources of variability that are not of interest, in order to minimize their effects on the observations.

.

In the example of taxa richness, it may be useful to sample each of several rivers, with a 1-m^2 kick net in early spring before the bulk of emergence. Many reaches are sampled in this example to examine and characterize the variability due to different reaches (the sampling unit). Each reach is sampled in the same way, in the same place, and in the same index period (time frame)

in an attempt to minimize variability due to habitat and season, which are not of interest in this particular study.

In the above example, taxa richness may vary with habitat, among rivers, and time of sampling (season, year), a fact that may be particularly important for sets of reference sites. If the spatial and temporal components of variability within rivers are large, then it is best to use either an index period sample or to estimate a composite from several determinations. For example, taxa richness may vary more between spring and fall samples within a river than among similar rivers within an ecoregion.

Measurement error is the result of methodological bias and errors: gear bias; improper use of gear or improper training; variability in use of gear; laboratory errors (chemical analysis errors); and natural variability that is not of interest and is not being sampled. Measurement error is minimized with methodological standardization; selection of cost-effective, low variability sampling methods; proper training of personnel; and quality assurance procedures designed to minimize methodological errors. Method performance standards (see Chapter 3) are designed to help ensure that these kinds of errors are consistently held to a minimum.

ANOVA is used to estimate measurement error. All multiple observations of a variable are used (from all streams with multiple observations), and streams are the primary effect factor. The root mean square error (RMSE) of the ANOVA is the estimated standard deviation of repeated observations within sites. A hypothesis test (F-test) is not of interest in this application because it tests the trivial hypothesis that sites are different from one another.

The estimated RMSE is used in the same way as standard deviation; in this case, it is an estimate of the standard deviation around a single point, and it is used to estimate a confidence interval around the point. The advantage is that a confidence interval can be estimated without replication at the site, because we are using the population estimate of standard deviations around single measurements.

Having a standard deviation, one can estimate confidence intervals around an index score or O/E score of a site (Fore 2004). For a single (non-repeated) sample, the confidence interval is:

$$CI = \pm Z_{\left(1-\frac{\alpha}{2}\right)} \frac{S_{rep}}{\sqrt{n_{rep}}}, \qquad \text{Equation 7}$$

where S_{rep} is the standard deviation calculated as RMSE with ANOVA, n_{rep} is the number of replicates at the site, and $Z_{(1-\alpha/2)}$ is the cumulative standard normal deviate (Z – score) for the chosen α. This approach makes three assumptions:

- measurement error is normally distributed,
- measurement error is not affected by site class or impairment, and

- the sample standard deviation of repeated measures is an unbiased estimate of population measurement error.

The same procedure can be used to estimate variability due to season and year, if sites are resampled in multiple seasons or years.

Natural variability that is not of interest for the questions being asked, but may affect ability to address these questions, should be estimated with the RMSE method. If the variance estimated from RMSE is unacceptably large (i.e., as large or larger than variance expected among sample units), then it is often necessary to alter the sampling protocol, usually by increasing sampling effort in some way, to further reduce the measurement error. Measurement error can be reduced by multiple observations at each sample unit (e.g., multiple Ponar casts at each sampling event, multiple observations in time during a growing season or index period, depth-integrated samples, or spatially integrated samples).

Spatial integration of sample material and compositing the material into a single sample is almost always more cost-effective than retaining separate, multiple observations. This is especially true for relatively costly laboratory analyses such as organic contaminants and benthic macroinvertebrates.

For quality assurance, some effort will always be required for repeated samples so that measurement error can be estimated from a subset of sites. Repeated measurement at 10% or more of sites is common among many monitoring programs, and is recommended (see Chapter 3).

8.3 Site-Specific Assessments

The next two sections deal primarily with the analytical approaches that can be used for site-specific and watershed assessments. The design considerations for these approaches were outlined in Chapter 3. Here we describe the analysis methods.

For site-specific assessments using before-after control-impact (BACI) type designs, the analytical approaches depend on which of the designs was used. In any of the analytical approaches, however, some attention to data preparation is necessary. Most of the BACI analysis approaches use a form of analysis of variance (ANOVA) or simpler parametric means comparison tests (e.g., t-test). As mentioned in the multimetric data preparation discussion above, variables that will be compared using parametric analyses must adhere to some basic assumptions. All the variables to be used, including individual metric, multimetric or O/E scores, should be explored visually for normality and equal variance. There are tests that can be used to examine these assumptions as well (e.g., Kolmogorov-Smirnov). The most important assumption is independence of observations. As mentioned above, as long as there is substantial spatial and temporal separation of sampling, independence generally should not be a problem.

The simplest test in site-assessment design is the t-test. The t-test can be used to compare two means or to compare a mean to some specific value (i.e., is the mean O/E score in samples below a discharge different from 1.0?). In the BACI designs, simple t-tests can be used to compare

pair-differences between the before and after periods. A significant t-test would suggest that mean differences changed after the treatment (e.g., impact, discharge location, or restoration activity) (Rathbun 1999, Smith 2002). The t-test can be performed using any standard packaged software and conceptual information is available in any introductory statistical text (Sokal and Rohlf 1995). Non-parametric versions of the t-test can also be used, the Mann-Whitney test being the most common. Again, these tests are explained in most texts.

In addition to the t-test, a simple ANOVA can be used to test difference in means between before and after (period) data or control and impact data (location). In this case, only two means are being compared, but sampling times are used to parse some of the variance of the model (Table 8-5). The significance is tested on one contrast alone. An extension of this simple comparison is when multiple sample sites exist either upstream and downstream, or before and after an impact. The simple ANOVA is extended by including a factor for sites, which are treated as replicates (Table 8-6). The presence of the site replicates affects the principal factor comparison (period or location) by attributing variance to the sampling location. In classic BACI designs, however, control sites are added and both before-after and control-impact contrasts are available, and the interaction term between BA and CI is the statistic of interest. ANOVA is also commonly used in this approach (Table 8-7).

The logical extension of the BACI model is to include multiple paired sampling times. The analysis is similar to repeated measures, and the ANOVA table for this BACI paired (BACIP) design adds a factor for the sampling times within each period, but the interaction is still the statistic of interest (Table 8-8). As noted above, paired samples between the control and impact site can be represented as differences between the two paired observations and the before and after period mean differences compared with a two-sample test (Stewart-Oaten et al. 1986). If a two sample t-test is used to compare differences, the analysis is similar to the interaction test (Underwood 1991, Smith et al.1992, Smith 2002). The final version includes the incorporation of multiple control streams as well as multiple sample times (Table 8-9, Underwood 1991, 1994). Once more, the interaction test is the statistic of interest; but some have argued that more individual contrasts can also be used, for example breaking the BA x CI sum of squares into variance associated with before (B x CI) and after (A x CI). Other extensions also exist and are discussed in Underwood (1992, 1994).

8.4 Watershed Assessments

The general focus of watershed assessments is to characterize resource condition across a watershed or a broad region. For example, such assessments are routinely performed for meeting 305(b) reporting requirements under the CWA and the design options were discussed in Chapter 3. The probabilistic designs favored for this approach lend themselves to a variety of analyses related to a number of assessment elements. This section discusses a few of these analytical options.

TABLE 8-5. An ANOVA table for the simple before-after model. A test for Location (control vs impact or upstream vs downstream) would be similar, but the location would be the principal treatment instead of Period. MS = mean squares; t_B and t_A are the number of observations before and after (Smith 2002).

Source	SS	df	F
Period: Before-After	SS_{BA}	1	MS_{BA}/MS_{times}
Sampling times	SS_{times}	$t_B + t_A-2$	
Total	SS_{Total}	$t_B + t_A-1$	

TABLE 8-6. ANOVA table for a similar design to Table 8-5, but with multiple sampling sites for each treatment. M indicates the number of sites (Smith 2002).

Source	SS	df	F
Period: Before-After	SS_{BA}	1	MS_{BA}/MS_{times}
Sampling times	SS_{times}	$t_B + t_A-2$	
Replicate sites	SSE	$(M-1)(t_B + t_A)$	
Total	SS_{Total}	$M(t_B + t_A)-1$	

TABLE 8-7. ANOVA table for the two-factor BACI design. N is the total number of observations, with multiple observations over time or space (Smith 2002).

Source	SS	df	F
Period: Before-After	SS_{BA}	1	
Location: Control-Impact	SS_{CI}	1	
Interaction: BA x CI	SS_{BACI}	1	MS_{BACI}/MS_E
Error	SSE	N-4	
Total	SS_{Total}	N-1	

TABLE 8-8. ANOVA table for the BACIP design (Smith 2002).

Source	SS	df	F
Period: Before-After	SS_{BA}	1	
Times within period	$SS_{t(BA)}$	$t_B + t_A-2$	
Location: Control-Impact	SS_{CI}	1	
Interaction: BA x CI	SS_{BACI}	1	MS_{BACI}/MS_E
Error	SSE	$t_B + t_A-2$	
Total	SS_{Total}	$2(t_B + t_A)-1$	

TABLE 8-9. ANOVA table for the asymmetrical BACI design with L-1 control sites and N observations (Smith 2002).

Source	SS	df	F
Period: Before-After	SS_{BA}	1	
Times within period	$SS_{t(BA)}$	$t_B + t_A - 2$	
Location: Control-Impact	SS_{CI}	L-1	
Interaction: BA x CI	SS_{BACI}	L-1	MS_{BACI}/MS_E
Error	SSE	(L-1) x $(t_B + t_A - 2)$	
Total	SS_{Total}	N-1	

In a truly random design, the estimate of average condition is fairly straightforward and is simply calculated as the overall mean ($\bar{y}$) of all the values. The variance of the mean is estimated as:

$$\text{var}(\bar{y}) = \frac{s^2}{n}, \qquad \text{Equation 8}$$

where s^2 is the sample variance (Rathbun 1999). Similarly, the proportion of river miles in a certain condition can be estimated from such designs as $\hat{p}$, the proportion of sample sites showing that condition with corresponding variance:

$$\text{var}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1} \text{ (Rathbun 1999).} \qquad \text{Equation 9}$$

For stratified random designs, individual strata means can be calculated; and an average overall mean condition across the entire study region can be calculated as a weighted average, where the percent of the resource within each strata is used as the weight. The region is split into *K* strata and the average condition for environmental variable $\bar{y}$ can be estimated as:

$$\bar{y} = \frac{1}{L}\sum_{h=1}^{K} L_h \cdot \bar{y}_h, \qquad \text{Equation 10}$$

where $\bar{y}_h$ is the sample mean of n_h observations in stratum *h* calculated as:

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}, \qquad \text{Equation 11}$$

L_h is the length of rivers in stratum *h*, and *L* is the total river length in the population of rivers (Rathbun 1999). The variance estimate for the regional average is:

$$\text{var}(\bar{y}) = \frac{1}{L^2}\sum_{h=1}^{K} L_h{}^2 \frac{s_h{}^2}{n_h}. \qquad \text{Equation 12}$$

The proportion of river miles in a given condition can be estimated as:

$$\hat{p} = \frac{1}{L}\sum_{h=1}^{K} L_h \cdot \hat{p}_h \text{ (Rathbun 1999),} \qquad \text{Equation 13}$$

where $\hat{p}_h$ is the proportion of sample stations from stratum *h* showing that condition. The variance associated with the measure is:

$$\text{var}(\hat{p}) = \frac{1}{L^2}\sum_{h=1}^{K} L_h{}^2 \frac{\hat{p}_h(1-\hat{p}_h)}{n_h - 1}. \qquad \text{Equation 14}$$

8.5 Gradient Designs

The design chapter discussed the use of gradient designs to identify trends in condition variables with source or stressor data (e.g., to assess biological condition under varying levels of urbanization). These designs primarily use regression and correlation analysis approaches. Only a brief discussion is given here, and interested users are directed to texts on regression and correlation analysis. Ordinary least squares regression and correlation are the simplest designs, where one is interested in exploring or predicting a particular dependent variable response given a level of some independent variable. Ideally, these data should all meet the requirements of standard parametric statistical analyses, and transformations should be used if these assumptions are violated. Data preparation is, therefore, also an important step in these analyses. It is strongly recommended that bivariate scatter plots be used to examine bivariate relationships before running correlation or simple linear regressions. These plots are valuable in exploring the strength and nature (linear or non-linear) of the relationships among variable pairs (Reckhow and Warren-Hicks 1997), and may recommend transformations worth exploring.

In reality, it is difficult to randomize all the sites, as one is often interested in reflecting the entire potential range in source or stressor levels. One potential solution for this is to use a validation set of data. Randomly selecting 10-20% of the available data and setting it aside, building the regression models and testing their accuracy with validation data is one option. Resampling approaches (e.g., bootstrapping or jackknifing) could also be used, especially if setting aside data is not an option (e.g., sample size issues). Again, it is important to guard against the over extension of regression models used in this way. There is a temptation to link correlation or regression as used here to causation. Technically, because of limited control over the independent variables, causation is a problematic concept. Correlation and regression certainly increase insight and can contribute to strength-of-evidence arguments, but when used as data mining tools, they can often lead to spurious relationships where causation is theoretically troubling (e.g., the ratio of percent row crop to percent evergreen land cover in the riparian zone does not cause a decrease in species diversity per se). Causal pathways still often need to be identified. There is great danger in packaged software that allows large batch correlation and regression modeling. Technically, running 100 correlations will lead to five significant results (at p = 0.05) from chance alone. Care must be taken to adjust the acceptable significance level for multiple unplanned tests (e.g., Bonferroni or Dunn-Šidák) and perhaps use this to better guide which relationships merit attention. In addition, a number of model diagnostics exist, though too

many to adequately cover in this document, but they are likely covered in most introductory statistical texts or regression analysis texts. Critical diagnostics include regression coefficients, residuals analysis, outlier analysis, and goodness of fit.

Multiple linear regression approaches can also be used. In these cases a dependent variable of interest is regressed against a number of independent variables. Single, combinatorial, or transformed independent variables can all be used. This powerful tool allows the user to compare effects of variables together and also to dissect the partial contribution of individual variables to the total response. As an example, one could explore the contribution of riparian canopy cover to the response of an index to land use alteration. In many cases, intact riparian zones contribute to higher index scores than those predicted for a given level of watershed disturbance (e.g., urban land cover) (Yoder et al. 1999). This can be assessed using partial residuals analysis in multiple linear regression.

Multiple regression models can either be forced (where a set of independent variables are used in the model) or variable selection procedures can be used where variables are added in the order with which they reduce the overall variance (forward selection, backward elimination, and stepwise approaches can be used). All possible model approaches can also be used, but run the same risk for any multiple tests approach (see above).

As with any approach, there are a number of pitfalls with multiple regression. It is very easy to generate significant multiple regression models as every added variable will reduce the error of the model. One risk is generating over-fit models (models that are unique to the modeled data, but not generally applicable). This can be avoided either using validation data or any number of tools that penalize a model for adding additional variables (e.g., Akaike's Information Criterion, AIC). Another big risk in multiple regression models is multicollinearity, or the inclusion of independent variables that are redundant. Multicollinear predictors dramatically impact the estimation of regression coefficients and may increase the risk of overfitting, which should be assessed. Removing highly correlated variables is encouraged, and diagnostics for identifying multicollinearity also exist in many software programs (e.g., variance inflation factor).

A variety of model diagnostics (in addition to those just described) exist for multiple regression; most of which are similar to those used in linear regression. They are also related to residuals analysis, outlier or leverage point analysis, and model fit. One unique diagnostic for multiple regression is partial residual analysis. Partial residual analysis examines the relationship between the response variable and a predictor when the effect of all other predictors is removed (i.e., already modeled). This approach allows the user to look at the unique contribution of individual predictors and can be done numerically and visually.

Exploratory pattern analysis across large gradients can also take advantage of the large number of multivariate statistical approaches. These methods (e.g., principal components analysis [PCA], detrended canonical correspondence analysis [DCCA], and non-metric multidimensional scaling) can be used to identify patterns in environmental stressors related to sources and to identify patterns in assemblage change across environmental gradients. These approaches are especially useful with large datasets containing many variables, like the ones being generated by many agencies. The approaches are designed to reduce the dimensionality of data to identify

prominent gradients. Users interested in multivariate statistical analyses should consult the array of resources available to guide these analyses (e.g., Manly 1994, Legendre and Legendre 1998, McCune and Grace 2002)

8.6 Reporting Results

This section briefly describes strategies for report writing that have been successful in assessment programs. The topic was dealt with in Barbour et al. (1999), and here we review important elements from that discussion. Reports should be tailored to the intended audience. Technical reports intended for fairly knowledgeable audiences can include greater detail on design and methodological description, greater flexibility in use of technical graphics that may require some sophistication to interpret, and more detailed discussion of technical issues. These reporting formats are likely familiar to most technical experts in any field and, for professional manuscripts, are dictated by the intended journal.

More frequently, however, assessment information is reaching a broader, less technical or non-technical audience including water resource managers and environmentally conscientious citizens. Communicating the condition of water resources and the potential impact of human activities on those resources is an important goal of resource monitoring (Karr and Chu 1999). Effective communication is vital for conveying technical information to non-technical audiences involved in important environmental decision-making. Reporting style and formats are important for assuring this is done accurately and efficiently, and a variety of resources are available to guide reporting (e.g., USEPA Office of Environmental Information).

8.6.1 Graphical Displays

The adage that "a picture is worth a thousand words" is no less true for conveying science than it is for conveying any other information. Well-designed, straightforward graphs can more effectively reveal patterns in biological response than strictly statistical tools, especially for non-technical audiences. Patterns, including outliers, may convey important information for both site assessment and diagnosis (Karr and Chu 1999). Some examples of useful graphical techniques are presented for specific program objectives:

- Classification – Graphs should easily convey strong differences among groups of sites within classes. Two common displays are bivariate scatter plots (Figure 8-8) from ordinations for clarity and cluster dendrograms, used to compare degree of separation of site groups based on sets of characteristics (Figure 8-9). Both are used to support classification decisions for building models.
- Problem identification and water resource status – Conveying information about the status of water resources of regional condition assessments is a critical task of assessment programs. This information needs to be conveyed easily and clearly. It also requires consolidating information from many samples. Pie charts (Figure 8-10), box and whisker plots (Figure 8-11), and bar charts are straightforward reporting tools for this job.

- Trend monitoring and assessment – Illustrating resource changes over space or time requires graphical displays that convey trends. Line graphs are ideal tools (Figure 8-12). Cumulative frequency curves are a bit more technical (Figure 8-13), but are also efficient ways to illustrate the percent of observations below some critical value.
- Causal diagnosis – Illustrating sources of impairment is not necessarily straightforward, as it often requires the evaluation of several variables simultaneously or in series. Indeed, the process of stressor diagnosis with environmental monitoring data is multi-faceted and challenging. However, certain graphical approaches do lend themselves to comparing diagnostic information. Bar charts, sun ray plots, and box-and-whisker plots are good options (Figures 8-14–8-16).

8.6.2 Report Format

Two common formats are used for assessment reports: summary reports for making management decisions and those designed for more technical audiences. The goal of each is to highlight the objectives, scientific process, results, and final assessment. However, the first format is designed for use by managers and can also be valuable as a public information tool. The latter format is better for technical review and dissemination of scientific results to an audience of technical peers.

The ecosummary is an example of the first format (Figure 8-17). The style is simple, efficiently documents results, and assists a non-technical audience in making informed decisions. These reports are similar to executive summaries in content (between 1-4 pages). Simple color graphics can be used to enhance the presentation of findings. The purpose of the study should be clear, as well as the results and take home message. A summary of data, as well as technical information, can be attached as subsequent pages or an easy link to the information provided.

The second format for reporting is the scientific report, which is generally structured following peer-reviewed journal formats. These generally are reviewed by colleagues or non-agency peers to validate the technical quality of the work. Standard formats include an abstract or summary, followed by an introduction highlighting the technical foundation and outlining the study objectives, a methods section, a results section, and a discussion and conclusion section. Citation of relevant technical literature is necessary to support the validity of both the design and interpretation of the work. Preparation of these reports likely requires more effort than the summary report. However, this report includes all the supporting information, and is a more substantial defense of the work. Research to be published in journals will have to adhere to individual journal requirements.

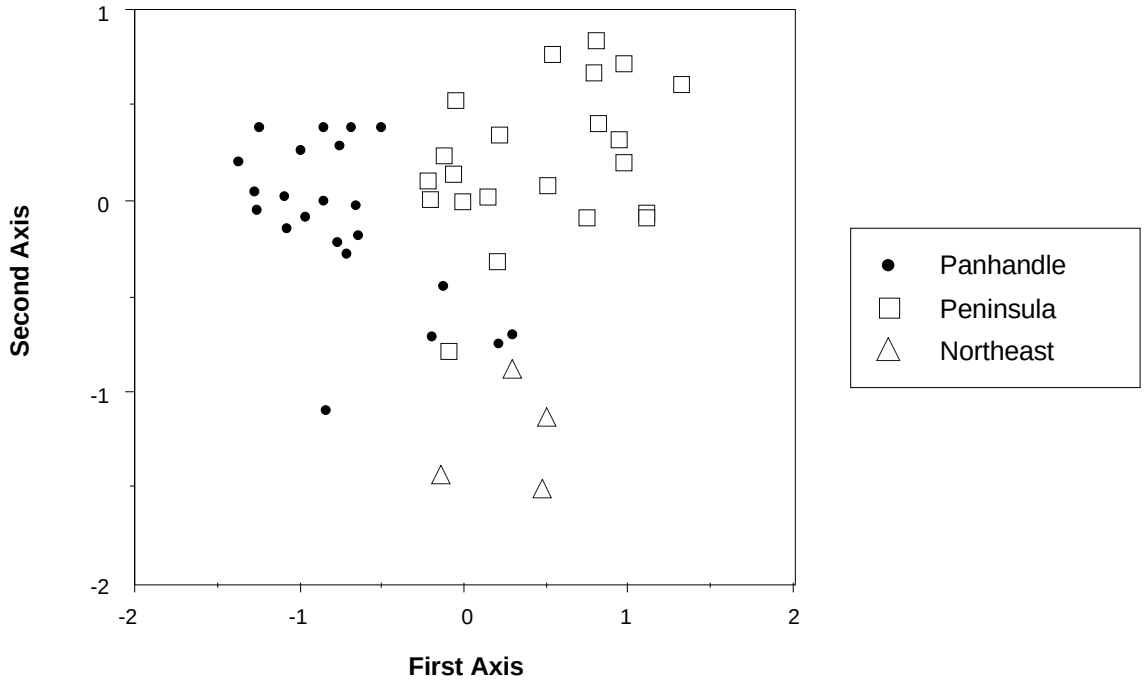


FIGURE 8-8. A bivariate scatter plot of an ordination used to support site classification. This figure for [Location] shows that grouping of sites based on taxonomic composition in ordination space reflects natural regional classes.

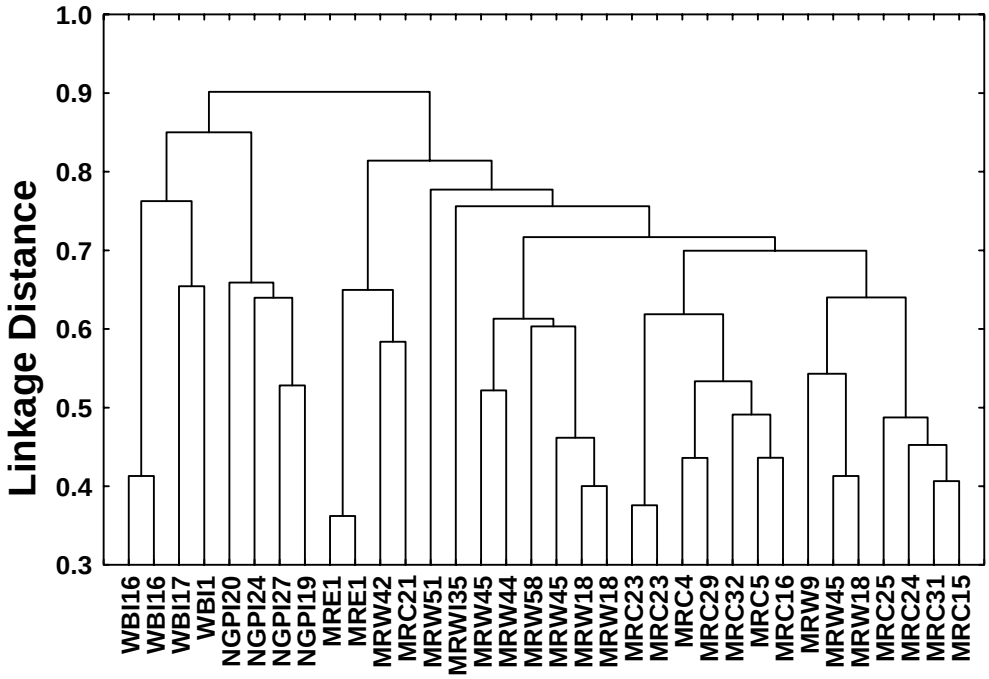


FIGURE 8-9. An example dendrogram, illustrating reference site clusters based on taxonomic composition. These figures are also used to support site classification.

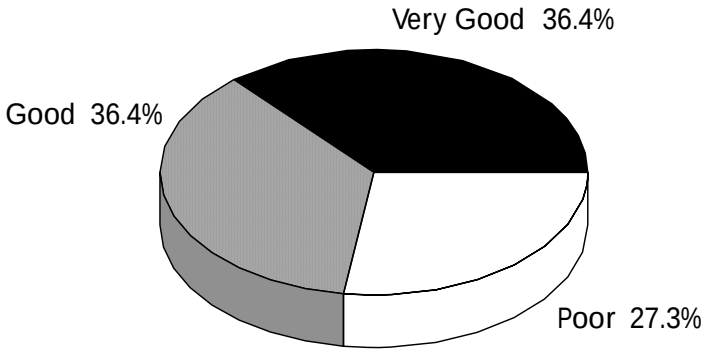


FIGURE 8-10. A pie chart, used to efficiently illustrate proportional information. This example shows the percent of stream miles in certain ecological condition categories within a watershed.

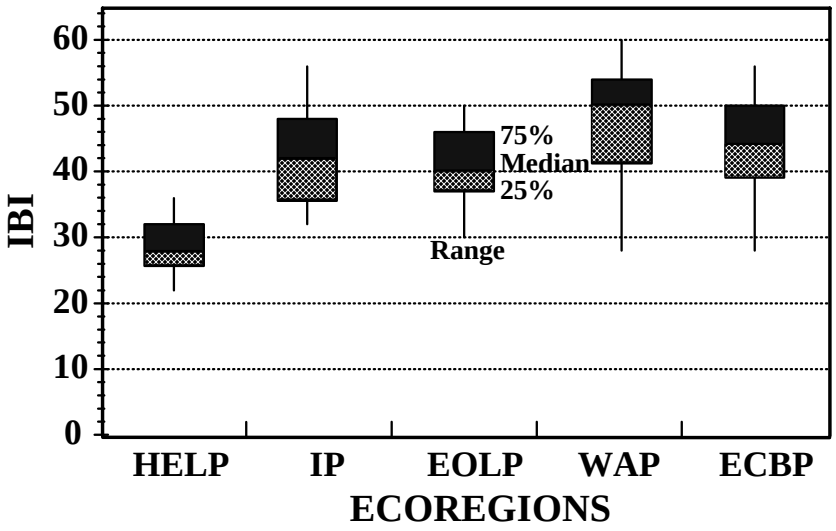


FIGURE 8-11. Box and whisker plots are used to illustrate differences in the distribution of values among different categories. Both central tendencies and a sense of variability can be conveyed. This particular figure illustrates differences in IBI scores among 5 ecoregions (contributed by Ohio EPA).

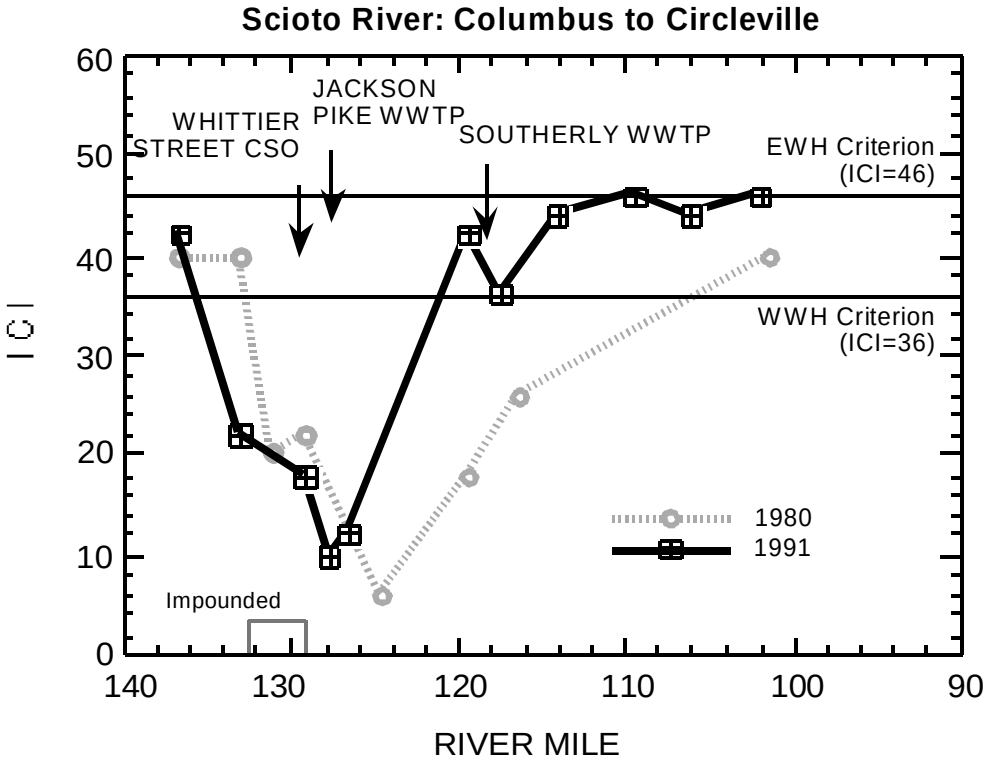


FIGURE 8-12. A line graph used to illustrate trends in the dependent variable relative to the independent variable. These are excellent tools for conveying temporal trends or trends along certain gradients. This example illustrates changes in a multimetric index along a river between two time periods (contributed by Ohio EPA).

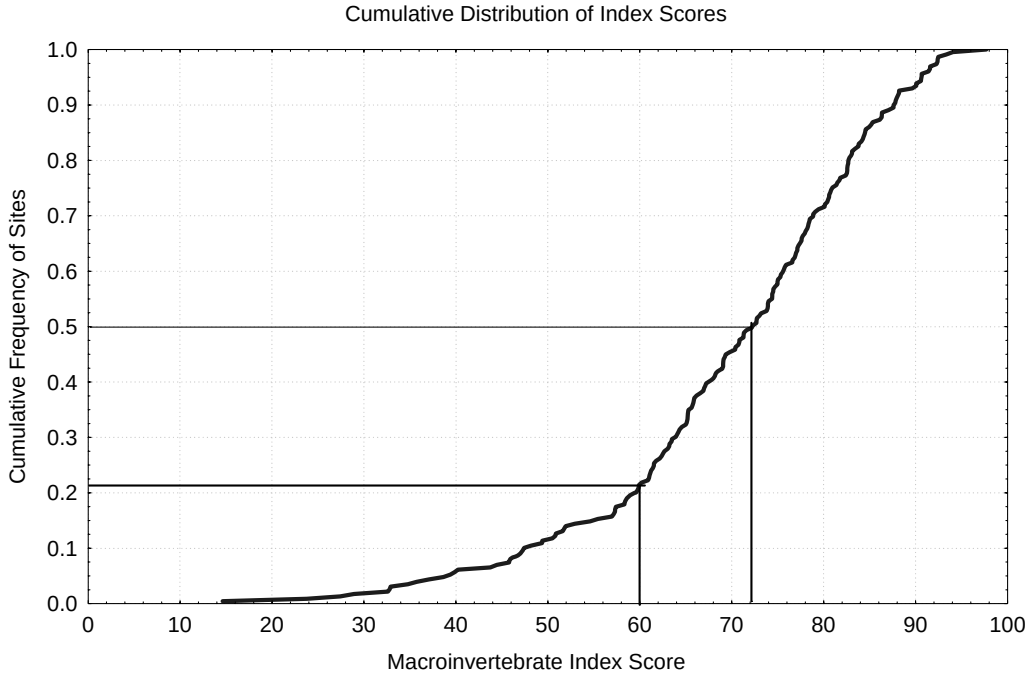


FIGURE 8-13. Cumulative frequency diagrams can be used to illustrate the ordered accumulation of observations from lowest to highest. These can be used to determine the percent of values exceeding any given value along the x-axis. This figure illustrates the median (50%) multimetric score and also indicates what percent of sites scored at or below 60 (21%).

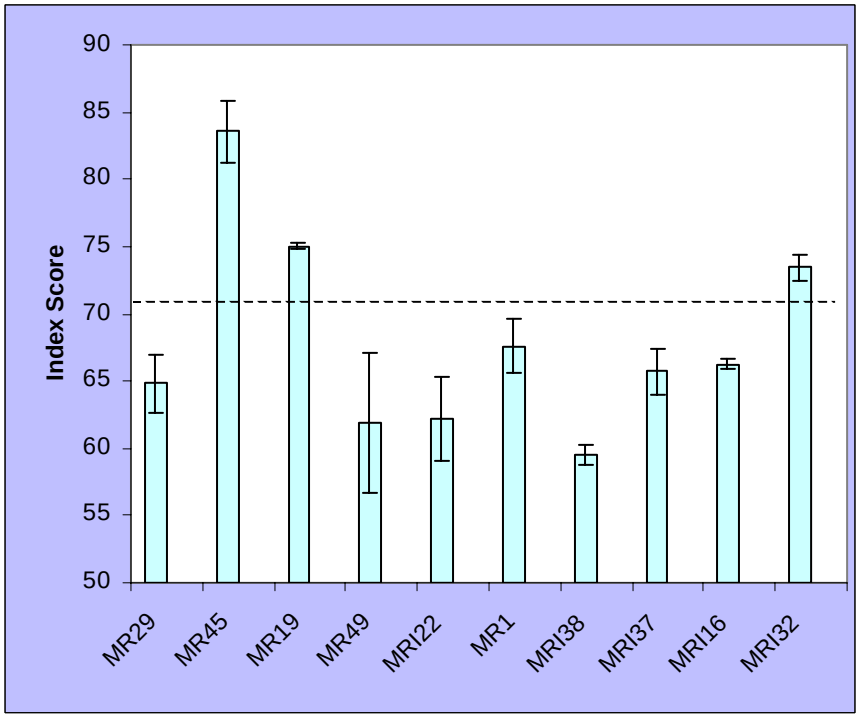


FIGURE 8-14. A bar chart used to display the magnitude and variance of values for individual elements. This format can be used to rank relative scores. This example illustrates average multimetric scores and standard errors for several watersheds.

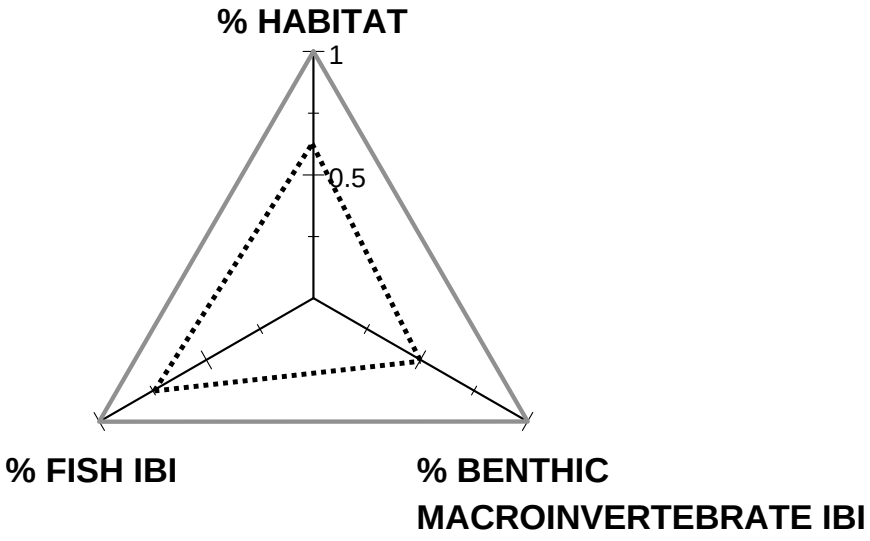


FIGURE 8-15. Sun ray plots are used to compare more than two endpoints simultaneously. Values can be scaled or compared to reference. This example simultaneously shows two multimetric indexes and a habitat score for a site relative to reference (1.0).

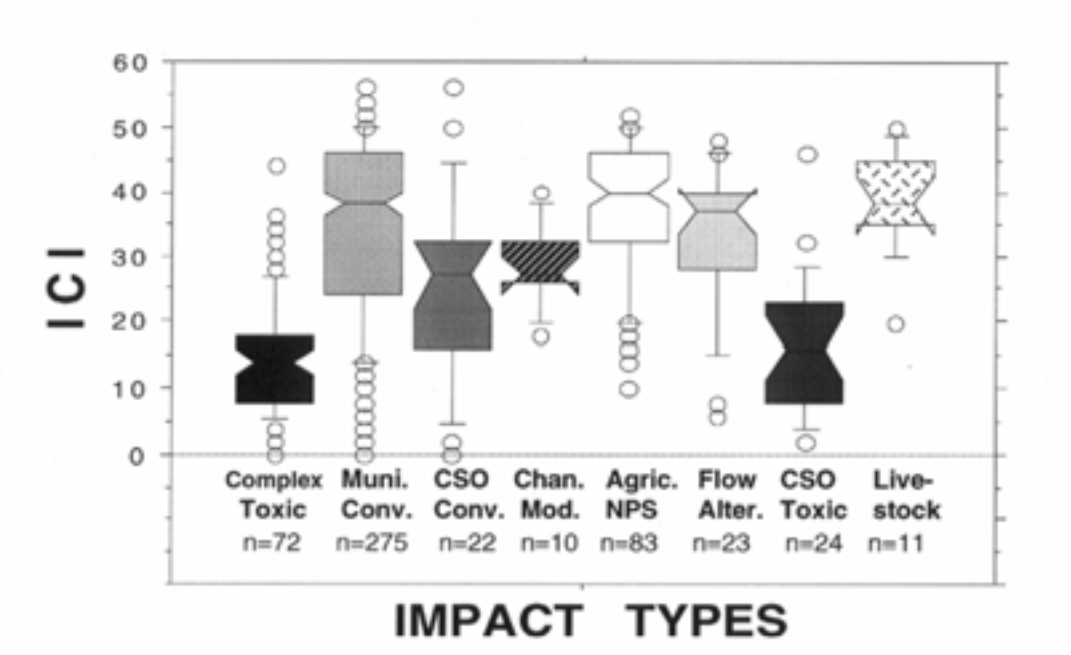


FIGURE 8-16. Box and whisker plots can also be used to illustrate the relative magnitude and variability associated with different variables on a common scale. This example illustrates multimetric values associated with different impacts (provided by Ohio EPA).

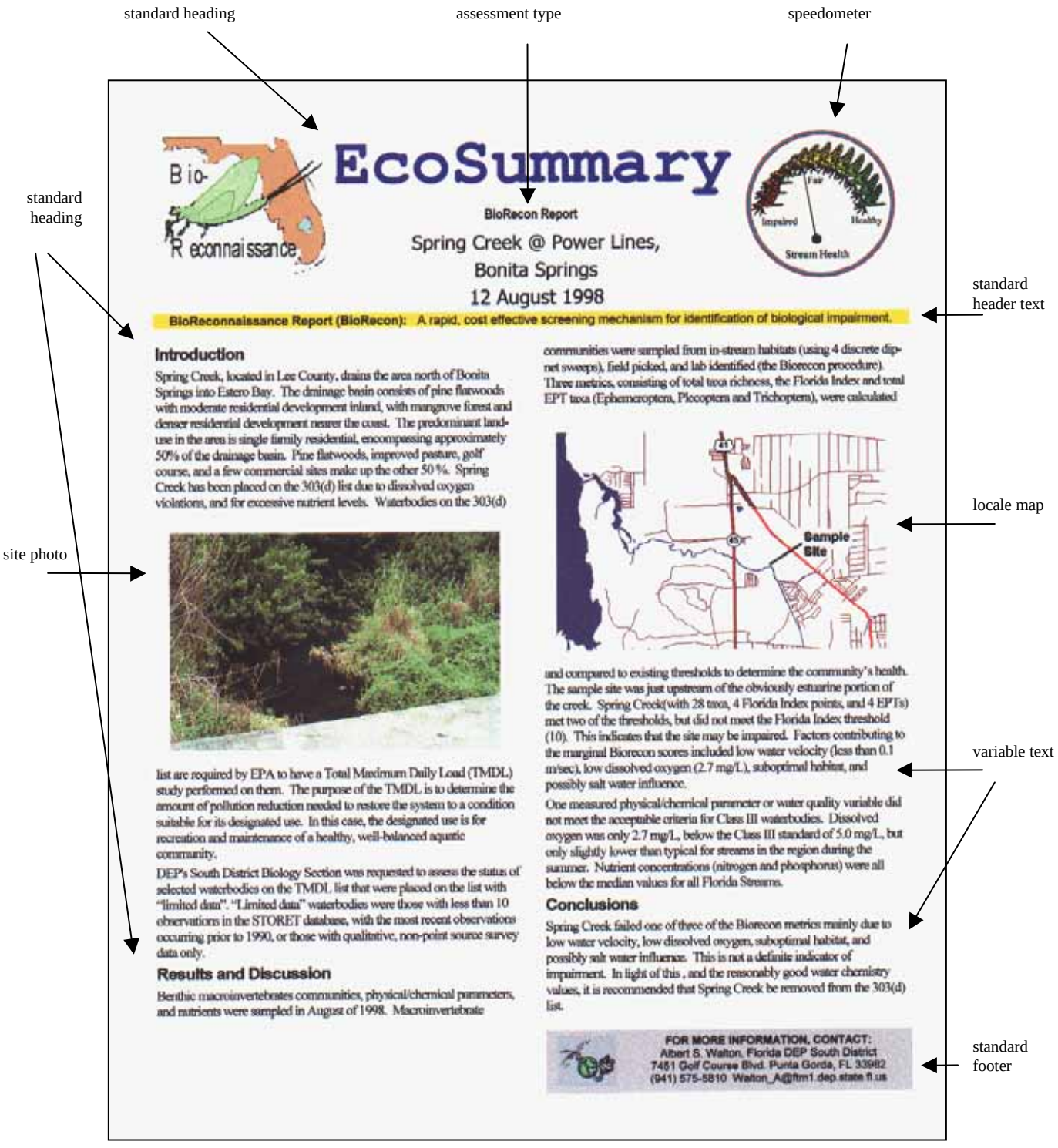


FIGURE 8-17. Florida Department of Environmental Protection Ecosummary – an example summary report.

Chapter 9.0 Literature Cited

Adams, S. M., A. M. Brown, and R. W. Goede. 1993. A quantitative health assessment index for rapid evaluation of fish condition in the field. Transactions of the American Fisheries Society 122:63-73.

Adams, S. M., and M. G. Ryon. 1994. A comparison of health assessment approaches for evaluating the effects of contaminant-related stress on fish populations. Journal of Aquatic Ecosystem Health 3:15-25.

Adams, S. M., K. D. Ham, M. S. Greeley, R. F. LeHew, D. E. Hinton, and C. F. Saylor. 1996. Downstream gradients in bioindicator responses: point source contaminant effects on fish health. Canadian Journal of Fisheries and Aquatic Sciences 53:2177-2187.

Allan, J. D. 1995. Stream ecology: structure and function of running waters. Chapman and Hall, London, UK.

Allan, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems Annual Review of Ecology, Evolution, and Systematics 35:257–84.

Anderson, J. B., and W. T. J. Mason. 1968. A comparison of benthic macroinvertebrates collected by dredge and basket samplers. Journal of the Water Pollution Control Federation 40:252-259.

Andrus, C. 2000. Experiences from the field: alcove sampling in the Willamette River, Oregon. In D. H. Landers, S. Allen-Gil, J. Reynolds, L. Carstensen, J. Sharber, P. Bayley, R. M. Hughes, and C. Andrus. New Perspectives in Electrofishing, EPA/600/R-99-108. US Environmental Protection Agency, Office of Research and Development, Washington, DC.

Angermeier, P. L., and J. R. Karr. 1984. Relationships between woody debris and fish habitat in a small warmwater stream. Transactions of the American Fisheries Society 11(3):716-726.

Angermeier P. L., and J. R. Karr. 1986. Applying an index of biotic integrity based on stream fish communities: considerations in sampling and interpretation. North American Journal of Fisheries Management 6:418-429.

Angermeier P. L., and R. A. Smogor. 1995. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. Canadian Journal of Fisheries and Aquatic Sciences 52:936-949.

Angradi, T. R. (editor). 2006. Environmental Monitoring and Assessment Program: Great River Ecosystems, Field Operations Manual. EPA/620/R-06/002. US Environmental Protection Agency, Washington, DC. (available at: http://www.epa.gov/emap/greatriver/fom.html.).

APHA (American Public Health Association). 1998. Standard methods for the examination of water and wastewater. 20th edition. American Public Health Association, Washington, DC.

Araujo, F. G., L. Fichberg, B. Carvalho, T. Pinto, and G. Peixoto. 2003. A preliminary index of biotic integrity for monitoring the condition of the Rio Paraiba do Sul, Southeast Brazil. Environmental Management 32(4):516-526.

Archibald, R. E. M. 1972. Diversity in some South African diatom associations and its relation to water quality. Water Research 6:1229-1238.

Bahls, L. L. 1992. Periphyton bioassessment methods for Montana streams. MT Department of Health and Environmental Services. Water Quality Bureau, Helena, Montana.

Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39:765-774.

Barbour, M. T., and J. Gerritsen. 1996. Subsampling of benthic samples: a defense of the fixed-count method. Journal of the North American Benthological Society 15:386-391.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185-211.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water, Washington, DC.

Barbour, M. T., J. B. Stribling, and J. R. Karr. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Pages 63-77 in W. S. Davis and T. P. Simon (editors). Biological assessment and criteria. Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Barbour, M. T., J. B. Stribling, and P. F. M. Verdonschot. 2006. (in press). The multihabitat approach of USEPA's rapid bioassessment protocols: Benthic macroinvertebrates. (Journal of the Spanish Association of Limnology) Limnetica 25(3-4).

Barbour, M. T., and C. O. Yoder. 2004. Technical guidelines: technical elements of a bioassessment program. Unpublished manuscript.

Bartsch L. A., W. B. Richardson, and T. J. Naimo. 1998. Sampling benthic macroinvertebrates in a large floodplain river: considerations of study design, sample size, and cost. Environmental Monitoring and Assessment 52:425–439.

Bayley, P. B. 1989. Aquatic environments in the Amazon basin, with an analysis of carbon sources, fish production, and yield. Pages 399-408 in D. P. Dodge (editor). Proceedings of the International Large River Symposium (LARS). Canadian Special Publication of Fisheries and Aquatic Sciences 106. Canadian Department of Fisheries and Oceans, Ottawa, Canada.

Bayley, P. B. 1995. Understanding large river -floodplain ecosystems. BioScience 45(3):153-158.

Beak, T. W., T. C. Griffing, and A. G. Appleby. 1973. Use of artificial substrate samplers to assess water pollution. In J. Cairns, and K. L. Dickson (editors). Biological methods for the assessment of water quality. American Society for Testing and Materials Spec. Technical Publication 528.

Beaumont, W. R. C., G. Peirson, and M. J. Lee. 2006. Factors affecting the characteristics and propagation of voltage gradient fields from electric fishing anodes. Fisheries Management & Ecology 13(1):47-52.

Benke, A. C. 2001. Importance of flood regime to invertebrate habitat in an unregulated river–floodplain ecosystem. Journal of the North American Benthological Society. 20:225-240.

Bell, D. 1973. The Coming of Post-Industrial Society. Basic Books, New York.

Benke, A. C., R. L. Henry III, D. M. Gillespie, and R. J. Hunter. 1985. Importance of snag habitat for animal production in southeastern streams. Fisheries 10(5):8-12.

Benke, A. C., and J. L. Meyer. 1988. Structure and function of a blackwater river in the southeastern USA. Internationale Vereinigung für Theoretische und Angewandte Limnologie, Verhandlungen 23:1209-1218.

Berger, W., H. McCarty, and R. K. Smith. 1996. Environmental laboratory data evaluation. Genium Publishing Corp, Amsterdam, New York.

Berkman, H. E., C. F. Rabeni, and T. P. Boyle. 1986. Biomonitors of stream quality in agricultural areas: fish vs invertebrates. Environmental Management 10:413-419.

Blocksom, K. A., and J. E. Flotemersch. 2005. Comparison of macroinvertebrate sampling methods for non-wadeable streams. Environmental Monitoring and Assessment 102:243-262.

Boivin T. G., D. Coates, D. Werle, and J. Baird. 2005. Remote sensing applications for information generation and monitoring for large river basin fisheries in tropical developing countries. Paper and poster prepared for the Second International Symposium on the Management of Large Rivers for Fisheries (February 2003), Phnom Penh, Cambodia.

Bryce, S. A., J. M. Omernik, and D. P. Larsen. 1999. Ecoregions: a geographic framework to guide characterization and ecosystem management. Environmental Practice 1:141-155.

Cao, Y., and C. P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. Journal of Applied Ecology 42:954-965.

Cao, Y., D. P. Larsen, and R. M. Hughes. 2001. Evaluating sampling sufficiency in fish assemblage surveys -- a similarity-based approach. Canadian Journal of Fisheries and Aquatic Sciences 58:1782-1793.

Cao, Y., D. P. Larsen, R. M. Hughes, P. L. Angermeier, and T. M. Patton. 2002, Sampling effort affects multivariate comparisons of stream assemblages. Journal of the North American Benthological Society 21(4):701-714.

Carbonneau, P. E., S. N. Lane, and N. E. Bergeron. 2004. Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. Water Resources Research 40, W07202, doi:10.1029/2003WR002759.

Carpenter, S., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. 1998. Nonpoint pollution of surface waters with phosphorous and nitrogen. Issues in Ecology, No. 3.

Carter, J. L., and V. H. Resh. 2001. After site selection and before data analysis: sampling, sorting and laboratory procedures used in stream benthic macroinvertebrate monitoring by USA state agencies. Journal of the North American Benthological Society 20(4):658-682.

Caton, L. R. 1991. Improved subsampling methods for the EPA rapid bioassessment benthic protocols. Bulletin of the North American Benthological Society 8:317-319.

Clark, M. J. R., and P. H. Whitfield. 1994. Conflicting perspectives about detection limits and about the censoring of environmental data. Water Resources Bulletin 30:1063-1075.

Clarke, R. 2000. Uncertainty in estimates of biological quality based on RIVPACS. Pages 39–54 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambelside, UK.

Clements, W. H. 1991. Characterization of stream benthic communities using substrate-filled trays: colonization, variability, and sampling selectivity. Journal of Freshwater Ecology 6:209-221.

Cochran, W. G. 1977. Sampling Techniques. 3rd Ed. John Wiley & Sons. New York.

Cormier, S. M., S. B. Norton, G. W. Suter, II, D. Altfater, and B. Counts. 2002. Determining the causes of impairment in the Little Scioto River, Ohio: part II. Characterization of causes. Environmental Toxicology and Chemistry 21:1125-1137.

Courtemanch, D. L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. Journal of the North American Benthological Society 15:381-385.

Craft, J. A., J. A. Stanford, and M. Pusch. 2002. Microbial respiration within a floodplain aquifer of a large gravel-bed river. Freshwater Biology 47(2):251-261.

Davies, P. E. 2000. Development of a national river bioassessment system (AUSRIVAS) in Australia. Pages 113-124 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of freshwaters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Davies, S. B., and S. K. Jackson. 2006. The Biological Condition Gradient: A conceptual model for interpreting detrimental change in aquatic ecosystems. Ecological Applications (in press).

Davis, W. S., and T. P. Simon (editors). 1995. Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Davis, W. S., B. D. Snyder, J. B. Stribling, and C. Stoughton. 1996. Summary of state biological assessment programs for streams and wadeable rivers. EPA 230-R-96-007. Office of Policy, Planning, and Evaluation, US Environmental Protection Agency. Washington, DC.

Descy, J. P. 1979. A new approach to water quality estimation using diatoms. Nova Hedwigia 64: 305-323.

De Shon, J. 1995. Development and application of the invertebrate community index (ICI). Pages 217-244 in W. S. Davis, and T. P. Simon (editors). Biological assessment and criteria. Lewis Publishers, Boca Raton, Florida.

Dettmers, J. M., S. Gutreuter, D. H. Wahl, and D. A. Soluk. 2001. Patterns and abundance in main channels of the upper Mississippi River system. Canadian Journal of Fisheries and Aquatic Sciences 58:933-942.

Diamond, J. M., M. T. Barbour, and J. B. Stribling. 1996. Characterizing and comparing bioassessment methods and their results: A perspective. Journal of the North American Benthological Society 15:713-727.

Diamond, J. M., J. C. Hall, D. M. Pattie, and D. Gruber. 1994. Use of an integrated monitoring approach to determine site-specific effluent limits. Water Environment Research 66:733-743

Dolloff, C. A., and M. L. Warren. 2003. Fish relationships with wood in small streams. Proceedings of the International Conference on Wood in World Rivers. American Fisheries Society Special Symposium 37:179-193.

Dumont, S. C. and J. A. Dennis. 1997. Comparison of day and night electrofishing in Texas reservoirs. North American Journal of Fisheries Management 17:939-946.

Dunne, T., and L. B. Leopold. 1978. Water in environmental planning. W. H. Freeman and Co., New York.

Dynesius, M., and C. Nilsson. 1994. Fragmentation and flow regulation of river systems in the northern third of the world. Science 266:753-762.

Eberhardt, L. L. 1976. Quantitative ecology and impact assessment. Journal of Environmental Management 4:27-70.

Edsall, T. A., T. E. Behrendt, G. Cholwek, J. W. Frey, G. W. Kennedy, and S. B. Smith. 1997. Use of remote sensing techniques to survey the physical habitat of large rivers. U.S. Geological Survey Great Lakes Science Center. Contribution 983.

Ellis, J. L. and D. C. Schneider. 1997. Evaluation of a gradient sampling design for environmental impact assessment. Environmental Monitoring and Assessment 48:157-172.

Emery, E. B., T. P. Simon, F. H. McCormick, P. L. Angermeier, J. E. DeShon, C. O. Yoder, R. E. Sanders, W. D. Pearson, G. D. Hickman, R. J. Reash, and J. A. Thomas. 2003. Development of a multimetric index for assessing the biological condition of the Ohio River. Transactions of the American Fisheries Society 132:791-808.

Emery, E. B., J. A. Thomas, M. Bagley, and T. R. Angradi. 2006. Fish. In T. R. Angradi (editor). Environmental monitoring and assessment program: great river ecosystems field operations manual. EPA 620-R-06-002. US Environmental Protection Agency, Washington, DC.

Faith, D. P., C. L. Humphrey, and P. L. Dostine. 1991. Statistical power and BACI designs in biological monitoring: comparative evaluation of measures of community dissimilarity based on benthic macroinvertebrate communities in Rockdale Mine Creek, Northern Territories, Australia. Australian Journal of Marine and Freshwater Research 42:589-602.

Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. BioScience 52:483-498.

Faux, R. N., B. A. McIntosh, D. J. Norton, and J. R. Sedell. 1998. Thermal remote sensing of stream temperatures. Annual report, US Environmental Protection Agency, USDA Forest Service, and Bureau of Land Management.

Feminella, J. W. 2000. Correspondence between stream macroinvertebrate assemblages and 4 ecoregions of the southeastern USA. Journal of the North American Benthological Society 19:442-461.

Ferraro, S. P., F. A. Cole, W. A. DeBen, and R. C. Schwartz, 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington. Canadian Journal of Fisheries and Aquatic Sciences 46:2157–2165.

Fitzpatrick, F. A., I. R. Waite, P. J. D'Arconte, M. R. Meador, M. A. Maupin, and M. E. Gurtz. 1998. Revised methods for characterizing stream habitat in the National Water-Quality Assessment Program. US Geological Survey Water-Resources Investigations Report 98-4052.

Fjerdingstad, E. 1950. The microflora of the River Molleaa with special reference to the relation of benthic algae to pollution. Folia Limnologica Scandanavica 5:1-123.

Fjerdingstad, E. 1965. Taxonomy and saprobic valency of benthic phytomicroorganisms. Internationale Revue der gesamten Hydrobiologie 50(4):475-604.

Flannagan, J. F., and D. M. Rosenberg. 1982. Types of artificial substrates used for sampling freshwater benthic macroinvertebrates. In J. J. Cairns (editor). Artificial substrates. Ann Arbor Science, Ann Arbor, Michigan.

Flotemersch, J. E., Autrey, B. C., and Cormier, S. M. 2001. Comparisons of boating and wading methods used to assess the status of flowing waters. EPA 600-R-00-108. US Environmental Protection Agency, Cincinnati, Ohio.

Flotemersch, J. E., and K. A. Blocksom. 2005 Electrofishing in boatable rivers: Does sampling design affect bioassessment metrics?. Environmental Monitoring and Assessment 102:263-283.

Flotemersch, J. E., K. A. Blocksom, J. J. Hutchens, and B. C. Autrey. 2004. Association among invertebrates and habitat indicators for large rivers in the Midwest: how sampling distance, point-sampling of habitat, and subsample size effect measures of large river macroinvertebrate assemblages. EPA-600-R-04-177. US Environmental Protection Agency, Cincinnati, Ohio.

Flotemersch, J. E., K. A. Blocksom, J. J. Hutchens, and B. C. Aturey. 2006. Development of a standardized large river bioassessment protocol (LR-BP) for macroinvertebrate assemblages. River Research and Applications 22:775-790.

Flotemersch, J. E., and D. C. Jackson. 2003. Seasonal foraging by channel catfish on terrestrially burrowing crayfish in a floodplain-river ecosystem. Ecohydrology and Hydrobiology 3(1):61-70.

Flotemersch, J. E., and D. C. Jackson. 2005. Importance of crayfish to channel catfish in a floodplain-river ecosystem, in L. H. Fredrickson, S. L. King and R. M. Kaminski (editor). Ecology and management of bottomland hardwood systems: The state of our understanding. University of Missouri-Columbia. Gaylord Memorial Laboratory Special Publication No. 10. Puxico, Missouri.

Fore, L. S. 2004. Development and testing of biomonitoring tools for macroinvertebrates in Florida streams. Prepared for the Florida Department of Environmental Protection, Tallahassee, Florida.

Fore, L. S., J. R. Karr, and L. L. Conquest. 1994. Statistical properties of an index of biological integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences 51:1077-1087.

Fore, L. S., J. R. Karr, and R. W. Wisseman. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15:212-231.

Frey, D.G. 1977. Biological integrity of water–a historical approach. Pages 127-140 in R. K. Ballentine and L. J. Guarria (editors). The integrity of water. US Environmental Protection Agency, Washington, DC.

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 10:199-214.

Furse, M. T., D. Moss, J. F. Wright, P. D. Armitage, and R. J. M. Gunn. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macroinvertebrate communities. Freshwater Biology 14:257-280.

Galat, D. L., and A. G. Frazier. 1996. Overview of river-floodplain ecology in the upper Mississippi River basin. In J.A. Kelmelis (editor). Science for floodplain management into the 21st century. Volume 3. Scientific Assessment and Strategy Team, US Government Printing Office, Washington, DC.

Galat, D. L., L. H. Fredrickson, D. D. Humburg, K. J. Bataille, J. R. Bodie, J. Dhorenwend, G. T. Gelwicks, J. E. Havel, D. L. Helmers, J. B. Hooker, J. R. M. Jones. 1998. Flooding to restore connectivity of regulated, large-river wetlands. BioScience 48:721-733.

Gale, W. F., and J. D. Thompson. 1974. Aids to benthic sampling by scuba divers in rivers. Limnology and Oceanography 19:1004-1007.

Gale, W. F., and J. D. Thompson. 1975. A suction sampler for quantitatively sampling benthos on rocky substrates in rivers. Transactions of the American Fisheries Society 104:398-405.

Gammon, J. R. 1973. The effect of thermal inputs on the populations of fish and macroinvertebrates in the Wabash River. Water Resources Research Center Technical Report 32, Purdue University, West Lafayette, Indiana.

Gammon, J. R. 1976. The fish populations of the middle 340 km of the Wabash River. Water Resources Research Center Technical Report 86, Purdue University, West Lafayette, Indiana.

Gammon, J. R., and T. P. Simon. 2000. Variation in a great river index of biotic integrity over a 20-year period. Hydrobiologia 422/423:291–304.

Ganasan, V., and R. M. Hughes. 1998. Application of an index of biological integrity (IBI) to fish assemblages of the rivers Khan and Kshipra (Madhya Pradesh), India. Freshwater Biology 40: 367–383.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge, UK.

Gerritsen, J., M. T. Barbour, and K. King. 2000a. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. Journal of the North American Benthological Society 19:487-496.

Gerritsen, J., Burton, J., and M. T. Barbour. 2000b. A stream condition index for West Virginia wadeable streams. Prepared by Tetra Tech, Inc. for the US Environmental Protection Agency. Region 3 Environmental Services Division, and Office of Science and Technology, Office of Water, US Environmental Protection Agency, Washington, DC.

Gerritsen, J., B. Jessup, and E. Leppo. 2002. Response of stream invertebrates to multiple anthropogenic stressors. Water Environment Research Foundation Technical Report 96-IRM-2.

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 (revised edition). EPA 822-B-96-001. US Environmental Protection Agency, Office of Water, Washington, DC.

Goede, R. W., and B. A. Barton. 1990. Organismic indices and an autopsy-based assessment as indicators of health and condition of fish. American Fisheries Society Symposium 8:93-108.

Gorman, O. T., and J. R. Karr. 1978. Habitat structure and stream fish communities. Ecology 59(3):507-515.

Grafe, C. S. (editor). 2002. Idaho river ecological assessment framework: an integrated approach. Idaho Department of Environmental Quality. Boise, Idaho.

Graf, W. L. 2001. Damage control: restoring the physical integrity of America's rivers. Annals of the Association of American Geographers 91:1-27.

Graham, S. P. 1986. Comparison of day versus night electrofishing efficiency on largemouth bass at O'Shaughnessy Reservoir. Service Note 579, Division of Wildlife, Ohio Department of Natural Resources, Columbus, Ohio.

Greenberg L, P. Svendsen, and A. Harby. 1996. Availability of microhabitats and their use by brown trout (Salmo trutta) and grayling (Thymallus thymallus) in the River Vojman, Sweden. Regulated Rivers: Research & Management 12:287–303.

Griffith, M.B., P.R. Kaufmann, A.T. Herlihy, and B.H. Hill. 2001. Analysis of macroinvertebrate assemblages in relationship to environmental gradients in Rocky Mountain streams. Ecological Applications 11: 489-505.

Griffith, M.B., B.H. Hill, A.T. Herlihy, and P.R. Kaufmann. 2002. Multivariate analysis of periphyton assemblages in relation to environmental gradients in Colorado Rocky Mountain streams. Journal of Phycology 38: 38-95.

Griffith, M.B., P. Husby, R.K. Hall, P.R. Kaufmann, and B.H. Hill. 2003. Analysis of macroinvertebrate assemblages in relation to environmental gradients among lotic habitats of California's Central Valley. Environmental Monitoring and Assessment 82(3): 281-309.

Growns, J., A. King, and F. Betts. 1999. The snag bug: a new method for sampling macroinvertebrate communities on large woody debris. Hydrobiologia 405:67-77.

Gutreuter, S., R. Burkhardt, and K. Lubinski. 1995. Long term resource monitoring program procedures: fish monitoring. LTRMP 95-P002-1. 42 pp. + Appendixes A-J. National Biological Service, Environmental Management Technical Center, Onalaska, Wisconsin, July 1995.

Hauer, F. R., and V. H. Resh. 1996. Benthic macroinvertebrates. In F. R. Hauer, and G. A. Lamberti (editors). Methods in stream ecology. Academic Press, New York.

Hawkins, C. P., and M. R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: implications for bioassessment. Journal of the North American Benthological Society 19:501-517.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000a. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456-1477.

Hawkins, C. P., R. H. Norris, J. Gerritsen, R. M. Hughes, S. K. Jackson, R. K. Johnson, and R. J. Stevenson. 2000b. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541-556.

Hayes, D. B., C. P. Ferreri, and W. W. Taylor. 1996. Active fish capture methods. Pages 193-220 in, B. R. Murphy and D. W. Willis (editors). Fisheries techniques. American Fisheries Society, Bethesda, Maryland.

Herlihy, A. T., and C. W. Hendricks. 2000. Water chemistry and microbiology. Chapter 5 in J. M. Lazorchak, B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). Environmental monitoring and assessment program - surface waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams. US Environmental Protection Agency, Cincinnati, Ohio.

Herzog, D. P., V. A. Barko, J. S. Scheibe, R. A. Hrabik, and D. E. Ostendorf. 2005. Efficacy of a benthic trawl for sampling small-bodied fishes in large river systems. North American Journal of Fisheries Management 25:594-603.

Hester, F. E., and J. S. Dendy. 1962. A multiple-plate sampler for aquatic macro-invertebrates. Transactions of the American Fisheries Society 91:420-421.

Hickman, G. D., and T. A. McDonough. 1996. Assessing the reservoir fish assemblage index: A potential measure of reservoir quality. Pages 85-97 in D. DeVries (editors). Reservoir Symposium – Multidimensional approaches to reservoir fisheries management. Reservoir Committee, Southern Division, American Fisheries Soceity, Bethesda, Maryland.

Hill, B. H., and A. T. Herlihy. 2000. 'Periphyton' in J. M. Lazorchak, B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). Environmental monitoring and assessment program – surface waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams. USEPA 620-R-00-007. US Environmental Protection Agency, Cincinnati, Ohio.

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.

Hill, B. H., A. T. Herlihy, P. R. Kaufmann, R. J. Stevenson, F. H. McCormick, and C. Burch Johnson. 2000. Use of periphyton assemblage data as an index of biotic integrity. Journal of the North American Benthological Society 19(1):50-67.

Hill, M. O. 1979. TWINSPAN - a FORTRAN program for arranging multivariate data in an ordered two way table by classification of the individuals and the attributes. Department of Ecology and Systematics, Cornell University, Ithaca, New York.

Hill, T. D., and D. W. Willis. 1994. Influence of water conductivity on pulsed AC and pulsed DC electrofishing catch rates for largemouth bass. North American Journal of Fisheries Management 14:202-207.

Hillebrand, H., C. D. Durselen, D. Kirschtel, U. Pollingher, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. Journal of Phycology 35:403-424.

Hilsenhoff, W. L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist 20:31-39.

Hocutt, C. H. 1981. Fish as indicators of biological integrity. Fisheries 6(6):28-30.

Hoffman, J. L. R. 2003. Susquehanna River pilot study: large river assessment project. Publication 228 Susquehanna River Basin Commission, Harrisburg, Pennsylvania.

Hubert, W. 1996. Passive capture techniques. Pages 157-192 in B. Murphy and D. Willis (editors). Fish Techniques. 2nd edition. American Fisheries Society, Bethesda, Maryland.

Huet, M. 1959. Profiles and biology of western European streams as related to fish management. Transactions of the American Fisheries Society 88:155-163.

Hughes, R. M., and J. R. Gammon. 1987. Longitudinal changes in fish assemblages and water quality in the Wilamette River, Oregon. Transactions of the American Fisheries Society 116:196–209.

Hughes, R. M., and A. T. Herlihy. 2006 (accepted). Electrofishing effort needed to estimate consistent IBI scores in raftable Oregon Rivers. Transactions of the American Fisheries Society.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, S. S. Intelmann, S.C. Corbett, M. C. Arbogast, and R. C. Hjort. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. North American Journal of Fisheries Management 22:1229-1240.

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.

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.

Hugueny, B., Camara, S., Samoura, B., Magassouba, M. 1996. Applying an index of biotic integrity based on fish communities in a west African river. Hydrobiologia 331:71-78.

Humphries, P., J. E. Growns, L. G. Serafina, J. H. Hawking, A. J. Chick, and P. S. Lake. 1998. Macroinvertebrate sampling methods for lowland Australian Rivers. Hydrobiologia 364:209-218.

Hurlbert, S. J. 1984. Pseudo-replication and the design of ecological field experiments. Ecological Monographs 54:187-211.

ITFM (Intergovernmental Task Force on Monitoring Water Quality). 1995. The strategy for improving water quality monitoring in the United States: Final report of the Intergovernmental Task Force on monitoring water quality. US Geological Survey, Reston, Virginia.

Jennings, M. J., L. S. Fore, and J. R. Karr. 1995. Biological monitoring of fish assemblages in Tennessee Valley reservoirs. Regulated Rivers 11:263-274.

Jessup, B., and J. Gerritsen. 2000. Development of a multimetric index for biological assessment of Idaho streams using benthic macroinvertebrates. Prepared for the Idaho Department of Environmental Quality, Boise, Idaho.

Jessup, B. K., J. B. Stribling, and J. Gerritsen. 2002. Further evaluation of the Wyoming stream integrity index. Prepared by Tetra Tech, Inc. for the US Environmental Protection Agency. Region 8 Wyoming Department of Environmental Quality, and US Environmental Protection Agency.

Johnson, B. R., K. A. Blocksom, J. E. Flotemersch, and B. Autrey. 2004 (draft). Development of large river bioassessment protocols (LR-BP) for macroinvertebrate assemblages, quality assurance project plan. US Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, Ohio.

Johnson, B. L., W. B. Richardson, and T. J. Naimo. 1995. Past, present, and future concepts in large river ecology. BioScience 45(3):134-141.

Jones, K. K., C. A. Simenstad, D. L. Higley, and D. L. Bottom. 1990. Community structure, distribution, and standing stock of benthos, epibenthos, and plankton in the Columbia River estuary. Progress in Oceanography 25(1):211-241.

Junk, W. J., P. B. Bayley, and R. E. Sparks. 1989. The flood pulse concept in river-floodplain systems, in Dodge, D. P. (editor) Proceedings of the International Large River Symposium, Canadian Special Publication for Fisheries and Aquatic Sciences 106:110-127.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6(6):21-27.

Karr, J. R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications 1:66-84.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, DC.

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, Illinois.

Karr, J. R., R. C. Heidinger, and E. H. Helmer. 1985. Effects of chlorine and ammonia from wastewater treatment facilities on biotic integrity. Journal of the Water Pollution Control Federation 57:912-915.

Kaufmann, P. R. (editor). 1993. Physical habitat. Pages 59-69 in R. M. Hughes (editor). Stream indicator and design workshop. EPA 600-R-93-138. US Environmental Protection Agency, Corvallis, Oregon.

Kaufmann, P. R. 2000. Physical habitat characterization - non-wadeable rivers. Chapter 6 in J. M. Lazorchak, B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). Environmental monitoring and assessment program - surface waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams. US Environmental Protection Agency, Cincinnati, Ohio.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Quantifying Physical Habitat in Wadeable Streams. EPA 620/R-99/003. US Environmental Protection Agency, Washington, D.C.

Kaufmann, R. R., and E. G. Robison. 1997. Physical habitat assessment. In J. L. Lazorchak, D. J. Flemm, and D. V. Peck (editors). Environmental monitoring and assessment program – surface waters: field operations and methods for measuring the ecological condition of wadeable streams. EPA 620-R-94-004F. US Environmental Protection Agency, Washington, DC.

Kedwards, T. J., S. J. Maud, and P. F. Chapman. 1999. Community-level analysis of ecotoxicological field studies: II. Replicated design studies. Environmental Toxicology and Chemistry 18:158-166.

Keith L. H. (editor). 1988. Principles of environmental sampling. ISBN 0-8412-1173-6. ACS professional reference book. American Chemical Society, Columbus, Ohio.

Kelly, M. 1988. Mining and the freshwater environment. Elsevier Applied Science, London, UK.

Kelly, M. G. 1998. Use of the trophic diatom index to monitor eutrophication in rivers. Water Research 32:236-242.

Kentucky DOW. 2002. Methods for assessing biological integrity of surface waters in Kentucky. Division of Water, Water Quality Branch, Natural Resource and Environmental Protection Cabinet, Frankfort, Kentucky.

Kerans, B. L., and J. R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. Ecological Applications 4:768-785.

King, R. S., M. E. Baker, D. F. Whigham, D. E. Weller, T. E. Jordan, P. F. Kazyak, and M. K. Hurd. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecological Applications 15:137-153.

Klemm, D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, W. T. Thoeny, M. B. Griffith, and W. S. Davis. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. Environmental Management 31(5): 656-669.

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. US Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio.

Kolkwitz, R., and K. Marsson. 1908. Ökologie der pfanzlichen Saprobien. Berichte der Deutshcen botanischen Gesellschaft 26A:505-519.

Knutson, M. G., and E. E. Klass. 1998. Floodplain forest loss and changes in forest community composition in the upper Mississippi River: a wildlife habitat at risk. Natural Areas Journal 18:138-150.

Krammer, K., and H. Lange-Bertalot. 1986. Bacillariophyceae, Part 2, Volume 1: Naviculaceae. In H. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer (editors). Freshwater Flora of Middle Europe. Gustav Fischer Publisher, New York.

Krammer, K., and H. Lange-Bertalot. 1988. Bacillariophyceae, Part 2, Volume 2: Bacillariaceae, Epithemiaceae, Surirellaceae. In H. Ettl,, J. Gerloff, H. Heynig, and D. Mollenhauer (editors). Freshwater Flora of Middle Europe. Gustav Fischer Publisher, New York.

Krammer, K., and H. Lange-Bertalot. 1991a. Bacillariophyceae, Part 2, Volume 3: Centrales, Fragilariaceae, Eunotiaceae. In H. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer (editors). Freshwater Flora of Middle Europe. Gustav Fischer Publisher, Stuttgart, Germany.

Krammer, K., and H. Lange-Bertalot. 1991b. Bacillariophyceae, Part 2, Volume 4: Achnanthaceae, Critical Supplement to Navicula (Lineolatae) and Gomphonema, Complete List of Literature for Volumes 1-4. In H. Ettl, G. Gartner, J. Gerloff, H. Heynig, and D. Mollenhauer (editors). Freshwater Flora of Middle Europe. Gustav Fischer Publisher, Stuttgart, Germany.

Larsen, D. P. 1997. Sample survey design issues for bioassessment of inland aquatic ecosystems. Human and Ecological Risk Assessment 3(6):979-991.

Lange-Bertalot, Horst. 1979. Pollution tolerance of diatoms as a criterion for water quality estimation. Nova Hedwigia 64:285-304.

Lazorchak, J. M., B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). 2000. Environmental monitoring and assessment program – surface waters: field operations and methods for measuring for measuring the ecological condition of non-wadeable rivers and streams. US Environmental Protection Agency, Cincinnati, Ohio.

Lazorchak, J. M., Klemm, D. J., and D. V. Peck (editors). 1998. Environmental Monitoring and Assessment Program -Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams. EPA/620/R-94/004F. US Environmental Protection Agency, Washington, D.C.

Legendre, P., and L. Legendre. 1998. Numerical ecology. 2nd English edition. Elsevier, New York.

Legleiter, C. J., D. A. Roberts, W. A. Marcus, and M. A. Fonstad. 2004. Passive optical remote sensing of river channel morphology and instream habitat: physical basis and feasibility. Remote Sensing of Environment 93:493– 510.

Lehtinen, R. M., N. D. Mundahl, and J. C. Madelczyk. 1997. Autumn use of woody snags by fishes in backwater and channel border habitats of a large river. Environmental Biology of Fishes 49:7-19.

Leland, H. V., and S. V. Fend. 1998. Benthic invertebrate distributions in the San Joaquin River, California, in relation to physical and chemical factors. Canadian Journal of Fisheries and Aquatic Sciences 55:1051-1067.

Leonard, P. M., and D. J. Orth. 1986. Application and testing of an index of biotic integrity in small, coolwater streams. Transactions of the American Fisheries Society 115:401-414.

Leopold, L. B., Wolman, M. G., and J. P. Mill. 1964. Fluvial processes in geomorphology: W.H. Freeman Press. San Francisco, California.

Li, J., A. Herlihy, W. Gerth, P. Kaufmann, S. Gregory, S. Urquhart, and D. Larsen. 2001. Variability in stream macroinvertebrates at multiple spatial scales. Freshwater Biology 46:87–97.

Liebmann, H. 1962. Handbuch der frischwasser- und abwasserbiologie. R. Oldenbourg, München, Germany.

Lisle, T. 1986. Effects of woody debris on anadromous salmonid habitat, Prince of Wales Island, southeast Alaska. North American Journal of Fisheries Management 6:538-550.

Lorenz, C. M., G. M. Van Dijk, A. G. M. Van Hattum, and W. P. Cofino. 1997. Concepts in river ecology: implications for indicator development. Regulated Rivers Research and Management 13:501-516.

Lowe, R. L. 1974. Environmental requirements and pollution tolerance of freshwater diatoms. EPA-670-4-74-005. US Environmental Protection Agency, Cincinnati, Ohio.

Lowe, R. L., and Y. Pan. 1996. Benthic algal communities as biological monitors. In R. J. Stevenson, M. L. Bothwell, and R. L. Lowe (editors). Algal ecology. Freshwater benthic ecosystems. Academic Press, San Diego, California.
Lowe-McConnell, R. H. 1975. Fish communities in tropical freshwaters. Their distribution, ecology, and evolution. Longman, London, UK.

Lubinski, K. S. 1993. A conceptual model of the upper Mississippi River system ecosystem. EMTC 93-T001. US Fish and Wildlife Service, Environmental Management Technical Center, Onalaska, Wisconsin.

Lubinski, K. S., and C. H. Theiling. 1999. Ecological status and trends of the upper Mississippi River system 1998: a report of the long term resource monitoring program (April 1999). LTRMP 99-T001. US Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin.

Lymburner, L., P. Hairsine, J. Walker, and A. Held. 2004. Riparian ecosystems and buffers: multi-scale structure, function, and management. AWRA Summer Specialty Conference Proceedings (June 28-30, 2004). Olympic Valley, California.

Lyons, J. 1992a. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. General Technical Report, NC-149. Forest Service, US Department of Agriculture, St. Paul, Minnesota.

Lyons, J. 1992b. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. North American Journal of Fisheries Management 12:198–203.

Lyons, J., R. R. Piette, and K. W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warmwater rivers. Transactions of the American Fisheries Society 130:1077–1094.

Lyons, J., L. Wang, and T. D. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. North American Journal of Fisheries Management 16:241-256.

Manly, B. J. F. 1994. Multivariate statistical methods: a primer. 2nd edition. Chapman and Hall, London, UK.

Manly, B. F. J. 2001. Statistics for environmental science and management. Chapman and Hall/CRC, Boca Raton, Florida.

Marchant, R., L. Barmuta, and B. C. Chessman. 1995. A preliminary study of the ordination and classification of macroinvertebrate communities from running waters in Victoria, Australia. Australian Journal of Marine Freshwater Research 45:945-962.

Maret, T. R., and D. S. Ott. 2004. Assessment of fish assemblages and minimum sampling effort required to determine biotic integrity of large rivers in Southern Idaho, 2002. US Geological Survey Water-Resources Investigations Report 03 –4274.

Maryland Department of Natural Resources. 1999. State of the streams: 1995-1997 Maryland biological stream survey results. CBWP-MANTA-EA-99-6. Maryland Department of Natural Resources, Annapolis, Maryland.

Mason, W. T., Jr. 1998. Macrobenthic monitoring in the Lower St. Johns River, Florida. Environmental Monitoring and Assessment 50:101-130.

Mason, W. T., C. I. Weber, P. A. Lewis, and E. C. Julian. 1973. Factors affecting the performance of basket and multiplate macroinvertebrate samplers. Freshwater Biology 3:409-436.

Matsushita, Y., and M. Shida. 2001. Comparison of catches between selective and conventional otter trawl for a coastal fishery. Reviews in Fisheries Science 9:33-42.

Matthews, R. A., P. F. Kondratieff, and A. L. Buikema, 1980. A field verification of the use of the autotrophic index in monitoring stress effects. Bulletin of Environmental Contamination and Toxicology 25:226-233.

McCormick, F. H., and R. M. Hughes. 2000. Aquatic vertebrates. Pages 10-1–10-14 in J. M. Lazorchak, B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). Environmental monitoring and assessment program – surface waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams. EPA 620-R-00-007. US Environmental Protection Agency, Cincinnati, Ohio.

McCormick, F. H., D. V. Peck, and D. P. Larsen. 2000. Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. Journal of the North American Benthological Society 19:385-404.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon.

McDonough, T. A., and G. D. Hickman. 1999. Reservoir fish assemblage index development: A tool for assessing ecological health in Tennessee Valley Authority impoundments. Pages 523-540 in T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Meador, M. R., T. F. Cuffney, and M. E. Gurtz. 1993. Methods for sampling fish communities as part of the National Water Quality Assessment Program. USGS/OFR 93-104. US Geological Survey, Washington, DC.

Meador, M. R., C. R. Hupp, T. F. Cuffney, and M. E. Gurtz. 1993. Methods for characterizing stream habitat as part of the National Water-Quality Assessment Program. U.S. Geological Survey, Open-File Report 93-408. Raleigh, North Carolina.

Mebane, C. A., T. R. Maret, and R. M. Hughes. 2003. An index of biological integrity (IBI) for Pacific Northwest rivers. Transactions of the American Fisheries Society 132:239-261.

Merritt, R. W, J. D. Allan, K. W. Cummins, K. J. Wessell, J. O. Wilhelm. 2005. Qualitative biological and habitat protocols for Michigan's non-wadeable rivers. Michigan Department of Environmental Quality, Lansing, Michigan.

Merritt, R., and K. W. Cummins. 2006. An Introduction to the Aquatic Insects of North America. 4th edition. Kendall/Hunt Publishing Company, Dubuque, Iowa.

Merritt, R. W., K. W. Cummins, and V. H. Resh. 1996. Design of aquatic insect studies: collecting, sampling, and rearing procedures. In R. W. Merritt, and K. W. Cummins (editors). An introduction to the aquatic insects of North America. 3rd edition. Kendall/Hunt Publishing, Dubuque, Iowa.

Mertes, L. A. K. 2002. Remote sensing of riverine landscapes. Freshwater Biology 47:799– 816.

Mertes, L. A. K., M. O. Smith, and J. B. Adams. 1993. Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images. Remote Sensing of Environment 43:281–301.

Messer, J. J., R. A. Linthurst, and W. S. Overton. 1991. An EPA program for monitoring ecological status and trends. Environmental Monitoring and Assessment 17:67-78.

Meyer W. B., B. L. Turner, II (editors). 1994. Changes in land-use and land-cover: a global perspective. Cambridge University Press, Cambridge, UK.

Meyer, J. L. 1990. A blackwater perspective on riverine ecosystems. BioScience 40(9):643-651.

Minns, C. K., V .W. Cairns, R .G. Randall, and J. E. Moore. 1994. An index of biotic integrity (IBI) for fish assemblages in the littoral zone of Great Lakes areas of concern. Canadian Journal of Fisheries and Aquatic Sciences 51: 1804-1822.

Minshall, G. W., R. C. Peterson, K. W. Cummins, T. L. Bott, J. R. Sedell, C. E. Cushing, and R. L. Vannote. 1983. Interbiome comparison of stream ecosystem dynamics. Ecological Monographs 51:1-25.

Mississippi DEQ (Mississippi Department of Environmental Quality). 2006. Development and Application of the Mississippi Benthic Index of Stream Quality (M-BISQ). June 30, 2003. Prepared by Tetra Tech, Inc., Owings Mills, Maryland for Mississippi Department of Environmental Quality, Office of Pollution Control, Jackson, Mississippi.

Montgomery, D. R. and J. M. Buffington. 1998. Channel responses, classification, and response. Pages 13-42 in R. J. Naiman and R. E. Bilby (editors). River Ecology and Management. Springer-Verlag, New York.

Morales, E. A., and M. Potapova. 2000. Third NAWQA workshop on harmonization of algal taxonomy, (May 2000). Patrick Center for Environmental Research, the Academy of Natural Sciences, Philadelphia, Pennsylvania.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41-52.

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. USGS Open-File Report 02-150. US Geological Survey, Reston, Virginia.

Mouw, J. E. B., J. A. Stanford, and P. B. Alaback. 2003. Influences of fluvial processes and hyporheic exchange on floodplain plant diversity and productivity. Oikos (submitted).

Muli, J. R. , and K. M. Mavuti. 2001. The benthic macrofauna community of Kenyan waters of Lake Victoria. Hydrobiologia 458:83-90.

Naiman, R. J. 1983. The annual pattern and spatial distribution of aquatic oxygen metabolism in boreal forest watersheds. Ecological Monograph 53:73-94.

Naiman, R. J. 1998. Biotic stream classification. Pages 97-119 in R. J. Naiman and R. E. Bilby (editors). River Ecology and Management. Springer-Verlag, New York.

Naiman, R. J. and R.E. Bilby (editors). 1998. River Ecology and Management: Lessons from the Pacific Coastal Ecoregion. Springer-Verlag, New York.

National Research Council. 1992. Restoration of aquatic ecosystems. National Academy Press, Washington, DC.

Norton, D. J., M. A. Flood, B. A. McIntosh, [and 14 others]. 1996. Modeling, monitoring, and restoring water quality and habitat in Pacific northwestern watersheds. Final report. Environmental Program, Government Applications Task Force, Central Intelligence Agency, Washington, DC.

Norton, S. B., S. M. Cormier, M. Smith, and R. C. Jones. 2000. Can biological assessment discriminate among types of stress? A case study for the eastern cornbelt plains ecoregion. Environmental Toxicology and Chemistry 19(4):1113-1119.

Novak, M. A., and R. W. Bode. 1992. Percent model affinity, a new measure of macroinvertebrate community composition. Journal of the North American Benthological Society 11(1):80-85.

Novotny, D. W. 1990. Electric fishing apparatus and electric fields in I. G. Cowx, I. G. and P. Lamarque (editors). Fishing with electricity, applications in freshwater fisheries management: Oxford, England, Fishing News Books, Blackwell Scientific Publications, Ltd.

Novotny, D. W., and G. R. Priegel. 1974. Electrofishing boats, improved designs, and operational guidelines to increase the effectiveness of boom shocker. Wisconsin Department of Natural Resources Technical Bulletin No. 73, Madison, Wisconsin.

NWQMC (National Water Quality Monitoring Council). 2006. Data elements for reporting water quality monitoring results for chemical, biological, toxicological, and microbiological analytes. Prepared by methods and data comparability board for the advisory committee on water information. Technical Report: NWQMC 05-03. US Geological Survey, Reston, Virginia.

Oberdorff, T., and R. M. Hughes. 1992. Modification of an index of biotic integrity based to characterize rivers in the Seine-Normandie Basin, France. Hydrobiologia 228:117-330.
Ohio EPA (Ohio Environmental Protection Agency). 1987. Biological criteria for the protection of aquatic life: volumes I-III. Ohio Environmental Protection Agency, Columbus, Ohio.

Ohio EPA (Ohio Environmental Protection Agency). 1989. Biological criteria for the protection of aquatic life. Vol. III. Standardized field sampling and laboratory methods for assessing fish sampling and macroinvertebrate communities. Ohio EPA, Division of Water Quality Monitoring and Assessment. Columbus, Ohio.

Ohio EPA (Ohio Environmental Protection Agency). 1999. Association between nutrients, habitat, and the aquatic biota in Ohio rivers and streams. Ohio Environmental Protection Agency Technical Bulletin MAS/1999-1-1. Monitoring and Assessment Section, Division of Surface Water, Ohio Environmental Protection Agency, Columbus, Ohio.

Omernik, J. M. 1995. Ecoregions: a spatial framework for environmental management. Pages 49-62 in W. S. Davis, and T. P. Simon (editors). Biological assessment and criteria. Lewis Publishers, Boca Raton, Florida.

Osenberg, C. W., P. J. Schmitt, S. J. Holbrook, K. E. Abu-Saba, and R. Flegal. 1994. Detection of environmental impacts: natural variability, effect size, and power analysis. Ecological Applications 4(1):16-30.

Overton, W. S., D. White, and D. L. Stevens, 1991. Design report for EMAP, the environmental monitoring and assessment program. EPA 600-3-91-053. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

Paller, M. H. 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. North American Journal of Fisheries Management 15:110-120.

Pan, Y., R. J. Stevenson, B. H. Hill, and A. T. Herlihy. 2000. Ecoregions and benthic diatom assemblages in Mid-Atlantic highlands streams, USA. Journal of the North American Benthological Society 19:518-540.

Patrick, R. 1968. Effect of invasive rate, species pool and size area on the structure of diatom community. Proceedings of the National Academy of Sciences 58:1323-1335.

Patrick, R. 1977. Ecology of freshwater diatoms and diatom communities. In D. Werner (editor). The biology of diatoms. Botanical Monographs Volume 13. Blackwell Blackwell Scientific Publications, Ltd.

Patrick, R., N. J. Hohn, J. Wallace. 1954. A new method for determining the pattern of the diatom flora. Notulae Naturae (Philadelphia) No. 259.

Patrick, R., and C. W. Reimer. 1966. The diatoms of the United States, exclusive of Alaska and Hawaii. The Academy of Natural Sciences of Philadelphia, Philadelphia, Pennsylvania.

Patrick, R., and C. W. Reimer. 1967. The diatoms of the United States. Volume 1, Monograph 13. The Academy of Natural Sciences of Philadelphia, Philadelphia, Pennsylvania.

Patrick, R., and C. W. Reimer. 1975. The diatoms of the United States, exclusive of Alaska and Hawaii. The Academy of Natural Sciences of Philadelphia, Philadelphia, Pennsylvania.

Patton, T. M., W. A. Hubert, F. J. Rahel, and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. North American Journal of Fisheries Management 20:394–398.

Paul, M. J., J. Gerritsen, and E. Leppo. 2002. Multivariate predictive invertebrate bioassessment models for Illinois streams. Draft Technical Report. Prepared by Tetra Tech, Inc. for Illinois Environmental Protection Agency, Springfield, Illinois.

Peck, D. V., D. K. Averill, A. T. Herlihy, R. M. Hughes, P. R. Kaugmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, M. R. Cappaert, T. Magee, and P. A. Monaco. 2006 (in press). Environmental Monitoring and Assessment Program – Surface Waters Western Pilot Study: Field Operations Manual for Non-Wadeable Rivers and Streams. EPA 620/R-06/xxx, U.S. Environmental Protection Agency, Washington, DC.

Penczak, T., and R. H. Mann. 1993. A preliminary evaluation of the effect of human activity on the fish populations of the Pilica River, central Poland. Polish Archives of Hydrobiology 40(1):101-115.

Petts, G.E. 1996. Water allocation to protect river ecosystems. Regulated Rivers: Research and Management. 12: 353-365.

Pitlo, J. 1992. An evaluation of the effect of a change in commercial harvest regulations on the channel catfish populations inhabiting the upper Mississippi River. Project F-109-R Des Moines. Federal Aid in Sport Fish Restoration, Iowa Department of Natural Resources, Des Moines, Iowa.

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. US Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

Poff, N. L., and J. D. Allan. 1995. Functional organization in stream fish assemblages in relation to hydrologic variability. Ecology 76:606-627.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The natural flow regime: a paradigm for river conservation and restoration. BioScience 47:769-784.

Poff, N. L., and J. V. Ward. 1990. Physical habitat template of lotic systems: recovery in the context of historical pattern of spatiotemporal heterogeneity. Environmental Management 14(5): 629-645.

Poole, G. C., J. A. Stanford, C. A. Frissell, and S. W. Running. 2002. Three-dimensional mapping of geomorphic controls on flood-plain hydrology and connectivity from aerial photos. Geomorphology 48(4):329-347.

Porter, S. G., T. F. Cuffney, M. E. Gurtz, and M. R. Meador. 1993. Methods for collecting algal samples as part of the national water-quality assessment program. Open-File Report 93-409. US Geological Survey.

Poulton, B. C., M. L. Wildhaber, C. S. Charbonneau, J. F. Fairchild, B. G. Mueller, and C. J. Schmitt. 2003. A longitudinal assessment of the aquatic macroinvertebrate community in the channelized lower Missouri River. Environmental Monitoring and Assessment 85:23-53.

Prescott, G. W. 1962. Algae of the western Great Lakes area. Wm. Brown, Dubuque, Iowa.

Rabeni, C. F., and E. K. Gibbs. 1978. Comparison of two methods used by divers for sampling benthic invertebrates in deep rivers. Journal of Fisheries Research Board of Canada 35:332-336.

Rankin, E. T. 1989. The qualitative habitat evaluation index (QHEI): rationale, methods, and application. Ohio EPA, Division of Water Quality Planning and Assessment, Columbus, Ohio.

Rathbun, S. L. 1999. Sampling design issues for section 305(b) water quality monitoring. Appendix G in Savannah River Basin REMAP: a demonstration of the usefulness of probability sampling for the purpose of estimating ecological condition in state monitoring programs. EPA 904-R-99-002. US Environmental Protection Agency Region 4, SESD, Athens, Georgia.

Raven, P. J., N. T. H. Holmes, F. H. Dawson, and M. Everard. 1998. Quality assessment using river habitat survey data. Aquatic Conservation: Marine and Freshwater Ecosystems 8:477-499.

Raven, P. J., N. T. H. Holmes, F. H. Dawson, P. J. A. Fox, M. Everard, I. R. Fozzard, and K. J. Romen. 1998. River habitat quality: the physical character of rivers and streams in the UK and Isle of Man. ISBN 1-873760-429. Environment Agency, Bristol, England.

Reash, R. J. 1999. Considerations for characterizing Midwestern large river habitats. Pages 463-474 in T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press. Boca Raton, Florida.

Reckhow, K. H., and W. Warren-Hicks. 1997. Biological criteria: technical guidance for survey design and statistical evaluation of biosurvey data. EPA 822-B-97-021. US Environmental Protection Agency, Office of Water, Washington, DC.

Reeves, G. H., L. E. Benda, K. M. Burnett, P. A. Bisson, and J. R. Sedell. 1995. A disturbance-based ecosystem approach to maintaining and restoring freshwater habitats of evolutionarily significant units of anadromous salmonids in the Pacific Northwest. Pages 334-349 in J. L. Nielsen (editor). Evolution and the aquatic ecosystem: defining unique units in population conservation. American Fisheries Society Symposium 17, Bethesda, Maryland.

Reimer, G. 1991. The ecological importance of floodplains for fish at the river March (Austria). Archives of Hydrobiology 121:355-363.

Reisner, M. 1986. Cadillac desert: the American west and its disappearing water. Penguin Books, New York.

Resh, V., and D. Rosenberg. 1984. The ecology of aquatic insects. Praeger, New York.

Reynolds, J. B. 1996. Electrofishing. Pages 221-253 in B. R. Murphy and D. W. Willis (editors). Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

Reynolds, L. A. T. Herlihy, P. R. Kaufmann, S. V. Gregory, and R. M. Hughes. 2003. Electrofishing effort requirements for assessing species richness and biotic integrity in Western Oregon streams. North American Journal of Fisheries Management 23:450-461.

Reynoldson, T. B. and J. F. Wright. 2000. The reference condition: problems and solutions. Pages 293-303 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Science 53(Supplement 1):295-311.

Richter, B. D., D. P. Braun, M. A. Mendelson, and L. L. Master. 1997. Threats to imperiled freshwater fauna. Conservation Biology 11:1081-1093.

Rogers, S. J., and T. W. Owens. 1995. Long Term Resource Monitoring Program Procedures: Vegetation monitoring. National Biological Service, Environmental Management Technical Center, Onalaska, Wisconsin, July 1995. LTRMP 95-P002-3. 10 pp. + Appendixes A-E.

Rosenberg, D. M., and V. H. Resh. 1982. The use of artificial substrates in the study of freshwater benthic macroinvertebrates. In J. J. Cairns (editor). Artificial substrates. Ann Arbor Science, Ann Arbor, Michigan.

Rosenberg, D. M., and V. H. Resh (editors). 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

Rosenzweig, M. 1995. Species diversity in space and time. Cambridge University Press. London, UK.

Roth, N. E., J. D. Allan, and D. L. Erickson. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landscape Ecology 11(3):141-156.

Round, F. E., R. M. Crawford, and D. G. Mann. 1990. The diatoms: biology and morphology of the genera. Cambridge University Press, Cambridge, UK.

Sabater, S., F. Sabater, and J. Armengol. 1988. Relationships between diatom assemblages and physicochemical variables in the River Ter (NE Spain). Internationale Revue der Gesamten Hydrobiologie 73:171-179.

Salsburg, D. S. 1985. The religion of statistics as practiced in medical journals. The American Statistician 39:220-223.

Salsburg, D. S. 1986. Statistics for toxicologists. New York, Marcel Dekker, Inc.
Sanders, R. E. 1991. Day versus night electrofishing catches from near-shore waters of the Ohio and Muskingum Rivers. Ohio Journal of Science 192:1–59.

Sanders, R. E., R. J. Miltner, C. O. Yoder, and E. T. Rankin. 1999. The use of external deformities, erosion, lesion, and tumors (DELT anomalies) in fish assemblages for characterizing aquatic resources: a case study of seven Ohio streams. Pages 225-248 in T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Schmitt, C. J., V. S. Blazer, G. M. Dethloff, D. E. Tillitt, T. S. Gross, W. L. Bryant, L. R. DeWeese, S. B. Smith, R. W. Goede, T. M. Bartish, and T. J. Kubiak. 1999. Biomonitoring of environmental status and trends (BEST) program: field procedures for assessing the exposure of fish to environmental contaminants. Information and Technology Report 1999-0007. Biological Resources Division, US Geological Survey.

Scott, M. T., and L. A. Nielsen. 1989. Young fish distribution in backwaters and main-channel borders of the Kanawha River, West Virginia. Journal of Fish Biology 35:21-27.

Sedell, J. R., J. E. Richey, and F. J. Swanson. 1989. The river continuum concept: a basis for the expected ecosystem behavior of very large rivers? Canadian Special Publication of Fisheries and Aquatic Sciences 106:110-127.

Sheaffer, W. A., and J. G. Nickum. 1986. Backwater areas as nursery habitats for fishes in Pool 13 of the upper Mississippi River. Hydrobiologia 136:131-140.

Sheehan, R. J., and J. L. Rasmussen. 1999. Large rivers. Pages 529–559 in C. Kohler, W. A. Hubert (editors.). Inland fisheries management in North America. 2nd edition American Fisheries Society. Bethesda, Maryland.

Shelton, L. R. 1994. Field guide for collecting and processing stream-water samples for the national water-quality assessment program. Open-File Report 94-455. US Geological Survey, Sacramento, California.

Shepard, J. 2001. The state of the world's rivers. Center for Global Environmental Education, St. Paul, Minnesota.

Simon, T. P. 1991. Development of index of biotic integrity expectations for the ecoregions of Indiana. I. Central Corn Belt Plain. EPA-905-9-91-025. US Environmental Protection Agency, Region V, Chicago, Illinois.

Simon, T. P. (editor). 1999. Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Simon, T. P., and J. Lyons. 1995. Application of the index of biotic integrity to evaluate water resources integrity in freshwater ecosystems. Pages 245–262 in W. S. Davis and T. P. Simon (editors). Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Simon, T. P., and. R. E. Sanders. 1999. Applying an index of biotic integrity based on great river fish communities: considerations in sampling and interpretation. Pages 475–505 in T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Simonson, T. D., J. Lyons, and P. D. Kanehl. 1994. Quantifying fish habitat in streams-transect spacing, sample size, and a proposed framework. North American Journal of Fisheries Management 14:607-615.

Simpson, J. C., and R. H. Norris. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125-142 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of freshwaters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Slack, K. V., R. F. Ferreira, and R. C. Averett. 1986. Comparison of four artificial substrates and the ponar grab for benthic invertebrate collection. Water Resources Bulletin 22:237-248.

Sladecek, V. 1973. System of water quality from the biological point of view. Archiv für Hydrobiologie – Advances in Limnology 7:1-218.

Smith, E. P. 2002. BACI design. Pages 141-148 in A. H. El-Shaarawi and W. W. Piegorsch, (editors). Encyclopedia of Environmetrics. John Wiley & Sons, Ltd, Chichester, UK.

Smith, E. P., D. R. Orvos, and J. J. Cairns. 1992. Impact assessment using the before-after-control-impact (BACI) model: concerns and comments. Canadian Journal of Fisheries and Aquatic Sciences 50:627-637.

Smith, F., S. Kulkarni, L. E. Myers, and M. J. Messner. 1988. Evaluating and presenting quality assurance sampling data. Chapter 10, pages 157-168 in L. H. Keith (editor). Principles of environmental sampling. ISBN 0-8412-1173-6. ACS professional reference book. American Chemical Society, Columbus, Ohio.

Smith, R. A., G. E. Schwartz, and R. B. Alexander. 1997. Regional interpretation of water quality monitoring data. Water Resources Research 33:2781-2798.

Smith, S. B., A. P. Donahue, R. J. Lipkin,V. S. Blazer, C. J. Schmitt, and R. W. Goede. 2002. Illustrated field guide for assessing external and internal anomalies in fish. Information and Technology Report, 2002-0007. US Geological Survey (available at: http://www.cerc.usgs.gov/pubs/center/pdfDocs/ITR_2002_0007.pdf).

Snyder, D. E. 2003. Electrofishing and its harmful effects on fish. Information and Technology Report. USGS/BRD/ITR—2003-0002. US Government Printing Office, Denver, Colorado.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd edition. W. H. Freeman and Co., New York.

Sparks, R. E., P. B. Bayley, S. L. Kohler, and L. L. Osborne. 1990. Disturbance and recovery of large floodplain rivers. Environmental Management 14:699-709.

Stalnaker, C. B., R. T. Milhous, and K. D. Bovee. 1989. Hydrology and hydraulics applied to fishery management in large rivers. Page 13 in D. P. Dodge (editor). Proceedings of the International Eastern Rivers and Mountains Network – Phase I: Large River Symposium. Canadian Special Publication of Fisheries and Aquatic Sciences 106.

StatSoft. 1994. STATISICA Mac version 4.1. StatSoft Inc., Tulsa, Oklahoma.

Steedman, R. J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. Canadian Journal of Fisheries and Aquatic Sciences 45:492-501.

Stevens, D. L. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8:167-95.

Stevens, D. L., and A. R. Olsen. 1991. Statistical issues in environmental monitoring and assessment. Draft document – prepared for publication in Proceedings of the Section on Statistics and the Environment, American Statistical Association, Alexandria, Virginia.

Stevens, D. L., and A. R. Olsen. 1999. Spatially restricted surveys over time for aquatic resources. Journal of Agricultural, Biological, and Environmental Statistics 4:415-28.

Stevenson, R. J. 1996. An introduction to algal ecology in freshwater benthic habitats. Pages 3-30 in R. J. Stevenson, M. L. Bothwell, R. L. Lowe (editors). Algal ecology. Academic Press, San Diego, California.

Stevenson, R. J., and L. Bahls. 1999. Periphyton protocols. In M. T. Barbour, J. Gerritsen, B. D. Snyder, J. B. Stribling (editors). Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. 2nd edition. EPA 841-B-99-002. Office of Water, US Environmental Protection Agency, Washington, DC.

Stevenson, R. J., and J. P. Smol. 2002. Use of algae in environmental assessments. In J. D. Wehr, and R. G. Sheath, eds. Freshwater Algae in North America: Classification and Ecology. Academic Press, San Diego, California.

Stewart, J. G., and V. A. Barko. 2005. New records of the crystal darter (Crystallaria asprella) in the middle Mississippi River. American Midland Naturalist 154:471-473.

Stewart-Oaten, A., and J. R. Bence. 2001. Temporal and spatial variation in environmental assessment. Ecological Monographs 71:305-339.

Stewart-Oaten, A., W. W. Murdoch, and K. R. Parker. 1986. Environmental impact assessment: pseudoreplication in time? Ecology 67:929-940.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005. An Ecological Assessment of Western Streams and Rivers. EPA 620/R-05/005, US Environmental Protection Agency, Washington, DC.

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.

Stokesbury, K. D. E., J. B. Bighy, and S. W. Ross. Selectivity and efficiency of two otter trawls used to assess estuarine fish and macroinvertebrate populations in North Carolina. Estuaries 22(4)882-888.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38:913-920.

Stribling, J. B., and S. R. Davie. 2005. Design of a long-term environmental monitoring program for the Lake Allatoona/Upper Etowah River watershed, Georgia. Prepared for the US Army Corps of Engineers (Mobile District) and the Lake Allatoona/Upper Etowah River watershed partnership. Mobile, Alabama. (Available from: http://allatoona.sam.usace.army.mil/ Documents.htm)

Stribling, J. B., Jessup, B. K., White, J. S., Boward, D., and Hurd, M. 1998. Development of a Benthic Index of Biotic Integrity for Maryland Streams. CBWP-MANTA-EA-98-3. Prepared by Tetra Tech, Inc., Owings Mills, MD and Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. Annapolis, Maryland.

Stribling, J. B., S. R. Moulton, II, and G. Lester. 2003. Determining the quality of taxonomic data. Journal of the North American Benthological Society 22(4):621-631.

Suter, G. W. 1996. Abuse of hypothesis testing statistics in ecological risk assessment. Human and Ecological Risk Assessment 2:331-347.

Suter, II, G. W., S. B. Norton, and S. Cormier. 2002. A method for inferring the causes of observed impairments in aquatic ecosystems. Environmental Toxicology and Chemistry 21:1101-1111.

Taylor, B. R., and Z. Kovats. 1995. Review of artificial substrates for benthos sample collection. Prepared for Canada Centre for Mineral and Energy Technology (CANMET), Ottawa, Ontario, Canada.

Taylor, C. N., and C. E. Kuyatt. 1994. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. NIST technical note 1297. National Institute of Standards and Technology, Department of Commerce, Washington, DC.

Taylor, J. K. 1988. Defining the accuracy, precision, and confidence limits of sample data. Pages 102-107 in L. H. Keith (editor). Principles of environmental sampling. ISBN 0-8412-1173-6. ACS professional reference book. American Chemical Society, Columbus, Ohio.

Thorp, J. H. 1992. Linkage between islands and benthos in the Ohio River, with implications for riverine management. Canadian Journal of Fisheries and Aquatic Sciences 49(9):1873-1882.

Thorp, J. H., and M. D. DeLong. 1994. The riverine productivity model: an heuristic view of carbon sources and organic processing in large river ecosystems. Oikos 70:305-308.

Tockner, K., and J. A. Stanford. 2002. Riverine flood plains: present state and future trends. Environmental Conservation 29:308-330.

Townsend, C. R., and A. G. Hildrew. 1976. Field experiments on the drifting, colonization and continuous redistribution of stream benthos. Journal of Animal Ecology 45:459-772.

Underwood, A. J. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Australian Journal of Marine and Freshwater Research 42:569-587.

Underwood, A. J. 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable world. Journal of Experimental Marine Biology and Ecology 161:145-178.

Underwood, A. J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecological Applications 4:3-15.

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant trends over time. Ecological Applications 8(2):246-257.

USEPA (US Environmental Protection Agency). 1990. Biological criteria: national program guidance for surface waters. EPA 440-5-90-004. Office of Water Regulations and Standards, US Environmental Protection Agency, Washington, DC.

USEPA (US Environmental Protection Agency). 1992. Procedures for initiating narrative biological criteria. EPA 822-B-92-002. Office of Science and Technology, US Environmental Protection Agency, Washington, DC.

USEPA (US Environmental Protection Agency). 1998. Lake and reservoir bioassessment and biocriteria technical guidance document. EPA 841-B-98-007. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US Environmental Protection Agency). 2000a. Stressor identification guidance document. EPA 822-B-00-025. US Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (US Environmental Protection Agency). 2000b. Guidance for the data quality objectives process. EPA QA/G-4. EPA 600-R-96-055. US Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (US 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. US Environmental Protection Agency, Washington, DC. (available at: http://www.epa.gov/bioindicators/html/program_summary.html).

USEPA (US Environmental Protection Agency). 2005. Use of biological information to tier designated aquatic life uses in state and tribal water quality standards. EPA 822-R-05-001. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US Environmental Protection Agency). 2006. Technical guidance for assessing performance and comparability of bioassessment methods and their results. US Environmental Protection Agency, Office of Water, Washington, DC (in preparation).

Van Dam, H., A. Mertens, and J. Sinkeldam. 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. Netherlands Journal of Aquatic Ecology 28(1):117-133.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. Journal of Agricultural, Biological, and Environmental Statistics 2:370-388.

Van Sickle, J., C. P. Hawkins, D. P. Larsen, and A. T. Herlihy. 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24(1):178-191.

Van Sickle, J., D. D. Huff, and C. P. Hawkins. 2006. Selecting discriminant function models for predicting the expected taxa richness of aquatic macroinvertebrates. Freshwater Biology 51:359-372.

Van Sickle, J., and R. M. Hughes. 2000. Classification strengths of ecoregion, catchments and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19:370-384.

Vincent, R. 1971. River electrofishing and fish population estimates. Progressive Fish Culturist 33(3):163–169.

Vinson, M. R., and C. P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15:392-399.

Vitousek P. M., J. D. Aber, R. W. Howarth, G. E. Likens, P. A. Matson, D. W. Schindler, W. H. Schlesinger, and D. G. Tilman. 1997. Human alteration of the global nitrogen cycle: sources and consequences. Ecological Applications 7:737-750.

Voelz, N. J., and J. V. McArthur. 2000. An exploration of factors influencing lotic insect species richness. Biodiversity and Conservation 9:1543-1570.

Wade, D. C., and S. B. Stalcup. 1987. Assessment of the sport fishery potential for the bear creek floatway: biological integrity of representative sites, 1986. Report No. TVA/ONRED/AWR-87/30. Tennessee Valley Authority, Muscle Shoals, Alabama.

Waite, I. R., A. T. Herlihy, D. P. Larsen, and D. J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic highlands, USA. Journal of the North American Benthological Society 19:429-441.

Wang, L., J. Lyons, and P. D. Kanehl. 1998. Development and evaluation of a habitat rating system for low-gradient Wisconsin streams. North American Journal of Fisheries Management 18(4):775-785.

Wang, L., J. Lyons, P. D. Kanehl, and R. Gatti. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22(6):6-12.

Ward, J. V. 1989. Riverine-wetland interactions. Pages 385-400 in R. R. Sharitz and J. W. Gibbons (editors). Freshwater Wetlands and Wildlife. DOE Symposium Series No. 61, CONF-860301, USDOE Office of Scientific and Technical Information, Oak Ridge, Tennessee.

Ward, J. V. 1998. Riverine landscapes: biodiversity patterns, disturbance regimes, and aquatic conservation. Biological Conservation 83:269-278.

Ward, J. V., F. Malard, and K. Tockner. 2001. Landscape ecology: a framework for integrating pattern and process in river corridors. Landscape Ecology 17:35-45.

Ward, J. V., and J. A. Stanford. 1983. The intermediate disturbance hypothesis: an explanation for biotic diversity patterns in lotic ecosystems. Pages 347-356 in T. D. Fontaine, III and S. M. Bartell (editors). Dynamics of lotic ecosystems. Ann Arbor Science, Ann Arbor, Michigan.

Ward, J. V., and J. A. Stanford. 1995. The serial discontinuity concept: extending the model to floodplain rivers. Regulated Rivers Research and Management 10:159-168.

Weber, C. I. (editor). 1973. Biological field and laboratory methods for measuring the quality of surface waters and effluents. EPA 670-4-73-001. Natural Environmental Research Center, Office of Research and Development, US Environmental Protection Agency Cincinnati, Ohio.
Wehr, J. D., and R. G. Sheath (editors). 2003. Freshwater algae of North America: ecology and classification. Academic Press, San Diego, California.

Welcomme, R. L. 1985. River fisheries. Food and Agriculture Organization Fisheries Technical Paper 262. Food and Agriculture Organization of the United Nations, Rome, Italy.

Welcomme, R. L. 1989. Floodplain fisheries management. Pages 209-233 in J. A. Gore, and G. E. Petts (editors). Alternatives in regulated river management. CRC Press, Boca Raton, Florida.

Welcomme, R. L., K. O. Winemiller, and I. G. Cowx. 2006. Fish environmental guilds as a tool for assessment of ecological condition of rivers. River Research and Applications 22:377-396.

Wetzel, R. G. 2001. Limnology: lake and river ecosystems. 3rd edition. Academic Press, New York.

Wetzel, R. L. 1979. Periphyton measurements and applications. In R. L. Weitzel (editor). Methods and measurements of periphyton communities. American Society of Testing and Material ASTM STP 690:3-33.

Whited, D. C., J. A. Stanford, and J. S. Kimball. 2003. Application of airborne multispectral digital imagery to characterize the riverine habitat. Internationale Vereinigung für Theoretische und Angewandte Limnologie 28(3):1373-1380.

Whittaker, R. H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecological Monographs 22:1-44.

Wildhaber, M. L., P. J. Lamberson, and D. L. Galat. 2003. A comparison of measures of riverbed form for evaluating distributions of benthic fishes. North American Journal of Fisheries Management 23:543-557.

Wiens, J. A. 2002. Riverine landscapes: taking landscape ecology into the water. Freshwater Biology 47:501-515.

Wilhelm, J. G. O. 2002. A habitat rating system for non-wadeable rivers of Michigan. Master's Thesis, School of Natural Resources and Environment, University of Michigan, Ann Arbor, Michigan.

Wilhelm, J. G. O., J. D. Allan, K. J. Wessell, R. W. Merritt, and K. W. Cummins. 2005. Habitat assessment of non-wadeable rivers in Michigan. Environmental Management 36:592-609.

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.

Wright, J. F. 2000. An introduction to RIVPACS. Pages 1-24 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of freshwaters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Wright, J. F., M. T. Furse, and P. D. Armitage. 1993. RIVPACS: a technique for evaluating the biological water quality of rivers in the UK. European Water Pollution Control 3:15-25.

Wright, J. F., D. Moss, P. D. Armitage, and M. T. Furse. 1984. A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. Freshwater Biology 14:221-256.

Wright, J. F., J. M. Winder, R. J. M. Gunn, J. H. Blackburn, K. L. Symes, and R. T. Clarke. 2000. Minor local effects of a River Thames power station on the macroinvertebrate fauna. Regulated Rivers: Research and Management 16(2):159-174.

Yin, Y., J. S. Winkelman, and H. A. Langrehr. 2000. Long term resource monitoring program procedures: aquatic vegetation monitoring. LTRMP Program Report 95-P002-7. US Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin.

Yoccuz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72:106-111.

Yoder, C. O., R. J. Miltner, and D. White. 1999. Assessing the status of aquatic life designated uses in urban and suburban watersheds. In Proceedings of the National Conference on Retrofit Opportunities for Water Resource Protection in Urban Environments. EPA-625-R-99-002.

Yoder, C. O., and E. T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109-144 in W. S. Davis and T. P Simon (editors). Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Yoder, C. O., and M. A. Smith. 1999. Using fish assemblages in a state biological assessment and criteria program: essential concepts and considerations. Pages 17-56 in T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Zar, J. H. 1999. Biostatistical Analysis. 4th edition. Prentice Hall, Upper Saddle River, New Jersey.

Glossary

Abney level	A type of clinometer.
Accuracy	The nearness of a measured value to a true value, or a specified analytical truth.
Algae	Nonvascular plants mostly living in water.
Alluvial	Having to do with soil deposited by a river of other running water.
Ambient Monitoring	Sampling and evaluation of receiving waters not necessarily associated with episodic perturbations.
Allochthonous	Organic matter that was produced outside the system (e.g., wood, leaves, berries, insects etc.).
Anadromous	Describes fish that live most of life in oceans or lakes and migrate to streams to spawn.
Antidegradation Statement	Statement that protects existing uses, prevents degradation of high quality waterbodies unless certain determinations are made, and which protects the quality of outstanding national resource waters.
Assemblage	An association of interacting populations of organisms in a given waterbody, for example, fish assemblage or a benthic macroinvertebrate assemblage.
Aquatic Life Use	A beneficial use designation in which the waterbody provides suitable habitat for survival and reproduction of desirable fish, shellfish, and other aquatic organisms; classifications specified in state water quality standards relating to the level of protection afforded to the resident biological community by the state agency.
Attribute	Measurable part or process of a biological system.
Autecology	The ecology of individual organisms and populations, including physiological ecology, animal behaviour, and population dynamics. Usually only one or two species are studied.
Autochthonous	Organic matter produced within the system (e.g., algae, macrophytes).
Beneficial Uses	Desirable uses that water quality should support. Examples are drinking water supply, primary contact recreation (such as swimming), and aquatic life support.
Benthic Macroinvertebrates or Benthos	Animals without backbones, living in or on the sediments, of a size large enough to be seen by the unaided eye and which can be retained by a U.S. Standard No. 30 sieve (28 meshes per inch, 0.595 mm openings). Also referred to as benthos, infauna, or macrobenthos.
Best Management Practice	An engineered structure or management activity, or combination of these that eliminates or reduces an adverse environmental effect of a pollutant.
Best attainable conditions (Reference)	See *Reference Condition.*
Biological Assessment or Bioassessment	An evaluation of the biological condition of a waterbody using surveys of the structure and function of a community of resident biota.

Biological Criteria or Biocriteria	(Scientific meaning) are quantified values representing the biological condition of a waterbody as measured by structure and function of the aquatic communities typically at reference condition. (Regulatory meaning) are narrative descriptions or numerical values of the structure and function of aquatic communities in a waterbody necessary to protect the designated aquatic life use, implemented in, or through water quality standards.
Biological Diversity or Biodiversity	Refers to the variety and variability among living organisms and the ecological complexes in which they occur. Diversity can be defined as the number of different items and their relative frequencies. For biological diversity, these items are organized at many levels, ranging from complete ecosystems to the biochemical structures that are the molecular basis of heredity. Thus, the term encompasses different ecosystems, species, and genes.
Biological Indicator or Bioindicator	An organism, species, assemblage, or community characteristic of a particular habitat, or indicative of a particular set of environmental conditions.
Biological index	A metric or set of metrics collected into a single score calibrated to reference conditions and used as a measure of biological condition.
Biological Integrity	The ability of an aquatic ecosystem to support and maintain a balanced, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of natural habitats within a region.
Biological monitoring or Biomonitoring	Use of a biological entity as a detector and its response as a measure to determine environmental conditions. Ambient biological surveys and toxicity tests are common biological monitoring methods.
Biological survey or Biosurvey	Collecting, processing, and analyzing a representative portion of the resident aquatic community to determine its structural and/or functional characteristics.
Bioregion	Any geographical region characterized by a distinctive flora and/or fauna.
Bog	A type of wetland that accumulates acidic peat, a deposit of dead plant material.
Chain-of-Custody	Process for ensuring that the "holder" of samples or other items is known at all times, and is documented in writing.
Classification	The grouping of entities based on similarity in common attributes.
Clean Water Act (CWA)	An act passed by the U.S. Congress to control water pollution (formally referred to as the Federal Water Pollution Control Act of 1972). Public Law 92-500, as amended. 33 U.S.C. 1251 et seq.
Clean Water Act 303(d)	This section of the Act requires states, territories, and authorized tribes to develop lists of impaired waters for which applicable water quality standards are not being met, even after point sources of pollution have installed the minimum required levels of pollution control technology. The law requires that these jurisdictions establish priority rankings for waters on the lists and develop TMDLs for these waters. States, territories, and authorized tribes are to submit their list of waters on April 1 in every even-numbered year.
Clean Water Act 305(b)	Biennial reporting requires description of the quality of the Nation's surface waters, evaluation of progress made in maintaining and restoring water quality, and description of the extent of remaining problems.
Clinometer	An instrument to measure elevation angles above horizontal.
Community	An association of interacting assemblages in a given waterbody, the biotic component of an ecosystem.

Cosmopolitan Species	Species with worldwide distribution or influence where there is suitable habitat.
Criteria	Limits on a particular pollutant or condition of a waterbody presumed to support or protect the designated use or uses of a waterbody. Criteria may be narrative or numeric.
Data entry and storage	The processes and structures for entering and archiving environmental data into a data management system.
Data quality	The magnitude of error associated with a particular dataset.
Data Quality Objectives (DQOs)	Qualitative/quantitative statements that clarify objectives, define appropriate data, and specify tolerable levels of decision error for monitoring programs. They and are used to determine the quality and quantity of data needed.
DELT Anomalies	Percentage of Deformities, Erosions (e.g., on fins or barbels), Lesions and Tumors on fish assemblages.
Densiometer	An instrument used to measure vegetative canopy closure.
Design objectives	Qualitative and/or quantitative statements that clarify the purpose of a specific study design.
Designated Uses	Those uses specified in water quality standards for each waterbody or segment whether or not they are being attained. **Designated Use Attainment:** Degree to which a stream is meeting its water quality designated use goals.
Diagnostic capability	The capacity, in qualitative/quantitative terms, of a process or measure to identify the status or cause of a particular stream condition.
Diatoms	Unicellular forms of algae that grow a silica shell that is preserved in underwater sediments after they die.
Disturbance	Any temporary change in the environment that causes a long-term change I ecosystem, community, or population structure.
Ecological attributes	Inherent qualities or characteristics of biological communities and their physical and chemical environments.
Ecological integrity	The condition of an unimpaired ecosystem as measured by combined chemical, physical (including physical habitat), and biological attributes. Ecosystems have integrity when they have their native components (plants, animals and other organisms) and processes (such as growth and reproduction) intact.
Ecoregion	A relatively homogeneous ecological area defined by similarity of climate, landform, soil, potential natural vegetation, hydrology, or other ecologically relevant variables.
Ecosystem-level functions	Processes performed by ecosystems, including, among other things, primary and secondary production; respiration; nutrient cycling; decomposition.

Error	Variability, deviation from the true value. **Random** – variance the magnitude and direction of which cannot be predicted; some random error can be reduced through spatial and temporal aspects of the overall sampling design. **Systematic (or measurement)** – variance resulting from application or misapplication of sampling and analysis methods; generally controllable through consistent application of QC methods.
Existing Uses	Those uses actually attained in a waterbody on or after November 28, 1975, whether or not they are included in the water quality standards (November 28, 1975 is the date on which EPA promulgated its first water quality standards regulation). Because an existing use has been attained, it cannot be removed unless uses are added that require more stringent criteria.
Fluvial	Having to do with flowing water; see also lotic.
Function	Processes required for normal performance of a biological system (may be applied to any level of biological organization).
Glacial	Having to do with glaciers; glaciers are large, long-lasting rivers of ice formed on land.
Heterotrophic	Obtaining organic matter from other organisms rather than synthesizing it from inorganic substrates.
Hyporheic Zone	Area below the streambed where water percolates through spaces between the rocks and cobbles. Also known as the interface between surface water and groundwater.
Historic conditions (Reference)	See *Reference condition*.
Historical Data	Data sets from previous studies, which can range from handwritten field notes to published journal articles.
Historically documented taxa	Taxa known to have been supported in a waterbody or region prior to enactment of the Clean Water Act, according to historical records compiled by state or federal agencies or published scientific literature.
Human Disturbance	Human activity that alters the natural state and can occur at or across many spatial and temporal scales.
Index of Biological/Biotic Integrity	An integrative expression of site condition across multiple metrics. An index of biological integrity is often composed of at least seven metrics.
Indicators	An environmental attribute whose presence or magnitude is indicative of specific environmental conditions.
Invasive species	A species whose presence in the environment causes economic or environmental harm or harm to human health. Native species or non-native species may show invasive traits, although this is rare for native species and relatively common for non-native species.
Least disturbed conditons (Reference)	See *Reference condition*.
Lentic	Aquatic ecosystem where water is non-flowing (e.g., pond or lake).
Life-history requirements	Environmental conditions necessary for completing life cycles (including, among other things, reproduction, growth, maturation, migration, dispersal).

Linear response	A statistical relationship where one factor changes with another factor in a way that can be characterized with a straight line equation.
Lithophils	Organisms that thrive on rocks or stones.
Lithopelagophils	Organisms that spawn in open gravelly areas and have no guarding behavior.
Littoral	Zone near the edge of a body of water; depending on the context, it can be used to signify near-shore areas to several cm in depth.
Lotic	Flowing waters (e.g., as in streams, rivers); see also fluvial.
Maintenance of populations	Sustained population persistence; associated with locally successful reproduction and growth.
Measurement Quality Objectives (MQOs)	Statements that define the specific measurement goals needed to meet the Data Quality Objectives (DQOs); they are quantitative thresholds or qualitative statements of performance characteristics In general, the MQOs do not specify the methods, but provide criteria for describing different aspects of data quality.
Metric	A calculated term or enumeration representing some aspect of biological assemblage, function, or other measurable aspect and is a characteristic of the biota that changes in some predictable way with increased human influence.
Minimally disturbed conditions (Reference)	See *Reference condition*.
Montane	Descriptor of a geographic area dominated by mountains
Multimetric Index	An index that combines indicators, or metrics, into a single index value. Each metric is tested and calibrated to a scale and transformed into a unitless score prior to being aggregated into a multimetric index. Both the index and metrics are useful in assessing and diagnosing ecological condition. See Index of Biotic Integrity.
Multiple linear regression	Attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data.
Multivariate Analysis	Statistical methods (e.g. ordination or discriminant analysis) for analyzing physical and biological community data using multiple variables.
Narrative Biocriteria	Written statements describing the structure and function of aquatic communities in a waterbody necessary to protect a designated aquatic life use.
Native	An original or indigenous inhabitant of a region; naturally present.
Non-detrimental effect	Does not displace native taxa.
Non-native or intentionally introduced species	With respect to a particular ecosystem, any species that is not found in that ecosystem. Species introduced or spread from one region of the U.S. to another outside their normal range are non-native or non-indigenous, as are species introduced from other continents.
Non-wadeable streams and rivers	River reaches where boats are always necessary to access sample points or its only occasionally necessary to pull boats through shallow areas.
Numeric Biocriteria	Specific quantitative measures of the structure and function of aquatic communities in a waterbody necessary to protect a designated aquatic life use.

Performance Based Measurement Systems (PBMS)	Set of processes wherein the data needs, mandates, or limitations of a program or project are specified, and serve as criteria for selecting appropriate methods to meet those needs in a cost-effective manner.
Performance characteristics	Quantitative and qualitative descriptors of data quality, such as precision, accuracy, bias, representativeness, or completeness. Can also include terms such as selectivity, interferences, or others; other terms may be unique to particular methods or indicators.
Performance evaluation	Assessment of the acceptability of a measurement system based on the quality of data it produces.
Periphyton	A broad organismal assemblage composed of attached algae, bacteria, their secretions, associated detritus, and various species of microinvertebrates.
Phytoplankton	Microscopic, unicellular algae that are not attached to surfaces but typically remain suspended in the water column in aquatic ecosystems.
Piscivore	Predatory fish that eats mainly other fish.
Polyphils	Organism with no specialized spawning requirements, behavior, or preferred habitat.
P/R	Ratio of photosynthesis to respiration in a system.
Precision	The nearness of 2 or more measures of the same property.
Presently Attained Uses	Those uses actually being attained in a waterbody at the present moment.
Probabilistic design	Study or sampling characteristic that has randomization as a key component.
Quality assurance (QA)	A management system to assure quality in products or measurement systems.
Quality control (QC)	Technical procedures to ensure a process or product meets predetermined data quality objectives.
Random (probability) sampling	Drawing a sample unit from a population such that every unit has an equal probability of selection.
Rapid Bioassessment Protocols	Cost-effective techniques used to survey and evaluate the aquatic community to detect aquatic life impairments and their relative severity.
Reach	A length of stream or river lying between breaks in channel slope, local side-slopes, valley floor width, riparian vegetation, and bank material (Frissell et al. 1986).
Reference condition	The condition that approximates natural, unimpacted conditions (biological, chemical, physical, etc.) for a waterbody. Reference condition (Biological Integrity) is best determined by collecting measurements at a number of sites in a similar waterbody class or region under undisturbed or minimally disturbed conditions (by human activity), if they exist. Since undisturbed or minimally disturbed conditions may be difficult or impossible to find, least disturbed conditions, combined with historical information, models or other methods may be used to approximate reference condition as long as the departure from natural or ideal is understood. Reference condition is used as a benchmark to determine how much other water bodies depart from this condition due to human disturbance. *Also see Historic conditions, Minimally disturbed conditions, Best attainable conditions, and Least disturbed conditions.*

	Best Attainable Condition: a condition that is equivalent to the hypothetical ecological condition of least disturbed sites where the best possible management practices are in use. This condition can be determined using techniques such as historical reconstruction, best ecological judgment and modeling, restoration experiments, or inference from data distributions.
	Historic Condition: physical, chemical, and biological conditions existing only in the historical record, in databases, reports, and literature; contribute to development of reference expectations.
	Least Disturbed Condition: the best available existing conditions with regard to physical, chemical, and biological characteristics or attributes of a waterbody within a class or region. These waters have the least amount of human disturbance in comparison to others within the waterbody class, region or basin. Least disturbed conditions can be readily found, but may depart significantly from natural, undisturbed conditions or minimally disturbed conditions. Least disturbed condition may change significantly over time as human disturbances change.
	Minimally Disturbed Condition: the physical, chemical, and biological conditions of a waterbody with very limited, or minimal, human disturbance in comparison to others within the waterbody class or region. Minimally disturbed conditions can change over time in response to natural processes.
Reference criteria	A set of quantitative or qualitative rules used to identify reference sites. Usually based on a set of landcover and physical /chemical measures.
Reference site	A site selected for comparison with sites being assessed. The type of sites selected and the type of comparative measures used will vary with the purpose of the comparisons. For the purposes of assessing the ecological condition of sites, a reference site is a specific locality on a waterbody that is undisturbed or minimally disturbed and is representative of the expected ecological integrity of other localities on the same waterbody or nearby waterbodies.
Refugia	Accessible microhabitats or regions within a stream reach or watershed where adequate conditions for organism survival are maintained during circumstances that threaten survival, e.g., drought, flood, temperature extremes, increased chemical stressors, habitat disturbance, etc.
Regional Reference Condition	A description of the chemical, physical, or biological condition based on an aggregation of data from reference sites that are representative of a waterbody type in an ecoregion, subecoregion, watershed, or political unit.
Representativeness	A qualitative performance characteristic stating how well a value depicts what it is intended to depict.
Restoration	The re-establishment of pre-disturbance aquatic functions and related physical, chemical, and biological characteristics.
Rheophils	Organisms that flourish in free-flowing water.
Riparian area	Terrestrial ecosystem along the banks of a stream or river representing a vegetational transition between upland communities and the river.
River Invertebrate Prediction and Classification System (RIVPACS)	A predictive method developed for use in the United Kingdom to assess water quality using a comparison of observed biological species distributions to those expected to occur based on a model derived from reference data.
Sample collection	The process of taking a representative environmental measure.

Sample processing	The set of procedural steps a sample is taken through from collection to data entry.
Sampling Reach	A linear portion of a river selected for sampling purposes. A sampling reach may be of fixed (e.g., 1000 m) or variable length (e.g., 40 times the wetted width). See Section 3.1.1 for discussion.
Segment	A portion of a river system flowing through a single bedrock type and bounded by tributary junctions of major waterfalls (Frissell et al. 1986).
Sensitive taxa	Intolerant to a given anthropogenic stress; first species affected by the specific stressor to which they are "sensitive" and the last to recover following restoration.
Sensitive or regionally endemic taxa	Taxa with restricted, geographically isolated distribution patterns (occurring only in a locale as opposed to a region), often due to unique life history requirements. May be long-lived, late maturing, low fecundity, of limited mobility, or require mutualist relation with other species. May be among listed E/T or special concern species. Predictability of occurrence often low, therefore, requires documented observation. Recorded occurrence may be highly dependent on sample methods, site selection and level of effort.
Sensitive - rare taxa	Naturally occur in low numbers relative to total population density but may make up large relative proportion of richness. May be ubiquitous in occurrence or may be restricted to certain micro-habitats, but because of low density, recorded occurrence is dependent on sample effort. Often stenothermic (having a narrow range of thermal tolerance) or cold-water obligates; commonly k-strategists (populations maintained at a fairly constant level; slower development; longer life-span). May have specialized food resource needs or feeding strategies. Generally intolerant to significant alteration of the physical or chemical environment; are often the first taxa observed to be lost from a community.
Sensitive - ubiquitous taxa	Ordinarily common and abundant in natural communities when conventional sample methods are used. Often having a broader range of thermal tolerance than Sensitive- Rare taxa. These are taxa that comprise a substantial portion of natural communities, and that often exhibit negative response (loss of population, richness) at mild pollution loads or habitat alteration.
Spatial and temporal ecosystem connectance	Access or linkage (in space/time) to materials, locations, and conditions required for maintenance of interacting populations of aquatic life; the opposite of fragmentation; necessary for metapopulation maintenance and natural flows of energy and nutrients across ecosystem boundaries.
Spatial coverage	The area over which something is observed, measured, analyzed, or reported.
Stressors	Any physical, chemical, hydrologic, or biological factors that adversely affect aquatic organisms.
Structure	Taxonomic and quantitative attributes of an assemblage or community, including species richness and relative abundance structurally & functionally redundant attributes of the system = characteristics, qualities, or processes that are represented or performed by more than one entity in a biological system.
Study Design	Overall set-up of the study that includes the site selection, methods, number of replicate samples, and intended analyses. Examples include: **Regional assessments -** those that assess the average condition of water resource quality across a broad region for status and trends monitoring.

	Site-specific assessments - assessments where the focus is a particular site or small set of sites – usually for the purpose of assessing the effects of a specific impact (e.g., effluent) or the effectiveness of a given intervention (e.g., restoration). **Gradient assessments –** assessments focused on determining the strength and direction of biological response to specific stressors.
Subcategorized Uses	States and Tribes may adopt subcategories of a use and set the appropriate criteria to reflect varying needs of such subcategories of uses, for instance, to differentiate between cold water and warm water fisheries.
Swamp	A wetland featuring a permanent inundation of large areas of land by shallow bodies of water.
Taxa	A grouping of organisms given a formal taxonomic name such as species, genus, family, etc.
Taxa of intermediate tolerance	Comprise a substantial portion of natural communities; may be r-strategists (early colonizers with rapid turn-over times; "boom/bust population characteristics). May be eurythermal (having a broad thermal tolerance range). May have generalist or facultative feeding strategies enabling utilization of relatively more diversified food types. Readily collected with conventional sample methods. May increase in number in waters with moderately increased organic resources and reduced competition but are intolerant of excessive pollution loads or habitat alteration.
Temporal coverage	The time period over which something is observed, measured, analyzed, or reported.
Thalweg	A line drawn to joint the lowest points along the entire length of a streambed.
Tolerant taxa	Comprise a low proportion of natural communities. Taxa often are tolerant of a broader range of environmental conditions and are thus resistant to a variety of pollution or habitat induced stress. They may increase in number (sometimes greatly) in the absence of competition. Commonly r-strategists (early colonizers with rapid turn-over times; "boom/bust" population characteristics), able to capitalize when stress conditions occur. Last survivors.
Tolerance Value	A number indicating the relative capacity of an organism to survive and reproduce in the presence of stressors.
Total Maximum Daily Load	The sum of the allowable loads of a single pollutant from all contributing point and nonpoint sources; calculation of the maximum amount of a pollutant a waterbody can receive and still meet water quality standards and an allocation of that amount to the pollutant's source.
Use Attainability Analysis	Structured scientific assessment of the physical, chemical, biological or economic factors affecting attainment of the uses of waterbodies.
Wadeable stream or river	A fluvial waterbody that can be waded and/or adequately sampled by wading.
Water Quality Standards	A law or regulation that consists of the designated use or uses of a waterbody, the narrative or numerical water quality criteria (including biocriteria) that are necessary to protect the use or uses of that particular waterbody, and an antidegradation policy.

Water Resource Management (Non-Regulatory)	Decisions on management activities relevant to a water resource such as problem identification, need for and placement of best management practices, pollution abatement actions, and effectiveness of program activity.
Zooplankton	Planktonic animals that range in size from microscopic rotifers to macroscopic jellyfish.