United States
Environmental Protection
Agency

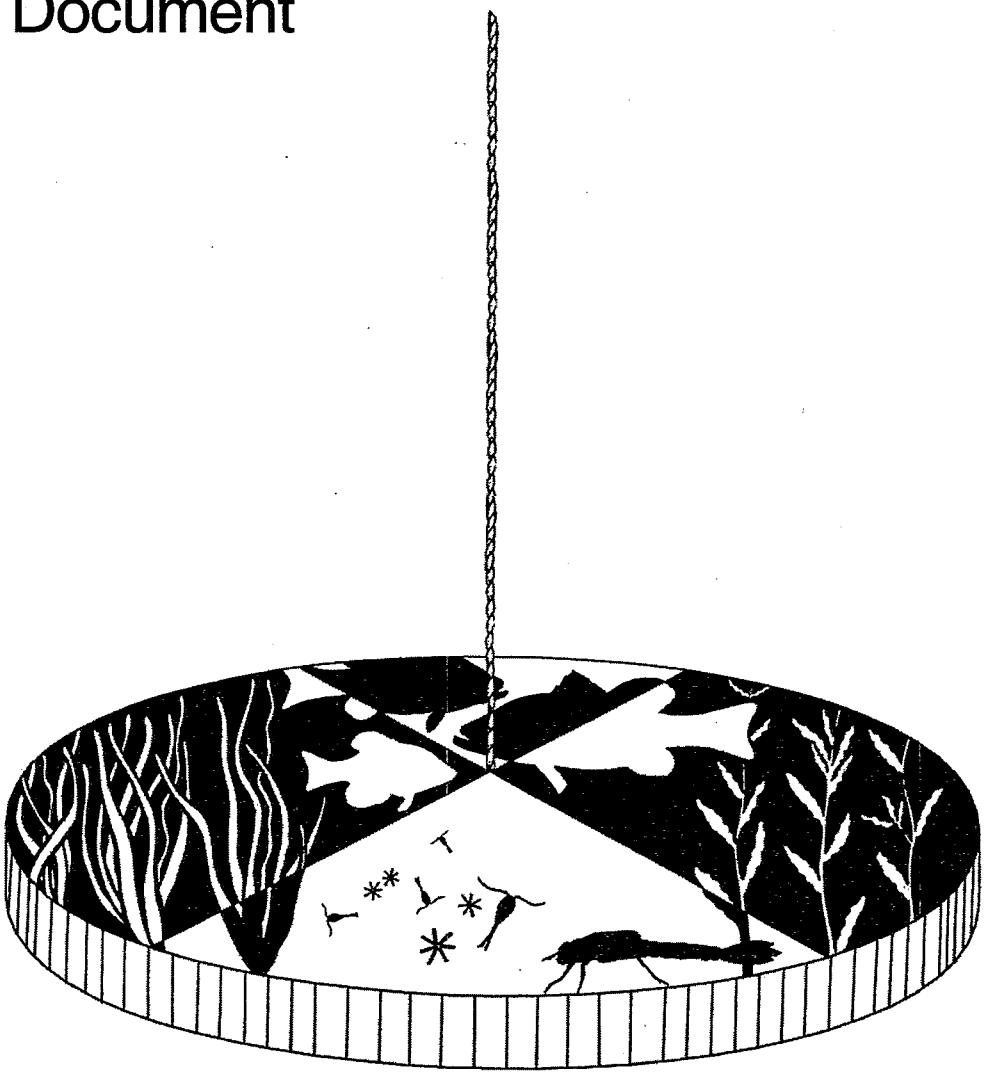
Office of Water
Washington, DC
(4503F)

August 1998
EPA 841-B-98-007

EPA

Lake and Reservoir Bioassessment and Biocriteria

Technical Guidance Document



Technical Guidance Document EPA 841-B-98-007

LAKE AND RESERVOIR BIOASSESSMENT AND BIOCRITERIA

TECHNICAL GUIDANCE DOCUMENT

August 1998

Office of Wetlands, Oceans, and Watersheds (4503F)
Office of Science and Technology (4304)
Office of Water
U.S. Environmental Protection Agency
Washington, DC

ACKNOWLEDGMENTS

Principal authors of this document are Jeroen Gerritsen (Tetra Tech, Inc.), Robert E. Carlson (Kent State University), Donald L. Dycus (Tennessee Valley Authority), Chris Faulkner (USEPA), George R. Gibson (USEPA), John Harcum (Tetra Tech, Inc.), and S. Abby Markowitz (Tetra Tech, Inc.).

Additional scientific, technical, editorial, and production contributions were made by Michael T. Barbour (Tetra Tech, Inc.), Allison M. Cauthorn (Tetra Tech, Inc.), Donald F. Charles (Patrick Center for Environmental Research), Joyce Lathrop-Davis (Tetra Tech, Inc.), Robert H. Kennedy (USACE), Edwin H. Liu (USEPA), Paula J. Proctor (Tetra Tech, Inc.), Susan Ratcliffe (USEPA), Amanda L. Richardson (Tetra Tech, Inc.), Blaine D. Snyder (Tetra Tech, Inc.), James B. Stribling (Tetra Tech, Inc.), Jeffrey S. White (Tetra Tech, Inc.), Sarah White (Tetra Tech, Inc.), Michel Schuster (Tetra Tech, Inc.), Emily Faalasli (Tetra Tech, Inc.), and Kelly Gathers (Tetra Tech, Inc.). This document was prepared by Tetra Tech, Inc.

Development of the protocols and this guidance document would not have been possible without the substantial contributions of the Lake Bioassessment Workgroup. Members of the Lake Bioassessment Workgroup included: Loren Bahls (Montana DHES), Michael T. Barbour (Tetra Tech, Inc.), Mary Belefski (USEPA), Mike Bira (USEPA), Don Bonneau (Iowa DNR), Roy Bouchard (Maine DEP), Dan Butler (Oklahoma Conservation Commission), Robert E. Carlson (Kent State University), Neil E. Carriker (Tennessee Valley Authority), Donald F. Charles (Patrick Center for Environmental Research), Dave Courtemanch (Maine DEP), Donald L. Dycus (Tennessee Valley Authority), Chris Faulkner (USEPA), Jeroen Gerritsen (Tetra Tech, Inc.), George R. Gibson (USEPA), David Hart (Patrick Center for Environmental Research), Steven Heiskary (MPCA), James Hulbert (Florida DEP), Susan Jackson (USEPA), Jack Jones (University of Missouri), Edwin Liu (USEPA), Ed Mills (Cornell University), Steve Paulsen (USEPA), Spencer Peterson (USEPA), Susan Ratcliffe (USEPA), Donna Sefton (USEPA), Tom Simon (USEPA), Kathy Stecker (South Carolina DHEC), and Richard S. Stemberger (Dartmouth College).

Cover artwork by Blaine Snyder of Tetra Tech, Inc.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Preface

This technical guidance document is based on the concept that bioassessment and biocriteria programs for lakes and reservoirs are interrelated and critical components of comprehensive water resource protection and management. The United States has approximately 40 million acres of lakes, ponds, and reservoirs. For the decade following the passage of the Clean Water Act in 1972, the Nation's lake acreage that experienced a decline in water quality was four times the acreage that experienced improvement (Johnson 1989). Managing, protecting, and restoring these waterbodies has been, and will continue to be, a challenge requiring the balancing of human and environmental health concerns with economic feasibility.

Our increased understanding of how lake systems function and respond to human activity has led to the recognition that environmental protection requires a holistic approach to lake management and protection. It has been necessary to expand our thinking in regard to lake monitoring approaches, incorporating biological assessments into traditional chemical and physical evaluations.

Section 101 of the Clean Water Act requires federal and state governments to "restore and maintain the chemical, physical and biological integrity of the Nation's waters." Natural, undisturbed aquatic ecosystems have high biological integrity, which is defined as "the condition of an aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by an evaluation of multiple attributes of the aquatic biota. Three critical components of biological integrity are that the biota is (1) the product of the evolutionary process for that locality, or site, (2) inclusive of a broad range of biological and ecological characteristics such as taxonomic richness and composition, trophic structure, and (3) is found in the study biogeographic region." (USEPA 1996a).

The goal of this guidance is to assist in protecting the ecological integrity of the Nation's lake and reservoir resources.

In 1992, the National Research Council of the National Academy of Sciences, calling for improved assessment programs to more effectively target lake restoration efforts, recommended the following (NRC 1992):

> *There is a great need for cost-effective, reliable indicators of ecosystem function, including those that will reflect long-term change and response to stress. Research on indicators should include traditional community and ecosystem measurements, paleoecological trend assessments, and remote sensing.*

Many natural resource agencies throughout the country have begun the process of developing and implementing biological assessment and criteria programs primarily for rivers and streams. This document is part of the effort to

advance the use of these strategies with regards to lakes and reservoirs, thereby fostering the development of credible and practical bioassessment programs.

The goal of this guidance is to assist in protecting the ecological integrity of the Nation's lake and reservoir resources. It does not address issues of human health assessments as these concerns are widely discussed in other technical documents and regulations. This guidance was developed through the experience of existing state, regional, and national lake monitoring programs. Several existing lake programs are used as case studies and examples throughout the document illustrating specific concepts or methods. It is important to remember that circumstances vary throughout the country and this document cannot specifically address every situation or experience.

This document is intended to provide managers and field biologists with functional methods and approaches for bioassessment and biocriteria.

The orientation of this document is toward practical decision making rather than research and its primary target audiences are state and tribal natural resource agencies. It is intended to provide managers and field biologists with functional methods and approaches that will facilitate the implementation of viable lake bioassessment and biocriteria programs that meet their needs and resources.

The methods, or protocols, presented here are organized in a tiered framework, ranging from trophic state surveys to more detailed bioassessment, allowing users flexibility in designing programs appropriate to their needs and resources. Procedures for program design, reference condition determination, field biosurveys, biocriteria development and data analysis are detailed. In addition, the document provides information on the application and effectiveness of lake bioassessment to existing EPA and state/tribal programs such as the Clean Lakes Program, 305(b) assessments, NPDES permitting, risk assessment, and watershed management. The appendices of the document include a glossary of terms, summaries of existing programs and protocols, detailed descriptions of biological assemblages, and procedures for statistical analysis of biological data.

The following is a summary of the information contained in each chapter:

Chapter 1: The Protection of Biological Integrity
This chapter introduces biological integrity, bioassessment and biocriteria as fundamental considerations in developing and implementing lake monitoring programs and discusses the relationship between these concepts and the Clean Water Act's goal of restoring and protecting the Nation's water resources. Chapter 1 provides a rationale for biomonitoring as an integral component of natural resource agency lake management and protection programs.

Chapter 2: Lake Biological Monitoring in USEPA, Local, State, Tribal, and Regional Protection and Management Programs
Monitoring is a vital element in natural resource protection programs. Chapter 2 summarizes the relationship of biological surveys and biocriteria to various programs in the Clean Water Act. The application of lake biomonitoring and the development of biocriteria in these programs play a critical role and can have significant benefits for natural resource agencies and their constituents. This chapter addresses where and how biomonitoring and biocriteria fit into these programs. In addition, this chapter explores some nonregulatory applications and benefits of biomonitoring programs.

Chapter 3: Overview of Bioassessment and Biocriteria
This chapter provides a sketch of the conceptual framework, application and approaches of bioassessment and biocriteria that are detailed in the remaining chapters.

Chapter 4: Selection and Characterization of Reference Conditions
Establishing reference conditions, which represent the best attainable conditions for lakes in a given region, lays the groundwork for the development of biomonitoring and biocriteria programs. The ecological health of a lake, as measured through biosurveys, is evaluated through comparison to the reference conditions. This chapter recommends and details an approach for designating and identifying reference conditions.

Chapter 5: Habitat Measurement
The evaluation of habitat provides essential clues as to the status of a lake's biological organisms. Chapter 5 discusses habitat, including both watershed and in-lake components, as an element of bioassessment programs.

Chapter 6: Biological Assemblages
This chapter describes the various biological organisms that are surveyed in lake bioassessment programs. Target assemblages were chosen primarily based on their ability to be sampled and analyzed in a cost-effective way and their use in existing programs.

Chapter 7: Tiered Sampling
Chapter 7 details an additive tiered approach to lake biosurveys that includes evaluation of habitat and biological assemblages, or organisms. The purpose of the tiered approach is to provide natural resource agencies a menu of assessment and protocol options that take into consideration varying levels of familiarity with biosurveys, regional needs, resource limitations, and regulatory requirements.

Chapter 8: Index Development
The final step toward functional bioassessment is the development of an index, comprised of the sum of a series of metrics, or measurement scores. The total index value of a test site is then compared to the index value for the reference condition. Chapter 8 provides an overview of procedures involved in selecting appropriate measurements and determining an index. The Tennessee Valley Authority's experience in developing metrics and indices is highlighted in this chapter as an example. (Appendix E provides more detailed discussions and examples of statistical methods used in data analysis.)

Chapter 9: Quality Assurance
This chapter discusses the various factors to consider in ensuring the reliability of monitoring and measurement data. Chapter 9 addresses quality assurance and control considerations for each step of the process including sampling design, field operations, laboratory operations, data analysis, and data reporting.

Chapter 10: Biocriteria Implementation
Chapter 10 discusses the characteristics of biocriteria and details the steps to implement a biocriteria program. Biocriteria provide natural resource agencies with a mechanism to protect the biological integrity of lakes and to establish aquatic life-use classifications. Issues of focus in this chapter include technical and resource considerations.

Appendix A: Glossary of Terms

Appendix B: Comparison of Existing Lakes Protocols

Appendix C: Paleolimnological Sampling

Appendix D: Biological Assemblages

Appendix E: Statistical Analysis Methods for Biological Assessment

Appendix F: Executive Summaries of State Pilot Studies

Appendix G: Literature Cited

In This Chapter...

- The Relationship Between Bioassessment and Biocriteria
- Uses of Bioassessment and Biocriteria

Chapter 1

The Protection of Biological Integrity

1.1 INTRODUCTION

Biological monitoring is integral to the measurement of the total ecological health of a waterbody and is becoming increasingly important in water quality monitoring and assessment. Historically, most natural resource programs have measured individual pollutants in the water column and sediments. Although such programs have effectively monitored and controlled point source discharges of nutrients and contaminants, their efforts to assess total ecological integrity, measured by combined chemical, physical (including habitat), and biological attributes, have been limited. Many surface waters have continued to deteriorate from nonpoint pollution, habitat modification, and other impacts of human activities (Karr 1991). For example, in the United States, the total lake acreage that deteriorated in quality from 1972 to 1982 was four times the acreage that improved (Johnson 1989).

This document describes a set of protocols for biological assessment of lakes and reservoirs relevant to issues of ecological integrity. It is not intended to address human health concerns as these issues have been addressed in previous

Around the country, various agencies use terms differently. This can lead to confusion when developing a guidance document intended for national use. Therefore, for the purposes of this document, the following terms are defined:

A ***biological survey (biosurvey)*** is the process of collecting and processing representative portions of a resident aquatic community to determine the community structure and function.

A ***biological assessment (bioassessment)*** is an evaluation of the biological condition of a waterbody that uses biosurveys and other direct measurements of resident biota in surface waters.

Biological monitoring (biomonitoring) is the use of a biological entity as a detector, and its response as a measure, to determine environmental conditions. Toxicity tests and biosurveys are common biomonitoring methods.

Biological criteria (biocriteria) are numeric values or narrative expressions that describe the reference biological condition of aquatic communities inhabiting waters of a given designated aquatic life use.

guidance documents. The protocols in this document are intended for use by local, state, tribal, and regional natural resource monitoring agencies, and they can be used in the implementation of biological criteria.

The document includes a general strategy for biocriteria development, identifies steps in the process, and provides technical guidance on how to complete each step, using the experience and knowledge of existing state, regional, and national surface water programs where appropriate. The protocols are tiered to allow flexibility in customizing individual monitoring programs according to the user's own requirements and available resources.

Biological monitoring is integral to the measurement of the total ecological health of a waterbody and is becoming increasingly important in water quality monitoring and assessment.

The multiple assemblage and multimetric assessment approach outlined here is designed to address elements and processes associated with community balance, trophic structure, and richness. This guidance is not intended to replace existing biological assessment or biocriteria programs. Rather, it can be used as a tool for developing new programs and/or enhancing current programs. Although not designed to "push the envelope" of lake bioassessment, this document was developed to provide methods that are technically credible, practical, and geared toward the genuine needs and resources of natural resource agencies.

1.2 THE CONCEPT OF BIOCRITERIA

Efforts to monitor human effects on waterbodies have ranged from 19th century physical observations of sediment and debris movement (Caper et al. 1983) to chemical metrics, currently the most commonly employed source of water quality criteria (USEPA 1992e). Investigators and resource managers, however, have long recognized that water column measurements reflect conditions only at the time of sampling.

As an important supplement to chemical sampling, biological measurements can reflect both current conditions and temporal changes in waterbodies, including the cumulative effects of successive disturbances. However, the development and widespread use of formal biological criteria have lagged behind chemical-specific, instream flow, and toxicity-based water quality criteria in waterbody management (USEPA 1985b, USEPA 1985c). Recent recommendations on monitoring strategies for aquatic resources have emphasized the need to accelerate the development of biological sampling as a regular part of surface water programs (USEPA 1987b, USEPA 1987c).

Biological criteria are benchmarks for water resource protection and management decision making. Expressed as numeric values or narrative expressions, they measure attainment of biological integrity. In turn, biological integrity describes the most robust aquatic community to be expected in a natural condition in a water resource relatively unaffected by human activities.

The development of biocriteria by natural resource agencies depends on the assessment of conditions at reference sites. Reference sites are not necessarily pristine, although they must exhibit only minimal impairment relative to the overall region of study. Based on biological sampling, or surveys, a bioassessment of multiple sites is done, resulting in values that represent the biological potential for waters in the region. The regional biological potential is then used to establish biocriteria. Biocriteria can then be used as a measuring stick for determining the status of test sites. The sites can be surveyed, scored, and compared to the established biocriteria.

Biocriteria supported by bioassessment serves several purposes in surface water programs. The use of biocriteria expands and improves water quality standards, helps identify impairment of beneficial uses, and helps set program priorities. The use of bioassessments to investigate impairment, evaluate the severity of problems, ascertain the causes of the problems, and determine appropriate remedial action is a step-by-step process. Decision criteria for ascertaining impairment are part of the implementation plan and the foundation for establishing biocriteria to determine beneficial use categories and assess subsequent management efforts. This should be followed by continued monitoring, improving the resource quality with each cycle. (See Figure 1-1.)

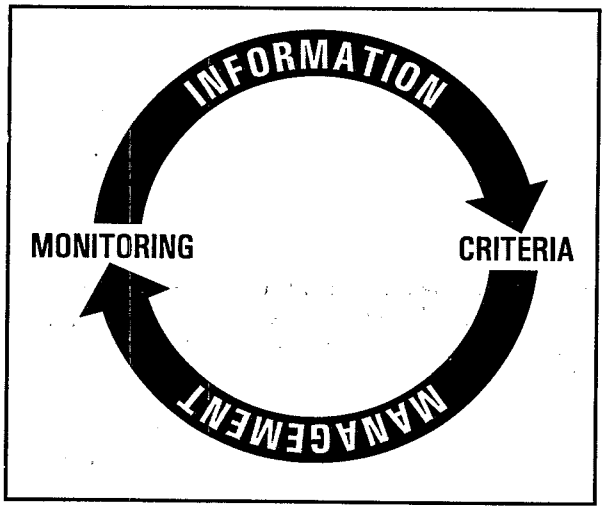


Figure 1-1. Interdependence of Environmental Monitoring and Environmental Criteria.

1.3 USES OF BIOASSESSMENTS AND BIOCRITERIA

By directly measuring the condition of the water resource at a site, surveys and assessments of resident biota are an important foundation in the derivation and maintenance of biocriteria and, thus, are a critical tool for natural resource agencies in protecting the quality of water resources. Biocriteria, in conjunction with surveys of aquatic assemblages, are useful for a variety of purposes including:

- Problem screening and identification, through the early detection of problems that other methods might fail to uncover or might underestimate.
- Assessing the effectiveness of implemented water resource management practices.
- Determining attainment of designated aquatic life uses.
- Refining aquatic life uses categories.
- Identifying impact sources.

Applications of bioassessments and biocriteria to specific USEPA, state, local, tribal, and regional management programs (such as under Clean Water Act sections 303, 305(b), 314, 319) are discussed in Chapter 2 of this document.

1.3.1 Screening and Identifying Problems

Monitoring of the resident biota can be used to identify and rank problem areas for further attention and dedication of resources. It can also serve as an early warning system to identify problems and to ensure against continued degradation.

Biological assessments can be used to establish priorities for remedial actions. Screening can be done on an individual lake to establish management priorities. Screening can also be used as a tool on a regional or statewide basis to determine programmatic priorities. For example, regional screening could determine whether nutrient controls, sediment controls, or toxic elimination should have the highest priority for improving regional surface water quality.

Monitoring of the resident biota can be used to identify and rank problem areas for further attention and dedication of resources.

1.3.2 Assessing Effectiveness of Management Practices

Bioassessments can be used to track the effectiveness of remediation measures. In managing nonpoint source pollution, the natural resource agency may initiate cooperative land use programs in a given area or install best management practices (BMPs) to improve the water resource. Both Nonpoint Source (NPS) and Clean Lakes Programs require monitoring of BMPs. Before-and-after bioassessments compared to the biocriteria "benchmark" make it possible to objectively evaluate the relative success of management by assessing actual biological community changes.

Before-and-after bioassessments compared to the biocriteria "benchmark" make it possible to objectively evaluate the relative success of management by assessing actual biological community changes.

While other management uses of biocriteria include reviewing the adequacy of NPDES permits, biocriteria are not recommended at this time for inclusion as NPDES permit limits.

Rather, they are ideal for assessing the adequacy of the permit to protect the resident biota. This can be done by comparing biosurvey results at the test site to the criteria established for that waterbody. Failure to meet the criteria suggests that the waterbody is not meeting its aquatic life use. One possible explanation is that the permit is not protective enough for the use class.

Monitoring the status and condition of resident communities over time is important to assess trends in the quality of the biota, whether to guard against further degradation or to measure relative improvement as a result of mitigation. Several natural resource agencies have established monitoring stations for conducting periodic biosurveys in streams as part of their biomonitoring programs. Very few natural resource agencies have initiated biological assessment for compliance monitoring in lakes.

1.3.3 Refining Aquatic Life Uses

Both classification and definition of designated uses of lakes and reservoirs are important in the planning, development, and use of biocriteria. Historical data from existing state efforts such as surface water classification and Clean Lakes Programs, along with additional field efforts, aid completion of these key planning steps. Information obtained through biological surveys can be used to explicitly describe each designated use.

Biocriteria relate directly to biological resource condition rather than surrogate concentrations of particular pollutants.

A designated use is a classification designated in state water quality standards for each waterbody or segment that defines the optimal purpose for that waterbody regardless of attainment status. The designated uses for lakes and reservoirs are usually defined by individual natural resource agencies and include such uses as drinking water, aquatic life, recreational use, industrial use, and agricultural use.

Use attainability—The potential for a waterbody to meet, reach, or develop to its optimal purpose or designated use.

Aquatic life uses—Classifications specified in state water quality standards for each waterbody or segment relating to the level of protection afforded to the resident biological community by the state agency.

General information on use designation can be found in *Biological Criteria: National Program Guidance for Surface Waters* (USEPA 1990a). Specific technical guidance for conducting use-attainability analyses is provided in *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses* (USEPA 1984).

Designated uses of waterbodies are formulated on, and in turn influence, the level of protection afforded the aquatic resource. Natural resource agencies establish standards appropriate to the protection of specific designated uses. For example, the designation outstanding waters is sometimes assigned to waterways which are located in undisturbed or minimally influenced watersheds and are characterized by aquatic communities that are deemed to be as naturally occurs (USEPA 1990a). Alternatively, other use designations may reflect preexisting land use patterns that prevent attainment of the highest quality waters. However, an observed downward trend does not justify lowered use designation.

1.3.4 Determining Attainment of Designated Use

Biological surveys and criteria are fundamental tools for assessing aquatic life use impairment. State water quality standards exist to define and protect designated uses conducive to overall water resource enhancement and preservation. Current biomonitoring tools used to judge nonattainment are not well-formulated in many instances. Consequently, many natural resource agencies rely exclusively or primarily on chemical-specific criteria to evaluate use impairment.

Biocriteria provide the only direct assessment of resource condition. and they are sensitive to a broader range of human influences on the watershed than are chemical criteria alone (Karr 1991, USEPA 1991b). By including biocriteria, a natural resource agency gains a much more complete assessment of the condition of the water resource. Biocriteria relate directly to biological resource condition rather than surrogate concentrations of particular pollutants.

Cumulative impacts on the biota can be measured, revealing synergistic degradation that may occur even though all specific permit conditions may be met. Similarly, this measure of the biotic community often reveals the sum total of effects over the entire year, not just at one point in time.

1.3.5 Identifying Causes of Impairment

The concept of measuring the attributes of aquatic communities in unimpacted areas for biocriteria was first developed for stream systems (Index of Biotic Integrity [IBI], Karr et al. 1986; Invertebrate Community Index [ICI], Ohio EPA 1987; Rapid Bioassessment Protocol [RBP], USEPA 1989b) Observed deviations from the unimpacted conditions are assumed to be indicative of impairment. Human-induced alterations affect biological integrity through their effects on five major classes of factors important to the aquatic biota (adapted from Karr et al. 1986):

- Energy base.
- Chemical constituents.
- Habitat structure.
- Hydrologic regimen.
- Biotic interactions.

These factors influence the aquatic biota and can adversely affect elements and processes that normally occur in a lake or reservoir. By specifically designing a survey to include all five of these elements, it is possible to address causality when a lake fails to meet its biocriteria. Such information will assist in diagnosing impaired sites and determining management actions, for example, distinguishing between impacts from toxic substances and disruption of habitat.

1.4 OTHER BIOCRITERIA REFERENCE DOCUMENTS

USEPA has developed technical guidance documents for implementing biocriteria in response to biocriteria development issues including legislative authority, steps in developing biocriteria, and the application of biocriteria to surface water management (USEPA 1990a). A reference guide to the technical literature pertaining to biocriteria has been developed to provide support interest from natural resource agencies (Table 1-1). This reference guide contains cross-references to technical papers that present concepts, approaches, and procedures necessary to implement habitat evaluations and biological surveys in the development and use of biocriteria.

In December 1990, a symposium on biological criteria provided a forum for discussing technical issues and guidance for the various waterbody types of the Nation's surface waters. The proceedings at this conference are presented in USEPA (1991b). The Agency has also developed guidance to help natural resource agencies initiate narrative biological criteria (USEPA 1992e).

Recently, the Agency issued a technical guidance document for biocriteria use in streams and small rivers (USEPA 1996a). Much of the approach and many of the issues addressed by the stream document serve as a template for developing biocriteria for other waterbody types, including lakes.

Table 1-1. Biocriteria reference documents.

Title	Document Citation
Biological Criteria: National Program Guidance for Surface Waters.	USEPA 1990. EPA-440/5-90-004. U.S. Environmental Protection Agency, Washington, DC.
Biological Assessment Methods, Biocriteria, and Biological Indicators: Bibliography of Selected Technical, Policy, and Regulatory Literature.	USEPA 1996. EPA-230-B-96-001. Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency, Washington, DC.
Biological Criteria: Research and Regulation. Proceedings of a Symposium.	USEPA 1991. EPA-440/5-91-005. U.S. Environmental Protection Agency, Office of Water, Health and Ecological Criteria Division, Washington, DC.
Procedures for Initiating Narrative Biological Criteria.	USEPA 1992. EPA-822-B-92-002. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.
Biological Criteria: Technical Guidance for Streams and Small Rivers.	USEPA 1996. EPA-822-B-96-001. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.
Summary of State Biological Assessment Programs for Streams and Rivers.	USEPA 1996. EPA-230-R-96-007. Office of Policy, Planning and Evaluation, U.S. Environmental Protection Agency, Washington, DC.

In This Chapter...

Chapter 2

Lake Biological Monitoring

in USEPA, Local, State, Tribal, and Regional Protection and Management Programs

A state monitoring program is the source of data for all other state resource management programs. It helps to identify water quality problems, identify waters needing total maximum daily loads (TMDLs), quantify loads, verify models, and evaluate the effectiveness of point and nonpoint source water quality controls. A state's monitoring program also serves as the backbone of its water quality programs. The biological monitoring protocols presented in this guidance document will strengthen a state's monitoring program. An effective and thorough water quality program can help to improve reporting (e.g., 305(b) reporting), increase the effectiveness of pollution prevention efforts, and document the progress of mitigation efforts.

Biological monitoring and the establishment of biocriteria provide scientifically sound and detailed descriptions of designated aquatic life use for waterbodies. Biocriteria are biological benchmarks for measuring the condition of aquatic biota. They help determine whether water quality goals are attained, set priorities, and evaluate the effectiveness of implemented controls and management actions. Developing and implementing biocriteria for lakes and reservoirs is complicated in some states because of a high level of human intervention on a significant percentage of lakes and reservoirs. Many lakes and reservoirs are managed by the states for different uses (e.g., drinking water, recreation, fishing). Several lake management practices mask natural conditions; for example, stocking of fish and periodic lowering of lake levels. In addition, entire regions of the country have no natural lakes but have abundant reservoirs, which do not have the same attributes as most natural lakes.

This section provides suggestions for the application of biological monitoring and biological criteria to lakes and reservoirs through existing state programs.

Despite the variability in lake conditions, performing biological monitoring and developing biocriteria for lakes have important benefits. This section provides suggestions for the application of biological monitoring and biological criteria to lakes and reservoirs through existing state programs (Table 2-1).

Table 2-1. Applications of lakes biological monitoring protocols and biocriteria.

Program	Biological Monitoring and Assessment	Biological Criteria
Section 305(b) Reporting	• Improving data for beneficial use assessment. • Improving water quality reporting.	• Identifying waters that are not achieving their aquatic life use support. • Defining an understandable endpoint in terms of "biological health" or "biological integrity" of waterbodies."
Section 314/Clean Lakes Program	• Assessing status of biological components of lake systems. • Measuring effects of ongoing restoration projects. • Measuring success of lake clean-up efforts and other mitigation activities. • Assessing lake trophic status and trends assessing biological trends. [Monitoring and sampling needs vary for each lake] [Clean Lakes Program Regulations monitoring components: algal pigments, algal genera, cell densities, algal cell volumes, limiting nutrients, macrophyte coverage, bacteria, and fish flesh analysis]	• Identifying lakes that are not attaining designated use (including aquatic life use) support. • Defining lake biological integrity based on a reference condition. • Identify impairments due to toxic substances.
Section 319/Nonpoint Source Program	• Evaluating nonpoint source impacts and sources. • Measuring site-specific ecosystem response to remediation or mitigation activities. • Assessing biological resource trends within watersheds.	• Determining effectiveness of nonpoint source controls.
Watershed Protection Approach	• Assessing biological resource trends within watersheds.	• Setting goals for watershed and regional planning.
TMDLs	• Identifying biological assemblage and habitat impairments that indicate nonattainment of water quality standards. • Documenting ecological/water quality response as a result of TMDL implementation. • Priority ranking waterbodies.	• Identifying water quality-limited waters that require TMDLs. • Establishing endpoints for TMDL development, i.e., measuring success.
NPDES Permitting	• Measuring improvement or lack of improvement of mitigation efforts. • Developing protocols that demonstrate relationship of biological metrics to effluent characteristics.	• Performing aquatic life use compliance monitoring. • Helping to verify that NPDES permit limits are resulting in achievement of state water quality standard.

Table 2-1. Applications of lakes biological monitoring protocols and biocriteria (continued).

Program	Biological Monitoring and Assessment	Biological Criteria
State Monitoring Programs	• Improving water quality reporting. • Documenting improvement or lack of improvement of mitigation efforts including lake clean-up efforts, TMDL application, NPDES efforts, nonpoint source pollution controls, etc. • Problem identification and trend assessment. • Prioritizing waterbodies.	• Measuring effectiveness of controls. • Performing watershed planning. • Performing regional planning.
Risk Assessment	• Providing data needed to estimate ecological risk to assessment endpoints.	• Development of an assessment or measurement endpoint.
Water Quality Criteria and Standards	• Developing data bases for lake phytoplankton, macroinvertebrates, fish plants, and other assemblages. • Developing indices that assess lake biota compared to reference.	• Identifying waterbodies that are not attaining aquatic life use support. • Refining aquatic life use classifications. • Developing site-specific standards.

2.1 SECTION 305(B) WATER QUALITY ASSESSMENT

Section 305(b) establishes a process for reporting information about the quality of the Nation's water resources (USEPA 1993c, USEPA 1994f). States, the District of Columbia, territories, and certain River Basin Commissions have developed programs to monitor surface and ground waters and to report the current status of water quality biennially to USEPA. Special grants are available for Native American groups to provide similar assessments of water quality on tribal lands. This information is compiled into a biennial National Water Quality Inventory report to Congress. The 305(b) reports are a major data source helping USEPA to:

- Determine the status of water quality. (Are the designated/beneficial uses being met?)
- Evaluate the causes of poor water quality and the relative contributions of pollution sources.
- Report on the activities under way to assess and restore water quality.
- Determine the effectiveness of control programs.
- Determine the workload remaining in restoring waters with poor quality and protecting threatened waters.

Use of biological assessment in 305(b) reports helps to define an understandable endpoint of relevance to society—the biological health and integrity of waterbodies. Many of the better known and widely reported pollution cleanup success stories have involved the recovery or reappearance of valued sport fish and other pollution intolerant species to systems from which they had disappeared (USEPA 1980b, USEPA 1985a). Improved coverage of biological integrity issues, based on monitoring protocols with clear bioassessment endpoints, will make the 305(b) reports more accessible and meaningful to many segments of the public.

Section 305(b) establishes a process for reporting information about the quality of the Nation's water resources.

The 305(b) process encourages monitoring and assessment for all lakes. The Clean Water Act Section 314 Clean Lakes Program outlines specific assessment or classification information for significant publicly owned lakes. Section

314(a)(2) of the CWA, as amended by the Water Quality Act of 1987, requires the states to submit a biennial assessment of their lake water quality as part of their 305(b) reports (USEPA 1993c). The specific elements of the assessment, as outlined in section 314(a)(10)(A-F), constitute the minimal requirements for approval and for subsequent grant assistance as required by section 314(a)(4). Each state report should reflect the status of lake water quality in the state, restoration/protection efforts, and trends in lake water quality. Each state should report the total number of significant publicly owned lakes and their acreage, the trophic status of each lake, control methods, restoration and rehabilitation efforts, the number of impaired and threatened lakes, acid effects on lakes, toxic effects on lakes, trends in lake water quality, and a description of the state's water quality standards that are applicable to lakes.

Biological monitoring can provide data that could augment several of the 305(b) reporting requirements. In particular, the following lake assessment activities and reporting requirements could be enhanced through the use of the biological monitoring information:

- Measuring the success of restoration and rehabilitation efforts when measured against reference conditions.
- Measuring the success of Clean Lakes Program projects.
- Developing and using lake water quality standards or, if water quality standards have not been developed for lakes, developing and using other biological measures to determine impaired or threatened status of lakes.
- Identifying lakes and lake acres affected by acidity or toxics and those with elevated levels of toxics.
- Identifying sources of acidity and toxic pollutants in lakes and estimating the number of affected lake acres attributed to each source.
- Identifying lake water quality trends, including trends in acidity, toxic pollutants, and their effects.

States are encouraged to develop integrated water quality strategies that include lake and reservoir management, restoration, and protection activities.

The Waterbody System (WBS) can generate many of the tables needed to report the required 305(b) summary data (USEPA 1994f). The Waterbody System can record general information on the types of monitoring protocols used in making assessments for specific lakes. Since WBS is intended as a data base of assessments, it does not have facilities for storing actual monitoring data or bioassessment metrics. Bioassessment information could, however, be entered in WBS comment fields.

2.2 SECTION 314 CLEAN LAKES PROGRAM

Historically, the Clean Lakes Program has been active in awarding grants for the study and restoration of publicly owned lakes. Under this program, states are encouraged to develop integrated water quality strategies that include lake and reservoir management, restoration, and protection activities.

The Clean Lakes Program regulations (40 CFR part 35, subpart H) list the primary components that could be monitored to characterize the biological component of a lake system, including algal pigments, algal genera, cell densities, algal cell volumes, limiting nutrients, macrophyte coverage (by species), bacteriological components, and fish flesh analysis. The regulations do not specifically require monitoring for macroinvertebrates. Whether a complete limnological investigation or some more focused set of investigations should be undertaken depends on the status of available baseline data and the problems affecting a particular lake.

Monitoring and sampling needs vary from lake to lake. For example, a lake program might do a more detailed benthic macroinvertebrate survey if dredging or restoration work involving the disturbance of sediments is planned. Even if this survey work is being done for dredging purposes only, it can aid in the formulation of an on-site reference. The use of a reference condition, whether it is developed by historical data or through a regional approach, can improve Clean Lakes projects by identifying biological impairments that were previously

unknown or not adequately documented based on chemical and physical monitoring data alone. In particular, biological monitoring will provide data to help accomplish the following:

- Determine the success of restoration and rehabilitation efforts when measured against reference conditions.
- Better characterize the biological component of the lake system.
- Measure aquatic life use support.
- Develop and use lake water quality standards, or develop and use other biological measurements to determine impairment or threatened status of lakes.
- Develop and update lake management plans.

All of the activities listed above can be partially achieved through the use of biological monitoring protocols in lake programs. They will lead to improved data for assessing beneficial uses and for improving both 305(b) and other grant reporting requirements.

2.3 SECTION 319 NONPOINT SOURCE PROGRAM

The 1987 Water Quality Act Amendments to the Clean Water Act added section 319, which established a national program to control nonpoint source (NPS) pollution. States assess their NPS pollution problems and submit these assessments to USEPA. The assessments include a list of "navigable waters within the state which, without additional action to control nonpoint sources of pollution, cannot reasonably be expected to attain or maintain applicable water quality standards or the goals and requirements of this Act." Other activities under the section 319 process require the identification of categories and subcategories of NPS pollution that contribute to the identification of impaired waters, descriptions of the procedures for identifying and implementing BMPs, control measures for reducing NPS pollution, and descriptions of state and local programs used to abate NPS pollution. Based on the assessments, states have prepared nonpoint source management programs, and USEPA grants are now available to assist in the implementation of approved state programs.

Biological assessment techniques can improve evaluations of nonpoint source pollution controls (or the combined effectiveness of current point and nonpoint source controls) by comparing biological integrity indicators before and after implementation of controls. Likewise, biocriteria can be used to measure site-specific ecosystem response to remediation or mitigation activities aimed at reducing nonpoint source pollution impacts or response to pollution prevention activities.

States assess their NPS pollution problems and submit these assessments to USEPA. The assessments include a list of "navigable waters within the state which, without additional action to control nonpoint sources of pollution, cannot reasonably be expected to attain or maintain applicable water quality standards or the goals and requirements of [section 319 of the Clean Water Act.]"

Several section 319 projects involve lake restoration (USEPA 1994f). Currently, biocriteria have not been developed for these lakes, but their use would greatly improve the ability of lake managers to focus their efforts. By providing a measuring tool, biocriteria can be key in identifying the most significant sources of a lake's pollutants. Minimum lake monitoring guidance for nonpoint source pollution assessment is being developed and will include biological protocols for lakes.

2.4 WATERSHED PROTECTION APPROACH

Since 1991, USEPA has been promoting the Watershed Protection Approach as a framework for meeting the Nation's remaining water resource challenges (USEPA 1994k). The agency's Office of Water has taken steps to reorient and coordinate point source, nonpoint source, lakes, wetlands, coastal, ground water, and drinking water programs in support of the watershed approach. USEPA has also promoted multi-organizational, multi-objective watershed management projects across the Nation.

The watershed approach is an integrated, holistic strategy for more effectively protecting and managing surface water and ground water resources and achieving broader environmental protection objectives using the naturally defined hydrologic unit (the watershed) as the integrating management unit. Thus, for a given watershed, the approach encompasses not only the water resource, such as a stream, river, lake, estuary, or aquifer, but all the land from which water drains to the resource. The watershed approach places emphasis on all aspects of water resource quality: physical (e.g., temperature, flow, mixing, habitat); chemical (e.g., conventional and toxic pollutants such as nutrients and pesticides); and biological (e.g., health and integrity of biotic communities, biodiversity).

The Clean Lakes Program (CLP) has been an important model for the Watershed Protection Approach and ecosystem management (USEPA 1994k). The CLP has been referred to as the quintessential watershed program because it has taken a holistic, place-based approach that uses sound science, involves stock holders, and forms partnerships for comprehensive, integrated action to protect and restore lake resources in the Nation. A newly developed Clean Lakes Program framework calls for better integration of the CLP with nonpoint source, water quality management, permitting, and other ecosystem protection activities.

The watershed approach is an integrated, holistic strategy for more effectively protecting and managing surface water and ground water resources and achieving broader environmental protection objectives using the naturally defined hydrologic unit (the watershed) as the integrating management unit.

2.5 SECTION 303(D) THE TMDL PROGRAM

The technical backbone of the Watershed Protection Approach is the process for total maximum daily loads (TMDL). TMDLs is a tool used to achieve applicable water quality standards. The TMDL process quantifies the loading capacity of a waterbody for a given stressor and ultimately provides a quantitative scheme for allocating loadings (or external inputs) among pollutant sources (USEPA 1994c). In doing so, the TMDL quantifies the relationships among sources, stressors, recommended controls, and water quality conditions. For example, a TMDL might mathematically show how a specified percent reduction of a pollutant is necessary to reach the pollutant concentration reflected in a water quality standard.

Section 303(d) of the CWA requires each state to establish, in accordance with its priority rankings, the total maximum daily load for each waterbody or reach identified by the state as failing to meet or not expected to meet water quality standards after imposition of technology-based controls.

In addition, TMDLs are vital elements of a growing number of state programs. For example, as more permits incorporate water quality-based effluent limits, TMDLs are becoming an increasingly important component of the point source control program.

TMDLs are suitable for nonchemical as well as chemical stressors (USEPA 1994c). These include all stressors that contribute to the failure to meet water quality standards, as well as any stressor that presently threatens but does not yet impair water quality. TMDLs are applicable to waterbodies impacted by both point and nonpoint sources. Some stressors, such as sediment deposition or physical alteration of instream habitat, might not clearly fit traditional concepts associated with chemical stressors and loadings. For these nonchemical stressors, it might sometimes be difficult to develop TMDLs because of limitations in the data or in the technical methods for analysis and modeling. In the case of nonpoint source TMDLs, another difficulty arises in that the CWA does not provide well-defined support for regulatory control actions as it does for point source controls, and controls based on another statutory authority might be necessary.

Because they directly measure the aquatic community's response to pollutants or stressors, biological surveys can provide compelling evidence of water quality impairment. Biological assessments and criteria address the cumulative impacts of all stressors, especially habitat degradation, loss of biological diversity, and nonpoint source pollution. Biological informa-

tion can help provide an ecologically based assessment of the status of a waterbody and thus can be used to decide which waterbodies need TMDLs (USEPA 1993c).

Incorporation of bioassessment data aids in the ranking process to target waters for TMDL development by allowing more accurate prioritization because of the direct link between bioassessment and ecological integrity (i.e., the condition of an unimpaired ecosystem as measured by combined chemical, physical, and biological attributes of surface waters (Barbour et al. 1992).

Finally, the TMDL process is a geographically based approach to preparing load and wasteload allocations for sources of stress that might impact waterbody integrity. The geographic nature of this process will be complemented and enhanced if ecological regionalization is applied as part of the bioassessment activities. Specifically, similarities among ecosystems can be grouped into ecoregions. The ecoregion concept provides a geographic framework for more efficient aquatic resource management.

2.6 SECTION 402 NPDES PERMITS AND INDIVIDUAL CONTROL STRATEGIES

All discrete sources of wastewater must obtain a National Pollutant Discharge Elimination System (NPDES) permit, which regulates the facility's discharge of pollutants. The approach to controlling and eliminating water pollution is focused on the pollutants determined to be harmful to receiving waters and on the sources of such pollutants. Authority for issuing NPDES permits is established under section 402 of the CWA (USEPA 1989a).

Point sources are generally divided into two types, industrial and municipal. Nationwide, there are approximately 50,000 industrial sources, which include commercial and manufacturing facilities. Municipal sources, also known as publicly owned treatment works (POTWs), number about 15,700 nationwide. Wastewater from municipal sources results from domestic wastewater discharged to POTWs, as well as the "indirect" discharge of industrial wastes to sewers.

USEPA does not recommend the use of biological criteria as the basis for deriving an effluent limit for an NPDES permit (USEPA 1994e). Unlike chemical-specific water quality criteria, biological criteria do not measure the concentrations or levels of chemical stressors. Instead, they directly measure the impacts of any and all stressors on the resident aquatic biota. Because of this, biological criteria do not definitively establish the causal relationship between a biological impact and its source. This is not to say that biological criteria have no role in the permitting process, now or in the future. Where appropriate, biological criteria can be used for assessment purposes within the NPDES process (USEPA 1996a). The criteria can provide information on the status of a waterbody where point sources might cause, or contribute to, a water quality problem. In conjunction with chemical water quality and whole-effluent toxicity data, biological criteria can be used to detect previously unmeasured chemical water quality problems and to evaluate the effectiveness of implemented controls.

Biological information can help provide an ecologically based assessment of the status of a waterbody and thus can be used to decide which waterbodies need TMDLs.

Some states have already demonstrated the usefulness of biological criteria under certain circumstances to indicate the need for additional or more stringent permit limits (e.g., sole-source discharge into a stream where there is no significant nonpoint source discharge, habitat degradation, or atmospheric deposition) (USEPA 1996a). In these situations, the biological findings triggered additional investigations to establish the cause-and-effect relationship and to determine the appropriate limits. In this manner, biological criteria support regulatory evaluations and decision making. Biological criteria can also be useful in monitoring highly variable or diffuse sources of pollution that are treated as point sources such as wet-weather discharges and stormwater runoff (USEPA 1996a). Traditional chemical water quality monitoring is not usually appropriate for these types of point source pollution, and a biological survey of their impact might be critical to evaluate these discharges and treatment measures effectively.

2.7 RISK ASSESSMENT

Ecological risk assessment is defined as "The process that evaluates the likelihood that adverse ecological effects may occur or are occurring as a result of exposure to one or more stressors" (USEPA 1992c). Risk management is a decision-making process that involves all the human-health and ecological assessment results, considered with political, legal, economic, and ethical values, to develop and enforce environmental standards, criteria, and regulations (Maughan 1993). Ecological risk assessment can be performed on an on-site basis or can be geographically based (i.e., watershed scale) to assess risks to ecologically valuable endpoints (USEPA 1996d).

Results of regional bioassessment studies can be used in watershed ecological risk assessments to develop regional empirical models of biological responses to stressors. Such models can then be used in a predictive mode, together with predicted exposure information, to predict risk due to stressors or to alternative management actions. Risks to biological resources are characterized, and sources of stress can be prioritized. Watershed risk managers can use such results for critical management decisions.

2.8 SECTION 303(C) USEPA WATER QUALITY CRITERIA AND STANDARDS

The water quality standards program, as envisioned in section 303(c) of the CWA, is a joint effort between the states and USEPA. The states have primary responsibility for setting, reviewing, revising, and enforcing water quality standards. USEPA develops regulations, policies, and guidance to help states implement the program and oversees states' activities to ensure that state-adopted standards are consistent with the requirements of the CWA and that water quality standards regulations (40 CFR Part 131) are met. USEPA has authority to review and approve or disapprove state standards and, where necessary, to promulgate federal water quality standards. A water quality standard defines the water quality goals of a waterbody, or a portion thereof, by designating the use or uses to be made of the water, setting criteria necessary to protect those uses, and preventing degradation of water quality through antidegradation provisions. States adopt water quality standards to protect public health or welfare, enhance the quality of water, and protect biological integrity.

Results of regional bioassessment studies can be used in watershed ecological risk assessments to develop regional empirical modes of biological responses to stressors.

Environmental stressors can be chemical, physical, or biological in nature, and likewise can impact the chemical, physical, and biological characteristics of an aquatic ecosystem. For example, the impact of a chemical stressor might be observed in impaired functioning or loss of a sensitive species and a change in community structure. The impact of a biological stressor, such as an introduced species, can result in a change in community structure through competition, predation, etc. Ultimately, the number or intensity of all stressors within an ecosystem will be evidenced by a change in the condition and function of the biotic community. The interactions among chemical, physical, and biological stressors and their compounding impacts emphasize the need to directly detect and assess actual water quality impairments of the biota.

Sections 303 and 304 of the CWA require states to protect biological integrity as part of their water quality standards. This can be accomplished, in part, through the development and use of biological criteria. As part of a state or tribal water quality standards program, biological criteria can provide scientifically sound and detailed descriptions of the designated aquatic life use for a specific waterbody or segment. They fulfill an important assessment function in water quality-based programs by establishing the biological benchmarks for (1) directly measuring the condition of the aquatic biota, (2) determining water quality goals and setting priorities, and (3) evaluating the effectiveness of implemented controls and management actions.

The challenge of evaluating effects from ecological stressors will best be met when the condition of the biota within an ecosystem can be assessed directly. Biological criteria for aquatic life will

help meet this need by allowing direct assessment of the condition of the biota that live either part or all of their lives in aquatic systems. These criteria (narrative or numeric) describe the expected biological condition of an aquatic community. They can be used as benchmarks to identify biological impairments and to help define ecosystem goals and endpoints. Biological criteria supplement traditional measurements (for example, as backup for hard-to-detect chemical problems) and will be particularly useful in assessing impairment due to nonpoint source pollution and nonchemical (e.g., physical and biological) stressors. Thus, biological criteria fulfill a function missing from USEPA's traditionally chemical-oriented approach to pollution control and abatement (USEPA 1996a).

Biological criteria can also be used to refine the aquatic life use classifications for a state. Each state develops its own designated use classification system based on the generic uses cited in the CWA, including protection and propagation of fish, shellfish, and wildlife. States frequently develop subcategories to refine and clarify designated use classes when several surface waters with distinct characteristics fit within the same use class or when waters do not fit well into any category; for example, cold-water versus warm-water habitat. As data are collected from biosurveys to develop a biological criteria program, analysis may reveal unique and consistent differences between aquatic communities that inhabit different waters with the same designated use. Therefore, measurable biological attributes can be used to refine aquatic life use or to separate one class into two or more subclasses.

2.9 OTHER USES

Although biological criteria and monitoring might be perceived in a regulatory context as one form of water quality management, they serve many other equally important functions, including the following:

Sections 303 and 304 of the CWA require states to protect biological integrity as part of their water quality standards. This can be accomplished, in part, through the development and use of biological criteria.

- Evaluating the effectiveness of management practices.
- Regional planning.
- Watershed planning.
- Determining management priorities for multiple waterbodies.
- Further classifying and qualifying relative water quality in a waterbody.
- Characterizing aquatic life that is at risk from various hazards.
- Providing a means to evaluate impacts that might not be protected by traditional risk assessment methods.

In This Chapter...

- Outline of the Biological Assessment and Criteria Process
- Application to Lakes

Chapter 3

Overview of Bioassessment and Biocriteria

3.1 CONCEPTUAL FRAMEWORK

The impact of human activities on lakes has been recognized for many centuries, and in the past 50 years, there has been more focus on the biological measurement of this impact. By 1950, the first index of aquatic species' tolerance to organic pollution, the "Saprobic System," was in use (see Hynes 1994 for review). More recently, indices such as the Hilsenhoff Biotic Index (HBI), which takes into account both organic pollution tolerance and the relative abundance of species (e.g., Hilsenhoff 1987), have been developed.

As modern ecologists recognized that human influences were reducing local and global biological diversity, the measurement of community structure (including species diversity and ecological roles) assumed increasing importance in evaluation of polluted sites. Indices to measure species diversity and distribution in a community (Pielou 1977) were developed, but achieved only limited use because their one-dimensional focus leads to high levels of uncertainty in assessment.

3.1.1 Multimetric Biological Assessment

The multiple attribute (or multimetric) approach, incorporating pollution tolerance, diversity, and ecological functions, was developed to more fully characterize the human impact on aquatic organisms. Karr (1981) and Karr et al. (1986) developed the fish Index of Biotic Integrity (IBI) and demonstrated that combinations of these attributes, or measurements, forming an index, provide valuable assessments of water resources.

The multimetric approach defines an array of measurements, each of which represents a measurable characteristic of the biological assemblage that changes in a predictable way with increased or decreased environmental stressors.

The multimetric approach defines an array of measurements, each of which represents a measurable characteristic of the biological assemblage that changes in a predictable way with increased or decreased environmental stressors (USEPA 1996a, USEPA 1997d). When integrated, a multimetric index functions as an overall indicator of biological condition. Each assemblage in the

aquatic community (for example, fish or algae) might have differing responses to pollution or degraded conditions. Thus, assessment methods that target multiple species and assemblages are capable of detecting a broad range of stresses and reflect the condition of a large segment of the ecosystem. However, there is not yet a complete understanding of how measurements respond, either quantitatively or qualitatively, to perturbation in general and to particular stresses.

Biological assessment of waterbodies depends on our ability to define, measure, and compare biological condition among similar systems.

To provide for an effective assessment, the variables selected to determine biological integrity should:

Be relevant to societal concerns—Biological measurements must be related to the properties of biotic systems that are of concern to society, such as native species, fish production, and biological diversity.

Be responsive to environmental stresses—Biological measurements and the measurements developed from them must be sensitive to environmental stress, and the response must be interpretable.

Have low uncertainty—Variability should be understood and measurement error should be controllable.

Be cost-effective—The cost incurred in measurement should be proportional to the value of the information obtained.

Be environmentally benign to measure—Sampling methods that disturb or alter habitats and organisms should be avoided.

Assessment of biological integrity typically focuses on a few broad but integral classes of ecological properties (e.g., Barbour et al. 1992, Karr 1991) that respond to anthropogenic impacts (e.g., Schindler 1988, Schindler et al. 1989), including:

Health—Individuals or populations.

Species structure and composition—The number and kinds of species in an assemblage. Species structure includes both diversity and the presence of pollution-tolerant species.

Trophic structure—The relative proportion of different feeding levels, such as filter feeders, scavengers, or predators.

System function—The productivity and material cycling of the system.

Multimetric assessment typically includes several measurements of at least three properties (species structure, trophic structure, and system function). Individual and population health measurements are used less often because they are not yet well developed for invertebrates and plants.

Biological assessment of waterbodies depends on our ability to define, measure, and compare biological condition among similar systems. Impairment of the waterbody is judged by its departure from the expected condition. This ability requires a functional definition of biological integrity as the condition of the aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by community structure and function (USEPA 1990a).

This definition of biological integrity makes the explicit assumption that natural, undisturbed systems are healthier than those changed by human activities. Because biological integrity is defined relative to unimpaired conditions, it must also be measured relative to those conditions. The four classes of ecological properties listed above are measurable relative to natural or unimpaired conditions.

Few waterbodies, however, are unimpacted. Minimally impaired waterbodies typically form the basis for defining reference conditions for biological assessment. Artificial lakes, such as reservoirs and impoundments, have no natural or "least disturbed" condition. Nevertheless, it is possible to define "most desirable" and "least desirable" conditions for artificial lakes.

3.1.2 Biological Assessment Process

The information of the biological variables is transformed to numeric scores, or rankings from good to poor. Such scores reduce the complexity

and uncertainty of multidimensional data for purposes of assessment, remediation, and communication of results to the public and decision makers. For example, managers might need to know whether a lake is in good condition, whether it needs to be watched more closely, or whether more intensive studies should be made to determine a course of action for restoration or remediation. Data analysis streamlines the information from the data to two or three dimensions that can be used in decision-making.

Multimetric biological indices are similar in concept to the common economic indices such as the Index of Leading Economic Indicators (Lahiri and Moore 1991). Both economic and biological indices are based on comparison to an operationally defined and measurable reference standard. In the economic indices, individual attributes are first standardized as a percentage of a baseline value, usually an annual average from a decade before (Green and Beckman 1992). The attribute scores are summed, and the sum is likewise expressed as a percentage of the index baseline. Standardization weights indicators equally and allows the use of indicators with different units (hours worked, persons unemployed, billions of dollars, etc.). In multimetric biological indices the metrics are standardized as a score compared to a reference standard. The basic procedural steps for biological assessment are as follows:

1. Sample the biological groups (assemblages) selected by the program, recording the relative abundance and other characteristics of each species.

2. Calculate chosen metrics using relative abundance and other measurements: for example, number of species, number of intolerant species, percent abundance of filter feeders.

3. Compare each to its expected value under reference conditions and assign a numeric score corresponding to good (similar to reference), fair (different from reference), or poor (substantially different from reference).

4. Sum the scores of all metrics of an assemblage to derive a total score for the assemblage.

5. Compare the total score to the biological criterion based in part on the expected total score under reference conditions.

In biological assessment, reference conditions are established by identifying least impaired reference sites, characterizing the biological condition of the reference sites, and setting thresholds for scoring the measurements. For reservoirs or in other instances where "best-quality" lakes are too few or not definable, an alternative is to select the highest quality conditions from among all lakes (TVA 1994).

Multimetric bioassessment is most effective when it is modified to specific regional conditions.

Multimetric bioassessment is most effective when it is modified to specific regional conditions. Bioassessment of streams has been successful when modified and calibrated regionally (e.g., Barbour et al. 1996a, Miller et al. 1988, Ohio EPA 1990). Success requires region-specific selection and calibration of measurements, as well as regional characterization of reference conditions. For example, submerged macrophytes are rare in rocky, high-elevation or high-latitude lakes and may be an inappropriate assemblage in such a region.

3.1.3 Biological Assessment in Ecological Risk Assessment

Ecological risk assessment "evaluates the likelihood that adverse ecological effects may occur or are occurring as a result of exposure to one or more stressors" (USEPA 1992c). Risk assessment is a process for organizing and analyzing data, information, assumptions, and uncertainties in order to examine the likelihood of adverse effects (USEPA 1996d). This process provides risk managers with a framework for explicitly considering available scientific information in conjunction with social, political, and economic factors when planning a course of action with environmental consequences.

Problem formulation is the foundation of risk assessment and depends on identification of assessment endpoints, development of conceptual models, and creation of an analysis plan.

Assessment endpoints are "explicit expressions of the actual environmental value that is to be protected" (USEPA 1992c). Assessment endpoints include both a valued ecological entity and an attribute of that entity that is potentially at risk (USEPA 1996d). For example, the fish community of a lake is an entity, and its overall similarity to native fish communities in undisturbed lakes could be the attribute for ecological risk assessment.

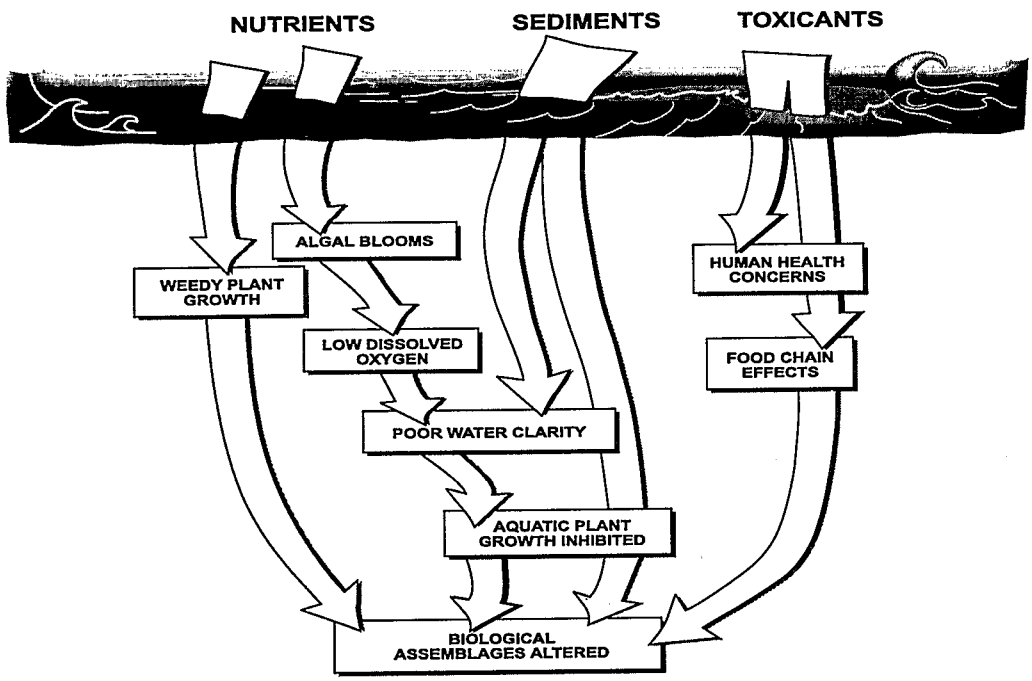
Biological assemblages and their attributes, as discussed in this and other biocriteria documents (e.g., USEPA 1996a), are clearly potential assessment endpoints for ecological risk assessments. Following risk assessment, a decision may be made to proceed with a management action. Monitoring can help determine if the desired result of the management action is achieved. Again, monitoring must include assessment endpoints, and established biocriteria can provide unambiguous ecological assessment endpoints.

Biological assessment emphasizes evaluation of both habitat and biota.

3.2 APPLICATION TO LAKES

Biological assessment emphasizes evaluation of both habitat and biota. As integrators of processes in their watersheds, lakes receive and retain matter and energy released throughout the watershed. Human activities are part of these processes and can affect a lake's habitat and biological community. The impact of human activities directly affects lake habitat and can alter the lake's physical-chemical environment (Figure 3-1). For example, contaminant discharges can affect the chemistry of both the water and the sediment. Agricultural and urban land uses in the watershed contribute sediment that affects the physical habitat. Humans can affect biological communities either directly by such activities as stocking and harvesting, or indirectly through impacts to the physical and chemical habitat of the biota.

Previous multimetric indices of lake quality have focused on lake condition compared to water quality standards, rather than on the actual biological condition of a lake compared to its regional potential. A multimetric index for



A Healthy Lake has clean water, balanced algal growth, adequate oxygen levels, and abundance and diversity of fish and bottom-dwelling invertebrates. Natural aquatic plants flourish in appropriate habitats, and bottom habitat is uncontaminated.

Figure 3-1. Effects of pollutants in lakes.

environmental quality of the Great Lakes used physical, chemical, biological, and toxicity variables (Steinhart et al. 1982). The Ohio EPA developed a multimetric assessment for inland lakes and reservoirs, the Ohio Lake Condition Index (LCI) (Davic and DeShon 1989), which was used to report lake condition for more than 300 public lakes in Ohio. The Ohio LCI consists of 14 metrics which represent biological, chemical, physical, and public perception of lake condition. Biological components in the Ohio LCI include fish IBI, macrophytes, phytoplankton chlorophyll, fecal coliform bacteria, and fish tissue contamination. Data are compared against water quality standards or general criteria to determine good, fair, or poor condition.

The Tennessee Valley Authority (TVA) developed biological assessment for its reservoirs that used a similar approach to the multimetric indices developed for stream assessment (Dycus and Meinert 1992, TVA 1994). TVA's assessment uses five indices based on benthic macroinvertebrates, fish, chlorophyll *a*, sediment quality, and dissolved oxygen. The macroinvertebrate and fish indices are multimetric.

The USEPA lake biological assessment procedure developed in this document may include up to seven biological assemblages: planktonic algae, attached algae, sedimented diatoms, aquatic plants, bottom-dwelling invertebrates, fish, and planktonic animals (Figure 3-2). Habitat scoring components include the watershed, nearshore zone, water chemistry, and sediment.

The proposed assessment of lake condition is accomplished with additive indices that integrate the habitat and biological scores. The process produces up to three habitat scores, and three or more biological index scores. The scores reduce the complexity of a lake to an understandable level for guiding appropriate remediation or other management actions.

3.2.1 Tiers for Sampling

Biological assessment of lakes is implemented in tiers corresponding to the level of effort required. Each suggested tier includes both biological and habitat components. The tiered approach for lake bioassessment developed here allows customization of the methodology to the

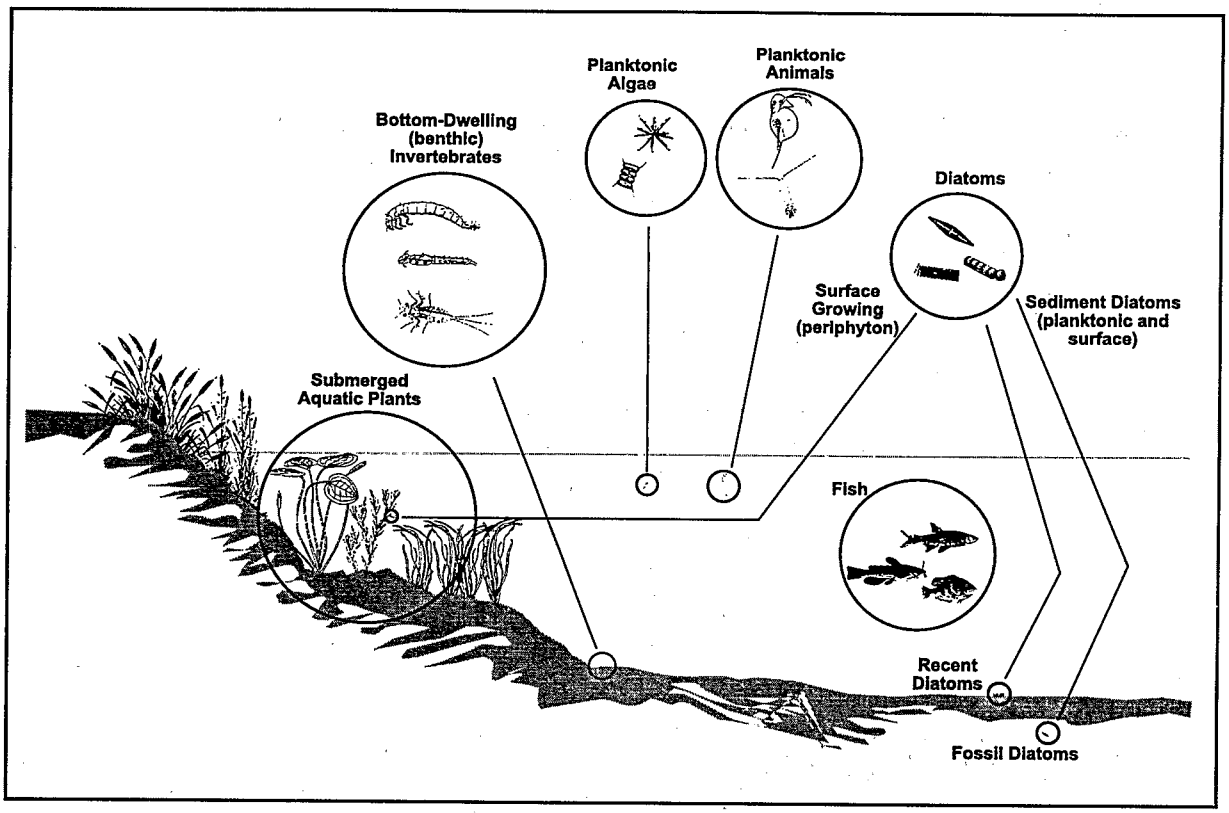


Figure 3-2. Biological assemblages used for lake assessments.

user's needs, questions, and resources available. Tier 1 focuses on sampling trophic state indicators, and Tier 2 focuses on sampling biological assemblages for composition and structure indicators (Figure 3-3 Table 3-1). Each tier is further divided into single- and multiple-visit sampling, A and B, respectively. Tier 1A and 1B are the same except that Tier 1B requires several samples during the growing season to obtain seasonal averages of chlorophyll *a* and nutrient concentrations.

Relevant lake classes must be determined by existing information and the professional judgment of scientists familiar with lakes of the region.

Tier 2A consists of biological assemblages that integrate lake conditions and are sampled during an index period. Tier 2B consists of assemblages with individuals that are short-lived, and hence do not integrate over time. Tier 2B assemblages are sampled repeatedly during the growing season to obtain seasonal averages.

Because chlorophyll and nutrient concentrations are highly variable, Tier 1A, which is sampled only during an index period, may fail to characterize an individual lake. Tier 1A is appropriate for characterizing a region or a class of lakes, especially if many lakes are to be sampled. For characterizing the trophic state of individual lakes with confidence, Tier 1B is preferred.

Both Tier 2A and 2B sample biological assemblages to estimate indicators of species structure, trophic structure, and function. Tier 2B requires multiple visits and analysis, but does not necessarily obtain better or more precise information than Tier 2A.

3.2.2 Classification of Lakes

Because there is tremendous variation in the physical, chemical, and biological characteristics of lakes nationwide, the first step in defining reference conditions is to classify lakes so that comparisons can be made within, not across, classes. Classification of natural lakes should reflect the inherent properties of lakes independent of human influence and therefore must be made on the basis of measurements that do not change as a result of human activities.

A second requirement of classification is that it should reflect differences in the biota of the classes. A deep lake might have a fish assemblage different from that of a shallow lake, and classification should distinguish between the two types of systems. Several lake classifications have been proposed (e.g., Hutchinson 1957, Leach and Herron 1992); however, only a handful of lake classes would be present in a single region. Relevant lake classes must be determined by existing information and the professional judgment of scientists familiar with lakes of the region.

3.2.3 Characterization of Reference Conditions

Five elements, detailed in Section 4.2, may be used to establish reference conditions for lake biological assessment:

- Biological survey of sites.
- Paleolimnology.

Table 3-1. Sampling tier summary.

Tier 1A	Trophic State Indices and macrophyte cover. Sampled once during index period. Inference limited to regional assessment.
Tier 1B	Trophic state indices and macrophyte cover. Sampled repeatedly during growing season.
Tier 2A	Tier 1 (1A or 1B) plus two or more integrating biological assemblages: macrophytes, macroinvertebrates, sedimented diatoms, fish. Sampled once during index period.
Tier 2B	Tier 1B plus two or more short-term biological assemblages: phytoplankton, zooplankton, periphyton. Sampled repeatedly during growing season.

- Evaluation of historical data.
- Prediction of expected conditions using models.
- Expert consensus.

Expert consensus is required for developing reference conditions. Reference conditions developed from empirical data are preferred: such as biosurveys, sites, paleolimnology, or historical data.

A biological survey provides the best current information about the biota for the system of concern as a real world reflection of biological integrity. This information is essential to determining the reference condition and subsequent biological criteria. There are two approaches for characterizing reference conditions from a biological survey:

Site based—Selection of minimally impaired or most natural sites in a region; or

Condition based—Setting reference conditions as the best available ambient biological conditions.

Paleolimnology is the microscopic examination of sediment cores to provide an accurate record of the relative abundance of certain organisms (primarily diatoms) over the history of natural lakes. The advantage of paleolimnology is that any lake with an accurate sedimentary record can be a reference site regardless of the severity of present-day pollution. Thus, a truly representative sample of lake reference sites can be drawn. With some exceptions, paleolimnology is generally not applicable to impoundments.

A panel of diverse regional experts involved in the determination of the reference condition and the derivation of the biocriteria is the best approach to thoroughly and objectively assimilate the above information. With a carefully selected and balanced panel, all of the nuances of the local ecology as well as the best interests of the jurisdiction can be equated to the designated uses of the waterbody in designing the most protective criteria possible. This approach also reduces the risk of making insufficiently informed decisions inherent in data interpretation by just one or a few like-minded people.

3.2.4 Reference Condition in Reservoirs

Throughout this document where differences between lakes and reservoirs dictate alternative methods, strategies, etc., an icon appears, directing the reader to reservoir-specific information.

The methodology described in this document is intended for both reservoirs and natural lakes. Because reservoirs are entirely artificial environments, "natural reference condition" has no meaning. Reservoirs, created by the damming of a stream, have characteristics of both rivers and lakes (Thornton 1990a). Reservoirs are divided into three zones (riverine, transitional, and lacustrine), which correspond to flowing, river-like conditions; transition to lake conditions; and nonflowing, lake-like conditions near the dam, respectively. With expected life spans ranging from one to several decades, reservoirs are more ephemeral than most natural lakes and have several physical characteristics not shared with natural lakes. The lakes most like reservoirs are those formed by natural dams in stream valleys (e.g., beaver dams, terminal moraines, landslides).

Because reservoirs are entirely artificial environments, "natural reference condition" has no meaning. Reservoirs, created by the damming of a stream, have characteristics of both rivers and lakes.

Reservoirs vary widely in physical characteristics of shape, size, and hydrology. They can range from small shallow impoundments, to deep storage reservoirs, to "run of the river" flow-through reservoirs on large rivers. They are built and managed for widely different purposes, including flood control, navigation, water storage, hydroelectric generation, gamefish production, and others. The management practices in turn affect both physical characteristics (water level variability, stratification) and biota (stocking of fish).

Although no "natural" reservoir reference conditions can exist, the operational determination of reference conditions for reservoirs is

the same as that for natural lakes. Reservoirs can be classified according to hydrology, morphometry, management objectives, and other factors. Age of the reservoir will be important in determining the assessment expectations of the reservoir.

Historical data are important because they provide insight to past conditions essential to knowing what may be achievable, especially for degraded or significantly altered systems.

Comparison of the historical record to present reference site data greatly expands the manager's perspective of the system. However, care must be exercised in making these comparisons when the objectives and survey methods have changed over time.

Ecological models may be used to identify water chemistry reference conditions for reservoirs or for other significantly altered waterbodies. Most reservoirs are less than 50 years old, and there is insufficient empirical evidence to document the expected condition of basins for all regions. Where documentation is available (historical data), extrapolation and model development help qualify the reference condition and may be the best way to derive and calibrate the biocriteria.

3.2.5 Metric Determination

Metrics are evaluated for relevance to biological assessment and for response to stress. Expected measurement values vary as a function of regional species pools, regional characteristics (climate, geology, soils, land use, regional scale barriers to colonization), and local site characteristics (habitat factors, including local barriers). A regional approach involving collaboration of neighboring jurisdictions will enhance characterization of reference conditions. Cross-state comparisons can be made more easily if common methods and measurements can be established among states.

A regional approach involving collaboration of neighboring jurisdiction will enhance characterization of reference conditions.

Metrics are typically calculated from data collected on single assemblages of lake biota, such as planktonic algae, zooplankton, fish, aquatic plants, and benthic invertebrates. The metrics might include counts, species identifications, ratios, and indices combining several data variables depending on the level of effort, or tier, of the survey.

3.2.6 Data Analysis

When performing bioassessment of lakes, individual metrics are assigned scores, usually a number corresponding to good, fair, or poor relative to the values of the measurements in reference conditions (Karr 1991, Karr et al. 1986). This serves to standardize the metrics on the same scale so they can be combined into an additive index. Measurement scores are summed to obtain an index score for each assemblage, such as an IBI or macroinvertebrate community score. Currently, each measurement is weighted equally in the summed index score.

Additive biological indices collapse a great deal of information into a single number. Yet they have been shown to be reliable in detecting impairment of aquatic systems (Fore et al. 1994, Fore et al. 1996, Wallace et al. 1996); they are simple to compute once criteria are established, and they are easily communicated to managers and the public (Gerritsen 1995).

Habitat component scores may give clues to the causes of impairments reflected in biological indices rated fair or poor. Habitat variables that are significantly different from reference conditions are identified as probable causes of impairment, warranting further investigation or remediation. This sort of bioassessment cannot establish cause of impairment; it can only separate probable from improbable causes of impairment. In any given bioassessment, several probable causes might be identified.

3.3 BIOCRITERIA

Biological data are used to help set biological criteria based on management needs and defined management classes. States may draft general narrative biocriteria early in their program—even before they have designated reference sites or refined their approach to biological surveys. This does not mean that having reference sites and a refined system for conducting surveys is unimportant; it means

that a biocriteria program begins with writing into law a statement of intent to protect and manage water resources predicated on an objective or benchmark, for example, "aquatic life shall be as naturally occurs."

When the objective to restore and protect the biological integrity of the water resources has been formally mandated, then the operational meaning of the statement and the identification of the agency responsible for developing the necessary procedures and regulations can be stipulated as the state's first steps toward the development of narrative and numeric biological criteria. The key point is that natural or minimally impaired water resource conditions become the criteria for judgement and management.

Although based on the same concept as narrative biocriteria, numeric biocriteria include discrete quantitative values that summarize the status of the biological community and describe the expected condition of this system for different designated water resource uses.

The key distinction between narrative biocriteria supported by a quantitative database and numeric biocriteria is the direct inclusion of a specific value or index in the numeric criteria. This index allows a level of specification to water resource evaluations and regulations not common to narrative criteria. To develop numeric criteria, the resident biota are sampled at minimally impaired sites to establish reference conditions. Attributes of the biota, such as species richness, presence or absence of indicator taxa, and distribution of trophic groups, help establish the normal range of the biological community as it would exist in unimpaired systems.

RESERVOIR

Case Study: Biological Assessment of Reservoirs by TVA

The Tennessee Valley Authority is currently using a multimetric biological assessment methodology on its reservoirs. The Tennessee River watershed drains portions of four ecoregions: Blue Ridge, Central Appalachian Ridge and Valley, Southwestern Appalachians, and Interior Plateau (Omernik 1987) (Figure 3-3). The Tennessee River begins at the confluence of the Holston and French Broad Rivers and receives drainage from the Ridge and Valley and Blue Ridge ecoregions. Downstream, the river drains a small portion of the Southwestern Appalachians and a large part of the Interior Plateau. The main stream carries water from two to four ecoregions. Therefore, dividing the main stream reservoirs by ecoregion does not contribute to a meaningful classification. Figure 3-3 illustrates that the tributary reservoirs can be easily divided by ecoregion. There are several reservoirs with watersheds entirely within the Blue Ridge and Ridge and Valley ecoregions. There is a third, and more dispersed, group of tributary reservoirs in the Interior Plateau.

Physical, chemical, and biological indicators were selected to provide information on the health or condition of habitats or ecological compartments. The open water or pelagic area was represented by physical and chemical characteristics of water (including chlorophyll) in midchannel. The shoreline or littoral area was evaluated by sampling the fish community. The bottom or benthic compartment was evaluated using two indicators: quality of surface sediments in midchannel (determined by chemical analysis of sediments) and examination of benthic macroinvertebrates from a transect across the full width of the sample area (including overbanks if present).

Three areas were selected for monitoring: the inflow area, generally riverine in nature, the transition zone or mid-reservoir area where water velocity decreases due to increased cross-sectional area, suspended materials begin to settle, and algal productivity increases due to increase water clarity; and the forebay, the lacustrine area near the dam. Overbanks, basically the floodplain which was inundated when the dam was built, were included in transition zone and forebay areas. Four large embayments (all with drainage areas greater than 500 square miles and surface areas greater than 4500 acres) were included in the Vital Signs Monitoring program. Ecosystem interactions within an embayment are mostly controlled by physical characteristics of the embayment and by activities and characteristics within the embayment watershed, usually with little influence from the main body of the reservoir (Meinert et al. 1992).

Sampling frequencies and index periods take into account the expected temporal variation for each indicator. Physical and chemical components vary significantly in the short term so they are monitored monthly from spring to fall. Biological indicators better integrate long-term variations and are sampled once each year. Fish assemblage sampling is conducted in autumn (September-November).

Initially, benthic macroinvertebrate sampling was conducted in early spring (February-April) to avoid aquatic insect emergence. The TVA experience showed that a late winter/early spring sampling period is not acceptable for benthic macroinvertebrates because results reflected conditions which occurred the previous year. This causes results from this indicator to be out of synch with the other four indicators. A late fall/early winter collection avoids problems resulting from early spring sampling.

The TVA case study is continued in subsequent chapters.

Case Study: Biological Assessment of Reservoirs by TVA (continued)

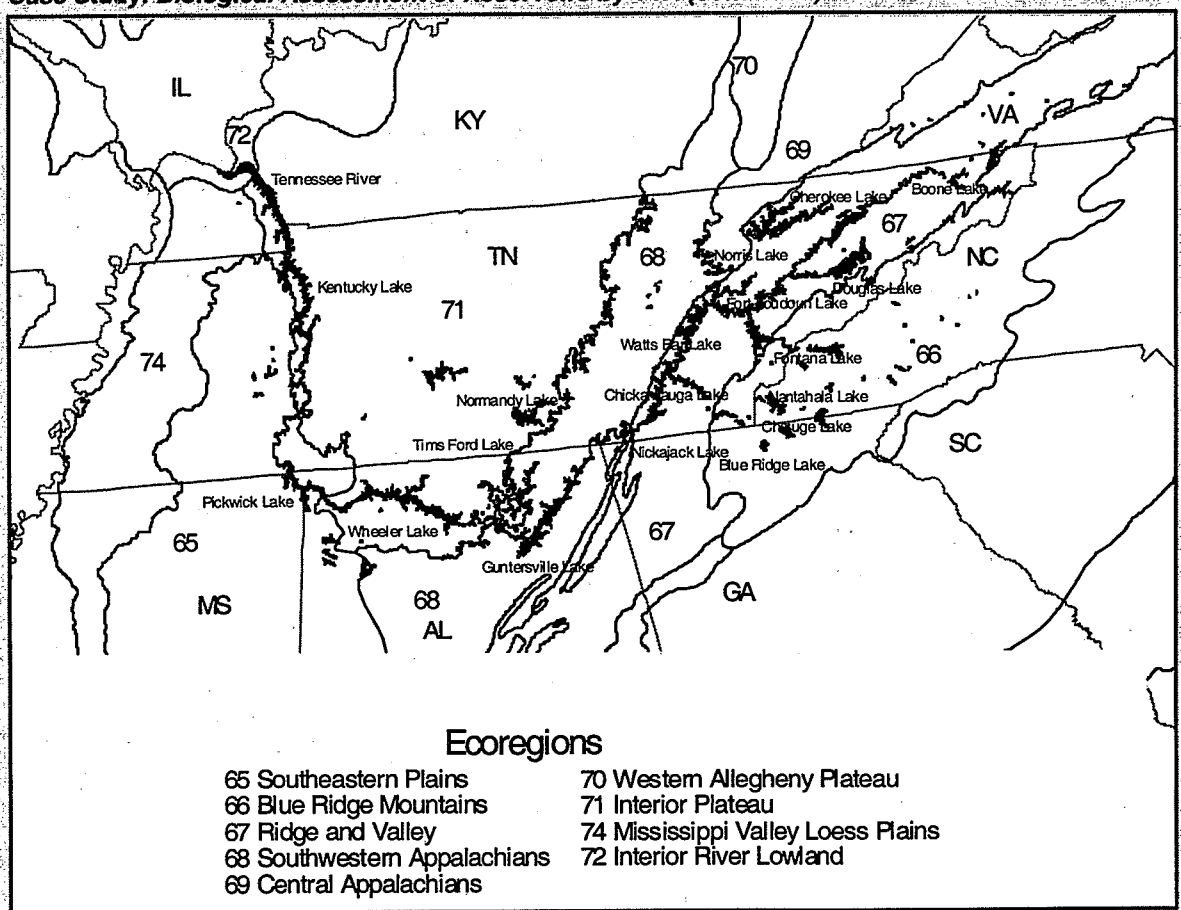


Figure 3-3. Distribution of TVA reservoirs in ecoregions.

In This Chapter...

- Preliminary Classification of Lakes
- Methods for Establishing Reference Conditions

Chapter 4

Selection and Characterization of Reference Conditions

Establishment of reference conditions is key to biological assessment and biocriteria programs. Reference conditions are a representation of the biotic potential for lakes in the absence of human activity or pollution. The attainment of aquatic life use is evaluated against the expectations of the reference condition as expressed in the biocriteria. Reference conditions are expectations on the status of biological communities under minimal anthropogenic disturbances and pollution. The expectations are usually based on the status of reference sites, which might be subject to anthropogenic influences. Ideally, reference sites are minimally impacted by human pollution and disturbance. The care that states use in selecting reference sites and developing reference condition parameters, together with the survey techniques employed, will bear directly on their ability to defensibly assess a waterbody. At a minimum, reference conditions should be identified for each of the lake classification categories developed for a state. As pointed out in Section 3.2.4, the definition of reference condition differs between natural lakes and artificial reservoirs.

The general sequence of reference condition characterization is to first assemble a panel of experts and make a preliminary classification of lake resources within a region. Following classification, sampling sites are selected, and habitat and biological data are obtained from those sites (either from existing data bases or from a survey). The preliminary classification is reconciled with the biological data to ensure that the final classification is biologically meaningful, and the reference conditions are characterized as part of the biocriteria development process.

At a minimum, reference conditions should be identified for each of the lake classification categories developed for a state.

4.1 REGIONALIZATION AND PRELIMINARY CLASSIFICATION

The regional differences in biological communities across the United States must be accounted for in the development of biological monitoring programs. This is done by comparing the biology of lakes to a regional reference condition. As biological conditions change across the country, the reference conditions will change also. To account for the regional differences in biological communities, and also for the differences that result from structural differences in biological

habitat (either natural or caused by human activities), USEPA recommends that states classify lakes into categories and that a reference condition should be developed for each of the lake categories. Biotic index comparisons can then be made within each category, and inappropriate biological comparisons between different classes will be avoided. Moreover, the aquatic life expectations of waterbodies are tempered by realistic regional expectations; there is no attempt to set a single numeric aquatic life designated use standard for the entire nation.

Lakes vary widely in size, shape, and ecological characteristics, and a single reference condition that applies to all lakes would be misleading.

Lakes vary widely in size, shape, and ecological characteristics, and a single reference condition that applies to all lakes would be misleading. The purpose of classification is to group similar lakes together: i.e., to prevent comparison of apples and oranges. By classifying lakes the variability of biological measures within classes is reduced and the variability among classes is maximized. Classification invariably involves professional judgment to arrive at a workable system that separates clearly different ecosystems, yet does not consider each lake a special case. The intent of classification is to identify groups of lakes that, under ideal conditions, would have comparable biological communities. As far as possible, classification should be restricted to those characteristics of lakes that are intrinsic, or natural, and not the result of human activities.

4.1.1 Definition of the Resource

Most large reservoirs, and some natural lakes, are on rivers and might be considered large pools in the rivers rather than lakes. At what point does a pool become a lake? For the purpose of lake bioassessment, it is when distinctly lake-like flora and fauna occur (i.e., phytoplankton and zooplankton). Phytoplankton require a water retention time of 3 days or more (Uhlmann 1971). Microzooplankton (e.g., rotifers) have generation times roughly twice that of phytoplankton cells; therefore, the minimum retention time for zooplankton to develop may be approximately 1 week.

For the purposes of bioassessment described here, a lake is any inland body of open water with some minimum surface area free of rooted vegetation and with an average hydraulic retention time of more than 7 days.

These characteristics distinguish lakes from small ponds and wetlands, and from riverine pools (natural or artificial) that retain their lotic character. The distinction between lake and small pond is arbitrary, and the minimum size for a waterbody to be considered a lake must be set by resource agencies. For practical reasons, this document does not explicitly consider emergent wetlands at the margins of lakes. Bioassessment methods for wetlands are being developed separately by USEPA and other agencies.

The unit of assessment and sampling (the sampling unit) is, most commonly, a definable, relatively self-contained basin of a lake. Most lakes have a single basin and thus will consist of a single sampling unit. Larger lakes, and especially reservoirs, have embayments, arms, and basins that are hydrologically isolated from the main body of the lake. Each isolated basin can be considered a separate sampling unit because of restricted water flow between basins. Large lakes can thus comprise several sampling units. Alternatively, a state may wish to define the sampling unit as an area or point in space (e.g., $1m^2$).

Most reservoirs are also divided into three zones—riverine, transitional, and lacustrine—to reflect differences among these zones (Thornton 1990b). Each zone is a separate sampling unit; in large reservoirs, zones might be represented in each major arm (TVA 1994).

4.1.2 Basic Rules

There is no single "best" classification, nor are resources available to determine all possible differences between all lakes in a region. The key to classification is practicality within the region or state in which it will be applied; local conditions determine the appropriate classes. Classification will depend on regional experts

familiar with the range of lake conditions in a region, as well as biological similarities and differences between the lakes. Ultimately, classification can be used to develop a predictive model of lake characteristics that affect the values of the biological metrics and indices in reference sites.

There are two fundamental approaches to classification, *a priori* and *a posteriori* (Conquest et al. 1994). The *a priori* approach consists of developing logical rules for classification based on observed patterns in characteristics of the objects. Thus, classifying lakes on ecoregion, surface area, and maximum depth would be an *a priori*, rule-based classification. The *a posteriori* approach develops groups from a data base of observations from the sites. The classification is restricted to those sites and variables in the data base and typically involves cluster analysis to develop the groups. The *a posteriori* approach is useful for exploratory analysis of a substantial data set, but it is not appropriate for operational assessment and management, where a site's class must be established from prior information (e.g., maps) before intensive data are collected. A few general rules for the development of *a priori* lake classification include:

- In *a priori* classification, lake characteristics that are readily affected by human activities or are a biological response to physical or chemical conditions should not be used as classification variables. Such responses might include trophic state, chlorophyll, or nutrient concentrations. For example, in the Northern Lakes and Forests ecoregion of Minnesota, lake trophic state is characteristically low whereas in the nearby Northern Glaciated Plains ecoregion, trophic state is relatively high (Heiskary 1989). The classification variable in this case is ecoregion, and trophic state is a response to ecoregion. A eutrophic lake in the Northern Lakes and Forests is considered impaired, but a eutrophic lake in the Northern Glaciated Plains is not considered impaired. Using trophic state as a classification variable could lead to misclassifications and inappropriate assessments.
- As shown in the example above, the best classification variables are those which are readily obtained from maps, bathymetric charts, or regional water characteristics, such as alkalinity or hardness.

4.1.3 Considerations for Reservoirs

Several differences between reservoirs and natural lakes affect the classification and interpretation of biological data (Thornton 1990a, Wetzel 1990):

Distribution—Reservoirs are most numerous in regions with few natural lakes: the nonglaciated parts of North America (except Florida) have the largest numbers of reservoirs (Thornton 1990a).

Form—The form or shape of the basin and watershed may be the most important distinction between natural and artificial lakes. Shape substantially influences the hydrology and water quality of reservoirs. Large reservoirs are drowned river valleys and tend to be long and deep with numerous embayments from tributaries. The watersheds of reservoirs are typically much larger than those of natural lakes and contribute greater sediment loads.

Most of the differences between reservoirs and natural lakes are resolved in classification of the lake resource.

Longitudinal gradient—Reservoirs have characteristics typical of both lakes and streams within the same basin. They are more like streams at the head where major tributaries enter and are more like lakes near the dam (Thornton 1990b).

Turbidity and loading—Reservoirs are typically more turbid, and they receive more nutrients and organic matter from their tributary streams than do most natural lakes.

Management—Reservoirs were built and are managed for specific purposes: hydro-power, irrigation, flood control, fisheries, and multiple uses. Management might include extreme water level fluctuations, fish stocking, and other effects not present in natural lakes.

Most of the differences between reservoirs and natural lakes are resolved in classification of the lake resource. The needs for which reservoirs were designed dictate many attributes of these waterbodies. Operational strategies can influence reservoir characteristics and resultant water quality (Kennedy and Walker 1990, Kennedy et al. 1985). The release of water from deep in the water column increases heat gain and the dissipation of materials accumulated in bottom waters (Martin and Arneson 1978, Wright 1967). Surface releases dissipate heat and retain materials. These and other operational differences can provide a basis for grouping reservoirs because reservoirs operated similarly can be expected to exhibit similar limnological responses, even when compared across large, heterogeneous regions.

4.1.4 Hierarchical Framework

This protocol is not intended to develop a classification scheme applicable to the entire United States. Overviews of global lake classification systems are in Hutchinson (1957) and in Leach and Herron (1992). Classification must be regional, and regional expertise must be used to determine those classification variables which are useful in a region.

A useful classification scheme is hierarchical, beginning at the highest (regional) level and stratifying as far as necessary (Conquest et al. 1994). The procedure is to classify lakes at the highest level (usually geographic), and then to continue stratification in the classification hierarchy to a reasonable point. Although several possible classification levels are outlined below, in practice, only one, or at most two, relevant levels would typically be used. Classification should be parsimonious to avoid proliferation of classes that do not contribute to assessment. One or two relevant levels of the hierarchy will yield the best classification scheme. The proposed hierarchical scheme below applies to both natural lakes and reservoirs.

Geographic Region—The geographic region (e.g., ecoregion, physiographic province) determines landscape-level features such as climate, topography, regional geology and soils, biogeography, and broad land use patterns. Ecoregions are based on geology, soils, geomorphology, dominant land uses, and natural vegetation (Hughes and Larsen 1988, Omernik 1987) and have been shown to account for variability of water quality and aquatic biota in several areas of the United States (e.g., Barbour et al. 1996a, Barbour et al. 1996b, Heiskary et al. 1987, Hughes et al. 1994, Ohio EPA 1987).

Because of the importance of geography in determining aquatic biota, the National Research Council's Aquatic Restoration Committee made the following recommendation (NRC 1992):

> *The committee believes that goals for restoration of lakes need to be realistic and should be based on the concept of expected conditions for individual ecoregions. Further development of project selection and evaluation techniques based on ecoregion concepts and refinement of ecoregion definitions and descriptions should be encouraged and supported by the U.S. Environmental Protection Agency.*

Many of the characteristics below that can be used as classification variables are often subsumed by ecoregion. For example, watersheds are often similar within ecoregions, having been formed by the regional geomorphology, and water quality characteristics such as alkalinity are determined by regional bedrock and soils. Within ecoregions, it might be sufficient to classify using only lake basin morphology (e.g., depth, area, development ratio): anthropogenic or natural origin; or management objective.

Anthropogenic Origin Reservoirs and other artificial lakes cannot have "natural" reference conditions. Therefore, reservoirs and natural lakes should be separated in developing reference expectations.

Watershed Characteristics—Watershed characteristics affect lake hydrology, sediment and nutrient loads, alkalinity, and dissolved solids. As noted above, many watershed characteristics are relatively uniform within an ecoregion and may not be necessary if ecoregions were the primary classification variable. Watershed characteristics that may be used as classification variables include:

- Lake drainage type (e.g., flowage, drainage, seepage, reservoir type).
- Land use.
- Watershed-to-lake area ratio (especially for reservoirs).
- Slope (especially for reservoirs).
- Soils and geology (erosiveness of soils).

Lake Basin Characteristics—Lake basin morphology influences lake hydrodynamics and lake responses to pollution. Characteristics of some reservoirs change with age, particularly regional shoaling and silting of aged reservoirs subject to high sediment loads (O'Brien 1990). Morphological metrics include:

- Depth (mean, maximum).
- Surface area.
- Bottom type and sediments.
- Shoreline development ratio (shoreline length: circumference of equal area circle).
- Age (of reservoirs).
- Epilimnetic/hypolimnetic discharge (reservoirs).

Lake Hydrology—Lake hydrology forms a basis for water quality. Mixing and circulation patterns influence nutrient retention and the development of hypoxia. Hydrological factors include:

- Retention time.
- Stratification and mixing.
- Circulation.
- Water level fluctuation and drawdown.

Characteristic Water Quality—Lakes can be classified by characteristic water types into categories, such as marl lakes, alkali lakes, ombrotrophic bog lakes, and others. Many water quality characteristics are relatively uniform within an ecoregion and as the result of regional, watershed, basin, and hydrologic characteristics. Water types are determined by the following water quality variables:

- Alkalinity.
- Salinity.
- Conductivity.
- Turbidity (Secchi depth, clarity, etc.).
- Color.
- Dissolved organic carbon (DOC).
- Dissolved inorganic carbon (DIC).

Classification must be regional, and regional expertise must be used to determine those classification variables which are useful in a region.

Human actions (e.g., discharges, land use) alter water quality, especially sediment and nutrient concentrations, but they can also affect alkalinity, salinity, conductivity, color, and DOC. Care must be taken that classification according to characteristic water types reflects natural conditions and not anthropogenic impacts. For example, if a lake is highly turbid due to poor land management practices, it should not be classified as highly turbid. Rather, it should be classified as it would have been in the absence of poor land use.

4.2 ESTABLISHING REFERENCE CONDITIONS

Five elements are used to establish lake reference conditions for biological monitoring and biological criteria: (1) expert consensus, (2) biological survey of sites, (3) paleolimnology, (4) evaluation of historical data, and (5) prediction of expected conditions using ecological models (Table 4-1).

4.2.1 Expert Consensus

Expert consensus is essential in supporting the information and data interpretation derived from the other approaches. It provides a balanced and comprehensive assessment of all of the information and promotes the optimum criteria when properly done. A panel of experts is assembled before any other steps are implemented, to guide the process and to select the best methods appropriate to the region for

Case Study: Selection of Candidate Reference Lakes

Florida has nearly 8,000 natural lakes larger than 10 acres. Owing to Florida's wet climate, flat topography, and abundant karst-dominated geomorphology, depressions are abundant and filled with water. In the process of developing bioassessment and biocriteria for Florida lakes, the Florida Department of Environmental Protection enlisted the help of USEPA geographers and academic limnologists to delineate lake ecoregions for the state. Forty-seven lake regions were identified (USEPA 1997c). These included regions with no natural lakes (only impoundments), regions with abundant lakes of a single type, heterogeneous regions with several lake types, and regions with ephemeral marsh lakes.

Several lake types were also identified including: sand ridge lakes, solution lakes, swamp lakes, riverine flowage lakes, marsh lakes, and others. After the lake regions had been identified, candidate reference lakes were selected in each region. Candidate reference lakes are representative and relatively least impacted within the lake region. In regions where all lakes are impacted (for example, the rapidly urbanizing area around Orlando, Florida), candidate reference lakes are those that are least impacted relative to the regional norm. Biologists and limnologists with regional and local expertise selected the candidate reference lakes. Following selection, candidate lakes were surveyed to determine lake type and to confirm that they were relatively least impacted. Reference sites were selected from the candidate sites and a full biological survey of the reference sites was conducted during Florida's lake index period (late summer/fall).

characterizing reference conditions. The panel should consist of skilled aquatic biologists, physical scientists, fisheries biologists, and natural resources managers.

In significantly disrupted areas where no candidate reference sites are acceptable, a form of this expert consensus is a workable alternative to establish reference expectations. Three or four biologists can be convened for each assemblage to be used in the assessment. Each expert should be familiar with the lakes of the region. Based on their collective expertise, they are asked to develop a description of the assemblage to be expected if the lakes were relatively unimpacted. This description, developed by consensus, will necessarily be more qualitative than quantitative, but will allow development of metrics and metric scoring.

The recommended empirical approach is to use a population of reference lakes to establish conditions that will be used to identify and calibrate metrics.

4.2.2 Biological Survey

The recommended empirical approach is to use a population of reference lakes to establish conditions that will be used to identify and calibrate metrics. Pairwise comparison of two lakes leads to the trivial conclusion that they are different (Hurlbert 1984). All monitoring sites, reference or impaired, can vary over time and space for natural reasons. A central measure from a composite of several reference sites is used to base expectations to account for natural variability and uncertainty. Statistically, this means that the status of a lake is judged by comparing the lake (the "test site") to a population of reference sites. In hypothesis-testing terminology, the null hypothesis examines whether the test lake is a member of the population of reference sites.

A critical requirement for the use of reference conditions in biocriteria is the USEPA antidegradation policy, which protects against incremental deterioration of waterbodies and reference conditions. An observed downward trend in reference sites cannot be used to justify relaxing reference expectations, reference conditions, and the associated biological criteria. Once established, biocriteria may only be refined in a positive direction in response to improved conditions.

To characterize reference conditions, surveys of both reference sites and known impaired sites are made for both biota and physical habitat. These data are needed to determine gradients of conditions (from best to impaired) for the purpose of measurement calibration and dis-

Table 4-1. Comparison of elements for characterizing reference conditions.

	Expert Consensus	Biological Survey	Paleolimnology	Historical Data	Predictive Models
Strengths	Guides and reviews other procedures May be used alone. Relatively inexpensive. Common sense and experience can be incorporated.	Yields obtainable, best current status. Any assemblages deemed important can be used. Two methods: - selected reference sites - best of abmient conditions	Yields historical time series for assemblages of diatoms, chrysophytes, and, to a lesser extent, some crustaceans and some insects. Can infer water quality.	Yields actual historical information on status. Inexpensive to obtain.	When data are insufficient. Works well for water quality.
Weaknesses	Qualitative descriptions of "ideal" assemblages. Might be unrealistic and not representative of a best attainable potential. Experts might have strong biases.	Even best sites subject to human impacts. Degraded sites might lower subsequent biocriteria.	Preservation of fish, invertebrates, macrophytes, and non-diatom algae is poor. Studies may require complex data analysis and interpretation by experts. Adequate sediment record may not exist in reservoirs.	Data might be limited. Studies likely were designed for different purposes; data might be inappropriate. Human impacts present in historical times were sometimes severe.	Extrapolation beyond known data and relationships is risky. Can be expensive.

crimination. The raw data must be evaluated within the ecological context (waterbody type and size, season, geographic location, and other elements) that defines what is expected for similar waterbodies.

Candidate metrics are developed from the key biological attributes, and the effects of stressors on specific metrics must be understood (USEPA 1996a). Those measurements that have a monotonic response to a gradient of conditions

There are two primary approaches for selecting or determining reference conditions using data from surveyed sites. The first approach uses selected best-quality sites as the basis for determining reference conditions. The second approach does not use reference sites, but draws its reference conditions directly from those found in a sample of many lakes of varying quality.

1. Selection of reference sites based on a prior definition of reference site criteria— *This approach is used when a sufficient number of lakes exist that are minimally impacted. Since nearly all lakes are affected by human activities to some degree, the lakes need not be pristine or unimpacted, but the level of impact must be minimal relative to lakes in the region. Reference sites are selected using local expert knowledge on candidate sites, mapped information such as land use and roads, and other existing data bases.*
2. Determination of reference conditions based on the best conditions found in a representative sample of lakes within a class— *This approach is used when few appropriate reference sites exist or when they cannot be suitably defined. A number of lakes within the class are surveyed, and the best conditions for each measurement are determined from the entire sample of lakes. These best conditions are then used as the reference for biological assessment within that lake class. This is the preferred approach for many large reservoirs and some exceptionally large or unusual lakes, where there are few other lakes of that class.*

(from unimpaired to heavily impaired) will be the best candidates for assessing biological impairment. Therefore, ambient sites other than reference sites should be surveyed as part of the data base. Selection and confirmation of the measurements must address the ability to differentiate between impaired and unimpaired sites.

Minimally Impaired Reference Sites

Reference sites must be carefully selected because they will be used as a benchmark against which test sites will be compared. The conditions at reference sites should represent the best range of minimally impaired conditions that can be achieved by similar lakes within the region. The reference sites must be representative of the region, and relatively least impacted compared to other lakes of the regions.

Sites that are undisturbed by human activities are ideal reference sites. However, land use practices and atmospheric pollution have so altered the landscape and quality of water resources nationally that truly undisturbed sites are rarely available. In fact, it can be argued that no unimpaired sites exist. Therefore, a criterion of "minimally impaired" must be used to determine the selection of reference sites. In regions where minimally impaired sites are significantly degraded, the search for suitable sites should be extended over a wider area.

Stringent criteria might require using park or preserve areas for reference lakes. Criteria for reference lakes will also pertain to the condition of the watershed, as well as the lake itself. If relatively unimpaired conditions do not occur in the region, the selection process could be modified to be more realistic and reflect attainable goals, such as the following:

Land use and natural vegetation—Natural vegetation has a positive effect on water quality and hydrological response of streams. Reference lakes should have at least some percentage of the watershed in natural vegetation.

Riparian zones—Zones of natural vegetation alongside the lakeshore and streams stabilize shorelines from erosion and contribute to the aquatic food source through allochthonous input. They also reduce nonpoint pollution by absorbing and neutralizing nutrients and contaminants. Watersheds of reference lakes should have at least some natural riparian zones regardless of land use.

Best management practices—Urban, industrial, suburban, and agricultural nonpoint source pollution can be reduced with successful best management practices (BMPs). Watersheds of reference lakes should have BMPs in place provided that the efficacy of the BMPs has been demonstrated.

Discharges—Absence or minimal level of permitted discharges (NPDES) into surface waters.

Management—Management actions, such as extreme water level fluctuations for hydropower or flood control, can significantly influence lake biota. Reference lakes should be only minimally impacted by management activities.

Predefined reference conditions for lakes have been used in Minnesota to determine ambient phosphorus criteria (Heiskary 1989). Maine uses a similar approach in regulating the water quality of streams and uses a reference standard of aquatic life as naturally occurs (Davies et al. 1993).

If a fixed definition of reference condition is deemed to be overly restrictive or an impractical ideal, then an empirical working definition is an alternative. For example, because natural conditions for reservoirs cannot be defined, the best existing conditions are used instead. This approach is also useful in ecoregions with little or no contiguous stands of natural vegetation

If all lakes in a region are significantly altered, it might not be possible to characterize reference conditions from ecoregional data. In this case, an alternative would be to use lakes from neighboring regions as reference sites if those lakes are deemed acceptable, by professional judgment, with respect to impact and overall comparability to the lakes of the affected region. This is one of the reasons why USEPA encourages interstate cooperation in monitoring and biocriteria development. If lakes from nearby regions cannot reasonably be considered reference sites, then reference conditions must be predicted or inferred from other information, including models and historical data. In designing such an approach, the consensus of a panel of regional experts helps ensure an objective and rational design.

remaining, such as in the agricultural Midwest. Choosing the best sites requires at least a representative survey (or better, a census) of lake watershed variables in the ecoregion. Individual lakes with the best conditions, such as the greatest percentage of forest or natural vegetation, the lowest percentages of agricultural and urban land use, etc., are chosen as reference sites.

Without antidegradation safeguards, the best available approach might allow continual deterioration. For example, construction and development in a lake watershed that is one of the "best" in a region might cause biological degradation of the lake. If the set of "best" lakes in the ecoregion have suffered similar degradation, they might still be the reference sites, but the new reference condition will be degraded relative to its earlier state. For example, Maine has a antidegradation policy that requires that lakes remain stable or improve in trophic state (Courtemanch et al. 1989, NALMS 1992). An effective antidegradation policy can promote continually improving conditions.

The selected reference lakes should be representative of each of the classes, and a sufficient number of lakes are then sampled to enable characterization of each class. A general "rule of thumb" for optimal sample size is 10-30 lakes per class, and each lake is a sampling unit (see Chapter 9 for estimating power and sample size). In regions where all lakes are impacted, the 10 to 30 relatively least impacted lakes of each class (e.g., ecoregion) are sampled, where "best" is determined by least anthropogenic disturbance or impacts, but not by most desirable biota. In regions where the population of unimpaired reference lakes is large, a stratified random sampling scheme (lakes in each class selected randomly) will yield an unbiased estimation of reference conditions.

"Stressed Reference Sites"—Effective metrics respond to environmental degradation and allow discrimination of impaired sites from the reference expectations. Metrics that do not respond are not useful in bioassessment. Response is determined by sampling a set of stressed sites in the same way as the reference sites—in effect, sampling a set of "stressed reference" sites. Lakes with known problems, such as nutrient loading, thermal pollution, toxic sediments, or urban land use, are good candidates for "stressed reference" sites. There should be several in each class or lake ecoregion for adequate tests of metric responses. Because impaired lakes are frequently objects of monitoring by natural resource agencies, data might already exist to test the biological metrics. However, the sampling methods for reference and impaired lakes should be comparable.

Sampling and Data Analysis—One or more of the recommended tiers of biological assemblages are sampled and identified. It is imperative that reference sampling include all assemblages that will be used in operational sampling and assessment. Sampling methods are described in Chapters 4 and 5; data analysis is described in Chapter 6.

Reference Conditions from Distributions of Biological Metrics

If sufficient minimally impaired reference sites do not exist or cannot be found, reference conditions can be selected from an entire population of sites. This approach is especially relevant for human-made impoundments and reservoirs, where no least-impaired systems exist, as well as for resources subject to strong and relatively uniform human impacts, such as lakes in large urbanized areas or in heavily agricultural regions. The approach was developed by Karr et al. (1986) for the Index of Biotic Integrity (IBI). It has since been applied to estuary assessment (Engle et al. 1994, Ranasinghe et al. 1994) and reservoir assessment (TVA 1994).

A representative sample of lakes is taken from the entire population. Sites that are known to be severely impaired may be excluded from the sample, if desired. The population distribution of each biological metric (Chapter 5) is determined, and the 95th percentile of each metric is taken as its reference value. The range from the minimum possible value (usually 0) to the reference value is trisected, and values in the top third of the trisected range are taken to be similar to reference conditions. Scoring of metrics is explained more fully in Chapter 6.

A central assumption of the population approach is that at least some sites in the population of lakes are in good condition, which will be reflected in the highest scores of the individual metrics. Because there is no independent definition of reference (independent of biological status), reference conditions defined in this way must be taken as interim

and subject to future reinterpretation. Again, antidegradation safeguards must be in place to prevent deterioration of the reference standard. Periodic examination of the reference standards for trends can detect deterioration or improvement. Strictly speaking, the distributional approach is circular because the reference biological conditions are characterized as the best of existing biological conditions, without consideration of impacts. This is necessary when reference criteria cannot be defined *a priori*, or when all lakes under consideration are equally impaired. The object of the method is to develop a measurement standard for assessment of lakes. Its validity must then rest on external confirmation of the response of metrics to stressors, usually from published or other independent studies.

Following the initial classification of the lakes in a region, biota are surveyed to determine those aspects of the classification that are relevant in explaining biological variability among lakes. The objective of the survey is to determine the final classification and to characterize the biota of each of the lake classes. Analysis of biological data includes testing classes developed in the initial classification, as well as aggregating classes as necessary to obtain a parsimonious classifica-

Case Study: Ecoregional Classification of Minnesota Lakes

Minnesota has over 12,000 lakes spread across diverse geographic areas. Previous studies had shown distinct regional patterns in lake productivity associated with regional differences in geology, vegetation, hydrology and land use (Heiskary and Wilson 1989). Four of the seven ecoregions in Minnesota (Omernik 1987) contain 98 percent of the lakes. These are the Northern Lakes and Forest (NLF), North Central Hardwood Forest (NCHF), Northern Glaciated Plains (NGP), and Western Corn Belt Plains (WCBP) (Figure 4-1). Minnesota has used environmental differences along with regional differences in lake uses to develop ecoregion-based frameworks for data analysis, developing monitoring strategies, assessing use patterns, and developing phosphorus goals and criteria (Heiskary 1989).

The Minnesota Pollution Control Agency (MPCA) and several other groups collected data on chlorophyll a concentrations and several water quality parameters (total phosphorus, total nitrogen, and Secchi transparency) in 90 reference lakes between 1985 and 1987. Secchi transparency data were collected mostly by volunteer participants in the Citizen Lake Monitoring Program. Reference lakes were chosen to represent minimally impacted sites within each ecoregion. Criteria used in selecting reference lakes included maximum depth, surface are, fishery classification, and recommendations from Minnesota Department of Natural Resources (DNR) (Heiskary and Wilson 1989). Lake morphometry had previously been examined. In addition to the reference lake data base, MPCA examined a statewide data base containing data collected by these same groups on approximately 1,400 lakes from 1977 to 1987.

Differences in morphology, chlorophyll a concentrations, total phosphorus, total nitrogen, and Secchi transparency were found among the 4 ecoregions in both studies. Lakes in the 2 forested ecoregions (NLF and NCHF) are deeper (median maximum depth 11 m) with slightly smaller surface areas (40 to 280 ha) than those in the plains ecoregions (NGP and WCBP). Lakes in the 2 plains ecoregions were typically shallow (median maximum depth 3 m) with larger surface areas (60 to 300 ha).

Box-and-whisker plots for chlorophyll a and water quality measurements in the reference lake study paralleled the morphological differences seen among the ecoregions (Heiskary and Wilson 1989). The 2 plains ecoregions had significantly higher chlorophyll a levels than either of the 2 forested ecoregions (Figure 4-2). Another biological parameter, ecological classification, also differs among the ecoregions. Ecological classification refers to the type of fish assemblage likely to be present if no fisheries management occurred. In the forested ecoregions, 37 percent to 48 percent of the lakes are classified as "basspanfish walleye" (Heiskary et al. 1987). Additionally, only the 2 forest ecoregions support any lakes classified as "walleye." Results of the statewide data base analysis showed these same trends. The results of these 2 data base analyses support the use of ecoregions in developing frameworks for data analysis, monitoring strategies, assessing use patterns, and developing phosphorus goals and criteria.

Case Study: Ecoregional Classification of Minnesota Lakes (continued)

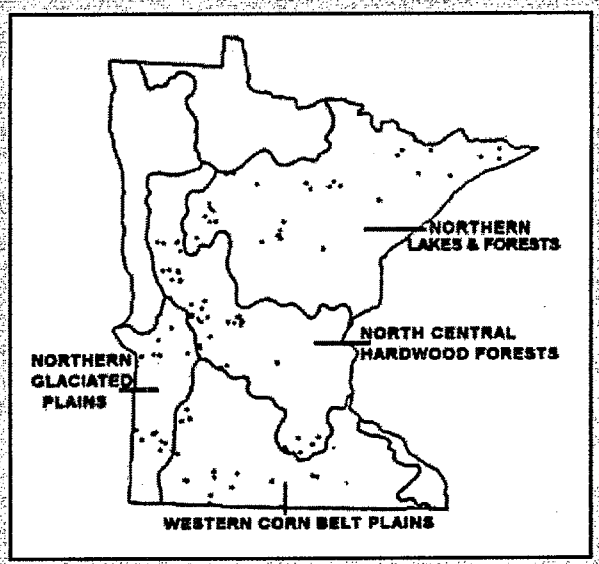


Figure 4-1. Minnesota ecoregions and sampled lakes. From Heiskary 1989.

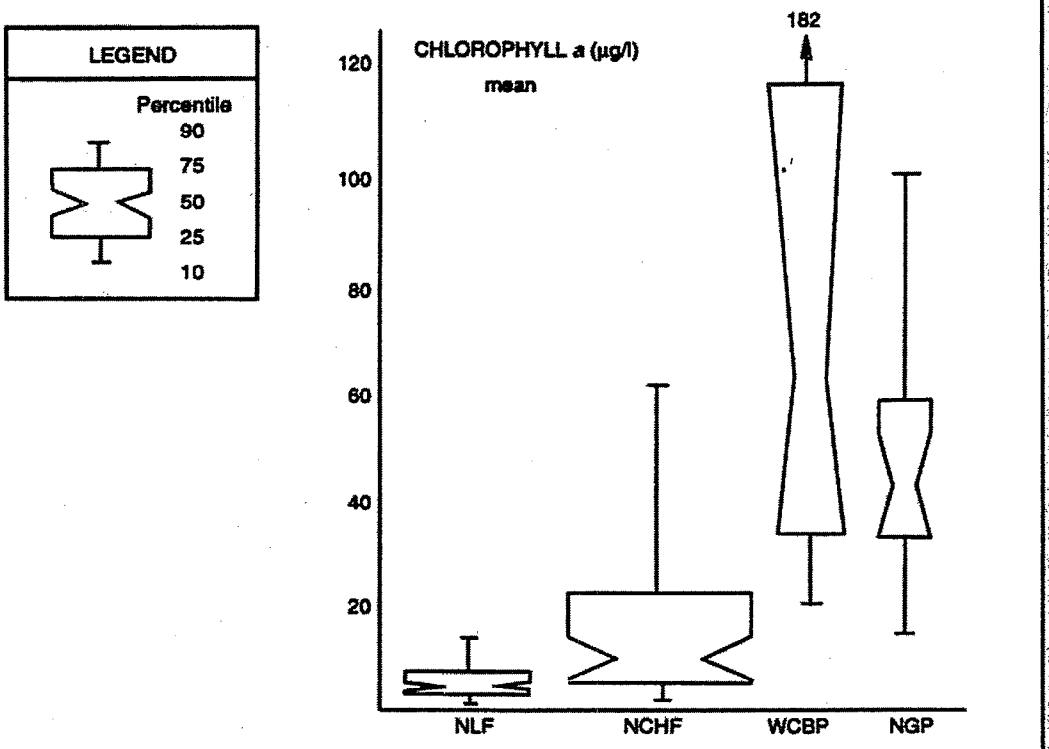


Figure 4-2. Chlorophyll *a* concentration of Minnesota reference lakes by ecoregion. Notches in box plots represent 95 percent confidence intervals of the medians. NLF = Northren Lakes and Forests; NCHF = North Central Hardwood Forests: WCBP = Western Corn Belt Plains; NGP = Northern Glaciated Plains. (From Heiskary 1989.)

tion that accounts for the greatest amount of biological variability. The survey may use existing data, although a new survey allows careful selection of reference sites representative of each of the classes of lakes.

4.2.3 Paleolimnology

An alternative to characterizing present-day reference conditions is to estimate historic or prehistoric pristine conditions. In many lakes, presettlement conditions can be inferred from fossil diatoms, chrysophytes, midge head capsules, cladoceran carapaces, and other remains preserved in lake sediments (e.g., Charles et al. 1994, Dixit et al. 1992). Fossil diatoms are established indicators of historical lake alkalinity, salinity, and trophic state (e.g., Hall and Smol 1992). Diatom frustules, composed of silica, are typically well preserved in lake sediments and easy to identify. However, remains of other organisms are problematic because of incomplete preservation.

Paleolimnological investigations can be performed in lakes in which identifiable remains are preserved, and the sediments can be dated to the period of interest (Charles et al. 1994). In some lakes, sediments are subject to scouring, resuspension, or periodic drying and are not suitable for coring. Most lakes have a quiescent depositional area in the deepest profundal waters, and these sediments receive material from both pelagic and littoral zones, as well as from the surrounding watershed. Reservoirs meeting the depositional criteria can also be analyzed in this way, yielding a history of the reservoir. However, historical conditions in a reservoir might or might not be a desired reference condition.

Paleolimnology can identify presettlement conditions (reference conditions) for an individual lake or for many lakes within a region (e.g., Cumming et al. 1992).

Design of paleolimnological studies to determine reference conditions can range from basic to complex. The simplest procedure is to analyze only the top and bottom of a sediment core, and to make a comparison of assemblages to determine if there has been a significant shift in taxa composition. If there is little difference, then there has probably been relatively little change in major ecological characteristics in the lake. If there are significant differences, then further investigation may be warranted, including quantitative inference of past water chemistry conditions (Charles and Smol 1994). The more informative approach is to analyze several sediment intervals from a sediment core that has been dated (usually Pb-210), and infer specific past conditions. This design leads to understanding of the magnitude, rate, and timing of change and can be related to specific watershed or in-lake events.

Using paleolimnology to characterize lake reference conditions requires selection of a time period for the reference. In general, the time period should be as close to the present as possible when anthropogenic impacts on the lakes were minimal. If there is concern that background conditions may have varied substantially, a few to several presettlement time periods could be analyzed to determine natural variability. In most cases this variability is relatively small compared with changes following European settlement.

The greatest advantage of paleolimnology is that a sample of reference sites can be selected without regard to present conditions in the lakes. Thus, there is usually no need to select "least-impaired" lakes because nearly all lakes in the selected reference period are least-impaired by definition. Reference sites are selected such that each lake class has at least 5 to 10 representative lakes. Reference sites should be representative of their respective class. Transitional, exceptional, or uncertain lakes should not be included in the reference sample.

The population approach to defining reference conditions means that a single site is never taken as a representative reference for an entire class. Similarly, the condition at only 1 time period of a single lake may not represent a reference for its present condition. Ecosystems are not constant in time, even in the absence of disturbance, and the condition of a single lake is likely to change in the course of a century. Therefore, samples of past conditions at several points in time are more likely to characterize reference conditions than a single sample.

Sampling and Data Analysis—Sediment diatoms are the recommended assemblage for paleolimnological determination of reference

conditions because preservation of frustules is excellent and identification is based solely on the frustules. Other assemblages (e.g., cladocerans, midges) are not recommended at this time because preservation is incomplete and identification of fragments is problematic. Cores are taken from the representative lakes and analyzed as described in Appendix C.

4.2.4 Historical Data

Some lakes have extensive historical data bases from the early to mid-20th century, typically on water quality, diatoms, zooplankton, or fish. However, historical data may not represent undisturbed conditions, and the biological data and auxiliary historical information should be examined carefully to ensure that the data actually represent conditions better than at present. Cultural eutrophication has occurred since neolithic peoples first settled on lakeshores, and in many American waterbodies cultural eutrophication was most pronounced in the 1950s and 1960s.

Historical data might not always be representative of lakes in a region because the lakes were selected for special reasons (e.g., unique lakes, near laboratory, site of water intake, etc.). Universities, municipal water supply departments, and other agencies are often good sources of long-term lake water quality data. It might be possible to augment present-day reference site data with historical data.

4.2.5 Modeling Approaches

Several modeling approaches can be used, including mathematical models (logical constructs following from first principles and assumptions), statistical models (built from observed relationships between variables), or a combination of the 2. The degree of complexity of mathematical models to predict reference conditions is potentially unlimited, with attendant increased costs and loss of predictive ability as complexity increases (Peters 1991). Mathematical models are complex and untestable hypotheses (Oreskes et al. 1994, Peters 1991). Nevertheless, models to predict water quality in rivers and reservoirs from first principles of physics and chemistry have been quite successful (e.g., Kennedy and Walker 1990).

The greatest advantage of paleolimnology is that a sample of reference sites can be selected without regard to present conditions in the lakes.

Statistical models can be fairly simple in formulation, such as the Vollenweider model, the Morphoedaphic Index, and others (Vighi and Chiaudani 1985, Vollenweider 1975, Mazumder 1994), to predict trophic status, but they require a sufficiently large data base to develop predictive relationships. If enough data exist to construct a statistical model, it is likely that there are lakes that can serve as reference sites.

RESERVOIR

Case Study: Reference Conditions - TVA Reservoirs

(For TVA's reservoir bioassessment, see Chapter 3.)

It was not possible to use the well-accepted approach of using least-impacted reference sites to determine characteristics or expectations of a reservoir since they are artificial systems. Other approaches must be used such as historical or preimpoundment conditions, predictive models, best observed conditions, or professional judgment. Preimpoundment conditions are clearly inappropriate. For the most part, models are of limited value for a large variety of indicators because of such great spatial and temporal variations within and between reservoirs. This leaves best observed conditions or professional judgement as the most viable alternatives for establishing appropriate reference conditions or expectations for reservoirs. TVA's experience has found use of best observed conditions using professional judgement as the best approach.

In using best observed conditions one assumes that, for the group of reservoirs to be compared, the range of observed values represents the range of expected conditions from good to poor for each community characteristic or metric included in the evaluation. Separation of reservoirs into appropriate classes was a critical step in developing reference conditions.

For dissolved oxygen (DO) and sediment quality, best observed conditions were not used; instead, ideal conditions were expected. That is, poor DO is unacceptable regardless of type of reservoir or dam operation. Sediments should not have high concentrations of metals, should have no or very low concentrations of pesticides, and should not pose a toxic threat to biota. In this situation, there is no need for classification because the same conditions are desired for all reservoirs.

For chlorophyll, benthos, and fish, the best observed conditions approach was used. For these, reservoirs were categorized because the same conditions do not exist for all reservoirs. The classification scheme that evolved for chlorophyll is actually a combination of two approaches: examination of the "natural" nutrient level in the watershed; and a conceptual/subjective decision as to the concentrations indicative of good, fair, and poor conditions. Two classes of reservoirs were developed: reservoirs draining nutrient-poor watersheds, primarily those in the Blue Ridge Ecoregion; and the mainstream reservoirs with their remaining tributary reservoirs.

For the benthic macroinvertebrate and fish assemblages, reservoirs were divided into four classes:

- *Reservoirs on the Tennessee River plus two navigable reservoirs on tributaries to the Tennessee River; this group of reservoirs has relatively short retention times and little winter drawdown.*
- *Reservoirs in the Blue Ridge Ecoregion.*
- *Reservoirs in the Ridge and Valley Ecoregion.*
- *Reservoirs in the Interior Plateau Ecoregion.*

In This Chapter...

- Watershed Activities
- In-Lake Water Quality
- Shorezone and Littoral Characteristics

Chapter 5

Habitat Measurement

Habitat measurement is used to assess the impacts of habitat on biota, and hence on the interpretation of changes in biota. Habitat must be taken into account to make accurate comparisons between ambient and reference conditions and to determine whether habitat might be a cause of impaired biota.

Human activities modify the watershed, with consequent effects on lake physicochemical and biological processes. Agricultural and urban land use affect nutrient, contaminant, and sediment loadings; and shorezone housing development can have a disproportionate influence on nutrient loadings compared with more distant parts of a lake watershed (Dillon et al. 1994). Shorezone development can also extend into the lake littoral zone with construction of docks, revetments, riprap, often leading to destruction of littoral wetlands and macrophytes.

The habitat experienced by aquatic organisms consists of the water and the substrate, including structure and constituent chemicals. For the purposes of this protocol, water quality is a component of habitat. In-lake habitat includes both the physical and chemical environment experienced by the biota, and is, in turn, influenced by the watershed through runoff and loadings. Habitat measurement seeks to identify the physical and chemical characteristics of the lake habitat—both natural and anthropogenic—that affect the biota of the lake.

Habitat measurement is used to assess the impacts of habitat on biota, and hence on the interpretation of changes in biota.

Habitat measurement, consisting of both watershed and in-lake observations, has two purposes. First, it helps in placing a lake into a category determined by a classification scheme. Second, it can help identify anthropogenic disturbances and exposure that might be responsible for biological degradation. Habitat measurement thus comprises two kinds of variables:

Classification variables—Those attributes intrinsic to the system and relatively unaffected by human activities (e.g., geology, soils, lake and watershed morphology).

Assessment variables—Those attributes which either are direct measures of human activity (e.g., land use, discharges) or are influenced by human activity (e.g., most water quality variables).

The classification variables are those which are not affected by human influence, and are primarily measures of the morphology and geology of the lake and watershed. The classification variables assist in placing the lake into one of the categories for which reference conditions have been determined. It is then possible to determine the deviation of conditions in the test lake from reference conditions, for both habitat and biological indicators.

The purpose of examining watershed parameters is to assist in classifying a lake.

Several habitat parameters are obtained or estimated from existing sources of information such as maps and Geographic Information Systems (GIS). The parameters include lake area, depth, shoreline length, watershed area, watershed slope, soil types, geology, and watershed land use.

The habitat measurement component of the field sampling program consists of in-lake physical and chemical measurements, as well as a shorezone habitat survey. The shorezone survey is based on the Environmental Monitoring and Assessment Program (EMAP) lake habitat assessment (USEPA 1994a, USEPA 1994b, USEPA 1993a).

5.1 WATERSHED HABITAT

5.1.1 Measurements

The purpose of examining watershed parameters is to assist in classifying a lake and to determine whether watershed conditions might account for observed biological status. A number of human practices in lake watersheds affect lake habitat through sediment loading, nutrient loading, contaminant loading, hydrologic changes, and direct habitat alteration (e.g., removal of wetlands). Any one human activity can influence several loading rates. For example, livestock management practices can affect both nutrient and sediment loads. Watershed parameters include both classification and assessment variables (Table 5-1). Most measures of morphology and land use can be obtained from USGS, state, or county data bases.

5.1.2 Watershed Metrics

Discharges—Data from permitted discharges can be used to develop direct estimates of point-source loadings into receiving waters, and they take into account the effects of sewage diversions and implemented control technologies. However, discharges cannot account for nonpoint sources.

Watershed Area—The quantity of runoff entering a lake is directly affected by the lakes watershed area. The ratio of lake watershed area to lake surface area affects sediment and nutrient loadings and retention time. Reservoirs with a small ratio are better able to support sport fish populations (Hill 1986). The ratio is especially important for reservoirs and flowage lakes, where its value can vary widely.

Land Use—Water quality, especially nutrient concentrations and turbidity, is strongly associated with land use. The most important land use variables are urban, agricultural, and forest land use, as percent of the watershed area. Also important is watershed road density (length per area), which can be an excellent predictor of trophic variables and chloride concentration (USEPA 1993a). More detailed breakdowns of land use classes (e.g., high-density urban, transportation, pasture, row crops, etc.) can be estimated for diagnostic investigation.

A detailed nonpoint source evaluation might be called for if more than one land use type appears to be a probable cause for impairment. A standard screening procedure (Schueler 1987) can be applied to estimate sediments, nutrients, and contaminants from both urban and nonurban sources. The screening procedure allows identification of the primary likely sources of impairment and hence a preliminary ranking of potential sources.

The land use variables are tabulated on a watershed-wide basis. This approach does not take into account the effects of distance from the receiving waters, riparian buffers, or best management practices (BMPs). Runoff and pollution of surface waters from agricultural land are highly variable, depending on slope, soil erosivity, tillage practices, distribution of rainfall, and the presence of riparian buffers and hedgerows (Schueler 1987). Taking into account riparian buffers and BMPs, together

with other watershed influences, would require a comprehensive runoff and loading model, and is beyond the scope of this guidance.

Population Density and Related Measurements—Nonagricultural pollution is the product of people and their activities; hence, population density is an excellent predictor of pollutant loadings. Population density is also strongly correlated with urban land use and discharges; therefore, simultaneous assessment with these collinear variables should be done with caution. Population density might be a more accurate indicator of total human activity than is land use, because population estimates are updated more frequently than land use data. The variables that most directly affect lake quality are discharges and the watershed impervious area. Nevertheless, population density may be a better single measurement.

5.2 IN-LAKE HABITAT

5.2.1 Measurements

Physical-chemical habitat measurement comprises several common measures of lake water quality and can point to water quality problems that are not observable at the coarser resolution of the entire watershed. It can also provide additional evidence for potential causes identified from the watershed or shoreline assessment. Physical and chemical parameters and the measurements derived from them are listed in Table 5-2.

Secchi Depth—Secchi depth, which has a long history as a lake assessment variable, is a simple and reliable measure of light transmittance and turbidity. It is used in various trophic indices, including Carlson's Trophic State Index (TSI) (Carlson 1977).

Nutrients—Water quality measurements can form the basis of several measurements (Table 5-2). Total phosphorus concentration forms part of Carlson's TSI (Carlson 1977) and is an important predictor of lake productivity in north temperate lakes (Vollenweider 1975). The nitrogen-to-phosphorus (N:P) ratio is used to predict the likelihood of cyanobacteria blooms (e.g., Smith 1983). Calculation of trophic state indices is given in Section 7.2.3.

Dissolved Oxygen—Dissolved oxygen (DO) is necessary for aquatic life, and most state water quality regulations include a standard for dissolved oxygen, usually expressed as the maximum amount of time that DO is allowed to fall below a critical value (typically 4 or 5mg/L). Several measurements have been developed for DO, including:

Population density is an excellent predictor of pollutant loadings.

- Index period DO measurement near bottom of lake.
- Depth from the surface at which DO falls below a threshold value (oxycline) (Scott et al. 1991).
- Annual or seasonal minimum value in hypolimnion or epilimnion.
- Annual or seasonal mean value in hypolimnion or epilimnion.
- Annual or seasonal percent time below a threshold DO value at the bottom of the lake (USEPA 1993a).
- Annual or seasonal mean water volume or percent of total volume below a threshold DO value (Dycus and Meinert 1992).

The first 2 measurements require only a single DO profile. Depth of the oxycline might be the most useful single-point DO measurement of a waterbody (Scott et al. 1991), provided that the observation is made when hypoxia is at its maximum annual extent (usually late summer). The remaining 4 measurements all require regular observations during a year or an index season. In general, estimates of time or volume below a threshold value are more precise and accurate than estimates of minimum values (USEPA 1993a).

5.3 SHOREZONE AND LITTORAL HABITAT

5.3.1 Measurements

The shorezone habitat assessment is important for identifying potential causes of impairment because many lakes are impacted by develop-

ment and land use on the shore. Because the lakeshore is the part of the watershed closest to the lake, shorezone land use has the largest potential impact on lake biological integrity. The shorezone assessment procedure is the same as that for watershed evaluation: shorezone habitat variables are compared to reference conditions and, if significantly different, are identified as probable causes of biological impairment.

EMAP Surface Waters has developed an extensive shorezone and littoral survey methodology to characterize riparian and littoral habitat (USEPA 1994a, USEPA 1993a, USEPA 1991e). The index period is late summer when vegetation is at its annual maximum. The riparian characterization consists of estimates of dominance of vegetation in canopy, understory, and groundcover; substrate type; bank angle; and dominance of human features (buildings, lawns, cultivation, etc.). Littoral characterization is done at a 10m distance from shore and includes depth, surface film, substrate, macrophyte cover, fish cover, and a summary habitat classification (USEPA 1993a). The shore of each lake is surveyed at 10 sites, and the frequency of disturbance is estimated for each lake from the survey data.

The shorezone and littoral assessment for lake biological surveys presented here is a modification of the EMAP shorezone assessment (Table 5-3) (USEPA 1994a).

Table 5-1. Watershed and basin habitat measurement and metrics.

	Measurements	Additional Metrics	Calculation	Indicator
Lake and Basin Morphology	Watershed drainage area.		Estimated from map contours.	Hydrology
	Lake surface area.		Map	
		Watershed: Lake area ratio.	Watershed area/lake area.	Sediment, nutrients.
	Shoreline length.	Shoreline development ratio.		Effect of riparian zone.
	Lake volume.		Estimated from Basin contours.	
	Maximum depth.		Measurement	Stratification potential.
		Mean depth.	Volume/surface area.	
		Mean basin slope.		
	Lake outflow.	Retention time.	Volume/outflow.	Eutrophication potential.
Land Use	% forest or natural vegetation.			Sediment, nutrients, hydrology.
	% agriculture.		GIS data base.	Sediment, nutrients, contaminants.
	% urban and residential.			Sediment, nutrients, contaminants, hydrology.
		Watershed impervious surface.	Estimate from land use.	Sediment, contaminants, hydrology.
	Population density.		U.S. Census, state or county.	Sediment, nutrients, contaminants, hydrology.
	Discharges		USEPA NPDES data base.	Nutrients, contaminants.
	Road density.		Maps, GIS.	Sediment, contaminants, hydrology.

Table 5-2. Physical and chemical measurements and metrics.

Measurements	Metrics	Calculation	Indicator
T Profile	Epilimnion temperature.	Mean from temperature profile.	
	Hypolimnion temperature.	Mean from temperature profile.	
	Metalimnion depth.	Inflection point of temperature profile.	
DO Profile	Epilimnion DO.	Mean from DO profile.	
	Hypolimnion DO.	Mean from DO profile.	
	Oxycline depth.	Depth at which DO falls below 2 mg/L.	DO problems.
	Hypoxic volume.	Volume of water with DO < 2 mg/L; annual or seasonal mean.	DO problems.
Secchi Depth (SD)	TSI (SD) = 60 - 14.41 ln (SD)		Transparency
Total N	TSI (N) = 54.45 + 14.43 ln (TN)		N enrichment.
Total P	TSI (P) = 4.15 + 14.42 ln (TP)		P enrichment.
	N:P ratio.	N concentration/P concentration (molar).	Enrichment
Silica			Depletion
Acid neutralizing capacity (ANC)	ANC		Sensitivity to acidifcation.
pH	pH		Acidity
Total Dissolved Solids (TDS)	TDS		Dissolved minerals.

5.3.2 Shorezone Metrics

Most shorezone measurements are means of the littoral and shorezone habitat metric values. The shorezone and littoral cover measurements are expressed as the mean of the values of all transects. The human influence measurements are different because they are based on presence or absence observations within the transects. These measurements are weighted, with each present observation receiving a score of 1 and each "adjacent" observation receiving a score of 1/2. The human influence score in each category is the mean of all transects. It is in the range of 0-1, with 0 reflecting no influence and 1 indicating that the influence (e.g., buildings) was found in every transect.

Table 5-3. Lakeshore habitat measurements and metrics (USEPA 1994a, EMAP Internal Report).

Habitat Measurement	Mean % Cover	Indicator
Bank Measurement - Rocky (%) - Soil (%) - Vegetation (%) - Other (%)	Mean % cover from shorezone habitat transects.	Bank Stability.
Bank Erosion (0-4)	0=none 4=severe erosion	
Riparian Vegetation Measurements - Canopy - (% cover) - Understory - (% cover) - Ground Cover - (% cover)	% cover of vegetation.	Disturbance
Human Influence Measurements Buildings - In-lake structures - Roads, railroads - Agriculture - Lawn - Dump or landfill	Influence score (mean score of transects). Presence/absence.	Human Influence.

In This Chapter...

- Description of Assemblages
- Response to Stress
- Discussion of Assemblage Analysis
- Level of Sampling Effort

Chapter 6

Biological Assemblages

The proposed biological sampling methodology is tiered, ranging from a trophic state assessment to detailed biosurveys. Many of the methods are based on those used in USEPA's Clean Lakes Program and Environmental Monitoring and Assessment Program (EMAP) lakes component.

Lake surveys require sampling of biological assemblages and habitat in one or more field visits. Several of the proposed lake biosurvey measurements are made from transects extending from the shore to the sublittoral habitat, and several other measurements are made from one or more stations in the pelagic region of the lake (Chapter 7). The integrated sampling scheme combines all sampling activities carried out on the transects and includes mid-lake sampling sites for pelagic samples. The number of transects, the number of sampling sites, the assemblages sampled, and the frequency of sampling vary among the survey tiers.

The study of any group of organisms will yield information on the status of their environment. The objective in selecting assemblages for lake bioassessment was to find assemblages that:

- Are unambiguously useful for biological assessment.
- Can be sampled and interpreted in a cost-effective way.
- Are consistent with the current mix of expertise in natural resource agencies.
- Can be easily converted to a multimetric index of the assemblage.

The recommended assemblages are phytoplankton, sedimented diatoms, submerged and floating aquatic macrophytes, crustacean zooplankton, benthic macroinvertebrates, fish, and periphyton. The discussion of each assemblage includes some estimates on the level of effort required for sampling. These are intended as general guidelines. Actual time and effort involved will depend on the specific expertise and resources available to individual agencies.

Emergent vegetation is not included as an assemblage in this document because methods for emergent plants are under development by USEPA and other agencies as part of the development of wetlands bioassessment methods. Several other potential assemblages were not considered because there was little information on their utility as environmental indicators for lakes. They included benthic meiofauna, protozoa, and bacteria. Background and rationale for the selected assemblages are presented in Appendix D.

6.1 PRIMARY PRODUCERS: TROPHIC STATE ASSESSMENT

Phytoplankton are the base of most lake food webs, and fish production is linked to phytoplankton primary production (Ryder et al. 1974). Excessive nutrient and organic inputs from human activities in lakes and their watersheds lead to eutrophication, characterized by increases in phytoplankton biomass, macrophyte biomass, nuisance algae blooms, loss of water clarity from increased primary production, and loss of oxygen in bottom waters. From a human perspective, eutrophication problems might include loss of aesthetic appeal, decreases in desirable gamefish, loss of accessibility due to increased macrophyte production, and increased cost of treating drinking water.

Trophic state is assessed with 4 Trophic State Indices (TSI)—chlorophyll *a*, Secchi depth, total nitrogen, and total phosphorus (Carlson 1977, Carlson and Simpson 1996)—and with Algal Growth Potential (AGP) (i.e., nutrient availability for algal growth). The chlorophyll TSI (Table 6-1) indicates whether algal biomass is low, medium, or high; the Secchi TSI indicates if algal growth may be limited by mineral turbidity; and the nutrient TSIs can indicate excess or limiting nutrient supply.

Level of Effort
Trophic state assessment is relatively inexpensive. Sample collection requires approximately 10 minutes on station and can be done by a single person. Filtration of chlorophyll samples requires another 10 to 15 minutes in the field. Chlorophyll, nutrients, and other water quality chemical analyses are standard and costs are well established in each region.

6.2 SUBMERGED MACROPHYTES

Macrophytes form an integral part of the littoral zone of many lakes, providing cover for fish and substrate for invertebrates. From a human perspective, overabundant macrophytes (or weedy conditions) can interfere with lake access by fouling equipment, interfering with recreational activities, and detracting from aesthetic appeal. A conspicuous lack of native macrophytes in habitats where they are expected to occur can result in reduced population of sport and forage fish and waterfowl (Crowder and Painter 1991). Potential macrophyte metrics are listed in Table 6-2.

Level of Effort
Submerged macrophyte analysis, including an estimate of total percent cover and identification of dominant species, requires approximately 1 to 2 hours in the field for a 300- to 500-acre lake. There is no laboratory analysis.

For the same size lake, macrophyte density or biomass measurements would require 2 to 4 hours in the field to collect samples, and to sort and weigh by species. Stem counts would likely require a longer time. Again, there is no laboratory analysis.

6.3 SEDIMENTED DIATOMS

Phytoplankton cells continually grow and die, and dead cells sink to the bottom. One group of algae, the diatoms, have shells (called frustules) made of silica (glass), which are preserved when the dead cells fall to the lake bottom. The preserved diatoms provide an integrated record

Table 6-1. Potential algal trophic state metrics.

Metric	Response to Stress
Chlorophyll concentration.	Elevated under eutrophication.
Chlorophyll TSI. TSI (Cl) = 30.6 + 9.81 ln (Chl)	Depressed under non-algal turbidity or toxicity (compared to Secchi and nutrient TSIs).
Algal growth potential (AGP).	Increases with nutrient concentration.

Table 6-2. Potential macrophyte metrics.

Metric	Response to Stress
Total vegetated area (% of littoral).	Substantially more or less than reference.
% exotics or weedy species.	More than reference.
No. of exotic species.	High
Density or biomass in vegetated areas.	Substantially more or less than reference.
No. of taxa.	Low
% dominant species (by weight).	High
Maximum depth of plant growth.	Reduced under enrichment, deeper under acidification.

of the diatom assemblage in the lake. A sample of the top 1 to 2cm of lake sediment contains a representative sample of diatoms from the most recent 1 to 3 years. Sedimented diatoms can be sampled at any time and will always yield a sample representative of the most recent years. Potential sedimented diatom metrics are listed in Table 6-3.

Level of Effort
Sedimented diatoms are sampled rapidly in the field. A sample requires approximately 1 hour to prepare and 2 to 4 hours to count and identify 300 to 500 cells.

6.4 BENTHIC MACROINVERTEBRATES

Benthic macroinvertebrates are long-term indicators of environmental quality; they integrate water, sediment, and habitat qualities (USEPA 1989b, USEPA 1990d). Macroinvertebrate species have sensitive life stages that respond to stress and integrate effects of both short-term and long-term environmental stressors. Classification of benthos according to their relative sensitivity to pollution and their functional feeding group level differentiates effects on ecological health in response to organic or toxic perturbations. Potential metrics are listed in Table 6-4.

Macroinvertebrates are sampled from the predominant substrate available in the sublittoral zone. The type of sampling gear will depend on the substrate being sampled: each substrate has its own optimal sampling gear (Chapter 7).

Level of Effort
A benthic sample, consisting of several grabs, requires 2 to 4 hours in the field. Sorting,

Table 6-3. Potential sediment diatom metrics.

Metric	Response to Stress
No. of taxa.	Reduced
Diversity indices (Shannon-Weiner, Simpson's, etc.).	Reduced
% dominant taxon.	Increased
% centric diatoms.	Reduced
Pollution tolerance indices (e.g., Lange-Bertalot; Bahls 1993).	Lower score under organic pollution.
% Nitszchia and Navicula (Bahls 1993).	Increased with sedimentation.
Indicator taxa (ecological categories).	Respond to specific stressors (acidity, salts, metals, eutrophication).
Disturbance index (Dixit and Smol 1994).	Increased

Table 6-4. Potential benthic macroinvertebrates metrics.

Metric	Response to Stress
No. of taxa.	Reduced
Mean number of individuals per taxon.	Substantially lower or higher.
% contribution of dominant taxon.	Elevated
Shannon-Wiener diversity.	Reduced
% intolerant species.	Reduced
% oligochaetes.	Elevated under organic enrichment.
ETO taxa (ephemeroptera, trichoptera, odonates).	Reduced under enrichment, DO, stress.
% non-insects.	Reduced
Crustacean + mollusc taxa.	Reduced under acid stress.
% crustaceans and molluscs.	Reduced under acid stress.
Tolerance indices (e.g., HBI [Hilsenhoff 1987]; Hulbert's Lake Condition Index [LCI] [Frydenborg et al. 1995]).	Reduced
% suspension feeders.	Reduced
% shredders.	Reduced under enrichment or in very large lakes.

counting, and identifying 100 organisms to species requires approximately 4 to 6 hours in the laboratory.

6.5 FISH

Fish assemblages include species that represent a variety of trophic levels (omnivores, herbivores, insectivores, planktivores, piscivores), and that exhibit a range of tolerance to water quality or habitat degradation. Fish are long-lived and integrate short-term temporal environmental changes, and also integrate effects of lower trophic levels (e.g., primary producers and benthic macroinvertebrates); thus, fish assemblage structure is reflective of integrated environmental health. Of all biological components of lakes, fish probably receive the greatest public attention because of sport and commercial fishing and attendant concerns regarding fish production success and safety for human consumption.

Fish are the most difficult and time consuming of all assemblages to sample; are wide-ranging and might not reflect local conditions in large lakes; and are actively and intensively managed by stocking and angling. Each feasible gear type suitable for their sampling in lakes is highly selective (USEPA 1994a, USEPA 1994b). Unbiased sampling methods such as explosives, rotenone, and draining a lake are generally too destructive. Because of lake fish assemblage sampling method bias, the use of a combination of more than one gear type is recommended (Chapter 7). Potential fish metrics are provided in Table 6-5. Among the most promising measurements are indicators of fish health (external gross pathology) and fish tissue contamination.

Level of Effort

Fish populations are generally nonrandomly distributed and clumped in response to habitat variables; therefore, the choice of sampling methods and equipment, index period, and sampling frequency depend upon waterbody physical characteristics and specific study objectives. An understanding of the attributes and/or biases of sampling equipment and methods used in fish assemblage surveys is essential in order to draw valid conclusions from the data. The relative labor intensity of fish sampling techniques varies greatly, depending on the specific method chosen and the abundance and diversity of the catch. For example,

passive techniques (e.g., trap nets, gill nets, etc.) generally require a deployment and capture cycle of many hours (e.g., overnight sets) to several days; and the processing of catches from either passive or active sampling techniques may require several hours depending on local abundances and method efficiencies.

6.6 PHYTOPLANKTON ASSEMBLAGE

Phytoplankton assemblage data, consisting of taxonomic identifications and abundances (relative or absolute) can be analyzed in two ways: by determining assemblage measurements based on species structure or by performing multivariate assemblage analysis. Potential phytoplankton metrics are listed in Table 6-6.

The recommended approach is to sample the phytoplankton assemblage and to count and identify cells to order or genus. Simplified field and laboratory procedures are possible for measurements based on higher taxonomic levels such as division or order. Identification to species is considered supplemental at this time because it is not clear that the information gained represents a substantial improvement over higher levels of taxonomy.

The phytoplankton assemblage requires 4 to 10 samples during the growing season to obtain a seasonal average of the phytoplankton assemblage. The exact number will need to be determined from preliminary or existing data sets.

Level of Effort

An integrated water column sample of phytoplankton requires approximately 10 minutes to collect. Laboratory identification and counting of 300 to 500 cells requires 1 to 4 hours in the laboratory.

6.7 ZOOPLANKTON

In most lakes, zooplankton are the central trophic link between primary producers and fish. Zooplankton are ubiquitous in all lakes and are quickly and easily sampled in the field. Zooplankton species richness is reduced under chemical stresses (Baker and Christensen

Table 6-5. Potential fish metrics.

Metric	Response to Stress
No. of taxa.	Reduced
No. of sunfish species.	Reduced
No. of sucker species	Reduced
No. of intolerant species.	Reduced
% tolerant individuals.	Increased
% piscivores.	Reduced
% ominivores.	Increased
% invertivores.	Reduced
% planktivores.	Increased
Reproductive	Reduced
Composition	Reduced
Total number of individuals.	Substantially different under stress.
Fish health (pathology) - Lesions and deformations - Histopathology	Reduced under severe organic pollution or contamination.
Tissue contaminants.	Elevated under contamination (e.g., mercury, organochlorines).

1991), and abundant large *Daphnia* are associated with clear lakes with healthy sport fish populations (Mazumder 1994). Trophic structure measurements require knowledge of feeding of zooplankton species—trophic links and complexity measures require the most detailed knowledge. Potential zooplankton metrics are shown in Table 6-7.

Zooplankton are sampled with vertical or oblique tows, using a plankton net equipped with a 7:1 reducing cone (DeBernardi 1984). The recommended approach is to sample 4 to 6 times during a growing season to obtain seasonal averages.

Level of Effort
A zooplankton sample can be collected in approximately 10 to 30 minutes in the field. Identification and counting of 100 to 200 organisms requires approximately 1 to 2 hours. Six samples in a growing season per lake thus requires six trips and 6 to 12 laboratory hours.

6.8 PERIPHYTON

Periphyton, the algae growing on solid substrates (rock, wood, sediment, macrophytes), have a long history of use in bioassessment of streams (Patrick 1949). Diatoms are often the group of choice among periphytic algae. Ecology of periphyton is much like other algal assemblages: they respond to nutrient enrichment; they are cropped by grazers; and their species composition is affected by pH, metal concentrations, trace elements, and contaminants. In addition, periphyton are affected by the physical and chemical characteristics of their substrate. Like phytoplankton, periphyton are subject to changing water chemistry and seasonal succession. Several sampling periods may be necessary to characterize lake periphyton.

Whereas periphyton have been used successfully in streams (Bahls 1993, Patrick 1949), their application as lake indicators is relatively new. Measurements of periphytic diatoms (Table 6-8) have shown promise for bioassessment, based on investigation of undisturbed reference lakes in Montana (Gerritsen and Bowman 1994), but actual responses to disturbance or pollution are as yet unknown.

Level of Effort
Analysis of a periphyton sample requires 2 to 6 hours, similar to diatoms and phytoplankton.

Table 6-6. Potential phytoplankton metrics.

Metric	Response to Stress
% cyanobacteria. % greens.	Elevated under eutrophication.
% diatoms. % chrysophytes.	Depressed under eutrophication.
% Anabaena, Aphanizomenon, Microcystis. % centric diatoms (of total diatoms). % pennate diatoms (of total diatoms). % colonial greens (Volvocales). % euglenophyta. % dinoflagellates.	Blue-green algae and colonial greens elevated under eutrophicaton.
No. of taxa.	Low under stress.
Diversity	Low under stress.
% dominance.	High under stress.
Lange-Bertalot index (pollution tolerance index, Bahls 1993).	Lower value under organic pollution.
Indicator taxa (presence or percentage).	Respond to specific stressors.

Table 6-7. Potential zooplankton metrics.

Metric	Response to Stress
% large *Daphnia* (> 1 mm).	Low under planktivorous fish predation.
No. of taxa.	Reduced under contamination or stress.
% dominance.	High under stress.
Trophic structure measurements - No. of trophic links - Complexity measures - % large predators - No. of predator species	Simplified trophic structure under stress.

Case Study: Florida Metric Selection and Index Development

Of 32 potential macroinvertebrate metrics examined, 9 were selected as candidate metrics for an invertebrate index for Florida lakes. Responsive metrics are shown in Figure 6-1. Most metrics have different values among the three lake types, with sandy-bottom lakes having the greatest macroinvertebrate number of taxa (and other related metrics), the greatest proportion of OET (Odonata, Ephemeroptera, Trichoptera) organisms, and the greatest proportions of filter feeders and surface gatherers (Fig. 6-1).

Several metrics were correlated with each other. The Shannon index was strongly correlated with total taxa and with dominance. Graphic examination of the relationships among the metrics showed that the Shannon-total taxa relationship was not entirely linear, and the Shannon-dominance relationship had large and asymmetric variance in the middle of the range (Fig. 6-2). Because of the variance and nonlinearity of the relationships, all of the candidate metrics were retained for inclusion in a potential index.

Reference and test lakes also differed in water column measurements (Fig. 6-3). Test lakes as a group had higher chlorophyll concentrations and reduced Secchi transparency, and higher total phosphorus than the corresponding reference lakes, showing increased trophic state in the test lakes. Total Kjeldahl nitrogen was higher in sandy and mud-muck test lakes, and algal growth potential was increased in the mud-muck lakes.

Eight invertebrate metrics appear responsive to lake stressors, and can be used in a lake invertebrate index (number of taxa, Shannon-Wiener diversity, Hulbert index, OET taxa, percent dominance, percent filterers, percent OET, and percent gatherers). Two trophic state indicators (Secchi depth and chlorophyll a concentration) also had characteristic values under reference conditions, and are best used in conjunction with an invertebrate index to determine status of a lake.

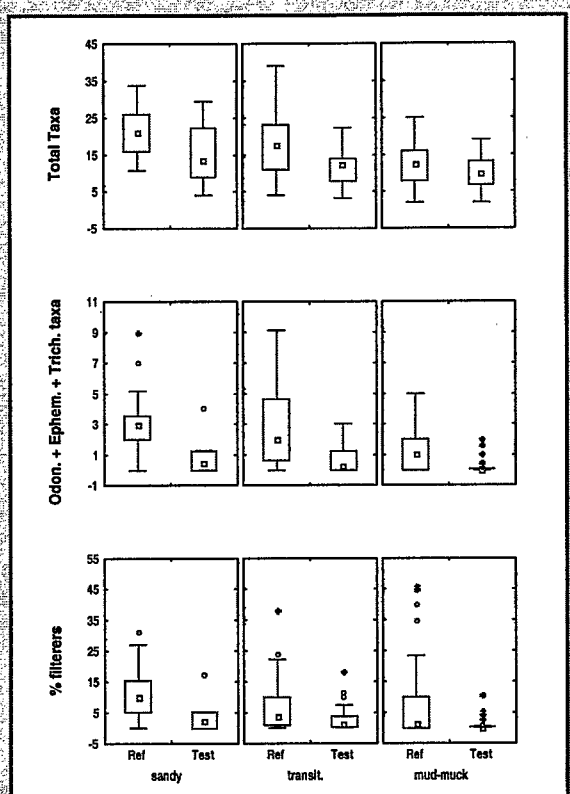


Figure 6-1. Metrics related to algal production, Florida lakes (from Florida DEP 1994). a: algal density (cells/ml) b: chlorophyll a c: chlorophyll TSI

Case Study: Florida Metric Selection and Index Development (continued)

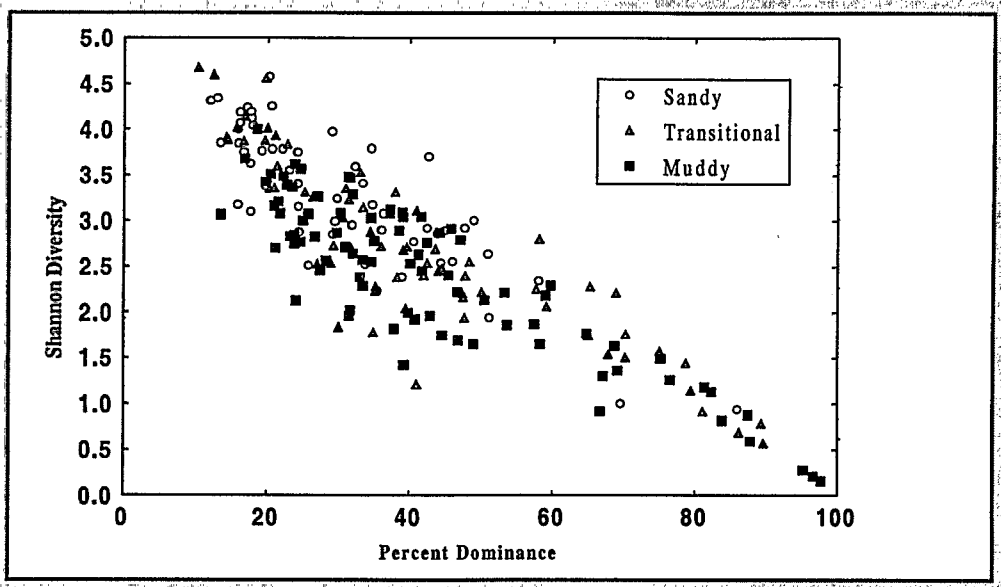


Figure 6-2. Relationship of two highly correlated attributes, Shannon diversity and percent dominance (from Gerritsen and White 1997).

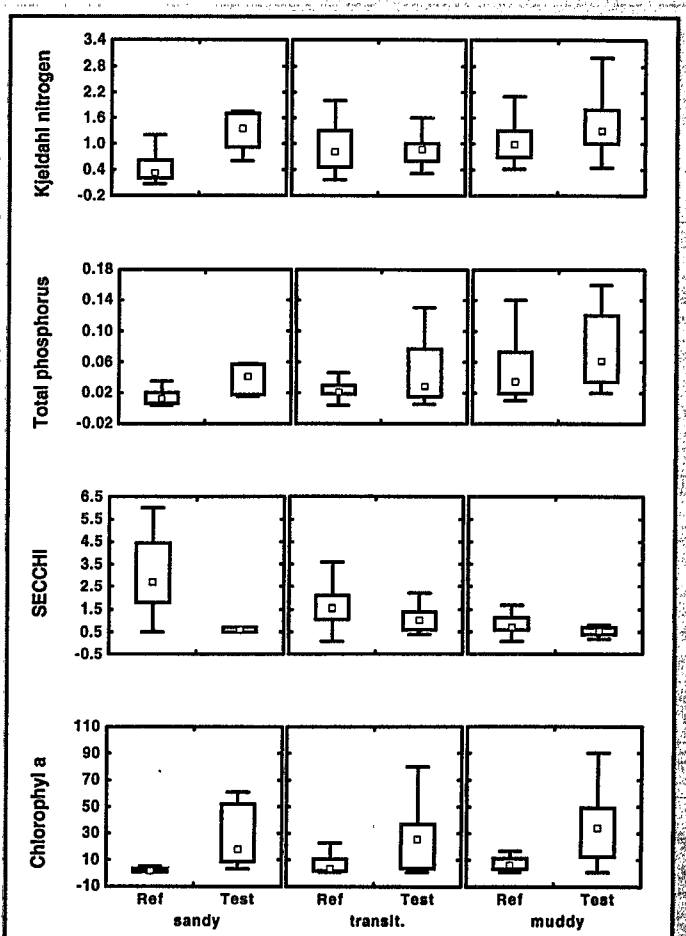


Figure 6-3. Distribution of nutrients, secchi depth, and chlorophyll *a* in Florida reference and test labs.

In This Chapter...

- Tier Structure
- Study Design Consideration

Chapter 7

Tiered Sampling

This chapter provides general guidance for designing a sampling program for lake bioassessment. Four sampling tiers are suggested options, and will need to be modified to meet the sampling objectives, project resources, and local conditions of individual programs.

Options for lake biological sampling include two sampling tiers, each with an "A" and "B" field component (Figure 7-1). There is also a desktop screening process, with no field sampling. Although sampling effort increases from Tier 1A to Tier 2B, quality of information is not necessarily related to sampling effort. Selection of a sampling tier must be based on the objectives of the biocriteria program. Tier 1 includes chlorophyll *a* and submerged macrophytes, and is consistent with Clean Lakes Program sampling. Tier 1 may be a single sample during a summer index period (Tier 1A) or monthly sampling during the growing season (Tier 1B). Tier 2A consists of assemblages that can be sampled a single time during the index period: submerged macrophytes, benthic macroinvertebrates, fish, or sedimented diatoms. Tier 2B consists of assemblages that are sampled several times during the growing season: phytoplankton, zooplankton, and periphyton. Both Tier 2A and Tier 2B require Tier 1 sampling. Although Tier 2A and 2B were developed as an "either or" choice, it is possible to perform both surveys as the assemblages sampled in each do not overlap. It should also be understood that although Tier 2B requires more effort and yields a greater quantity of data, due to multiple site visits, it does not necessarily produce better data than Tier 2A. A supplemental habitat assessment that includes diagnostic elements (as detailed in Chapter 5) can be added to any of the tiers.

7.1 DESKTOP SCREENING

The desktop screening assessment involves documentation of existing data without any observations in the field (Table 7-1). No assessment can be better than the data that go into it; therefore, desktop screening alone might be unreliable. Its use should be limited to planning for more detailed monitoring and assessment. It incorporates cost and time efficiencies, allowing evaluation of a large number of sites, and identifying potentially affected areas for further investigation using higher tiers. Information is obtained from land use data

Options for lake biological sampling include two sampling tiers each with an "A" and "B" field component.

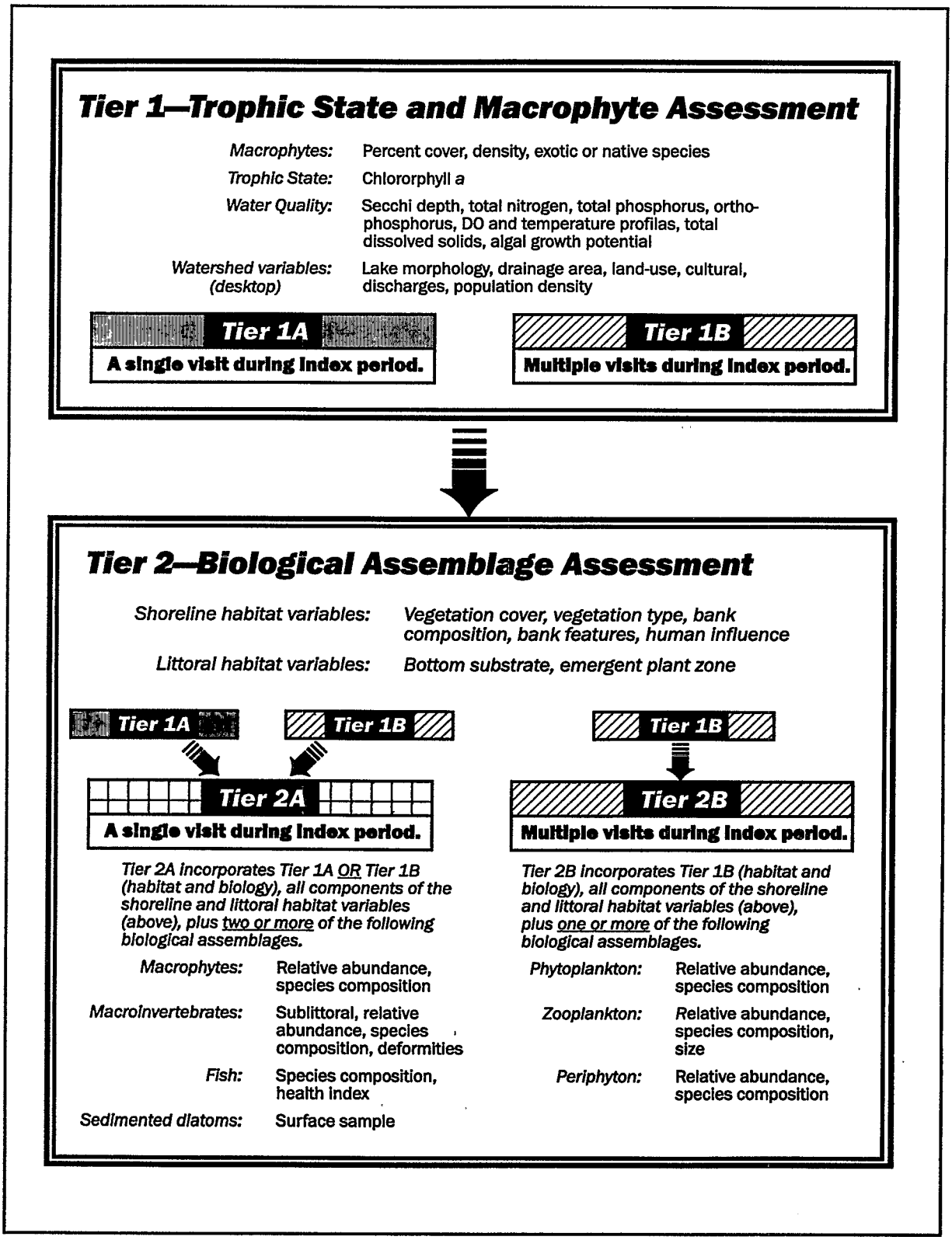


Figure 7-1. Tiered Sampling Structure.

and from a questionnaire to identify known problems in a lake (Table 7-1).

The questionnaire identifies existing known problems in lakes, but does not address new problems. An example questionnaire (Figure 7-2) is modeled after one for stream bioassessment (USEPA 1989b). Potential recipients of the questionnaire include regional biologists from natural resource agencies, the Cooperative Extension Service (CES), and academic biologists. Land use, NPDES, and population density data will identify lakes likely to have problems requiring further attention (primarily from eutrophication), but will not estimate biological impairment in the lakes. Components of desktop screening include the following:

Land Use—Land use information indicates the relative level of anthropogenic stresses in a lake watershed, especially nonpoint sources of pollutants. Many states estimate land use from satellite images.

Discharges—USEPA maintains a data base of NPDES discharges and their receiving waters.

Algae—Questions on the history of nuisance algal blooms and perceived problems with high turbidity due to algae are included in the questionnaire (Figure 7-2).

Macrophyte Survey—Local professionals knowledgeable of the macrophytes in the lake(s) are canvassed for existing data and information (Figure 7-2). The questionnaire can provide the following information:

- Extent of coverage.
- Dominant species.
- Past and present characteristics of the macrophyte assemblage.
- Factors believed to be limiting or expanding the spread of macrophytes.
- Past or present management practices used for control of macrophytes.

Fish Assemblage—Local professionals knowledgeable about fish assemblages can provide the following information:

- Expected condition of the fish assemblage.
- Likelihood of improvement and degradation.

Table 7-1. Desktop screening assessment.

	Component	Data Collection	Responds to or Indicator of
Watershed Components	1. Watershed land use, NPDES.	Maps, existing database, questionnaire. GIS databases, e.g., EPA Reach File; EPA BASINS; Census Bureau TIGER; USGS Land Use, Land Cover.	Identification of potential point and nonpoint source eutrophication, toxicity problems.
Biological Components	1. Algal production - Bloom history	Questionnaire	Identification of perceived problems (eutrophication).
	2. Plant assemblage - Macrophyte cover - Extent (% available habitat) - Density (% cover) - Known weed problems	Questionnaire	Identification of perceived problems (weeds, exotic plants, loss of native plants).
	3. Fish assemblage - Fishery problems	Questionnaire	Identification of perceived problems (species imbalance, exotic species, overfishing, overstocking, diseased).

LAKE BIOASSESSMENT QUESTIONNAIRE

This questionnaire is part of an effort to assess the biological health or integrity of lakes of this region. Our principal focus is on the biotic health of the designated waterbody as indicated by its biological assemblages and watershed use. You were selected to participate in the study because of your expertise in one or more of these areas and your knowledge of the waterbody identified in this questionnaire.

Please complete all statements. If you feel that you cannot complete the questionnaire but are aware of someone who is familiar with the waterbody, please give this person's name, address, and telephone number in the space provided below.

_____

_____

_____

Waterbody name: _____

Waterbody location (see map):
State _____ County _____ Long/Lat _____Ecoregion _____

Lake size _____ acres **or** (circle one): <10 acres, 10-100 acres, 100-1000 acres, 1000-10,000 acres, >10,000 acres

SECTION A - OVERALL ASSESSMENT

(Instructions: Answer questions 1 - 4 using the following scale. Answer by circling only **one** score for each question).

Score	Description
5	Species composition, age classes, and trophic structure comparable to non (or minimally) impacted sites of similar waterbody size in that ecoregion.
4	Species richness somewhat reduced by loss of some intolerant species; young of the year of top carnivores rare; less than optimal abundances, age distributions, and trophic structure for waterbody size and ecoregion.
3	Intolerant species absent, considerably fewer species and individuals than expected for that waterbody size and ecoregion, older age classes of top carnivores rare, trophic structure skewed toward omnivory.
2	Dominated by highly tolerant species, omnivores, and habitat generalists; top carnivores rare or absent; older age classes of all but tolerant species rare; diseased fish and anomalies relatively common for that waterbody size and ecoregion.
1	Few individuals and species present, mostly tolerant species and small individuals, diseased fish and anomalies abundant compared to other similar-sized waterbodies in ecoregion.
0	No fish.

Figure 7-2. Example of desktop screening questionnaire.

1. Circle the score that best describes your impression of the current condition of the waterbody.

   5 4 3 2 1 0

2. Classify the condition of the lake 10 years ago.

   5 4 3 2 1 0

3. Given present trends, what score will be representative of lake conditions 10 years from now?

   5 4 3 2 1 0

4. If the major human-caused limiting factors were eliminated, how would the lake be rated 10 years from now?

   5 4 3 2 1 0

Subsection A.1 - WATER QUALITY

(Instructions: Complete subsections A.1 - A.4 by **circling** the single most appropriate limiting factor and probable cause. If there is more than one limiting factor and cause, please **rank** them accordingly (by assigning a "1" for the primary factor and cause, "2" for the secondary factor and cause, etc.).

Limiting Factor	Rank
Temperature too high	____
Temperature too low	____
Turbidity	____
Salinity	____
Dissolved oxygen	____
Gas supersaturation	____
pH too acidic	____
pH too basic	____
Nutrient deficiency	____
Nutrient surplus	____
Toxic substances	____
Excessive water level fluctuation	____
Other (specify below)	____
_____	
Not limiting	____

Probable Cause	Rank
Quality of tributaries	____
In-lake processes	____
Point source discharge	____
Industrial	____
Municipal	____
Combined sewer	____
Mining	____
Upstream dam release	____
Nonpoint source discharge	____
Individual sewage	____
Urban runoff	____
Landfill leachate	____
Construction	____
Agriculture	____
Feedlot	____
Grazing	____
Silviculture	____
Mining	____
Dam surface release	____
Shorezone disturbance	____
Natural	____
Unknown	____
Other (specify below)	____
_____	

Figure 7-2. Example of desktop screening questionnaire (continued).

Subsection A.2 - HABITAT STRUCTURE

Limiting Factor	Rank
Excessive siltation	
Insufficient structure	
Insufficient shallows	
Insufficient macrophytes	
Excessive macrophytes	
Insufficient concealment	
Insufficient reproductive habitat	
Other (specify below)	
_____	
Not limiting	

Probable Cause	Rank
Agriculture	
Silviculture	
Mining	
Grazing	
Dam	
Diversion	
Channelization	
Snagging	
Natural	
Aquatic weed management	
Unknown	
Other (specify below)	
_____	

Subsection A.3 - FISH COMMUNITY

Limiting Factor	Rank
Overharvest	
Underharvest	
Fish stocking	
Non-native species	
Migration barrier	
Tainting	
Food limited	
Habitat	
Fish kills	
Other (specify below)	
_____	
Not limiting	

Probable Cause	Rank
Aquarists	
Point source	
Nonpoint source	
Natural	
Unknown	
Management	
State agency	
Federal agency	
Weed Control	
Other (specify below)	
_____	

Subsection A.4 - MAJOR LIMITING FACTOR

Limiting Factor	Rank
Water quality	
Water quantity	
Habitat structure	
Fish community	
Other (specify below)	
_____	

Figure 7-2. Example of desktop screening questionnaire (continued).

Subsection A.5 - ALGAE

(Instructions: Please provide short answers to questions 1-7, as appropriate).

1. Is there a presence and history of nuisance algae blooms? _____

2. Have algae blooms resulted in fish kills or other adverse changes to the fish community? _____

3. Has algae caused odor problems or taste problems in drinking water? _____

4. Have algae blooms deterred swimmers or affected other forms of contact recreation? _____

5. Are there other problems caused by algae blooms; and if so, what are they? _____

6. What is the source of your information? _____

7. Are there other sources of information that the agency should be aware of such as fishery records and grey literature studies? _____

SECTION B - AQUATIC MACROPHYTE COMMUNITY

(Instructions: Answer questions 1 - 3 using the following scale. Circle only one score for each question).

Score	Description
3	Extent and cover are comparable to non (or minimally) impacted sites of similar waterbody size in that ecoregion.
2	Macrophyte beds appear weedy. The extent and/or cover are greater than non (or minimally) impacted sites. The dominant species are those found in highly eutrophic waters.
1	Few macrophytes found compared to non (or minimally) impacted sites. Macrophytes that are found are usually exotics and are tolerant of a wide range of water quality conditions and/or fluctuations.
0	No macrophytes.

1. Circle the score that best describes your impression of the current macrophyte conditions of the lake.

   3 2 1 0

2. Classify the macrophyte conditions of the lake 10 years ago.

   3 2 1 0

3. Given the present trends, what score will be representative of lake conditions 10 years from now?

   3 2 1 0

Figure 7-2. Example of desktop screening questionnaire (continued).

Subsection B.1 - FACTORS EFFECTING MACROPHYTES

(Instructions: Complete subsection by **circling** the single most appropriate limiting factor and probable cause. If there is more than one limiting factor and cause, please **rank** them accordingly (by assigning a "1" for the primary factor and cause, "2" for the secondary factor and cause, etc.).

Limiting Factor	Rank	Probable Cause	Rank
Grass carp introduction	____	Aquarists	____
Exotic species	____	Point source	____
Excessive siltation	____	Nonpoint source	____
Drawdowns	____	Natural	____
Weed control	____	Unknown	____
Shoreline cleanup	____	Management	____
Excessive epiphytes	____	State agency	____
Excessive turbidity	____	Federal agency	____
Insufficient shallows	____	Fisherman	____
Elevation or latitude	____	Other (specify below)	____
Macrophyte beds are expanding	____	_____	
Other (specify below)	____		
_____			
Not limiting	____		

Subsection B.2 - MACROPHYTE EXTENT AND SPECIES

(Instructions: Please provide short answers to questions 1 - 4, as appropriate).

1. What is the extent of macrophyte coverage in the photic zone? _____

2. What are the dominant species? _____

3. What is the source of your information on macrophytes? _____

4. Are there other sources of information on the macrophyte community in this waterbody that the agency should be aware of such as management reports or grey literature studies? _____

SECTION C - WATERSHED CHARACTERISTICS AND LAND USE

(Instructions: Please provide short answers to questions 1 - 11, as appropriate).

1.	Watershed size	____ acres	5.	Urban	____%
2.	Elevation difference [watershed divide to lake surface]	____ft	4.	Agricultural	____%
3.	Forest or natural vegetation	____	6.	Suburban/residential	____%

7. Human population density in lake watershed _____

8. Number of dischargers within the watershed (e.g., NPDES permits) _____

9. What is the source of your information on the watershed? _____

10. Are there other sources of information on the watershed and surrounding land use that the agency should be aware of such as grey literature or land use planning documents? _____

Figure 7-2. Example of desktop screening questionnaire (continued).

- Major limiting factors.
- Water quality
- Habitat availability
- Management, harvest, or mortality

Desktop Integration

Based on responses to the questionnaire, perceived levels of impairment can be judged from the three biological assemblages: algae, macrophytes, and fish. The three evaluations are kept separate. Perception of a problem, or a substantial departure from expected conditions, earns a rating of "impaired" for the respective assemblage. The land use information is used to identify potential stressors on a lake.

7.2 TIER 1: TROPHIC STATE AND MACROPHYTES

Tier 1 requires sampling of primary producers to assess trophic state and aquatic macrophytes. It can be done with a single visit during an index period when the objective is a synoptic survey and screening of many lakes (Tier 1A). Tier 1A is only appropriate for regional assessments—it

Table 7-2. Tier 1: Trophic state and macrophyte sampling.

	Component	Data Collection		Responds to or Indicator of
		Tier 1A	Tier 1B	
Habitat Components	1. Watershed land use, population, NPDES.	Maps, existing database, questionnaire. GIS databases, e.g., EPA Reach File; EPA BASINS; Census Bureau TIGER; USGS Land Use, Land Cover. Desktop screening habitat.		Potential causes.
	2. In-lake physical habitat maximum depth area inflow.	Maps or survey (single visit).		Potential causes.
	3. Water Quality - DO, temperature profile - pH, alkalinity, conductivity - Secchi depth - Total dissolved solids - Nutrient concentration - Algal growth potential	Single index period. Surface or integrated.	Multiple visits. Water column sample.	DO problems, eutrophication, stratification, acidification, turbidity.
Biological Components	4. Algal chlorophyll *a* concentration.	Single visit chlorophyll sample from 0.5m. Surface integrated water sample.	Multiple visits.	Eutrophication
	5a. Submerged macrophytes - % of available habitat with macrophytes - dominant species	Single visit, aerial photos if possible; otherwise, estimate from shorezone survey. Identify dominant species.	Multiple visits.	Eutrophication, herbicides, exotics.

cannot be used to assess single lakes. More precise estimates for single lakes can be made with Tier 1B, comprising several sampling visits to determine growing season averages. Tier 1 consists of the Desktop Screening land use survey, lake physical habitat, water chemistry (dissolved oxygen, nutrient concentrations, conductivity, alkalinity, pH), Secchi depth, chlorophyll *a* concentration, and a submerged macrophyte survey (Table 7-2). The survey enables:

Design of a sampling program inevitably requires compromises to answer the intended questions in a reasonable time and at a reasonable cost.

- Identification of trophic state based on chlorophyll *a* concentration, nutrient concentration, and Secchi depth.
- Detection of weed problems or loss of aquatic macrophytes.
- Detection of midsummer oxygen stress.

7.2.1 Sampling Frequency for TSI Variables (Tier 1A vs. Tier 1B)

Tier 1A consists of sampling during an index period, typically mid to late summer for trophic state variables (e.g., chlorophyll *a*, Secchi depth, nutrients). Tier 1A is adequate for characterization of lakes in a region, when many lakes must be sampled to develop the characterization and assessment. Tier 1A will yield a good characterization of a region or a population of lakes, but precise characterization of individual lakes, for site-specific management, will require Tier 1B, with more frequent sampling. Tier 1B takes into account the changes in chlorophyll and nutrients that can occur in a short time and is used to estimate seasonal averages of the variables by sampling several times during the growing season. Trophic State Indices (TSI) are calculated from the seasonal average estimates of chlorophyll, Secchi depth, and nutrients. The number of sampling visits required depends on the temporal variation in the lake and the desired precision of the estimated seasonal average. Monthly sampling appears to be adequate for most purposes (Knowlton and Jones 1989).

7.2.2 Sample Locations for Trophic State Measurements

Design of a sampling program inevitably requires compromises to answer the intended questions in a reasonable time and at a reasonable cost. In lake biosurveys, the unit of interest (sampling unit) may be the whole lake, a lake basin, a tributary arm, or an embayment. In some situations, the unit of interest may be an area of the lake receiving discharges or runoff. The object of sampling is to characterize the sampling units with sufficient precision and accuracy to meet the needs of the program.

Sample sites are selected to be representative of the lake. Single sites have traditionally been located in the middle of the lake, usually over the deepest area. For unbiased characterization, multiple sites should be selected randomly. Sampling may be stratified by zones, e.g., littoral, pelagic, and inflows; or riverine, transitional, and lacustrine (Figure 7-3). Estimation of mean values for the whole lake should be weighted by the relative area or volume of each zone. Figure 7-4 shows an example of sampling locations for all tiers in a relatively simple lake (natural or impoundment).

Lakes may be characterized by single or multiple sample sites in each lake, depending on the objectives of the survey.

Single sample site

If the objective is to characterize a large population of lakes, as in a statewide survey, then a single sample per lake is most cost-effective. A single site is typically chosen as the midpoint of the central basin of the lake, and is usually sufficient to prioritize lakes within a region.

Large riverine reservoirs have known gradients of nutrients and productivity from the river inflow to the dam (Kennedy and Walker 1990), and a single site is not appropriate. Large reservoirs would require a minimum of three sites, corresponding to the riverine, transitional, and lacustrine zones, respectively (Figure 7-3).

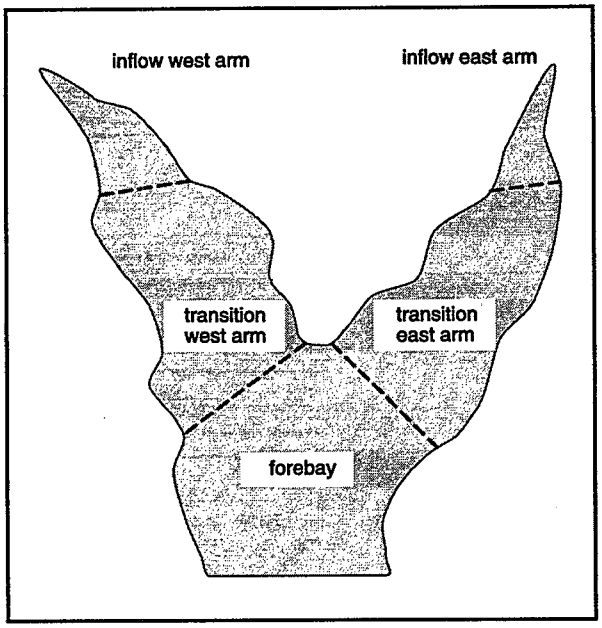


Figure 7-3. Sampling zones in large or complex lakes (large reservoirs, multi-basin lakes).

Multiple sites
If turbidity, nutrients, and algae are known to be variable across the surface of a lake, then multiple sample sites are required (Figure 7-4).

RESERVOIR

If gradients are known to occur, as in many large reservoirs, then sampling should be stratified by zones. For example, in a reservoir one could define the three reservoir zones (riverine, transitional, lacustrine) as sampling strata, and take two or more samples from each zone.

The exact number of sampling sites in a lake or lake zone is determined by the spatial variability of nutrients, turbidity, and chlorophyll; and the desired precision. In general, within a basin or reservoir zone, variation in time is larger than variation in space (Knowlton and Jones 1989). Thus, chlorophyll sampled 2 weeks apart may differ by several fold, but samples on the same day 500m apart are likely to differ much less.

If precise characterization of individual lakes is an objective of the biological survey then it is more cost-effective to sample repeatedly during the growing season (Tier 1B) than to sample multiple sites at a single time (Tier 1A).

Composite samples
Composite samples are taken from several sites in a lake or lake zone, and combined into a single sample for laboratory analysis. For example, water samples may be taken from four sites in a lake, and poured into a single clean bucket. The composite sample is subsampled for chlorophyll *a* and nutrients. Secchi depth temperature, and DO are measured at each of the four sites. Care must be taken that the methods and volume sampled are the same at each site. Composite samples characterize the lake better than a single sample and they save laboratory analysis costs. The principal disadvantage of composite samples is that they do not allow estimation of spatial variability within a lake.

7.2.3 Trophic State

The Tier 1 Trophic State Indices (TSI) are estimated from Secchi depth, chlorophyll *a*, and nutrient concentrations. Field methods for Secchi transparency and chlorophyll *a* are outlined below and summarized in Table 7-4.

Secchi Depth (SD)
Secchi depth is a measure of transparency. Turbidity caused by suspended sediments and algae decreases Secchi depth.

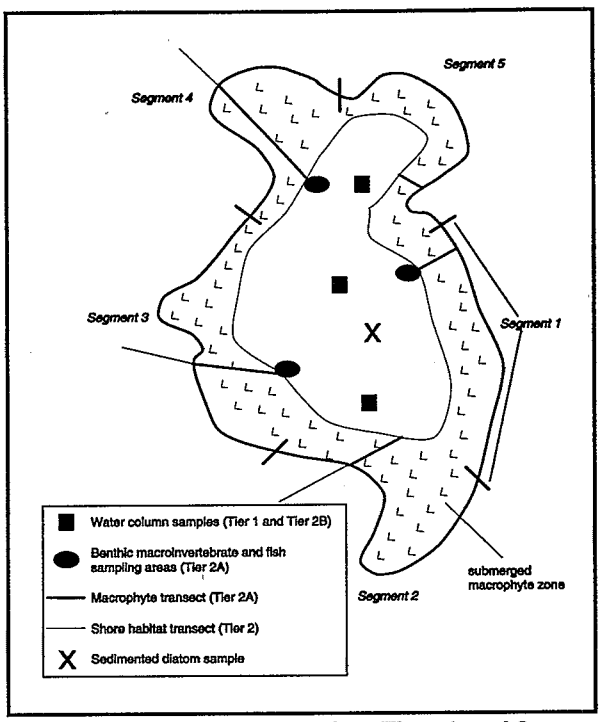


Figure 7-4. Integrated sampling, Tiers 1 and 2.

Sampling Location—Secchi disk transparency can be measured at one or more representative locations.

Frequency—Tiers 1A and 2A: single determination, midsummer. Tiers 1B and 2B: 6 to 10 samples during the growing season (e.g., March through October).

Sampling Procedure—Readings are obtained with a 20cm plastic or metal Secchi disk that is divided into black and white quadrants on a nonstretchable line, calibrated in centimeters. The disk is lowered into the water until it disappears from view, then is raised slowly to the point where it reappears. Secchi depth is the average of the two depths.

Observations are made from the sunny side of the boat or dock, during midday, without sunglasses, and as close as possible to the water in order to reduce glare.

Data Analysis—Secchi depth can be used in determining trophic state along with chlorophyll *a*.

Chlorophyll

Chlorophyll *a* sampling and analysis follow standard protocols (USEPA 1994a, USEPA 1994b).

Presampling—Samples must be collected in a clean container, without using acid washes or phosphorus detergents. Before sample collection, bottles and collectors should always be double or triple rinsed with the lake water to be sampled.

Sample Location—One location or several representative locations for composite sample.

Frequency—The same as Secchi depth.

Depth—Chlorophyll *a* concentration may be estimated from surface samples taken at 0.5m, from integrated epilimnion samples, or from integrated water column samples. Half-meter surface samples require the least equipment and can be taken by hand; epilimnion and integrated water column samples are taken with a flexible hose.

Sampling Procedure—Surface sample, 0.5m. A rinsed, l-liter sample bottle is inverted and held at depth (arm's length) by hand, turned up to fill, and brought to the surface.

Hose sample—A flexible hose is an easy method to obtain an integrated sample over the whole water column or over a defined portion, such as the epilimnion. The weighted end of a plastic hose is lowered to a given depth. The upper end is stoppered or clamped at the surface, and the weighted end is hauled to the surface with an attached line. The hose is emptied into a clean sample bucket, and chlorophyll and chemical subsamples can be drawn from the integrated sample. The hose may be lowered to 1m above the bottom for a water column sample, to the metalimnion, to twice Secchi depth as an estimate of the photic zone, or to a fixed depth (e.g., 5m). Each standard depth method has its own advantages and disadvantages (Carlson and Simpson 1996). Consistency of sampling method is more important than selecting the "best" standard depth.

Water samples are filtered for chlorophyll *a* extraction. A "rule of thumb" for the quantity to filter is 100ml for every foot of Secchi depth (330ml for every meter; D. Canfield, personal communication). Samples are vacuum-filtered on glass-fiber paper, and the filter papers are stored frozen in the dark. Detailed instructions

Table 7-3. Sampling summary for chlorophyll, water quality, and phytoplankton.

Habitat	Open water, 1 to 5 sites per lake or lake stratum.
Sampling Gear	Hand-held bottle or flexible hose.
Index Period	Single mid-season sample (Tier 1A) or monthly samples during growing season (Tier 1B, Tier 2B).
Sampling	Bottle: invert bottle at arm's length depth (0.5 m); turn. Uphose: lower open weighted hose through water column to predetermined depth, stopper, and haul up.
Analysis	Chlorophyll and water quality: standard methods. Phytoplankton: filter or settle and identify 300 to 500 cells to genus.

for filtering and analysis are in APHA (1992) and USEPA (1994a, 1994b).

Water Chemistry

Samples of water for chemical analysis are collected in the same manner as chlorophyll samples. Sampler bottles should be cleaned in a phosphate-free detergent prior to use and rinsed two to three times in lake water in the field. Samples may need to be preserved or filtered in the field depending upon which chemicals are to be analyzed.

Dissolved Oxygen and Temperature Profiles—A dissolved oxygen/temperature electrode (EPA Method 360.1) is used to measure both dissolved oxygen and temperature. Using the electrode, dissolved oxygen and temperature may be measured at 0.5m intervals to produce dissolved oxygen and temperature profiles. Dissolved oxygen electrodes should be calibrated against standard chemical titration methods before and after field use.

pH, Alkalinity and Acid Neutralizing Capacity—A calibrated pH meter may be used to determine pH. Acid neutralizing capacity is important to the ability of a waterbody to resist changes in pH due to addition of acid and is based upon the alkalinity of the water and dissociated organic compounds present. Carbonates, bicarbonates and hydroxides are the major contributors to alkalinity which is determined using calorimetric titration methods (APHA 1992). For more precise determination of acid neutralizing capacity, the Gran plot method is used (USEPA 1987a).

Total Dissolved Solids (TDS)—Total dissolved solids consist of inorganic solutes such as nitrates, sulfates, and carbonates, and organic substances dissolved in water (APHA 1992). TDS is measured by first filtering a measured volume of sample water through a filter, and weighing the dried residue. See APHA (1992) for specific methods.

Algal Growth Potential Test (AGPT)—Because nutrients are not always present in a form available to algae, direct chemical measurements may not be predictive of the actual potential for algal growth. The Algal Growth Potential Test (also know as a biostimulation study, APHA 1992) was developed to directly measure in a standardized way the potential of waters to support algal growth.

Total Nitrogen—Nitrogen is an important plant nutrient and may serve as a limiting factor in some waters, especially subtropical lakes. Total nitrogen is a combination of nitrate/nitrite nitrogen and total Kjeldahl nitrogen (organic and reduced nitrogen). Total Kjeldahl nitrogen is measured using a digestion technique that converts organic nitrogen to ammonia and includes any other ammonia present in the sample. Nitrate plus nitrite is measured with standard colorimeter methods (APHA 1992).

Trophic state determinations provide a method for determining whether increased nutrients or sediments (loading) are causing changes in a lake.

Total Phosphorus—Phosphorus is a limiting nutrient in many fresh waters. Total phosphorus can be analyzed using the automated procedure outlined in USEPA Method 365.1.

Estimation of Trophic State

Trophic state determinations provide a method for determining whether increased nutrients or sediments (loading) are causing changes in a lake. Carlson's TSI uses Secchi depth, chlorophyll *a*, and total phosphorus, each producing an independent measure of trophic state (Carlson 1977). Index values range from approximately 0 (ultraoligotrophic) to 100 (hypereutrophic). The index is scaled so that TSI = 0 represents a Secchi transparency of 64 m. Each halving of transparency represents an increase of 10 TSI units. For example, TSI of 50 represents a transparency of 2m, the approximate division between oligotrophic and eutrophic lakes (USEPA 1990b). A TSI is calculated from each of Secchi depth (SD), chlorophyll concentration (Chl), and total phosphorus concentration (TP) (Carlson 1977, Carlson and Simpson 1996).

$$TSI(Chl) = 30.6 + 9.81\ \ln(Chl)$$

$$TSI(TP) = 4.15 + 14.421\ \ln(TP)$$

$$TSI(SD) = 60 - 14.41\ \ln(SD)$$

Trophic state indices are used to infer trophic state of a lake and whether algal growth is nutrient limited or light limited. If the three indices are approximately equal, then phosphorus limits algal growth. If the three are not equal, then other interpretations exist (Carlson

and Simpson 1996). A trophic state index has also been developed for total nitrogen (TN) (Kratzer and Brezonik 1981, Carlson 1992):

$$TSI(TN) = 54.45 + 14.43\ \ln(TN)$$

For a more complete discussion of trophic state indices and their interpretation, see Carlson (1992) and Carlson and Simpson (1996).

7.2.4 Aquatic Macrophytes

The Tier 1 macrophyte survey is a visual estimate of percent cover of submerged and floating macrophytes in shallow water, and identification of the most dominant species and weedy or exotic species. The survey can be done with aerial photographs (if available); a visual whole-lake survey in small lakes, or examination of transects in large lakes. Three to ten transects should be sufficient for most lakes or embayments too large to survey in their entirety. Large lakes with known differences within the lake should be sampled by lake zone; for example, the shallow riverine zones of a reservoir may have greater macrophyte cover than the lacustrine zone.

To avoid bias, transects should be selected randomly within each lake zone. A method of selecting transects is to divide the shore into equal segments (corresponding to the number of transects). A point is selected randomly in each segment as the starting point for transects. Transects are perpendicular to shore to deeper water.

Total vegetative cover is estimated visually. The presence of algae mats and epiphytes should be noted. Cover might be difficult to estimate in turbid waters. Vegetation samples may be collected with a rake and total abundance estimated from the material raked in (ordinal scale: sparse, moderate, abundant). The most dominant species, and any weedy or exotic species, are identified.

7.3 TIER 2A: BIOLOGICAL ASSEMBLAGE ASSESSMENT

Tier 2A sampling requires two or more lake biotic assemblages: macrophytes, sedimented diatoms, fish, or macrobenthos (Table 7-4). Tier 1 variables, including DO, chlorophyll *a*, and Secchi depth, are also critic components of the Tier 2A survey. Tier 2A may be built on either Tier 1A or 1B. Macrophytes are the easiest of these assemblages to identify and count in the field (using wet weight instead of relative abundance). Sedimented diatoms are also relatively easy to sample, although identification and enumeration must be done in the laboratory. The choice of which plant assemblage to sample clearly depends on the importance of the assemblage in lakes of the region—diatoms would be the choice in regions where macrophytes are minor components of the lake system.

The habitat components of the Tier 2A survey build on the desktop screening and Tier 1 habitat assessment and also include a semi-quantitative shore zone habitat evaluation (Table 5-3). Tier 2A requires estimates of shorezone land use, riparian vegetation, emergent macrophyte extent and cover, and floating macrophyte extent and cover at several transects from the shore.

The Tier 2A faunal component consists of the benthic macroinvertebrates. Macroinvertebrates are sampled from the sublittoral zone, below the floating macrophyte zone, yet above the thermocline to avoid sampling predominantly anoxic areas. Tier 2A sampling typically consists of a single visit during an index period. Benthic macroinvertebrates may optionally be sampled more frequently to obtain growing season averages. Macrophytes are best sampled mid- to late in the growing season when plant biomass is near its annual maximum. Sedimented diatoms, which represent sedimentation of at least a year or more, may be sampled at any time.

Tier 2A allows more precise detection and identification of problems and potential causes than Tier 1, as well as detection of biological effects on the biotic assemblages selected for assessment.

7.3.1 Tier 2A: Transect Sampling

Establish Transects

Tier 2A sampling of macrophytes and benthic macroinvertebrates, and the shorezone habitat are surveyed along 3 to 10 transects perpendicular to the shoreline (Figure 7-4). Transects

Table 7-4. Tier 2A: Routine biological sampling.

	Component	Data Collection	Responds to or Indicator of
Habitat Components	1. Watershed land use, population, NPDES.	Desktop screening habitat.	
	2. Lake physical.	Tier 1 habitat.	
	3. Shorezone habitat assessment.	3-10 transects: - land use - bank stability - riparian vegetation - emergent vegetation	
	4. Water quality DO seasonal or annual mean, % depth-time Mean pH, alkalinity Secchi depth.	Tier 1 water quality (1A or 1B).	Trophic state, turbidity.
Biological Components	5. Algal chlorophyll *a*.	Tier 1 chlorophyll (1A or 1B).	Trophic state.
	6, 7. Assemblages (minimum 2):		
	a. Macrophyte species.	2-3 samples from transects; identify plants to species and weigh cumulative sample of each species, or count stems.	Trophic state, exotics, herbicides.
	b. Macrobenthos	Sublittoral surface sediment grab at end of each transect; identify to lowest practical level, 100-200 organisms.	DO, siltation, toxicity, productivity.
	c. Fish assemblages.	Littoral electrofishing sample at the end of each transect; sublittoral netting; identify to species, enumerate, weigh, and record incidence of external anomalies.	DO, toxicity, productivity.
	d. Sediment diatoms.	Surface sediment grab in deepest part of lake; identify to species and variety.	Nutrient enrichment, toxicity.

are the same as the Tier 1 macrophyte transects: the lake (or lake zone) shoreline is divided into equal length segments corresponding to the number of transects, and a transect start point is randomly selected in each segment.

7.3.2 Shorezone Measurements

Each transect is extended visually on the lake shorezone, and the condition of the shorezone is determined. Shorezone measurements include riparian vegetation cover estimates, lake bank substrate and erosion, and human modifications. Figure 7-5 is an example scoring sheet for habitat measurements showing how the variables are scored.

7.3.3 Aquatic Macrophytes

Tier 2A macrophyte sampling is more systematic and detailed than Tier 1. The objective is to obtain relative abundances of macrophyte taxa to develop assemblage measurements. Relative

abundance can be estimated by stem counts (number of stems of each species) or biomass. Biomass is preferred because a stem does not correspond to an individual plant, and biomass is a good indicator of species dominance in the habitat. An alternative to relative abundance is scoring presence and absence of species in quadrat.

One to four macrophyte sampling locations are established on each transect within depth zones between shorezone emergent and the unvegetated, sublittoral bottom. For example, location may be identified in 0-1m depth, 1-2m, 2-3m, and 3-4m depth (Weber et al. 1995).

Stem counts—May be done with the transect method, by counting stems touching a line held on the transect. Stems may also be counted in quadrants, where all stems within a 1/4 m^2 quadrat are counted and identified. Stem counts may require diving in water deeper than 1 m. One or more sampling stations (for quadrat sampling) are selected on each transect between the emergent macrophyte zone and the deepest extent of submerged macrophytes.

Biomass Sampling—The easiest method to estimate macrophyte biomass is with an aquatic weed rake (Table 7-5). At each station on the transect, an aquatic weed rake or thatching rake is dragged a set distance (e.g., 1m) to sample vegetation (Trebitz et al. 1993). Plants from all stations on the lake are identified and sorted by species, and the total of each species collected is weighed (wet weight) to obtain estimates of biomass and proportion of biomass of each species. Algae mats and epiphytic growth on leaves and stems are described. Voucher specimens of each species should be kept for complete identification and for permanent record. Depth is sounded at the lakeward edge of submerged vegetation.

Tier 2A macrophyte sampling is more systematic and detailed than Tier 1. The objective is to obtain relative abundances of macrophyte taxa to develop assemblage measurements.

Aquatic weed rakes are biased against macrophytes that can slip through the tines of the rake. Therefore, a more accurate estimate of biomass would be to clip all plants in the quadrat for wet weight determination. Clip plots would require diving or snorkeling in water more than 1m deep. Biomass can be estimated more accurately by drying the sorted plant material for dry weight determination, at the cost of additional processing.

The weed rake and wet weight determination is likely to be the most cost-effective methodology for most purposes. Although it undersamples certain species, it is likely to be consistent enough to use for biological surveys, as long as the same sampling methodology is used in all lakes.

Presence-Absence—Instead of estimating biomass, species can be scored for presence or absence within quadrants (Weber et al. 1995). Each sampling location along a transect is divided into four quadrants. Each quadrat is sampled with the rake, and each species receives one point for every quadrat in which it occurs.

7.3.4 Macroinvertebrates

The macroinvertebrate assemblage beyond the macrophyte zone is sampled with gear appropriate to the bottom type and depth (e.g., Ponar, Ekman grab sampler, dome sampler); and the assemblage is identified and characterized (Table 7-7).

Sampling Period—Two sampling periods have been identified either the most stressful period (usually late summer) or a period after recruitment (usually early spring) but before major emergence of adult insects.

Sampling Location—Along transects, the sublittoral habitat is recommended as the most appropriate habitat for sampling due to its relatively stable nature.

Sampling Gear—The type of gear will depend on the substrate being sampled (Table 7-7). A standard mesh size of 595 μm (No. 30 mesh) is required.

Sample Replication—To characterize the macroinvertebrate assemblage, multiple grabs are taken from several sites. Each transect ends in a macroinvertebrate sample site, and two to three grabs are taken at each site. Grabs may be composited into a single composite sample.

Table 7-5. Sampling summary for submerged macrophytes.

Habitat	Littoral zone.
Sampling Gear.	Tier 1: none. Tier 2A double-headed rake on chain (Trebitz et al. 1993), or 1m^2 quadrats and diving gear.
Index Period.	Late summer. (Macrophytes are sampled once regardless of tiers.)
Sampling	Tier 1: Estimate of area covered by macrophytes. Tier 2A: 2-3 semiquantitative rake samples to determine relative biomass of species; on randomly placed transects perpendicular to shore. Alternative: 1-3 randomly tossed quadrats on each transect, then stem count and identification of each species in quadrat.
Analysis	Tier 1: Dominant species identifed, % estimated. Tier 2A: All species identified, relative abundance of each estimated from wet weight or stem count (Trebitz et al. 1993).

Sample Processing—To process the sample, organisms are removed from sticks, rocks, and similar size objects. The remainder of the sample is placed in a tub and mixed into a fine, uniform slurry. After mixing, the slurry is sieved using a U.S. No. 30 sieve (595 um) to remove organic and mineral material. The benthic organisms are retained by the sieve, which can be emptied into a light-color, gridded sorting tray. Grid cells are selected at random and sorted until at least a 100-organism subsample has been counted and identified to the appropriate taxonomic level. The last grid cell is sorted completely until all organisms from the grid are identified to the lowest practical level. Further description of sorting is presented in EPA/440/4-89-001 (USEPA 1989b).

7.3.5 Fish Assemblage

Fish assemblages can be sampled by electrofishing in and/or beyond the macrophyte zone. Sampling effort for fish should be kept constant between transects. Electrofishing is generally the single most cost-effective sampling

Table 7-6. Sampling summary for benthic macroinvertebrates.

Habitat	Preferred: sublittoral. Alternative: profundal (if hypoxia is rare).
Sampling Gear.	Regionally most appropriate for substrate (Table 7-8); 595mm mesh (No. 30 sieve).
Index Period.	Regionally most appropriate. Preferred: Late summer (most stressful; most regions). Alternative: Early spring; winter (subtropical lakes).
Sampling	Lakewide composite samples of 2-3 grabs at each of 3-5 sublittoral sites (7 to 15 grabs total) or keep sites as replicates if an in-lake variance estimate will be used in assessment.
Analysis	Preferred: lowest practical taxonomic level, 100-organism subsample. Alternatives: more than 100 organisms.

Lakeshore Habitat Measurements and Metrics					
Lake Name: Date of Visit: Visit #:					
Lake ID: Team ID:					
Riparian Vegetation Measurements Transect ID:					
Areal Coverage Categories : 0 = Absent, 1 = Sparse (< 10%), 2 = Moderate(10-40%) 3 = Heavy (40 to 75%), 4 = Very Heavy (>75%)					
Canopy (%) cover					
Understory (%) cover					
Ground Cover (%) cover					
Barren (%) cover					
Bank Measurement					
Rocky (%)					
Soil (%)					
Vegetated (%)					
Other (%)					
Bank Erosion Score (0-4): 0 = None, 4 = Severe					
Human Influence Measurements 0 = Absent, 1 = Present within Transect 0.5 = Observed Adjacent to or Behind Transect					
Buildings					
In-Lake Structure					
Roads, Railroads					
Agriculture					
Lawn					
Dump or Landfill					

Figure 7-5. Example scoring sheet for shorezone habitat.

Case Study: Florida sampling methods

In 1995, FDEP adopted a new sampling protocol to obtain more representative samples of each lake, in part based on results from the earlier samples. Lakes greater than 1000 acres are divided into two or more basins, usually by separating at constriction points or between bathymetrically identifiable basins (Fig. 7-6). The 2-4m sublittoral zone of each lake basin is divided into 12 equal segments, and a grab is taken in each segment with a Petite Ponar or Ekman sampler (0.02m²) (Fig. 7-6). Positions of segments and sampling sites are estimated by eye in the field. The 12 grabs are combined into a single composite sample, which is randomly subsampled to a count of 100 organisms, identified to the lowest practical taxonomic level. Basins (in lakes greater than 1000 acres) are retained as separate sample units. Lakes smaller than 1000 acres are represented by a single 100-organism sample. A second grab sample is taken at each of the 12 stations for sediments, which are likewise combined into a single representative sample.

In fixed organism subsampling, a targeted number of organisms (typically 100 to 500) is identified. If fixed organism subsampling for benthos is conducted in an unbiased manner using a random selection method, the resulting information on richness and relative abundance is comparable among samples. For benthic samples, the targeted number is reached by randomly choosing several fractions or "grids" from a pan; all organisms enclosed within the grids are sorted to avoid bias toward large and easily seen individuals. Ideally, several (four or more) grids are sorted to ensure proper representation.

Surface and bottom water chemistry samples, and phytoplankton samples, are taken near the center of each lake. Observations included field measurements and laboratory analyses, and identification of phytoplankton to genus.

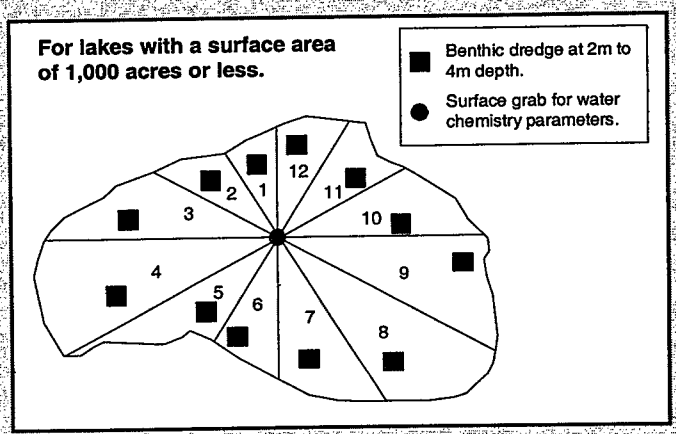


Figure 7-6. Florida lakes sampling scheme. (The lake is divided into 12 approximately equal segments. A Ponar grab is taken from each segment, at a random location in the 2 to 4 meter depth zone. Water chemistry, chlorophyll, and Secchi depth are measured from the center of the lake.)

Table 7-7. Benthic macroinvertebrate sampling gear appropriate for major substrate types.

Substrate	Gear Types
Submerged aquatic vegetation.	Dip net.
Rocks, gravel.	Diver operated dome sampler.
Sand	Peterson, Van Veen grabs.
Mud	Ponar, Ekman grabs.
Clay	Peterson, Van Veen grabs.

Case Study: TVA Benthic Macroinvertebrates Collection Methods

Benthic macroinvertebrate assemblage samples were collected in the spring (March and April) at 69 locations on 30 TVA reservoirs. Sample locations were selected in each of the forebay, mid-reservoir, and inflow areas, corresponding to lacustrine, transitional, and marine conditions, respectively (Figure 7-3). At each sample location, a line-of-sight transect was established across the width of the reservoir, and one Ponar grab sample collected at 10 equally spaced locations along this transect. When rocky substrates were encountered, a Peterson dredge was used. Care was taken to collect samples only from the permanently wetted bottom portion of the reservoir (i.e., below the elevation of the minimum winter pool level). Samples were washed in the field, transferred to a labeled collection jar, and fixed with 10 percent buffered formalin solution. Samples were sorted and identified in the field, to the lowest practical taxon, typically genus, and reported as number per square meter.

To assess the reproducibility of benthic macroinvertebrate sampling results, replicate samples were collected at 13 of the 69 sampling locations in 1994, with all types of reservoir locations (i.e., forebay, transition zone, embayment and inflow) included. At each of the replicate sampling locations, the sampling protocol involved collection of a first set of 10 samples, leaving the sampling location, and then returning as near as possible to the original transect site (on the same day) and repeating the collection of a second (replicate) set of 10 samples. Results from sets of replicate samples were evaluated for reproducibility.

method for fish (Scott et al. 1992) but it is not effective in deep water. If deep water fish are an endpoint of concern, then gill nets, fish traps, or trawls can be used. A combination of nets and electrofishing often provide a more representative sample of the fish assemblage; however, multiple methods translate to a substantial cost for field effort. A variety of nets may be used to sample littoral and sublittoral areas. Fish sampling methodologies are further outlined in EPA 600/R-92/111 (USEPA 1992b) and Table 7-8.

Sampling Procedures

Electrofishing—Multiple habitats are selected in littoral areas for electrofishing. Habitat distinctions are based on substrate (e.g., rocks, sand, clay) and on available cover (e.g., vegetation, woody debris).

Nets—A variety of nets are used to sample littoral and sublittoral areas. It is recommended that trapping nets (gill nets, trammel nets, fyke nets, trapnets) be set for 2 to 5 days with collection once or twice a day.

Table 7-8. Sampling summary for fish assemblage.

Habitat	Littoral and sublittoral zones.
Sampling Gear	Boat electrofisher (for available microhabitats within shallow littoral areas). Experimental gill nets (extended for littoral to sublittoral zones).
Index Period	Regionally most appropriate. Preferred: Late summer - early fall. Alternative: Early spring; winter (subtropical lakes).
Sampling	Littoral electrofishing sample reach of shoreline at the end of each transect. All microhabitats sampled within each measured littoral reach. Experimental gill nets (five panel nets) set perpendicular to shore at the end of each transect, extending from littoral to sublittoral zones.
Analysis	Preferred: All specimens identified to species, enumerated, measured, weighed, and examined for incidence of external anomalies. Alternative: Abundant species (e.g., greater than 50 individuals per sample) may be subsampled, measured, weighed, and data extrapolated for the species total.

- Gill nets or trammel nets are set in littoral areas, perpendicular to shore, and usually extend into sublittoral areas. To reduce size selectivity, an experimental gill net consisting of panels of five different mesh sizes is commonly used. Smaller mesh size (0.5in) is used in shallow areas and up to 2-2.5in mesh farther out.
- Fyke nets, trap nets, and fish traps can be used in shallow areas.
- Trawl and sonar can be used to sample pelagic areas.

Sample Processing—Fish samples are processed as recommended in the RBP manual EPA 440/4-89-001 (USEPA 1989b). Sampling duration and area or distance sampled are recorded in order to determine level of effort. Specimens are identified to species, then total numbers and weights, and the incidence of external anomalies is recorded for each. Voucher specimens of each species from each site are preserved in a 10 percent formalin solution, in a labeled jar. The voucher collections are placed with the state ichthyological museum to confirm identifications and to constitute a biological record. This is especially important for uncommon species, for species requiring verification of identification, and for documentation of new distribution records. If kept in a live well, most fish can be identified and counted in the field by trained personnel and returned to the lake alive. Additional information on field methods is presented in Karr et al. (1986) and EPA 600/R-92/111 (1992b).

Case Study TVA Fish Collection Methods

Shoreline electrofishing samples were collected during daylight hours from inflow, transition, and five forebay zones of most reservoirs from September to mid-November (Figure 7-3). Only one or two zones were sampled on reservoirs where zones were indistinguishable. No inflow zones were sampled in tributary reservoirs.

A total of 15 electrofishing transects, each covering 300m of shoreline, was collected from each of the sampled zones. All habitats were sampled in proportion to their occurrence in the zone. Where conditions permitted experimental gill nets were set overnight in each reservoir zone. Excessive current prevented use of gill nets in mainstream inflow areas. In forebay and transition zones, nets were set in all habitat types, alternating mesh sizes toward the shoreline between sets.

Total length (mm) and weight (g) were obtained for all sport species and channel catfish. Remaining species captured were enumerated prior to release. During electrofishing, fish observed but not captured were included if positive identification could be made and counts were estimated when high densities of identifiable fish were encountered. Young-of-year fish were counted separately and, as in stream IBI calculations (Karr 1981), were excluded from proportional and abundance metrics (due to sampling inefficiencies for the age group). Only fish examined closely to obtain length and weight measurements were inspected externally for signs of disease, parasites, and anomalies.

7.3.6 Sedimented Diatoms

Diatom frustules are preserved in lake sediments that are not disturbed or resuspended. Field sampling for sediment diatoms can be relatively fast. Field methods outlined below and

Table 7-9. Sampling summary for sedimented diatoms.

Habitat	Mid-lake, deep depositional area.
Sampling Gear	Grab (surface only). Corer (paleolimnology).
Index Period	None. Samples may be taken annually, biennially, triennially, etc.
Sampling	Single sample in mid-lake.
Analysis	Samples are divested of organic matter and 300-500 diatom frustules are identified to lowest pratical level.

in Table 7-10 are similar to those used in EMAP (USEPA 1994b).

Sample Location—Sediment samples are obtained from or near the deepest area of the lake. A single core sample is sufficient (Charles et al. 1994).

Sampling and Analysis—Sediment diatoms can be sampled with a corer that is able to reliably sample and retain the top 1 cm of sediment. The top 1 cm of sediment is carefully removed from the sampler and kept at 4°C in a plastic bag. Diatom samples are prepared, enumerated, and identified following the procedure from the EMAP manual (USEPA 1994b).

7.4 TIER 2B: SHORT-TERM INDICATORS (REPEATED SAMPLING)

Tier 2B consists of phytoplankton and zooplankton sampling in addition to Tier 1B sampling, and is conducted at the same sites and times as Tier 1B (Table 7-10). Sampling frequency may range from three samples during the growing season to monthly samples, depending on the objectives of the program. The number of sampling sites is the same as Tier 1B, and samples may be composited among the sample sites to economize laboratory effort, if within-lake spatial variability is not an issue.

7.4.1 Phytoplankton

Phytoplankton are subsampled from the same water sample collected for chlorophyll and nutrients in the Tier 1 sampling protocol. The water sample may be a surface sample or an epilimnion or photic zone hose sample. The large sample is mixed thoroughly before subsamples are taken from it.

A sample of 150 to 500ml is sufficient for phytoplankton. The phytoplankton sample is preserved in the field with Lugol's solution (APHA 1992). Cells are identified and counted using the Utermohl method on an inverted microscope, or by filtration onto a membrane filter (APHA 1992). The Utermohl method requires settling chambers and an inverted microscope, and the filter method requires a compound microscope and filtering apparatus.

7.4.2 Zooplankton

Sampling Procedure—Zooplankton are sampled with a vertical tow at the same sites as phytoplankton, trophic state, and water quality (Table 7-11). Nets of 118μm mesh and 30 cm diameter will sample most crustacean zooplankton. The net should be equipped with a cone to prevent spill and escape of active organisms. Zooplankton are anesthetized with carbonated water, and preserved in 4 percent formaldehyde. After fixing, long-term storage should be in 70 percent ethanol.

Analysis—The sample is split until 100 to 200 organisms remain in the subsample. Zooplankton are identified to genus; equipment includes dissecting microscope and keys. Lengths of *Daphnia* are recorded.

7.4.3 Periphyton

Periphyton should be sampled several times during the growing season: certain species might be dominant depending on the time of year. Field methods are outlined below and summarized in Table 7-12 (after Bahls 1993).

Sampling Location—A minimum of two random sampling points along each transect is suggested; a determination of greater sampling effort should be based on lake size and professional judgment.

Sample Collection—Collection can be from natural or artificial substrates depending on the preference of the investigation team or agencies. Natural substrates include rocks, logs, macrophytes, and mud. A composite sample of three to five substrates (e.g., fist-sized rocks) is obtained from each sample site. The area scraped from each substrate should be approximately equal. Use a pocket-knife or similar tool for scraping solid substrates. A spoon or large-bore eyedropper can be used for lifting microalgae from mud or silt substrates. Macroalgae can be picked by hand. Epiphytic algae can be dislodged from macroalgae, moss, and aquatic macrophytes by placing a portion of the higher plant in the sample container and

Table 7-10. Tier 2B: Water column biological sampling.

	Component	Data Collection	Responds to or Indicator of
Habitat Components	1. Watershed land use, population, NPDES.	Tier 0 habitat.	
	2. Lake physical.	Tier 1 habitat.	
	3. Shorezone habitat assessment.	3-10 transects: - land use - bank stability - riparian vagetation - emergent vegetation	
	4. Water quality DO seasonal or annual mean, % depth-time Mean pH, alkalinity Secchi depth	Tier 1B water quality (seasonal average).	Trophic state, turbidity.
Biological Components	5. Algal chlorophyll *a*.	Tier 1B chlorophyll (seasonal average).	Trophic state.
	6, 7. Assemblages (minimum 2):		
	a. Phytoplankton	Surface samples (0.5 m) or integrated samples (hose) identify to genus; count 100-500 cells.	Trophic state acidity, metals, water column toxicity.
	b. Zooplankton	Vertical tows; identify to genus; count 100-200 organisms, measure cladocerans.	Trophic state, contamination, trophic imbalance.
	c. Periphyton		

shaking vigorously. The moss or macrophyte is then removed and discarded (Bahls 1993).

Sample Preservation—Preserve samples in watertight, unbreakable jars. Water is added from the sample site to cover the sample; then enough Lugol's solution is added to impart a reddish-brown tint. Artificial substrates can be preserved intact in a suitable container or scraped in the field.

Sample Preparation—Extracellular organic matter is decomposed by oxidation, leaving only the diatom shells (frustules) as described in APHA (1992). Using the cleaned diatoms (frustules), a permanent mount is prepared and a proportional count is made of 300 to 500 cells (APHA 1992). Counts for each species are divided by the total count and multiplied by 100 to obtain percent relative abundance (PRA).

7.5 DIAGNOSTIC HABITAT SURVEY

More detailed habitat procedures allow monitoring agencies to focus on specific water and

Table 7-11. Sampling summary for crustacean zooplankton.

Habitat	Open water, 1 to 5 sites in lake.
Sampling Gear	Plankton net, 300mm (12in) mouth; 118mm mesh.
Index Period	Mid-summer index.
Sampling	Single vertical tow through water column from 0.5m above bottom to surface.
Analysis	Tier 1: Identify to species, measure *Daphnia*. 100 to 200 organisms.

Table 7-12. Sampling summary for periphytic diatoms.

Habitat	Rock, wood, silt, macrophyte substrates, 0.5 to 1 m depth (wading depth).
Sampling Gear	Spatula, toothbrush for scraping cells from substrates, eyedropper or spoon for lifting (Bahls 1993). Samples preserved in Lugol's solution.
Index Period	Preferred: mid-summer. Alternative: growing season, average of 7 to 10 samples.
Sampling	3-5 substrates (rock, wood, sand, mud, macrophytes) are sampled in the proportion of their occurrence at 3-5 sites around the lake. Single composite sample from all substrates and sites.
Analysis	300-500 diatom frustules are identified to species and enumerated.

sediment quality problems in a lake, and specific land use practices in the watershed, for identification of probable cause of impairment (Table 7-13). Supplemental habitat components may include: a detailed watershed assessment (soils and geology, detailed land use, agricultural practices); a stream assessment for migratory fish habitat; additional water quality analysis (nutrients, contaminants); and sediment quality (sediment grain size, sediment organic carbon, contaminants, toxicity).

Tiers 2A and 2B will allow detection of effects of toxic substances on the respective biological assemblages, but will not provide positive identification of toxicity as a probable cause of impairment. Positive identification of contamination and toxicity as a probable cause will require the supplemental survey, particularly habitat contaminant analysis and toxicity assays. The detailed land use measurements in the habitat assessment allow identification of more specific nonpoint source probable causes of impairment. The tiers allow detection of biological effects on at least two assemblages, and hence detection of effects at multiple levels (including cascades of effects).

7.5.1 Watershed and Shorezone Components

The diagnostic habitat survey is similar to the Tier 1 and Tier 2 habitat survey but evaluates more detailed components. In searching for probable causes of impairment, land use is broken down into more detailed land use categories, including high- and low-density residential, industrial and commercial transportation, cropland, pasture, orchard, mines, etc. If agriculture is thought to contribute to impairment, then the dominant agricultural practices should be documented, as well as their distribution in the watershed. If the fish assemblage shows impairment (particularly migratory fish), then fish spawning habitat in inflowing streams can be evaluated.

7.5.2 Sediment Analyses

The Sediment Classification Methods Compendium (USEPA 1992f) discusses various aspects of sediment analyses including sample collection and handling, quality assurance/quality control issues, and toxicity testing. In addition, this guide furnishes references for specific methods.

Sampling

There are three main types of devices used to collect sediment samples. The choice of sampler to be used for a particular study depends upon the nature of the sample needed. Grab samplers and core samplers can be used in toxicity testing and in evaluating chemical and physical properties of the sediment. Additionally, cores can be used in evaluating historical sediment records.

Equipment should be thoroughly cleaned between samples to prevent cross contamination. In some cases, preservation methods such as pH control or addition of chemical preservatives will need to be done. Standard methods for sample handling can be found in ASTM (1990).

Sediment Particle Size

Sediment particle size is measured using stacks of different sized sieves. The sediment to be analyzed is first heated to dryness. Samples may need to be stored cold, frozen, or preserved.

Table 7-13. Supplemental components.

Component	Data Collection	Responds to or Indicator of
1. Watershed - Soil and bedrock characteristics - Hydrology - Agricultural practices - Detailed land use categories (roads, mines, impervious surface, cropland, pasture, etc.)	Maps; survey of state and county agencies.	Physical classification. Probable cause.
2. Shore - Migrating fish spawning habitat	Tributary stream habitat survey.	Disturbance, habitat destruction.
3. Sediment quality - Toxicity, contaminants, total organic carbon, particle size	Annual grab in depositional environment (deepest point).	Exposure to toxics, contaminants.

Then a known weight of dried sediment is poured into a stack of sieves of different sizes to separate the particles. Each size fraction is then weighed and expressed as a percentage of the total dry sample weight.

Sediment Contamination

Chemical analyses that can be measured include metals, polyaromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), pesticides, and volatile and semivolatile organic pollutants. Metals are typically measured using atomic absorption spectrophotometry. Other constituents should be analyzed using USEPA approved methods (USEPA 1991f, ASTM 1990). Although it is not a contaminant, total organic content (TOC) should also be analyzed since it is an important indicator of the bioavailability of nonionic hydrophobic organic pollutants. Likewise, acid volatile sulfides (AVS) are important in determining the bioavailability of metals.

Sediment Toxicity Evaluation

Several approaches are recognized by USEPA for evaluating sediment toxicity. These approaches may be used separately or in combination to provide evidence of toxicity and to generate sediment quality criteria. (USEPA 1994j).

Whole (bulk) sediment toxicity testing is a method of evaluating the level of toxicity of a sediment sample. Typically, test organisms are exposed to sediment for 10 to 14 days. End-points used are growth and survival. The most often used organisms in freshwater sediment toxicity tests are the amphipod *Hyalella azteca* and larvae of the midge *Chironomus tentans*. Other organisms that have been tested include other benthic infauna such as the mayfly *Hexagenia* spp; and the worms *Tubifex tubifex* and *Lumbriculus variegatus*; and two cladocerans, *Daphnia magna* and *Ceriodaphnia dubia*. Results of exposure to contaminated sediments is compared with control (uncontaminated) sediments (USEPA 1994j, ASTM 1998, PSEP 1995, Environment Canada 1994).

In This Chapter...

- *Characterization*
- *Metric Selection*
- *Index Development*

Chapter 8

Index Development

8.1 OVERVIEW

The approach taken here for development of an index for assessment is called the multimetric approach. Biological attributes, or metrics, are calculated from the measurements. A score is assigned to each metric corresponding to its deviation from the expected value in reference sites. The multimetric index is the sum of all the metric scores. A separate index is developed for each assemblage sample (e.g., macrophytes, benthic invertebrates, fish).

The multimetric approach has been successfully applied to assessment of stream fish assemblages (Karr 1981, Karr 1991, Karr et al. 1986) and stream invertebrate assemblages (Ohio EPA 1987, USEPA 1989b, Barbour et al. 1995, Yoder and Rankin 1995). The approach appears to be statistically robust (Fore et al. 1994) and is straightforward to apply. Alternative methods of analysis and assessment are discussed in Appendix E.

Development of a multimetric index is the final step toward operational bioassessment. Three steps are necessary for development of an index: characterization of reference conditions, evaluation and final selection of metrics, and multimetric index building.

The basis of the multimetric approach is comparison of a metric to an expected (reference) distribution of values and a judgement of whether the value is within the expected range. Each metric is given an ordinal score of 5, 3, or 1, depending on whether it is similar to reference values (within the expected range), is somewhat different, or is very different, respectively (Figure 8-1). The expected range is usually expressed as a percentile of the reference distribution. Two methods of scoring are commonly used. The first is based on a lower percentile of a representative sample of reference sites (Figure 8-1a). The second method is used if predetermined reference conditions are not definable or if there are too few reference sites, and it is preferred for defining reference conditions for reservoirs. In the first method (Figure 8-1a), the 25th percentile of the reference site distribution is often used as the dividing line between optimal (similar to reference) and less than optimal. In the second method (Figure 8-1b), the 95th percentile of the entire population distribution is often used as the reference mark for trisecting metric values (e.g., Karr et al. 1986).

Development of a multimetric index is the final step toward operational bioassessment.

The index consists of the sum of all metric scores, and the total index value of a site is compared to the distribution of index values in reference conditions. Development of an index thus requires characterization of reference conditions to obtain the distributions of metric values, final selection of metrics based on metric response to stressors, and, finally, characterization of the index distribution in reference conditions.

Selection of metrics and development of a multimetric index requires a test data set composed of reference sites and nonreference (test) sites. The best sites may be impaired or may simply not meet the criteria for reference sites. Ideally, the test sites should include at least some lakes that are severely impaired by different stressors. If, for example, all test sites are eutrophic lakes, then the response of metrics to other stressors cannot be determined. Reference condition characterization uses only the reference site data—metric evaluation and index development use both reference and test site data.

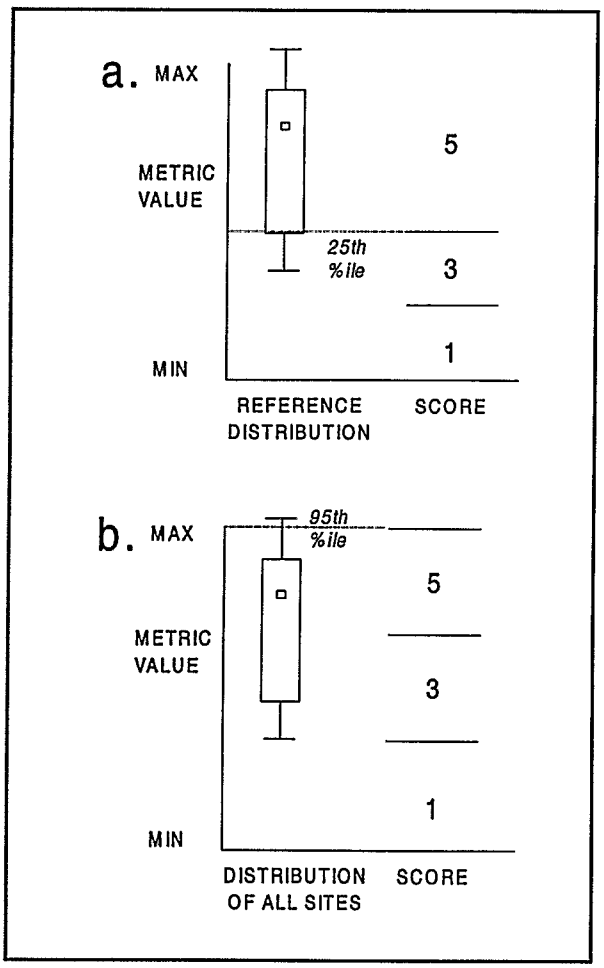


Figure 8-1. Basis of bioassessment scores — unimpaired reference sites; population distribution.

8.2 CHARACTERIZATION OF REFERENCE CONDITION

The objective of reference characterization is to finalize the classification of the reference sites and to describe (characterize) each of the lake categories in terms of metrics and other descriptive variables.

Several statistical tools can assist in the classification of sites, but there is no one set procedure. If the preliminary classification is relatively certain (based on well-developed prior knowledge and professional judgment, and graphical analysis of metrics) followed by necessary modifications and tests of the resultant classification, is usually sufficient to finalize the classification. If the preliminary classification is less certain, it might be necessary to develop a classification from the data, using one of several classification methods. These methods include cluster analysis and several ordination methods (e.g., principal components analysis, correspondence analysis, multidimensional scaling; Appendix E). Ordination is also useful for visualizing alternative *a priori* classification schemes.

8.2.1 Graphical Analysis

A key analysis method for biological metrics is graphical displays using box-and-whisker plots (e.g., Figure 8-1). In the form used here, the central point is the median value of the variable; the box shows the 25th and 75th percentiles (interquartile range); and the whiskers show the minimum to the maximum values (range). A common alternative is whisker extending to values within the "inner fence" (see Tukey 1977 for explanation); this method also plots outliers. Box-and-whisker plots are simple, straightforward, and powerful, and the interquartile ranges are used to evaluate whether there is a real difference between two areas and whether a metric is a good candidate for use in assess-

ment. Graphing the data should always be a first step in data analysis.

Statistical methods used by biologists are frequently tests of whether two or more populations have different means using t-tests, analysis of variance, or various nonparametric methods. However, the fundamental problem of biological assessment is not to determine whether two populations (or samples) have a different mean, but to determine whether an individual site (lake) is a member of the least-impaired reference population. If it is not, then a second question is how far it has deviated from that reference. Therefore, biological assessment requires the entire distribution of a metric, which is effectively displayed with a box-and-whisker plot.

In operational bioassessment, metric values below the lower quartile of reference conditions are typically judged impaired (e.g., Ohio EPA 1990). The actual percentile chosen (25, 10, or 5) is arbitrary and reflects the amount of uncertainty a monitoring program can tolerate.

8.2.2 Characterization

The preliminary classification is refined through inspection of plotted data (graphical analysis), professional judgment, and statistical tests of final classification hypotheses. First, the values and distribution of metrics are compared among ecoregion or lake type. Regions that appear to be similar to each other can be lumped together for final classification. For two regions to be lumped, most of the metric distributions must be similar. In addition to box plots of metrics, it is also useful to examine scatter plots of selected metrics and habitat variables such as lake size, salinity, or alkalinity. The number of taxa in a waterbody is often dependent on its size, for example, large lakes have more zooplankton species than small lakes (Dodson 1992). Salinity also influences the number of species found in aquatic systems, as do pH and alkalinity.

Refining the Classification

In sampling fish from reservoirs of the Tennessee Valley Authority, the number of fish species was found to vary by reservoir class and ecoregion (Hickman and McDonough 1996). Figure 8-2 (after Hickman and McDonough 1996) shows the number of fish species in different parts of four groups of TVA reservoirs. First, the number of fish taxa is relatively homogenous between forebay, transition, and inflow zones (Figure 8-2). The reservoir types differ in number of fish species, with the mainstream reservoirs having the most species, and the Blue Ridge reservoirs being relatively depauperate. Based on number of species, the Interior Plateau reservoirs are not significantly different from Ridge and Valley reservoirs, and TVA reservoirs could be considered to be in three groups (dotted lines). However, on the basis of other considerations, TVA has kept Interior Plateau reservoirs

The fundamental problem of biological assessment is not to determine whether two populations have a different mean, but to determine whether an individual site is a member of the least-impaired reference population.

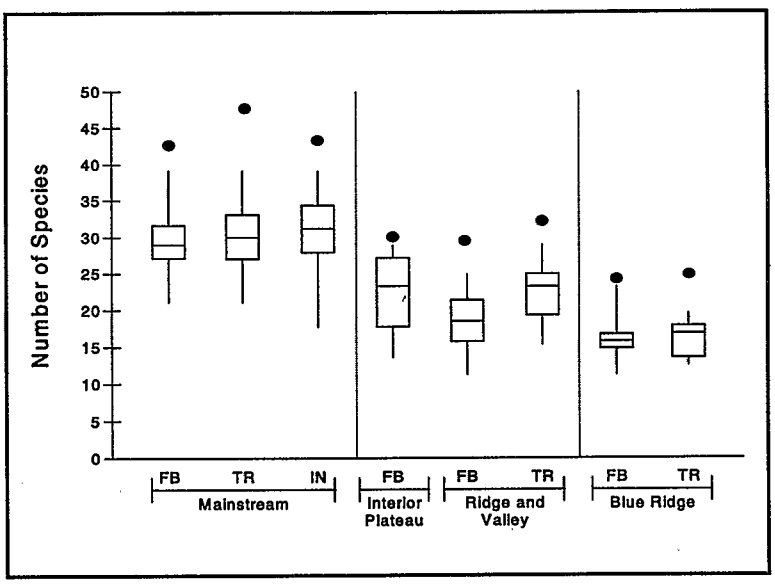


Figure 8-2. Species richness in TVA reservoirs (redrawn from Hickman and McDonough 1996.) Four reservoir classes are shown (mainstream, Interior Plateau, Ridge and Valley, and Blue Ridge). Dashed lines delineate three classes based on species richness alone. FB = forebay; TR = transition; IN = inflow.

separate from Ridge and Valley reservoirs.

Refining the Classification-Covariates
Certain physical or chemical attributes can have a strong influence on biological metrics, especially number of taxa metrics. The most important of the physical-chemical attributes to test are lake size, salinity (in arid regions), and alkalinity or pH. The example (Figure 8-3) shows number of taxa of benthic macroinvertebrates as a function of salinity in the littoral zone of Montana lakes and wetlands (Stribling et al. 1995). Finding a relationship as in Figure 8-3 requires adjusting reference expectations as a function of the covariate salinity in this case.

8.3 INDEX DEVELOPMENT

Following classification and characterization of reference conditions, metrics are evaluated for suitability in a multimetric index. Suitable metrics are those that respond in a predictable way to stressors on the system and that have low noise or variability.

8.3.1 Metric Variability

Metrics that are too variable within the reference sites are unlikely to be effective for assessment. A measure of metric variability is the ratio of the interquartile range to the distance between the lower quartile and the minimum possible value of the metric.

In operational bioassessment, metric values below the lower quartile of reference conditions are typically judged as not meeting reference expectations (e.g., Ohio EPA 1990). The range from 0 to the lower quartile can be termed a "scope for detection." For those metrics with low values under reference conditions and high values under impaired conditions, the scope for detection is the range from the 75th percentile to the maximum possible value (e.g., 100 percent) (Figure 8-4). The larger the scope for detection, compared to the interquartile range, the easier it will be to detect deviation from the reference condition. The "interquartile coefficient" is thus defined here as the ratio of the interquartile range to the scope for detection. The interquartile coefficient is analogous to the coefficient of variation and is used the same

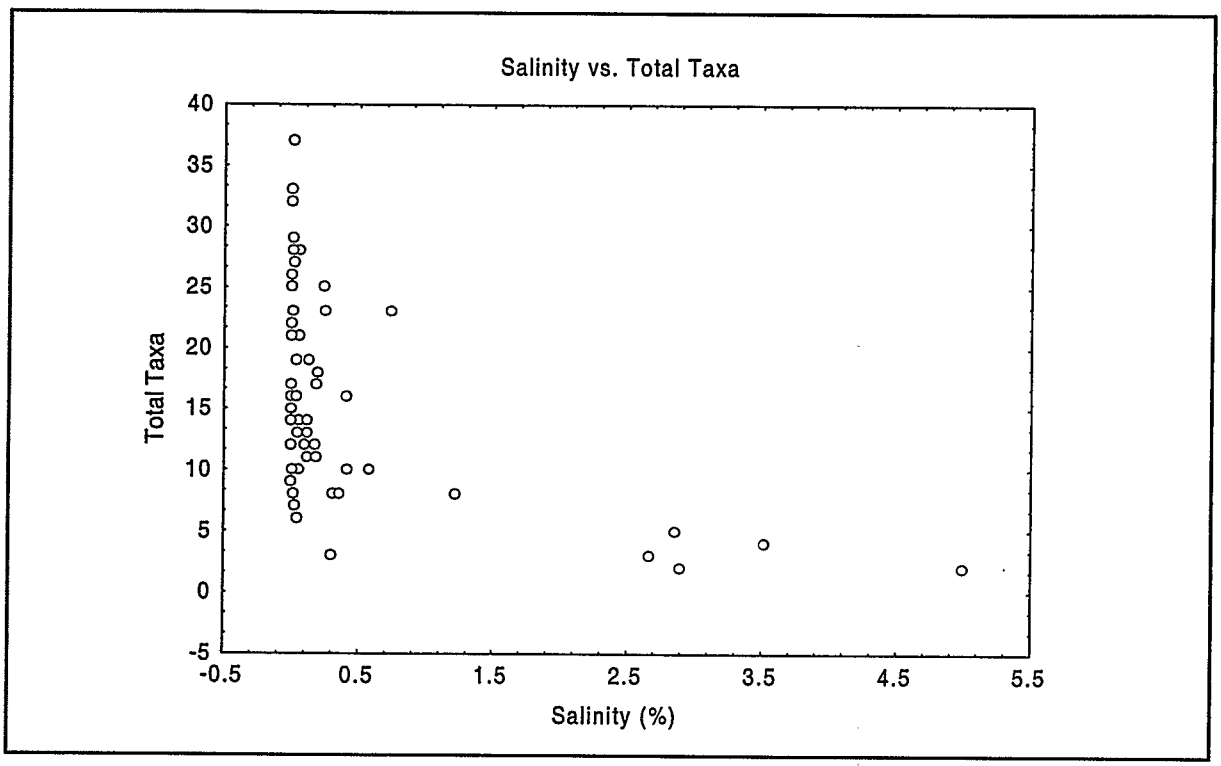


Figure 8-3. Benthic macroinvertebrate taxa richness in littoral zone of Montana lakes and wetlands.

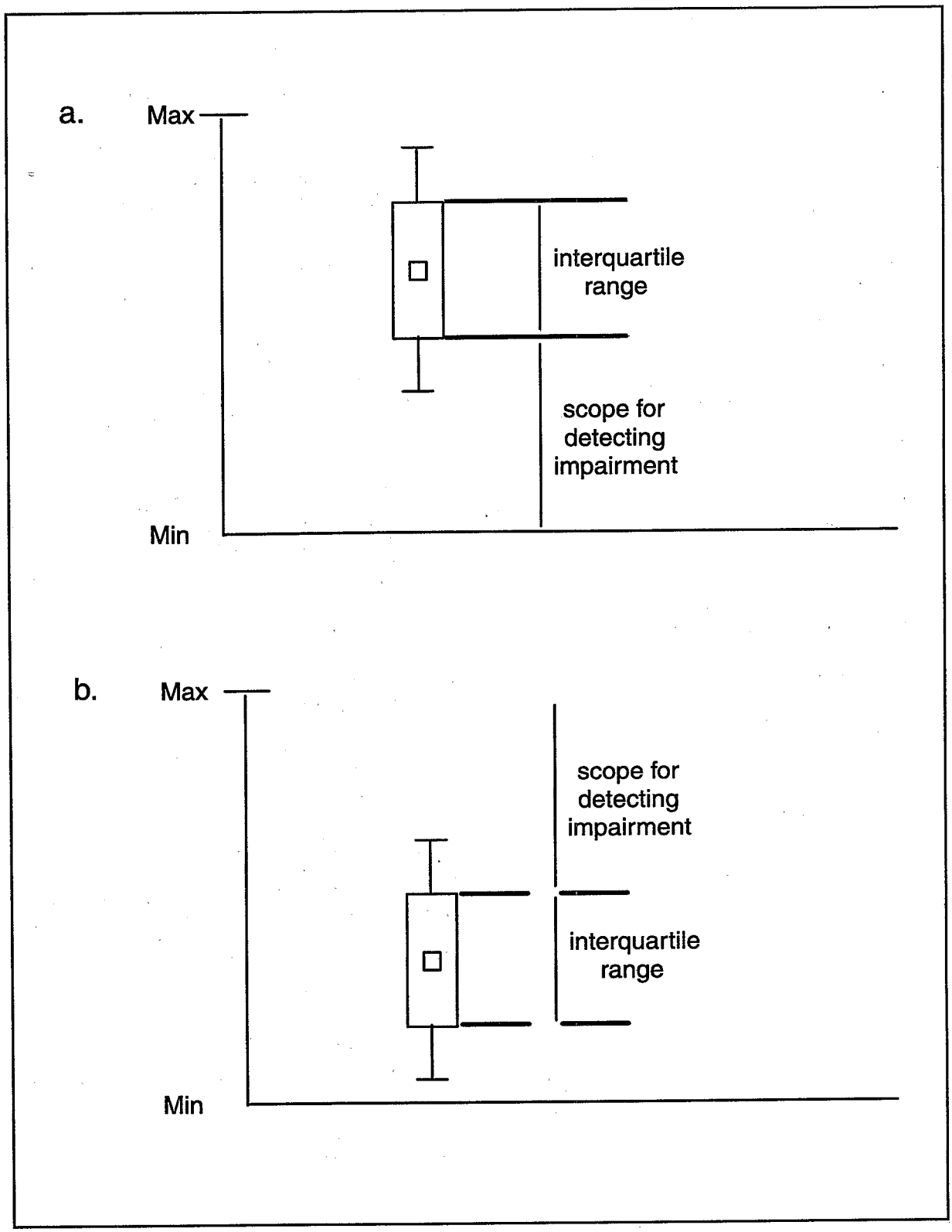


Figure 8-4. Assessing candidate metrics. a. Metrics that have high values under unimpaired conditions. b. Metrics that have low values under unimpaired conditions.

way, but it is bidirectional and uses percentiles in the same way that assessment uses percentiles. In general, an interquartile coefficient greater than 1 indicates excessive variability of a metric.

8.3.2 Metric Response

Response of metrics to stresses is evaluated by comparison of reference sites to test sites. The simplest comparison is using box-and-whisker plots of the metric distribution in reference and test sites (Figure 8-5). Alternatively, it may be possible to develop an empirical model of metric response to stressors. Several approaches are available including multiple regression, canonical correlation, canonical correspondence analysis, and log-linear models (Ludwig and Reynolds 1988, Jongman et al. 1987). For multivariate model building, refer to the above references or any statistical software package—it will not be outlined further in this document.

Metrics are judged responsive if there are significant differences in central tendency or in variance between reference and test sites (Figure 8-5). If the test sites are known to be impaired, then the mean or median values should be significantly different (Figure 8-5). If the test sites are simply lakes that do not meet reference criteria (i.e., they might be a mix of impaired and unimpaired lakes; shown as "unknown test sites" in Figure 8-5), then the variance in the test sites should be larger than that in the reference sites.

Variability and Uncertainty

Variability in values of measurements and metrics results in uncertainty of the assessment. Uncertainty can be reduced by increasing the sampling effort (repeated measurement) to obtain a better estimate of the mean value. This is especially important for the measurements that are the most variable: chlorophyll, nutrient concentration phytoplankton and zooplankton. Algal abundance and biomass may vary tenfold within the growing season (i.e., Wetzel 1975, Hecky and Kling 1981). A tenfold change in chlorophyll corresponds to 22.6 points in the TSI range, a substantial change.

Because of this variability, Tier 1A is unreliable for assessment of an individual lake and Tier 2A is recommended. Tier 1A is appropriate for assessing a class of lakes or a region, to answer questions such as: what is the status of lakes in the region, or how many lakes are oligotrophic?

As long as many lakes are sampled, the effect of errors in individual lakes is reduced in the evaluation of all lakes.

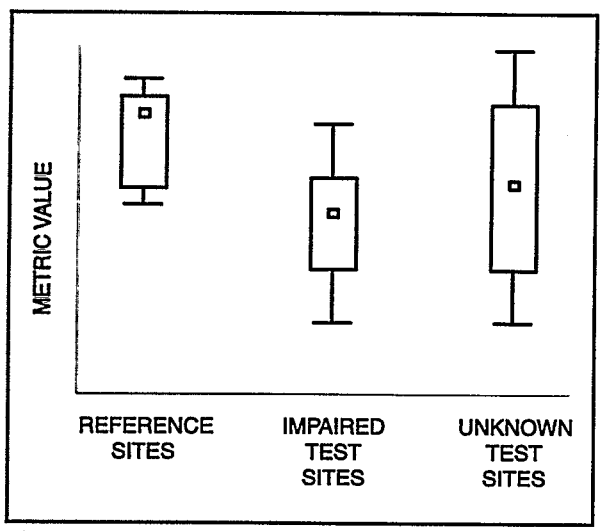


Figure 8-5. Responsiveness of metrics. A large difference between reference and impaired test sites indicates a responsive metric. Unknown sites are a mixture of impaired and unimpaired sites.

Metrics that are responsive to known or unknown stresses are retained for index development. Finally, responsive metrics are evaluated for redundancy. A metric that is highly correlated with another metric might not contribute new information to the assessment. Pairs of metrics with correlation coefficients greater than 0.9 should be examined carefully to determine whether both metrics are necessary. Often, strongly correlated metrics are calculated from the same raw data, or their method of calculation ensures correlation. For example, Shannon-Wiener diversity and percent abundance of the dominant taxon are strongly correlated in any data set.

A correlation alone (say, r >0.6) is not sufficient to eliminate one of a pair of correlated metrics. Some metrics might be sensitive only at severe or moderate stress; others might be sensitive across the entire range of stresses (Karr 1991). These would all contribute information, in spite of strong correlation. A scatterplot of correlated metrics is examined; if there is an apparent

nonlinear or curved relationship, then both should be retained. If the points all fall close to a straight line, then one of the metrics can be safely eliminated.

8.3.3 Scoring and Index Development

Combining unlike measurements is possible only when the values have been standardized by a transformation through which measurements become unitless (Schuster and Zuuring 1986). Standardization of these measurements into a logical progression of scores is the typical means for comparing and interpreting unlike metric values.

Two methods are commonly used for scoring metrics, which are based on the metric distribution in defined reference sites or in the population of sites, respectively. Each metric is given a score of 1, 3, or 5, corresponding to impaired, intermediate, or unimpaired biota, respectively (Figure 8-1).

Bisection scoring—(Figure 8-1a) Based on a lower percentile of the reference distribution; for example, the 25th percentile (Barbour et al. 1996b). In this method, values above the 25th percentile are considered unimpaired (similar to reference conditions) and values below the 25th percentile are considered impaired to some degree. The range from 0 to the 25th percentile is bisected, with values in the top half receiving a score of 3 and those in the bottom half receiving a score of 1 (Figure 8-1a).

Trisection scoring—(Figure 8-1b) Based on the 95th percentile of the population distribution (Karr et al. 1986). Metric values from 0 (or the lowest possible value) to the 95th percentile are trisected; values in the top one-third receive a 5, values in the middle third receive a 3, and values in the bottom third receive a 1 (most impaired).

The scoring method should reflect how well the reference sites represent unimpaired conditions. If reference sites are unimpaired and considered to be representative, bisection is recommended (Figure 8-1a). This method assumes that the reference sites are representative of relatively unimpaired conditions and that the metric distribution reflects natural variation of the metric. A value above the cutoff is then assumed to be similar to reference conditions. The lower quartile (25th percentile) is most frequently taken as the cutoff (e.g., Barbour et al. 1996b).

Two methods are commonly used for scoring metrics which are based on the metric distribution in defined reference sites or in the population of sites respectively.

The trisection method (Figure 8-1b) is best for scoring in regions where impacts might be so pervasive that nearly all reference sites are thought to be impacted or for assessment of reservoirs where reference sites cannot be defined. In trisection, it is assumed that at least some reference lakes attain an excellent value for the metric, but that many reference lakes are impaired and hence the lower limit of the reference distribution is not known. The 95th percentile is taken as the "best" value, and the range is trisected below it (Figure 8-1b).

Choice of scoring method should be based on confidence in the reference sites, rather than on the method that will produce the most conservative or most liberal scoring. If confidence is high that reference sites are representative of relatively unimpaired conditions, then the lower percentile cutoff and bisection are preferred. If confidence is low, then trisection below the 95th percentile is preferred.

If covariates such as lake size determine metric values, then the scoring should be adjusted for the covariates. Reference data are plotted as in Figure 8-6, and a locally weighted estimate is made of the appropriate percentile (95th or 25th) and the range below it is trisected or bisected accordingly. Figure 8-6 shows total zooplankton taxa in North American lakes ranging in size from $4m^2$ to nearly $10^{11}m^2$ (Lake Superior) (Dodson 1992). Few state assessment programs are likely to include lakes smaller than 10^4m^2 (1ha; 2.47 acres), nor larger than 10^9m^2 ($1000km^2$; 247,000 acres). In this example, considering only the middle range from 10^4m^2 to 10^9m^2, the slope is not apparent and adjusting for lake area would not be necessary.

The index is the sum of the scores of the selected metrics. The number of metrics in an index affect the variability of the index—those with more metrics tend to be less variable (Karr 1991). Index values are evaluated by compari-

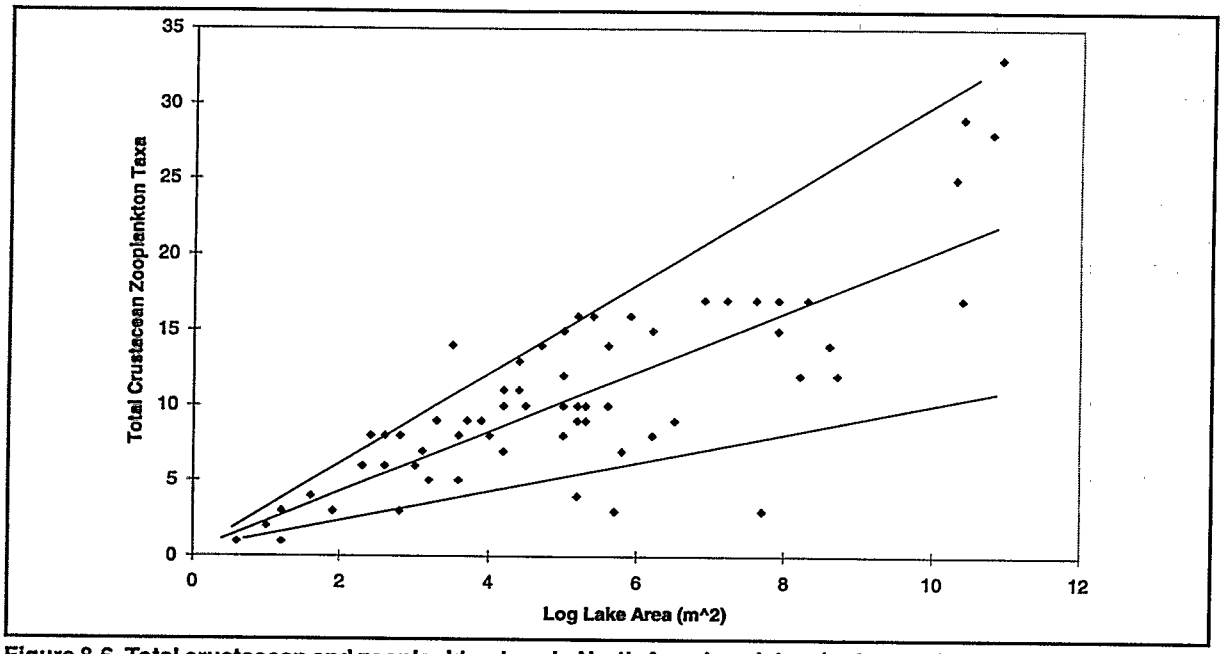


Figure 8-6. Total crustacean and zooplankton taxa in North American lakes (redrawn after Dodson 1992). If metrics show a relationship such as this with area, elevation, or some other physical covariate, then reference expectations must be adjusted to the covariate. The three lines show one possible method for scoring. In practice, most state assessment programs are not likely to span 10 orders of magnitude in lake area.

son to index values of the reference sites. Even the best reference sites do not receive perfect scores of the index. The final index scores are compared to the distribution of scores in the reference sites. Criteria for assessment are based on the distribution of index scores in reference sites. Those that correspond to the range of index values in reference sites support life use; those that are clearly below index values in reference sites do not support life use. Following appropriate review and revision, they can be established as biocriteria.

Choice of scoring method should be based on confidence in the reference sites rather than on the method that will produce the most conservative or most liberal scoring.

8.4 LAKE TIER INDICES

An index is calculated for each assemblage sampled. Each tier has three to six indices, which should all be reported. The indices can be summed into an overall lake index, which can be used to report overall condition but would not reveal the condition of the component assemblages. Indices within each tier might or might not be multimetric; Tier 1 indices are primarily single metrics, whereas indices of Tiers 2A and 2B might be composed of 3 to 12 metrics.

8.4.1 Tier 1

Tier 1 assessment consists of trophic state algal growth potential and macrophyte indices. Three TSI (chlorophyll, Secchi depth, and total phosphorus) are recommended; the fourth (total nitrogen) is also recommended in regions where nitrogen is suspected to be a limiting nutrient for algal growth. The TSI and AGPT are scored as metrics for their similarity to reference conditions, and the scores are summed for a "Trophic Reference Index."

The trophic metrics are unique in that they may be scored lower if their values are substantially higher as well as lower than reference values. For example, an unproductive (oligotrophic) lake in a region where lakes are expected to be productive (mesotrophic) would be given a lower score.

Tier 1 has two or more submerged macrophyte metrics, percent cover of macrophytes, and

dominance of exotic species. More metrics can be developed if macrophyte species are identified and relative abundances are estimated. Percent cover is scored by comparison to reference expectations, but dominance of exotic species is rated 5 if none are present, 3 if exotics are subdominant, and 1 if exotics are dominant. The two macrophyte metrics are summed for the Tier 1 macrophyte index.

Lakes are assessed from the scores of the two Tier 1 indices. Tier 1A and Tier 1B use the same metrics and indices; Tier 1B trophic metrics are estimated from seasonal mean measurements. Biocriteria can be established for further investigation or remedial action, based on the scores.

8.4.2 Tier 2A

Tier 2A assessment may consist of three to five indices:

- Trophic reference index of either Tier 1A or 1B.
- Macrophyte index of Tier 1 or a more detailed Tier 2A macrophyte index.
- Benthic macroinvertebrate index.
- Fish assemblage index.
- Sedimented diatom index.

The macroinvertebrate, fish, and diatom indices are developed from metrics as described in Chapter 6.

8.4.3 Tier 2B

Tier 2B consists of three to five indices:

- Trophic reference index of Tier 1B (seasonal averages).
- Macrophyte index of Tier 1.
- Phytoplankton index.
- Zooplankton index.
- Periphyton index.

The phytoplankton, zooplankton, and periphyton indices are developed from metrics as described in Chapter 6.

Case Study: TVA Scoring Criteria and Index Development

Chlorophyll *The classification scheme used to develop expectations for chlorophyll in Tennessee Valley reservoirs was based on the "natural" nutrient level in a watershed. Professional judgment was used to select concentrations considered indicative of good, fair, and poor conditions. Reservoirs were placed into one of two classes for chlorophyll expectations: those expected to be oligotrophic because they are in watersheds with naturally low nutrient concentrations, and those expected to be mesotrophic because they are in watersheds which naturally have greater nutrient availability. The reservoirs expected to be oligotrophic are those in the Blue Ridge Ecoregion. The remaining reservoirs, both mainstream reservoirs and tributary reservoirs, are expected to be mesotrophic.*

The range of concentrations selected to represent good, fair, and poor conditions is much lower for reservoirs in nutrient-poor areas (e.g., Blue Ridge) than for the other reservoirs. For reservoirs expected to be mesotrophic, the concern is that chlorophyll levels not become too great because of the associated undesirable conditions—dense algal blooms, poor water clarity, low DO, and noxious blue-green algae. Conversely, in cases where sufficient nutrients are available but chlorophyll concentrations remain low, there is likely something hindering this natural process. This is the reason for identifying a minimum level for the "good" range of expectations for mesotrophic reservoirs.

Sediment Quality *The sediment quality scoring criteria uses sediment chemical analyses for ammonia, heavy metals, pesticides, and PCBs.*

Benthic Macroinvertebrates *Seven assemblage characteristics (or metrics) were selected to evaluate the benthic macroinvertebrate assemblage. Six of the metrics are an average of the 10 samples taken at each site.*

1. *Number of taxa.*
2. *EPT taxa.*

Case Study: TVA Scoring Criteria and Index Development (Continued)

3. *Percent of samples with long-lived species.*
4. *Proportion as Tubificidae.*
5. *Proportion as two dominant taxa.*
6. *Total abundance excluding Chironomidae and Tubificidae.*
7. *Percentage of samples with no organisms present.*

Scoring criteria for each of the seven metrics were developed using the 5 years of Vital Signs monitoring data (1994-1996). Scoring ranges were developed as follows:

- *Individual criteria were developed for each type of sampling location (forebay, transition zone/mid-reservoir, embayment and inflow) for each of the four classes of reservoirs.*
- *Results from the 10 samples along a transect for each sample year were combined (averaged for most metrics) and outliers deleted.*
- *The range of average values was then trisected; with the upper one-third of the range representing desirable conditions assigned a value of 5 (good), the middle one-third assigned a 3 (fair), and the lower one-third representing undesirable conditions assigned a 1 (poor).*

Professional judgment and supplementary statistical analyses were used to adjust the cutoffs for each range as appropriate. Sample results at each site were compared with these criteria for each metric and assigned the rating described above—5 = good; 3 = fair; 1 = poor if they fell within the top, middle, or bottom group, respectively. Numerical ratings for the seven metrics were then summed. This resulted in a minimum score of 7 if all metrics at a site were poor, and a maximum score of 35 if all metrics were good.

Reservoir Fish Assemblage Index *The current RFAI uses 12 fish assemblage metrics from five general categories, including:*

Species Richness and Composition

1. *Total number of species.*
2. *Number of piscivore species.*
3. *Number of sunfish species.*
4. *Number of sucker species—suckers are also insectivorous but inhabit the pelagic and more riverine sections of reservoirs.*
5. *Number of intolerant species.*
6. *Percentage of tolerant individuals (excluding young-of-year).*
7. *Percentage of dominance by one species.*

Trophic Composition

8. *Percentage of individuals as omnivorous.*
9. *Percentage of individuals as insectivorous.*

Reproductive Composition

10. *Number of lithophilic spawning species.*

Abundance

11. *Total catch per unit effort (number of individuals).*

Fish Health

12. *Percentage of individuals with anomalies (diseases, lesions, tumors, external parasites, deformities, blindness, and natural hybridization).*

Establishing scoring criteria (reference conditions) by trisecting observed conditions requires a substantial data base for each class of reservoir and assumes the data base contains reservoirs with conditions ranging from poor to good for each metric. The smaller the number of reservoirs within a class, the less likely these assumptions can be met and the greater the need for sound professional judgment based on extensive knowledge of the reservoir assemblages being studied.

Because some reservoir classes contained relatively few reservoirs, the approach used to develop scoring criteria for RFAI was to include all sampling results from Vital Signs monitoring (1990-1994). A slightly different approach was used for species richness metrics than for abundance and proportional metrics. For species richness metrics, a list was made of all species collected from comparable locations within a reservoir class from 1990 to 1994. This species list was adjusted using inferences of experienced biologists knowledgeable of the reservoir system, resident fish species, susceptibility of

Case Study: TVA Scoring Criteria and Index Development (Continued)

each species to collection methods being used, and effects of human-induced impacts on these species. This effort resulted in a list of the maximum number of species expected to occur at a sampling location and be captured by collection devices in use. Given that samples are collected once each year, this maximum number of species would not be expected to be represented in that one collection. Therefore, the range from 0 to 95 percent of the maximum was trisected to provide the three scoring ranges (good, fair, and poor). Although 95 percent of the maximum number of species at a site would not be expected to be collected in one sampling event, this "high" expectation was adopted to keep these metrics conservative in light of potential uncertainties introduced by relying heavily on professional judgement.

Scoring criteria for proportional metrics and the abundance metric were determined by trisecting observed ranges after omitting outliers. Next, cutoff points between the three ranges were adjusted based on examination of frequency distributions of observed data for each metric along with professional judgment. In some cases, the narrow range of observed conditions required further adjustment based on knowledge of metric responses to human-induced impacts observed in other reservoir classes. Scoring criteria for the fish health metric are those described by Karr et al. (1986).

To develop metric scores for number of taxa, reproductive composition, and fish health metrics, electrofishing and experimental gill net sampling results were pooled prior to scoring. For abundance and proportional metrics, electrofishing and gill netting results were scored separately, then the two scores averaged to arrive at a final metric value. These scoring criteria separated sites into three categories assumed to represent relative degrees of degradation. Sample results are compared to these reference conditions and assigned a corresponding value: good = 5, fair = 3, and poor = 1.

Overall Assessment

To arrive at an overall health evaluation for a reservoir, the sum of the ratings from all sites are totaled, divided by the maximum potential ratings for that reservoir, and expressed as a percentage. For example, for a small reservoir with only one sample side, the health evaluation would be 20% (all five indicators rated poor—1 for a total score of 5 divided by the maximum possible total of 25) and the maximum would be 100% (all five indicators rated good—5). This same range of 20 to 100 percent applies to all reservoirs regardless of the number of sample sites, and the same calculation process is used.

The next step is to divide the 20 to 100 percent scoring range into categories representing good, fair, and poor ecological health conditions. This has been achieved as follows:

1. Results are plotted and examined for apparent groupings.
2. Groupings are compared to known, a priori conditions (focusing on reservoirs with known poor conditions), and good-fair and fair-poor boundaries are established subjectively.
3. The groupings are compared to a trisection of the overall scoring range. A scoring range is adjusted up or down a few percentage points to ensure a reservoir with known conditions falls within the appropriate category. This is done only in circumstances where a nominal adjustment is necessary.

These methods have been in use for 6 years. Each year slight modifications are made in the original evaluation process and the numerical scoring criteria for each of the five ecological health indicators (Table 8-1) based on experience gained from working with this process, review of the evaluation scheme by other professionals, and results of another year of monitoring. As a result, scoring ranges have changed slightly over the years. Low DO and poor benthos quality contributed most to poor scores among tributary reservoirs in 1994 (Figure 8-7). Reservoir health ratings also differed among ecoregions (Figure 8-8), with run-of-river reservoirs typically scoring highest.

Case Study: TVA Scoring Criteria and Index Development (Continued)

Table 8-1. Example of TVA's computational method for evaluation of reservoirs: Wilson Reservoir 1994 (run-of-the-river reservoir).

Aquatic Health Indicators	Observations*		Ratings*	
	Forebay	Inflow	Forebay	Inflow
Dissolved Oxygen		Tailrace DOs	3.5 (fair)	5 (good)
- Less than 2 mg/L (summer avg.)				
- % of X-sectional area	0.4[5]			
- % of X-sectional bottom length	10.7 [2]**	No		
- Less than 5 mg/L at 1.5 m (Yes/No)	No			
Chlorophyll a mg/L		No samples	3 (fair)	No rating
- Summertime average	13.5	--		
- Maximum concentration	30.0	--		
Sediment Quality		No samples	3 (fair)	No rating
- Toxicity	Rating = 1			
- *Ceriodaphnia* survival	Yes-0% sur.			
- Rotifer survival	Yes-30% sur.			
- Chemistry	Rating = 5			
- Metals/NH_3/pesticides	None			
Benthic Assemblage			2 (poor)	5 (excellent)
- Dominance	1	5		
- Tubificidae	5	3		
- Chironomidae	1	5		
- EPT	1	3		
- Long-lived	1	5		
- No. of taxa	1	5		
- Zero in sample	5	5		
- Non-tolerant density	1	3		
Total	20	34		
Fish Assemblage			4 (good)	3 (fair)
- RFAI	45	40		
Sample Location Sum.			15.5 of 25	13 of 15
Reservoir Sum.			28.5 of 40 (71%)	
Overall Reservoir Evaluation.			**"fair" (yellow)**	

* No samples taken from transition zone

** DO was 0 mg/L on the bottom in forebay

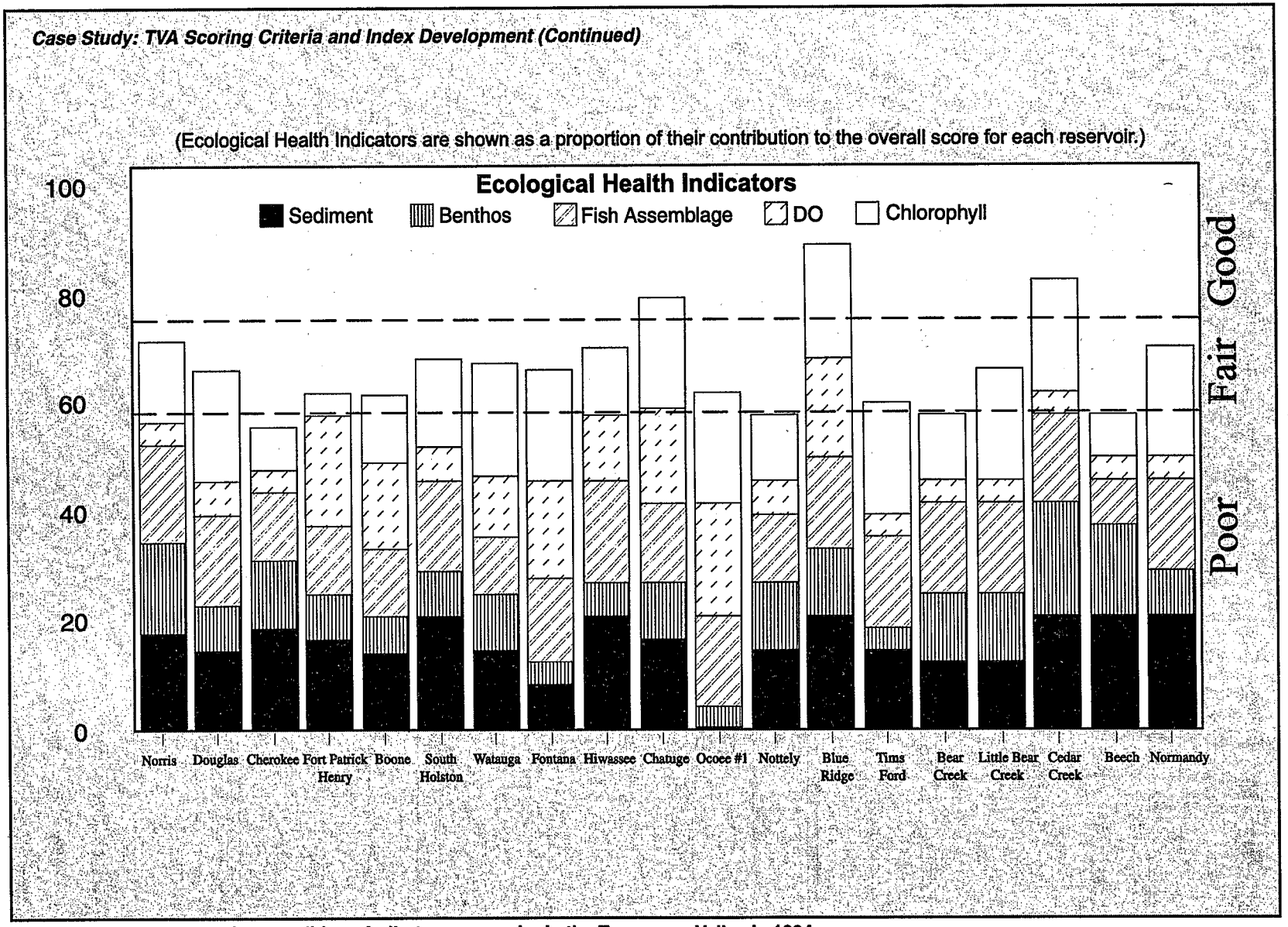


Figure 8-7. Overall ecological condition of tributary reservoirs in the Tennessee Valley in 1994.

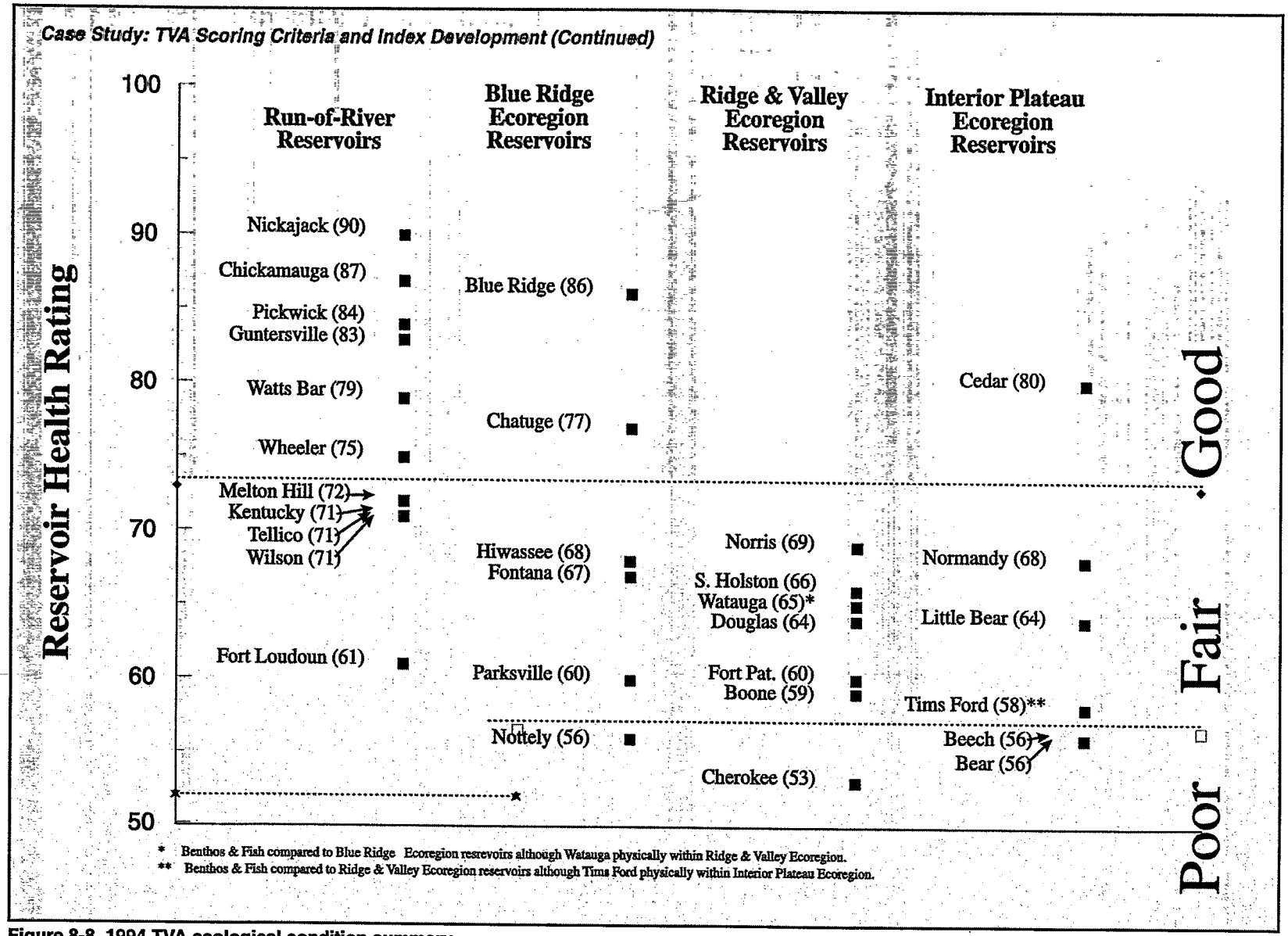


Figure 8-8. 1994 TVA ecological condition summary.

In This Chapter...

- Program Design
- Sampling Design
- Precision
- Quality Assurance
- Operational Quality Control

Chapter 9

Quality Assurance: Design, Precision and Management

Quality assurance (QA) is an integrated program for ensuring the reliability of monitoring and measurement data and includes quality control. Quality control (QC) refers to operational procedures for obtaining prescribed standards of performance in the monitoring and measurement process. Specific QC elements can be developed for most, if not all, project activities. All project activities, from sampling (data collection) and laboratory analysis to statistical analysis and reporting, are potential error sources (Peters 1988). Because error is cumulative and can significantly affect the results of a project, all possible efforts must be made to control it. Therefore, quality assurance is a continuous process that should be implemented throughout the entire development and operation of a program.

The purpose of an overall quality assurance project plan (QAPP), containing specific QC elements and activities, is to minimize—and when possible eliminate—the potential for error. Additionally, there are objective mechanisms for evaluating activities relative to pre-established measurement quality objectives and other project goals. The appropriateness of the investigator's methods and procedures and the quality of the data to be obtained must be ensured before the results can be accepted and used in decision making. QA is accomplished through:

- Program design.
- Investigator training.
- Standardized data gathering and processing procedures.
- Verification of data reproducibility.
- Instrument calibration and maintenance.

As outlined below, QA requirements apply to all activities in an ecological study. More detailed guidance and examples for QA activities should be obtained from USEPA (1994d, 1995, 1996c); more general guidance is outlined by USEPA (1993b).

Quality assurance is a continuous process that should be implemented throughout the entire development and operation of a program.

9.1 PROGRAM DESIGN

A central component of QA is overall study design, which includes formulation of questions and hypotheses, experimental design, and development of analysis approaches. The classical approach by which scientists plan research consists of the following steps:

- Statement of the problem to be resolved.
- Formulation of alternative hypotheses that will explain the phenomena or, in the case of problems that do not involve elaboration of processes, formulation of specific research questions.
- Establishment of boundaries within which to resolve the problem.
- Formulation of an experimental or study design that will falsify one or more hypotheses or answer the specific research questions.
- Establishment of uncertainty limits including setting acceptable probabilities of Type I and Type II errors for statistical hypothesis testing.
- Optimization of the study design including power analysis of the statistical design.

A central component of QA is overall study design, which includes formulation of questions and hypotheses, experimental design, and development of analysis approaches.

Experimental advances in basic sciences have not included the last two steps because uncertainty limits were inappropriate or unknown. Examination of experimental advances also reveals that a high degree of creativity and insight is required to formulate hypotheses and study designs; no formal planning process or "cookbook" can guarantee creativity and insight. Nevertheless, documentation of the planning process and a complete explanation of the conceptual framework help others evaluate the validity of scientific and technical achievements.

9.1.1 Specifying the Questions

The first task in developing a sampling and assessment program is to determine, and be able to state in simple fashion, the principal questions that the sampling program will answer.

The first task in developing a sampling and assessment program is to determine, and be able to state in simple fashion, the principal questions that the sampling program will answer. Questions may or may not be framed as hypotheses to test, depending on program objectives. For example, suppose that a sampling program objective is to establish reference conditions for biological criteria for lakes in state Y. Typically, the initial objectives of a survey designed to develop criteria are to identify and characterize classes of reference lakes. Initial questions may then include:

- Should state Y's minimally disturbed lakes be divided into two or more classes that differ in biological characteristics and dynamics?
- What are the physical, chemical, and relevant biotic characteristics of each of the lake classes?

After state Y's monitoring and assessment program has developed biological criteria, new questions need to be developed that encompass assessments of individual lakes, groups of lakes, or lakes of an entire region or state. Specific questions may include:

- Is lake Z similar to reference lakes of its class (unimpaired), or is it different from reference lakes (altered or impaired)?
- Overall, what is the status of lakes in state Y? How many (or what percentage) lakes are similar to reference conditions? How many lakes are impaired?
- Has lake Z changed over a certain period? Has it improved or deteriorated?
- Overall, have lakes in state Y improved or deteriorated over a certain period? Have individual lakes improved? Are more lakes similar to reference conditions now than some time ago?

Finally, resource managers often wish to determine the relationships among variables, that is, to develop predictive, empirical (statistical) models that can be used to design management responses to perceived problems. Examples of specific questions include:

- Can trophic state of a lake be predicted by areal phosphorus loading rate (e.g., Vollenweider 1968)?
- Can the biota of a lake be predicted by watershed land use (e.g., Dillon et al. 1994)?

These same models (e.g., analysis of variance, regression) are also used to help develop hypotheses on causal relationships between stressors and responses of systems. Establishing cause requires manipulative experiments, and since surveys and monitoring programs preclude experimental investigations, inference of causal relations will not be considered here. Often, there is enough experimental evidence available from other studies so that additional causal experiments are not necessary and would be superfluous (e.g., current knowledge of nutrients and trophic state generally makes it unnecessary to "prove" experimentally which nutrients are limiting).

9.1.2 Specifying the Population and Sample Unit

Sampling is statistically expressed as a sample from a population of objects. In some cases, the population is finite, countable, and easy to specify, e.g., all lakes in state Y, where each lake is a single member of the population. In other cases, the population is more difficult to specify and may be infinite, e.g., lake waters of state Y, where any location in any lake defines a potential member of the population (Thompson 1992). Sampling units may be natural units (entire lakes, cobbles in a littoral zone), or they may be arbitrary (plot, quadrat, sampling gear area or volume) (Pielou 1977). Finite populations may be sampled with corresponding natural sample units, but often the sample unit (like a lake) is too large to measure in its entirety, and it must be characterized with one or more second stage samples of the sampling gear (bottles, benthic grabs, quadrats, etc.)

In most sampling designs, each sample unit is assumed to be independent of other sample units. The objective of sampling is to best characterize individual sample units in order to estimate some attributes (e.g., number of taxa, DO) and the statistical parameters (e.g., mean, median, variance, percentiles) of a population of sample units. The objective of the analysis is to be able to say something (estimate) about the population. It is critical to distinguish between making an inference about a population of many lakes (e.g., "Reservoirs in the Blue Ridge are deep and oligotrophic") versus an inference about a single lake (e.g., "Lake Z has fewer fish species than unimpaired reference lakes"). These two kinds of inferences require different sampling designs: the first requires independent observations of many lakes and does not require repeated observations within sample units (pseudoreplication) (Hurlbert 1984); while the second often does require repeated observations within a lake. Table 9-1 depicts some examples of sample units and populations.

The objective of sampling is to best characterize individual sample units in order to estimate some attributes and the statistical parameters of a population of sample units.

9.1.3 Specifying the Reporting Unit

Finally, it is necessary to specify the units for which results will be reported. Usually, these units are the population (e.g., all lakes), but often subpopulations (e.g., lakes within a given lake district) and even individual locations (e.g., lakes of special interest) can be used. Subpopulations, or strata, are more homogeneous than

Table 9-1. Examples of sample units and populations.

Sample Unit	Sample Population	Infinite or Finite Population
A point in a specific lake.	All points in the lake	Infinite
A point in *any* lake of a state or region.	Total surface area or volume in a state or region	Infinite
A lake or a definable subbasin of a lake as a single unit. *(NOTE: Because lakes are most often discrete environments, this is likely to be the most common sample unit)*	All lakes in a state or region	Finite

the entire population, and are separated to facilitate comparison among them (see Section 9.2.1). In order to help develop the sampling plan, it is useful to create hypothetical statements of results in the way that they will be reported, for example:

- *Status of a place:* Lake Z is degraded.
- *Status of a region:* 20% of the lake area in state Y has an elevated trophic state, above reference expectations; *or* 20% of lakes in state Y have an elevated trophic state.
- *Trends at a place:* Benthic species richness in lake Z has decreased by 20% since 1980.
- *Trends of a region:* Average lake trophic state in state Y has increased by 20% since 1980; *or* Average benthic index values in 20% of lakes of state Y have increased by 15% or more since 1980.
- *Relationships among variables:* 50% increase of P loading above natural background is associated with decline in number of taxa of benthic macroinvertebrates, below reference expectations; *or* Lakes receiving runoff from large impervious parking lots have 50% greater probability of elevated trophic state above reference than lakes not receiving such runoff.

Specification of reporting units helps to focus the study design on relevant questions. Alternative designs can be examined for their ability to address the questions within the specified reporting units. Elements of the design that are not relevant to questions and reporting units are identified as superfluous.

9.2 SAMPLING DESIGN

9.2.1 Sources of Variability

Variability of data justifies the existence of statistics. Variability has many possible sources. The intent of sampling designs is to collect a representative sample of the population. For bioassessment, we also wish to (1)minimize variability due to uncontrolled measurement error and, (2)characterize and partition the natural variability. For example, we may stratify lakes by soil phosphorus content of the surrounding watersheds (e.g., Rohm et al. 1995) so that lakes within a soil P class may be likely to have similar water column total P concentrations. Typically, we stratify so that observations (sample units) from the same stratum will be more similar to each other than to sample units in other strata.

Variability of data justifies the existence of statistics.

When sampling lakes we often measure something (say, chlorophyll concentrations) at single points in space and time (center of the lake, 2m depth, 10 AM on 2 July). If we make the same measurement at a different place (littoral zone, 1 m) 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 our ability to accurately measure the quantity we are interested in. Measurement error can be affected by sampling gear, instrumentation, errors in proper adherence to field and laboratory protocols, and the choice of methods used in making determinations. 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 have influence on the central questions and reporting units.
3. Control other sources of variability that are not of interest and thus minimize their effects in the observations.

In our example of chlorophyll concentrations, we may want to sample each of several lakes in the deepest part, with a vertically integrated pump sample taken in early spring before stratification appears. Many lakes are sampled in order to examine and characterize the variability due to different lakes (the sampling unit). Each lake is sampled in the same way, in the same place, and in the same time frame in an attempt to minimize variability due to location, depth, and season, which are not of interest in this particular study.

In the above example, chlorophyll concentrations vary with location within a lake, among

lakes, and time of sampling (day, season, year). If the spatial and temporal components of variability within lakes are large, then it is best to use either an index period sample or to estimate a composite from several determinations. For example, measurements of chlorophyll concentrations typically vary more between spring and fall samples within a lake than they do between lakes. Therefore, lake chlorophyll concentrations are often estimated as a growing season average, taken from several determinations (for instance, monthly) during the growing season.

In analyses, especially hypothesis testing, multiple determinations within lakes may be a form of pseudoreplication (Hurlbert 1984), and should be used with caution. If the hypothesis refers to a *single lake* (e.g., chlorophyll concentration of lake Z is higher than a biocriterion), multiple determinations are often necessary for the test. If the hypothesis refers to *many lakes* (e.g., lakes in state Y have elevated chlorophyll compared to state Q), multiple determinations within lakes are pseudoreplication if they are used as independent observations in the test, rendering the test invalid (Hurlbert 1984). If multiple determinations for each lake are used to calculate a single seasonal mean or median, which is then used as an independent observation for the hypothesis test, there is no pseudoreplication. Repeated measurement designs—analysis of variance (ANOVA–ANalysis Of VAriance) or regression—can be used (e.g., Underwood 1994) as a single analysis that takes into account multiple determinations. These methods estimate means of repeated measures to maintain independence.

A less costly alternative to multiple measures in space is to use spatially composite determinations. In nutrient or chlorophyll determinations, a water column pumped sample, where the pump hose is lowered through the water column, is an example of a spatially composite determination. Benthic macroinvertebrates are often sampled with spatial composite determinations. For example, benthic macroinvertebrates in Atlantic Coastal Plain streams are typically sampled by 20 sweeps of a dip net in multiple habitats, and composited into a single sample (e.g. USEPA 1997b, Barbour et al. 1996a, Barbour et al. 1996b, Roth et al. 1997). Benthic sampling of Florida lakes is a composite of 12 Petite Ponar grabs made throughout the sublittoral zone of a lake or a sample unit (Gerritsen and White 1997) (see Florida case study in this chapter).

Multiple observations within a sample unit (e.g., within a lake) should not be considered independent observations unless they are taken to examine an explanatory variable of interest, such as effects of depth, lake zone, season, or year. The principal use of multiple measurements is to estimate measurement error, that is, the variability we should expect when a single determination is made in a lake.

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

Measurement error is the result of methodological biases 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.

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

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 above. If the variance estimated from RMSE is unacceptably large (i.e., as larges 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. Mea-

surement 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.

Sampling design is the selection of a part of a population in order to observe the attributes of interest, so that the values of those attributes can be estimated for the whole population.

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. The Florida invertebrate and TVA fish methodologies include the compositing of multiple sampler casts into a single sample, which is then counted and identified.

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

9.2.2 Alternative Sampling Designs

Sampling design is the selection of a part of a population in order to observe the attributes of interest, so that the values of those attributes can be estimated for the whole population. Classical sampling design makes assumptions about the variables of interest; in particular, it assumes that the values are fixed (but unknown) for each member of the population, until that member is observed (Thompson 1992). This assumption is perfectly reasonable for some variables, say, length, weight, and sex of members of an animal population, but it seems less reasonable for more dynamic variables such as nutrient concentrations, loadings, or chlorophyll concentrations of lakes. Designs that assume that the observed variables are themselves random variables are model-based designs, where prior knowledge or assumptions (a model) are used to select sample units.

The most basic probability-based design is simple random sampling, where all possible sample units in the population have the same probability of being selected.

9.2.3 Probability-based Designs (Random Sampling)

The most basic probability-based design is simple random sampling, where all possible sample units in the population have the same probability of being selected, that is, all possible combinations of n sample units have the equal probability of selection from among the N units in the population. If the population N is finite and not excessively large, a list can be made of the N units, and a sample of n units is randomly selected from the list. This is termed *list frame sampling*. If the population is very large or infinite (such as locations in a lake), one can select a set of n random (x,y) coordinates for the sample.

All sample combinations are equally likely in simple random sampling. There is no assurance that the sample actually selected will be representative of the population. Other unbiased sampling designs that attempt to acquire a more representative sample include stratified, systematic, multistage, and adaptive designs. In stratified sampling, the population is subdivided or partitioned into strata, and each stratum is sampled separately. Typically, partitioning is done so as to make each stratum more homogeneous than the overall population. For example, lakes could be stratified by ecoregion. Systematic sampling is the methodical selection of every kth unit of the population from one or more randomly selected starting units, and ensures that samples are not clumped in one region of the sample space. Multistage sampling requires selection of a sample of primary units, such as fields or hydrologic units, and then selection of secondary sample units such as plots or lakes within each primary unit in the first stage sample.

Estimation of statistical parameters requires weighting of the data with inclusion probabilities (the probability that a given unit of the popula-

tion will be in the sample) specified by the sampling design. In simple random sampling, inclusion probabilities are by definition equal, and no corrections are necessary. Stratified sampling requires weighting by the inclusion probabilities of each stratum. Unbiased estimators have been developed for specific sampling designs, and can be found in sampling textbooks, such as Thompson (1992).

9.2.4 Model-based Designs

Use of probability-based sampling designs may miss relationships among variables (models), especially if there is a regression-type relationship between an explanatory and a response variable. As an example, elucidation of lake response to phosphorus loading with the Vollenweider model (e.g., Dillon and Rigler 1974) required a range of trophic states from ultraoligotrophic to hypereutrophic. A random sample of lakes is not likely to capture the entire range (i.e., there would be a large cluster of mesotrophic lakes with few at high or low ends of the trophic scale), and the random sample may be biased with respect to the regression model.

In model-based designs, sites are selected based on prior knowledge of auxiliary variables, such as estimated phosphorus loading, lake depth, elevation, etc. Model-based designs may preclude an unbiased estimate of the population (e.g., regional trophic state), unless the model can be demonstrated to be robust and predictive. The population value is then predicted from the model and from prior knowledge of the auxiliary (predictive) variables.

Identifying and sampling selected least stressed reference sites to develop an index is an example of samples for a model. The model is the index (e.g., IBI) and the responses of its component metrics. Reference sites alone cannot later be used for unbiased estimation of the biological status of lakes. Ideally, it may be possible to specify a design that allows both unbiased estimation of a population and index or model development. Statisticians should be consulted in developing the sample design for a biocriteria and biological monitoring program. However, managers should be aware that there is strong disagreement among statistical schools of thought on the subject of sampling design.

9.3 EVALUATION OF STATISTICAL POWER

A principal aspect of probability sampling is determining how many samples will be required to achieve the monitoring goals and what is the probability of making an incorrect decision based on the monitoring results. The primary tool for conducting these analyses is statistical power analysis. Evaluating statistical power is key to developing data quality criteria and performance specifications for decision making (USEPA 1996c) as well as evaluating the performance of existing monitoring programs (USEPA 1992d). Power analysis provides an evaluation of the ability to detect statistically significant differences in a measured monitoring variable. The importance of this analysis can be seen by examining the possible outcomes of a statistical test. The null hypothesis (H_o) is the root of hypothesis testing. Traditionally, null hypotheses are statements of no change, no effect, or no difference. For example, the mean abundance at a test site is equal to the mean abundance of the reference sites. The alternative hypothesis (H_a) is counter to H_o, traditionally being statements of change, effect, or difference. Upon rejecting H_o, H_a would be accepted.

Statisticians should be consulted in developing the sample design for a biocriteria and biological monitoring program; however, managers should be aware that there is strong disagreement among statistical schools of thought on the subject of sampling design.

The two types of decision errors that could be made in hypothesis testing are depicted in Table 9-2. A Type I error (i.e., false positive) occurs when H_o is rejected although H_o is really true. A Type II error (i.e., false negative) occurs when H_o is accepted although H_o is really false. The magnitude of a Type I error is represented by α and the magnitude of a Type II error is represented by β. Decision errors are the result of measure-

Evaluating statistical power is key to developing data quality criteria and performance specifications for decision making as well as evaluating the performance of existing monitoring programs.

Table 9-2. Errors in hypothesis testing.

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

ment and sampling design errors that were described in Section 9.2.1. A proper balance between sampling and measurement errors should be maintained because accuracy limits effective sample size and vice versa (Blalock, 1979).

9.3.1 Comparison of Significance Level and Power

Regardless of the statistical test chosen for analyzing the data, the analyst must select the significance level of the test. That is, the analyst must determine what error level is acceptable. The probability of making a Type I error is equal to the significance level (α) of the test and is selected by the data analyst. In many cases, managers or analysts define 1-α to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99 percent), although there have been environmental applications where 1-α has been set to 0.80. Selecting a 95 percent confidence level implies that the analyst will reject the H_o when H_o is really true (i.e., a false positive) 5 percent of the time.

A Type I error (i.e., false positive) occurs when H_o is rejected although H_o is really true. A Type II error (i.e., false negative) occurs when H_o is accepted although H_o is really false.

Type II error depends on the significance level, sample size, number of replicates, variability, and which alternative hypothesis is true. The power of a test (1-β) is defined as the probability of correctly rejecting H_o when H_o is false. In general, for a fixed sample size, α and β vary inversely. Power can be increased (β can be reduced) by increasing the sample size or number of replicates. Figure 9-1 illustrates this relationship. Suppose the interest is in testing whether there is a significant difference between the means from two independent random samples. As the difference in the two sample means increases (as indicated on the *x*-axis), the probability of rejecting H_o, the power, increases. If the real difference between the two sample means is zero, the probability of rejecting H_o is equal to the significance level, α. Figure 1A shows the general relationship between α and β if α is changed. Figure 1B shows the relationship between α and β if the sample size is increased. The tradition of 95% confidence (α = 0.05) is entirely arbitrary; there is no scientific requirement that confidence be set at 95%. Indeed, for environmental protection, power is at least as important—and possibly more important—than confidence (Peterman 1990, Fairweather 1991).

9.3.2 Basic Assumptions

Usually, several assumptions regarding data distribution and variability must be made to determine the sample size. Applying any of the equations described in this chapter is difficult when no historical data set exists to quantify initial estimates of proportions, standard deviations, means, or coefficients of variation. To estimate these parameters, Cochran (1963) recommends four sources:

- Existing information on the same population or a similar population.
- A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of the second step. Use data from both steps to estimate the final precision of the characteristic(s) sampled.

- A "pilot study" on a "convenient" or "meaningful" subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because often the pilot sample is not representative of the entire population to be sampled.
- Informed judgment, or an educated guess.

For evaluating existing programs, proportions, standard deviations, means, etc. would be estimated from actual data.

Some assumptions might result in sample size estimates that are too high or too low. Depending on the sampling cost and cost for not sampling enough data, it must be decided whether to make conservative or "best-value" assumptions. Because of the fixed mobilization costs, it is probably cheaper to collect a few extra samples the first time than to realize later that additional data are needed. In most cases, the analyst should probably consider evaluating a range of assumptions regarding the impact of sample size and overall program cost. USEPA recommends that if the analyst lacks a background in statistics, he/she should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

9.3.3 Simple Comparison of Proportions and Means from Two Samples

The proportion (e.g., percent dominant taxon) or mean (e.g., mean number of EPT taxa) of two data sets data sets can be compared with a number of statistical tests including the parametric two-sample *t*-test, the nonparametric Mann-Whitney test, and two-sample test for proportions (USEPA 1996c). In this case, two independent random samples are taken and a hypothesis test is used to determine whether there has been a significant change. To compute sample sizes for comparing two proportions, p_1 and p_2, it is necessary to provide a best estimate for p_1 and p_2, as well as specifying the significance level and power (1-β). Recall that power is equal to the probability of rejecting H_o when H_o is false. Given this information, the analyst substitutes these values into the following equation (Snedecor and Cochran, 1980)

$$n_o = (Z_a + Z_{2b})^2 \frac{(p_1 + p_2 q_2)}{(p_2 - p_1)^2}$$

Equation 1.

where Z and $Z_{2\beta}$ correspond to the normal deviate. Common values of $(Z + Z_{2\beta})^2$ are summarized in Table 9-3. To account for p_1 and p_2 being estimated, t could be substituted for Z. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 1; (2) round n_o up to the next highest integer, f; and (3) multiply n_o by $(f+3)/(f+1)$ to derive the final estimate of n.

EPA recommends that if the analyst lacks a background in statistics, he/she should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

To compare the mean from two random

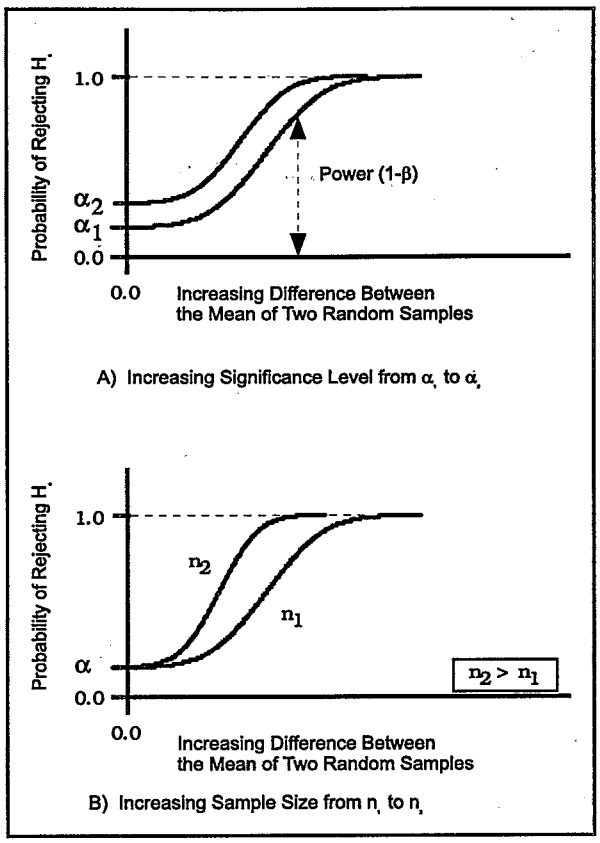


Figure 9-1. Illustration of significance (α) and power (1-β).

Table 9-3. Common values of $(Z_\alpha + Z_{2\beta})^2$ for estimating sample size for use with Equations 1 and 2 (Snedecor and Cochran 1980).

Power, 1-β	α for One-sided Test			α for Two-sided Test		
	0.01	0.05	0.10	0.01	0.05	0.10
0.80	10.04	6.18	4.51	11.68	7.85	6.18
0.85	11.31	7.19	5.37	13.05	8.98	7.19
0.90	13.02	8.56	6.57	14.88	10.51	8.56
0.95	15.77	10.82	8.56	17.81	12.99	10.82
0.99	21.65	15.77	13.02	24.03	18.37	15.77

samples to detect a change of (i.e., x_{-2}-x_{-1}), the following equation is used:

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(s_1^2 + s_2^2)}{\delta^2}$$

Equation 2.

where s_1 and s_2 are standard deviation of samples 1 and 2.

Common values of $(Z + Z_{2\beta})^2$ are summarized in Table 9-3. To account for s_1 and s_2 being estimated, Z should be replaced with t. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 2; (2) round n_o up to the next highest integer, f; and (3) multiply n_o by $(f+3)/(f+1)$ to derive the final estimate of n.

For large sample sizes or samples that are normally distributed, symmetric confidence intervals for the mean are appropriate.

A special case of Equation 2 arises for biocriteria, when we compare the mean of a sample from a lake to determine if the value is below some set limit, that is, if the lake is impaired or below a reference threshold. The threshold is fixed by previous investigations and decisions, and is not a random variable. We ask now whether we can detect a change of (i.e., C-x_{-1}), where C is the biocriteria limit:

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(s_1^2)}{\delta^2}$$

Equation 3.

In Equation 3, Z is most often one-tailed, because the concern is only whether the value is below the threshold.

9.3.4 Sample Size Calculations for Means and Proportions

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

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

Equation 4.

where $\bar{x}$ is the sample mean and s^2 is the sample variance.

This equation is appropriate if the samples are normally distributed or the sample size is greater than 30 (Wonnacott and Wonnacott, 1972), although Helsel and Hirsch (1992) suggest that highly skewed data might require more than 100 observations.

Although several approaches exist to estimate confidence levels for any percentile, many rely on assuming a normal or lognormal distribution. The approach presented here (Conover, 1980) for more than 20 observations does not

Example Sample Size Calculations for Comparing Proportions and Population Means

Example 1—Sample size calculation for comparing proportions

To detect a difference in proportions of 0.20 with a two-sided test, α equal to 0.05, 1-β equal to 0.90, and an estimate of p, and p_2 equal to 0.4 and 0.6, n_o is computed from Equation 1 as

$$n_o = 10.51\frac{[(0.4)(0.6) + (0.6)(0.4)]}{(0.6 - 0.4)^2} = 126.1$$

Rounding 126.1 to the next highest integer, f is equal to 127, and n is computed as 126.1 x 130/128 or 128.1. Therefore 129 samples in each random sample, or 258 total samples, are needed to detect a difference in proportions of 0.2. Since these are proportions, the result means that the total count in the sample must be at least 129. For example, to detect the above difference in the proportion of dominant taxon (e.g., benthic macroinvertebrates or fish) of two lakes, at least 129 individuals must be counted and identified in each lake.

The example illustrates that a statistically significant difference can be easily detected in proportions if sufficient individuals are sampled. However, it is doubtful that a difference between 40% and 60% in dominant taxon is biologically meaningful.

Example 2—Sample size calculation for comparing population means

To detect a difference of 20 in mean abundance with a two-sided test. The standard deviation, s, was estimated as 30 for both samples based on previous studies; was selected as 0.05; and 1-β was selected as 0.90. Substituting these values into Equation 2 yields

$$n_o = 10.51\frac{(30^2 + 30^2)}{20^2} = 47.3$$

Rounding 47.3 to the next highest integer, f is equal to 48, and n is computed as 47.3 x 51/49 or 49.2. Therefore 50 samples in each random sample, or 100 total samples, are needed to detect a difference of 20.

rely on these assumptions. Conover (1980) also provides a procedure for smaller sample sizes.

To calculate the confidence interval corresponding to the median, lower quartile, or upper quartile, the following procedure is used.

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

$$X_1 \leq \cdots \leq X_r \leq \cdots \leq X_p \leq \cdots \leq X_s \leq \cdots X_n$$

   where x_p corresponds to the median (i.e., p=0.5), lower quartile (i.e., p=0.25), or upper quartile (i.e., p=0.75).

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

$$r^* = np - Z_{a/2}(np(1-p))^0$$

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

Equation 5.

where $Z_{\alpha/2}$ is selected from a normal distribution table.

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

It can be seen from Equation 5 that estimation of medians or quartiles from small samples can result in large confidence intervals for the estimate. For example, the 90% confidence interval for the lower quartile of a sample of n=10 covers the first five observations. A sample of less than 10 observations would have a confidence interval extending below the smallest observation. This is the reasoning behind a general "rule of thumb" that estimation of reference conditions should be based on a sample of 10 or more sites, if at all possible.

Estimation of reference conditions should be based on a sample of 10 or more sites, if at all possible.

Case Study: Optimization of Benthic Sampling in Florida Lakes

To optimize its lake sampling protocols, Florida DEP performed a pilot study on 9 lakes. Each lake was sampled with twelve petite Ponar grabs (0.02 m²) distributed approximately equidistant in the sublittoral zone of the lake (2-4 m depth; Fig. 7-6). Each grab was kept separate in laboratory identification and enumeration. The lakes spanned a wide range in benthic macroinvertebrate diversity and abundance (Table 9-4), from 7 to 63 taxa, and 228 to 3540 organisms in 0.24 m² sampled. In seven of the nine lakes, the number of taxa continued to increase with sampling effort, and did not reach an asymptote with twelve Ponar samples.

To illustrate the effects of compositing sample casts, each sample of 12 grabs was composited into 2 replicate samples of 6 casts, so that each sample consisted of alternate casts (Fig. 7-6). This yielded 2 alternative sampling protocols: 12 Ponar replicates for each lake, and two replicates of 6 Ponars each. 4 candidate metrics were calculated: number of taxa (cumulative for composited samples), percent dominance, sensitive taxa (ephemeroptera, trichoptera, odonata), and log abundance. Standard deviation of each metric, as measurement error in determining the "true" value for each lake, was estimated with the root mean square error (RMSE) from an analysis of variance (Table 9-5).

All metrics had a lower coefficient of variation (CV) in the composited protocol than in the uncomposited, showing the advantages of compositing multiple deployments of small sample gear such as Ponars. Composited samples reduce costs because fewer jars and records are required, and sampling time is reduced some. Laboratory analysis can be reduced by subsampling a fixed number of organisms (e.g., 100, 200, or 300) from the composite sample for identification. It has been shown with the same Florida data (Barbour and Gerritsen 1996) that subsampling a fixed number of organisms (100 or more) yields adequate estimates of number of taxa, which are actually more precise than taxa density (total taxa in a fixed area or volume) (Hurlbert 1971). Subsamples that are larger than the target number can be reduced computationally by rarefaction (Hurlbert 1971, Vinson and Hawkins 1996, Barbour and Gerritsen 1996). Based on these results, Florida DEP adopted the following sampling protocol for lake benthic invertebrates:

- *12 Ponars randomly deployed in 12 segments of the 2-4m depth zone of lakes less than 1000 acres.*
- *Ponar casts are composited into a single sample and sieved through a 500 m mesh screen.*
- *A subsample of 100 benthic macroinvertebrates is sorted and identified to the lowest practical taxonomic level.*
- *The sampling protocol is duplicated at approximately 10% of sites to estimate measurement error.*

Table 9-4. Number of taxa and individuals in 12 cumulative Ponar samples from 9 Florida lakes.

Lake	Cumulative taxa	Cumulative individuals
Overstreet	63	768
Post	54	454
Camel	54	3540
Logan	42	1649
Mic	34	2828
Oche	31	1849
Del	16	228
Pickett	9	370
Adams	7	495

Table 9-5. Comparison of two sample processing protocols, Florida lakes.

	mean of 12 Ponars				mean of 2 samples of 6 composited Ponars			
	Population mean (9 lakes)	Range (9 lakes)	s.d. (individual lake)	CV (average lake)	Population mean (9 lakes)	Range (9 lakes)	s.d. (individual lake)	CV (average lake)
No. of taxa	8.85	2-19	3.62	40.9%	25.7	5.5-44.5	4.36	16.9%
% dominance	58.8%	40%-96%	14.8%	25.2%	50.4%	16%-96%	8.9%	17.7%
Sensitive taxa (ETO)	0.39	0-1.7	0.628	161%	1.6	0-5.5	1.27	79.4%
Total indiv (ln)	4.13	2.78-5.60	0.717	17.4%	6.12	4.68-7.48	0.145	2.4%

Case Study: Estimation of Power for TVA Fish Samples

TVA samples reservoir fish, benthic macroinvertebrates, water column chlorophyll, dissolved oxygen, and sediment contamination to rate the overall health of its reservoirs. 5 indices are calculated, one for each indicator group. Measurements are duplicated at selected reservoirs to obtain estimates of variability.

In 1996, fish sampling was repeated at seven reservoirs. The TVA Reservoir Fish Assembly Index (RFAI) is composed of 12 metrics (see Chapter 8). Ranges of metric values in 1996 (for all reservoirs) and metric standard deviations (from multiple determinations at single reservoirs are given in Table 9-6.

From the standard deviation of the RFAI score, we can estimate the number of samples required to detect differences among lakes.

1. Difference between two lakes (or between two sampling times within a lake)

To detect a difference of 10 in mean RFAI score with a two-sided test. The standard deviation, s, was estimated as 4.027 for both samples (Table 9-6); α was selected as 0.05; and 1-β was selected as 0.80. Substituting these values into Equation 2 yields

$$n_o = 7.85\frac{(4.027^2 + 4.027^2)}{10^2} = 2.54$$

Rounding 2.54 to the next highest integer, f is equal to 3, and n is computed as 2.54 x 6/4 or 3.82. Therefore, 4 samples in each reservoir, or 8 total samples, are needed to detect a difference of 10 in RFAI score between two reservoirs, with a probability of 0.80 of finding a true difference.

2. Test whether a lake is below a threshold (biocriteria)

To detect a difference of 10 in mean RFAI score below a threshold. The same standard deviation estimate is used as above (4.027; Table 9-6); α and 1-β were selected as 0.05 and 0.80, respectively, but is now one-sided. Substituting these values into Equation 2 yields:

$$n_o = 6.18\frac{4.027^2}{10^2} = 1.002$$

Rounding 1.002 to the next highest integer, f is equal to 2, and n is computed as 1.002 x 5/3 or 1.67. Therefore, 2 samples are needed to detect a difference of 10 in RFAI score below a threshold, with a probability of 0.80 of finding a true difference. If the effect size, or distance below the threshold, were increased to 15, then the required sample size would be 1. Thus if we find an RFAI value from a single unreplicated sample to be 15 points below a threshold, then we would expect that replication would not change a conclusion that the reservoir RFAI is below the threshold, 95% of the time. This example shows the potential value of adaptive sampling strategies, where a decision to increase sampling effort is based on the value of the first replicate. If the index value is very far below a threshold, there is no need to replicate. As the index value approaches the threshold, sampling effort needs to increase in order to make a decision at the prescribed power and significance. At some point, the sampling effort becomes so costly that judgement is reserved; i.e., no decision is made.

Table 9-6. Minimum and maximum values, and standard deviations of repeated measures, of reservoir fish metrics and the RFAI.

Metric	Minimum (all reservoirs, 1990-96)	Maximum (all)	*s* of repeated measures (n=7)
Total taxa	12	47	1.389
Piscivore species			1.309
Sunfish species			0.756
Sucker species			0.463
Intolerant species			0.463
Percent tolerant			0.118
Percent dominant			0.122
Percent omnivores			0.118
Percent insectivores			0.141
Simple lithophil species			0.463
Total individuals			15.677
Percent anomalies			0.0051
RFAI index score	18	58	4.027

9.4 MANAGEMENT

9.4.1 Personnel

Trained and experienced biologists should be available to provide thorough evaluations, provide support for various activities, and serve as QC checks. They should have training and experience commensurate with the needs of the program. At least one staff member should be familiar with establishing a QA framework. QA programs should document personnel responsibilities and duties and clearly delineate project organization and lines of communication (USEPA 1995). A time line illustrating completion dates for major project milestones or other tasks can be a tremendously useful tool to track project organization and progress.

Protocols should be developed for designing a data base and for screening, archiving, and documenting data.

9.4.2 Resources

Laboratory facilities, adequate field equipment, supplies, and services should be in place and operationally consistent with the designed purposes of the program so that high-quality environmental data can be generated and processed in an efficient and cost-effective manner (USEPA 1992b). Adequate taxonomic references and scientific literature should be available to support laboratory work, data processing, and interpretation.

For the field operations aspect of an ecological study, the major QC elements are instrument calibration and maintenance, crew training and evaluation, field equipment, sample handling, and additional effort checks.

9.5 OPERATIONAL QUALITY CONTROL

Protocols should be developed for designing a data base and for screening, archiving, and documenting data. Data screening identifies questionable data based on expected values and obvious outliers. Screening is especially important if data are gathered from a variety of sources and the original investigators and data sheets are no longer available. The following text box defines the qualitative and quantitative data characteristics that are most often used to describe data quality.

Six qualitative and quantitative data characteristics usually employed to describe data quality:

1. *Precision*—The level of agreement among repeated measurements of the same characteristic.
2. *Accuracy*—The level of agreement between the true and the measured value, where the divergence between the two is referred to as bias.
3. *Representativeness—The degree to which the collected data accurately reflect the true system or population.*
4. *Completeness*—The amount of data collected compared to the amount expected under ideal conditions.
5. *Comparability—The degree to which data from one source can be compared to other, similar sources.*
6. *Measurability*—The degree to which measured data exceed the detection limits of the analytical methodologies employed; often a function of the sensitivity of instrumentation.

These measurement quality indicators require *a priori* consideration and definition before the data collection begins. Taken collectively, they provide a summary characterization of the data quality needed for a particular environmental decision. Duplication of approximately 10 percent of the total sampling effort is a common level for operational QC. Replication of samples at a randomly selected subset of field sites (usually, 10 % of the total number is considered appropriate) is used to estimate precision, and representativeness of the samples and the methods; splitting samples into subsamples can be used to check precision of the methodology, and reprocessing of finished samples is used to check accuracy of laboratory operations.

9.5.1 Field Operations

For the field operations aspect of an ecological study, the major QC elements are instrument calibration and maintenance, crew training and evaluation, field equipment, sample handling, and additional effort checks. The potential errors in field operations range from personnel deficiencies to equipment problems. Field notes are integral to the documentation of activities and can be used to help locate potential recording errors. Training is one of the most important QC elements for field operations. Establishment and maintenance of a voucher specimen collection should be considered for biological data. Transcription errors during data entry can be reduced with double data entry. Table 9-7 gives examples of QC elements for field and laboratory activities.

9.5.2 Laboratory Operations

The QC elements in laboratory operations include sorting and verification, taxonomy, duplicate processing, archival procedures, training, and data handling. Potential error sources associated with sample processing are best controlled by staff training. Controlling taxonomic error requires well-trained staff with expertise to verify identifications. Counting error and sorting efficiency are usually the most prominent error considerations; they can be controlled by training and by duplicate processing, sorting, and verification procedures. See Table 9-7 for examples for QC elements for laboratory activities.

9.5.3 Data Analysis

Errors can occur if inappropriate statistics are used to analyze the data. Undetected errors in the data base or programming can be disastrous to interpretation. Problems in managing the data base can occur if steps are not taken to oversee the data handling, analysis, and summarization. The use of standardized computer software for data base management and data analysis can minimize errors associated with tabulation and statistical analysis. A final consideration is the possible misinterpretation of the findings. These potential errors are best controlled by qualified staff and adequate training.

9.5.4 Reporting

QC in reporting includes training, peer review, and the use of a technical editor and standard formats. The use of obscure language can often mislead the reader. Peer review and review by a technical editor are essential to the development of a sound scientific document.

Table 9-7. Example QC elements for field and laboratory activities

Project Activity	QC Element	Evaluation Mechanism
Field Sampling	Replicated samples at 10 percent of sites by same field crew.	Calculate relative percent difference (RPD) of index value or individual metric score
	Replicated samples at one to two of total sites by different field crew using same methods.	Calculate RPDs as above; use to evaluate consistency and bias.
Physical Habitat Assessment (Qualitative)	Ensure appropriate training and experience of operators; multiple observers.	Resume or other documentation of experience; discuss and resolve differences in interpretation.
Physical Habitat Assessment (Quantitative)	Replicated measurements at 10 percent of sites.	Calculate RPDs between replicate measurements; compare to preestablished precision objectives.
Laboratory: Sample Sorting	Sample residue checked for missed specimens to estimate sorting efficiency; check completed by separate lab staff.	Calculate percent recovery; compare to preestablished goals.
Laboratory: Sample Tracking	Logbook with record of all sample information.	Not applicable.
Laboratory: Taxonomic Identification	Independent identification and/or verification by specialist; ensure appropriate and current taxonomic literature available; adequate training and experience in invertebrate identifications; reference collection; exchange selected samples/specimens between taxonomists.	Calculate percent error; compare to preestablished goals.
Data Management	Proofreading; accuracy of transcription.	All transcribed data entries compared by hand to previous form—handwritten raw data, previously computer-generated tables, or data reports.
Data Analysis	Hand-check of reduced data.	For computer-assisted data reduction, approximately 10 percent of reduced data recalculated by hand from raw data to ensure integrity of computer algorithm.
	Appropriate statistics; training.	Review by statistician or personnel with statistical training.

In This Chapter...

- Effective Biocriteria
- Cost and Design Trade-Offs
- Cooperation for Cost-Effective Programs

Chapter 10

Biocriteria Implementation

10.1 CHARACTERISTICS OF EFFECTIVE BIOCRITERIA

Development of narrative or numeric biocriteria depends on the premise that biota provide a sensitive screening tool for measuring the condition of a water resource. Properly defined biocriteria can be used to protect the biological integrity of waterbodies and establish aquatic life use classifications.

Following the development of biocriteria, sites are evaluated to determine how well they meet the biocriteria or whether they have been significantly degraded. This determination is made by comparing the aquatic biota at potentially disturbed sites to the biocriteria, which are in turn based on minimally impaired reference conditions. The greater the anthropogenic impact in a watershed, the greater the impairment of the water resource. A corollary is that drainage basins not subject to anthropogenic impacts contain natural communities of aquatic organisms that reflect unimpaired conditions. These assumptions provide the scientific basis for formulating hypotheses about impairments—departures from the natural condition that result from human disturbances.

The establishment of formal biocriteria warrants careful consideration of planning, management, and regulatory goals. Effective biocriteria function to:

- Provide for scientifically sound evaluations.
- Protect the most sensitive biological value.
- Support and strive for protection of chemical, physical, and biological integrity.

Properly defined biocriteria can be used to protect the biological integrity of waterbodies and establish aquatic life use classifications.

Generally, optimal biocriteria share several common characteristics:

- They include specific assemblage characteristics required for attainment of designated use.
- They are clearly written and easily understood.
- They adhere to the philosophy and policy of antidegradation of water resource quality.
- They are defensible in a court of law.

In addition, biocriteria should be written to consider the best attainable condition at a site.

Overly stringent criteria that are unlikely to be achieved serve little purpose. Similarly, biocriteria that support a degraded biological condition defeat the intent of the Clean Water Act. Well-designed biocriteria are set at levels sensitive to anthropogenic impacts; they are not set so high that sites that have reached their full potential are considered in nonattainment or so low that unacceptably impaired sites are scored as meeting the criteria. It will be difficult to determine the full potential of a given lake. Balanced biocriteria will allow multiple uses to be considered so that any conflicting uses are evaluated at the outset. The best balance is achieved by developing biocriteria that closely represent the natural biota, protect against further degradation, and stimulate restoration of degraded sites.

The best balance is achieved by developing biocriteria that closely represent the natural biota, protect against further degradation, and stimulate restoration of degraded sites.

Several kinds of biocriteria are possible, and both narrative and numeric biocriteria have been effectively implemented. Narrative biocriteria consist of statements such as "aquatic life as it naturally occurs" or "changes in species composition may occur, but structure and function of the aquatic community must be maintained." Numeric values, such as measurements of community structure and function, can also serve as biocriteria as such or as quantitative refinements of narrative biocriteria. To account for a measure's natural variability in a healthy environment, the numeric criterion should be a defined range rather than a single number. Numeric criteria may also combine several such values in an index. Regardless of which kind is chosen, biocriteria should be both quantitatively based and supported by effective implementation guidelines and adequate capabilities including people, resources, methods, historical data, and management support. Additional general guidance regarding the writing of biocriteria is provided in EPA 440/5-90-004 (1990a) and EPA 822-B-92-002 (1992e).

10.2 STEPS TO IMPLEMENTATION

The first phase in a biocriteria program is the development of narrative biological criteria (USEPA 1992e). These criteria are essentially statements incorporated into water laws and regulations to formally consider the fate and status of aquatic biological communities. These statements of intent should include the following objectives:

1. Support the goals of the Clean Water Act to provide for the protection and propagation of fish, shellfish, and wildlife, and to restore and maintain the chemical, physical, and biological integrity of the Nation's waters.

2. Protect the most natural biological community possible by emphasizing the protection of its most sensitive components.

3. Refer to specific community characteristics that must be present for the waterbody to meet a particular designated use; for example, natural diverse systems with their respective communities or taxa indicated.

4. Include measures of community characteristics, based on sound scientific principles, that are quantifiable and written to protect or enhance the designated use.

5. In no case should impacts degrading existing uses or the biological integrity of the waters be authorized.

The use of multiple measures, or metrics, to develop biocriteria is a systematic process involving discrete steps. The process includes site classification (Chapter 4), a biological survey, evaluation of metrics with aggregation into indices (where indicated), formulation of biocriteria, and monitoring and assessment. The conceptual model for processing biological data into a biocriteria framework is adapted from EPA 822-B-96-001(USEPA 1996a) and summarized in Table 10-1. The process is as follows:

Step 1: Preliminary Classification of the Resource—The first decision that a resource agency must make is to determine the resource classes to which biocriteria will apply. Successful classification will result in less variation within a class, leading to more refined characterization of the reference condition and, there-

fore, to criteria with better resolution in detecting impairment. The preliminary classification should be based on lake characteristics that are not subject to pollution or disturbance, such as size, depth, morphology, or characteristics of the lake watershed.

Multijurisdictional collaboration is encouraged so that common methods and metrics can be established among states or other monitoring entities, and common reference conditions for multijurisdictional ecoregions can be characterized.

A set of reference sites are selected for each resource class; the reference sites are those least impacted by human influence, and they are characteristic of the resource class.

Step 2: Biological Survey—To determine the discriminatory power of the metrics within a lake class, the best-quality sites available, as well as those known to be impaired, are surveyed for biota and physical habitat. The use of standardized field collection methods allows a better interpretation of the raw data than does the use of a conglomeration of techniques.

Table 10-1. Sequential progression of the biocriteria process.

Step 1	Preliminary Classification to Determine Reference Conditions and Regional Ecological Expectations - Resource classification - Determination of best representative sites (reference sites representative of class categories)
Step 2	Characterization of Reference Condition - Historical data - Survey of reference sites and selected test sites. - Applicable models - Expert consensus
Step 3	Final Classification - Test preliminary classification - Revise if necessary
Step 4	Metric Evaluation and Index Development - Data analysis (data summaries) - Testing and validation of metrics by resource class - Evaluation of metrics for effectiveness in detecting impairment - Aggregation of metrics into index - Selection of biological endpoints
Step 5	Biocriteria Development - Adjustment by physical and chemical covariates - Adjustment by designated aquatic life use
Step 6	Implementation of Monitoring and Assessment Program - Determination of temporal variability of reference sites - Identification of problems
Step 7	Protective or Remedial Management Action Initiate - Programs to preserve exceptional waters - Implement management practices to restore the biota of degraded waters and to identify and address the causes of this degradation
Step 8	Continual Monitoring and Periodic Review of References and Criteria - Biological surveys continue to assess efficiency of management efforts - Evaluate potential changes in reference condition and adjust biocriteria as management is accomplished

Step 3: Final Classification—The preliminary classification is tested with biological data to determine whether it is reflected in the biota. If necessary, the classification is revised.

Any characterization of a reference condition should allow for the variability in biological data by using measures of central tendency and variability. Statewide or broader characterization of reference condition can be expected to exhibit high variance. The goal of classification is to minimize variability within classes by allowing the variability to be attributed to differences among classes.

Biocriteria may be based on an aggregated index, or established for several biological metrics and adjusted by aquatic life uses. The component information and data should always be retained.

Step 4: Metric Evaluation and Index Development—Potential metrics that have ecological relevance are identified in this step. Metrics are then evaluated for the ability to differentiate between impaired and nonimpaired sites. Values from various scales of measurement are transformed to scores, which are normally incorporated into an index, such as an Index of Biological Integrity (IBI) or an invertebrate index, which in turn becomes part of the final assessment. Metrics may also be used individually as indicators of biological condition in the overall assessment.

Step 5: Biocriteria Development—Biocriteria may be based on an aggregated index, or established for several biological metrics and adjusted by aquatic life uses. The component information and data should always be retained so future indexes or improvements in initial indexes can be calibrated with the data, and continuity of information preserved over time.

For example, a biocriterion for "Class A" lakes might be "a biotic index greater than the 25th percentile of least-impacted reference conditions." A "Class A" lake would be rated impaired if its biotic index fell below the 25th percentile of reference condition.

Step 6: Implementation of the Monitoring and Assessment Program—Use of biocriteria requires an operational monitoring and assessment program for two primary reasons: assessment of

Outline of Evaluation Criteria for Bioassessment Programs.

1. *Development of quality assurance and quality control bioassessment program plans.*
2. *Careful preparation of data quality objectives (DQOs) and design of field and laboratory studies to ensure the collection of representative data that will enable the biologists to achieve the objective of their program.*
3. *Preparation of standard operating procedures (SOPs) for field and laboratory methods.*
4. *Staff with adequate training and experience; division of labor within the program that permits specialization.*
5. *Use of approved methodology, use of technically defensible methodology if approved methodology is not available.*
   - *A. Sample collection*
   - *B. Sample processing*
   - *C. Organism identification*
   - *D. Counting*
   - *E. Biomass measurements*
   - *F. Data analysis and interpretation*
6. *Adequate space and physical facilities.*
7. *Adequate state-of-the-art field equipment, laboratory instrumentation, and supplies.*
8. *Adequate safety procedures.*
9. *Use of replication in sample collection and analysis to determine the precision.*
10. *Frequent calibration of field and laboratory instruments; log book documentation.*
11. *Chain-of-custody procedures for proper sample identification, handling, and logging to prevent misidentification and intermixing of samples.*
12. *Development and use of a taxonomic reference library for identifying specimens to the lowest possible taxonomic level.*
13. *Development and use of a reference specimen collection and use of outside experts to solve difficult problems in specimen identification.*
14. *Careful editing of data before they are placed in a computer file or used in reports.*
15. *Use of appropriate statistical analyses and other methods of data evaluation and interpretation.*

potentially impaired test sites and continued monitoring of selected reference sites to determine seasonal and annual variability and trends. A biocriteria program is the basis for a representative sampling program to determine statewide status and trends of the resource. The resources required to initiate a monitoring and assessment program are presented in the text box entitled "Outline of Evaluation Criteria for Bioassessment Programs."

Step 7: Protective and Remedial Management Action—The purpose of the entire process is to improve the water resource quality. Where problems have been identified through this effort, land use changes, discharges, abatements, and in-lake use adjustments are part of the management response. This may be done to improve degraded lakes or reservoirs or to protect exceptionally good ones from future damage. It should be recognized that implementing management action is potentially a multi-year process.

Step 8: Continual Monitoring and Periodic Reviews—The biocriteria-biomonitoring effort is designed to be a continuing process. Progress is expected but failures must be documented so monitoring and management efforts can be improved. The process progressively improves water resources by cycling back through the sequence.

10.3 TECHNICAL CONSIDERATIONS

The technical design of a biocriteria program affects the program's total cost. The sampling and analysis effort and data storage are two major cost elements of a biocriteria program. An optimal design balances the information needs of the monitoring agency with the cost of obtaining the information.

10.3.1 Taxonomic Level

Assemblages in Tiers 2A and 2B are identified to the lowest practical taxonomic level. Species level identification can be time consuming, especially for phytoplankton and benthic macroinvertebrates, and identification to family or genus might be more cost-effective.

10.3.2 Subsampling

Consistency of sampling methods and effort is critical in bioassessment. A sample is usually subsampled, in a random manner, to obtain a reasonable number of organisms for identification and enumeration (typically 100 to 500). Using fewer than 100 organisms might yield unreliable results, whereas using more than 500 is not cost-effective.

Taxonomic richness metrics, such as total taxa, diversity indices, and number of orders are sensitive to sample size. These values increase asymptotically with subsample size up to 500 organisms. Percent composition metrics (e.g., feeding groups, higher tax metrics) are less sensitive to subsample size; that is, the precision of an estimate for percent composition does not improve with subsamples greater than 100 organisms.

An optimal design balances the information needs of the monitoring agency with the cost of obtaining the information.

In order to control for the effects of sample size, it is critical that the methods are consistent in the number of organisms identified. For example, if the target subsample is 100 organisms, then subsamples smaller than 80 organisms should be rejected and subsamples larger than 120 organisms should be reduced mathematically by rarefaction (Hurlbert 1971) to make them comparable.

10.3.3 Spatial Variability and Replication

Replicating field samples by repeated measurements at a site is integral to biological surveys. These analyses have typically tested for significant differences between upstream and downstream pairs of sites. Significant differences were inferred to be due to discharges. However, Hurlbert (1984) pointed out that treatment of multiple measurements as replicates to infer cause is incorrect use of statistical inference. He pointed out that the site is the sampling unit, and repeated measurement of a sampling unit is not replication. True replication is achieved by replicating independent sampling units.

Repeated measurements, however, do have benefits, which must be weighed against the

cost. Repeated measurements are used to estimate measurement error, which is variability among measurements at the same site. Measurement error is due to spatial and temporal variability within a lake as well as actual errors made in sampling and analysis. It may be necessary to determine whether the measurement methodology adequately characterizes the site, and to determine the precision of metrics and indices (Fore et al. 1994). If measurement error is too large, it may be reduced by repeated measurements at a site or by a change of methodology (sample more microhabitats for a larger composite sample; increase subsample size). If the measurement error is acceptable, it is necessary only to take repeated measurements (replicates) for quality assurance at randomly selected sites (typically 10 percent of all sites). The QA replicates are then used to estimate measurement error.

Repeated measurements are used to sample more microhabitats at each site because the spatial distribution of organisms at a site can be patchy and a single measurement might not represent the composition of the assemblage. This usually results in a better estimate of the assemblage at the site. Since the site is the sampling unit, the measurement methodology can be altered to reduce measurement error. This is usually done by sampling multiple locations or habitat types, with several deployments of the specified sampling gear, and combining the hauls into a single composite sample. With a composite sample, a single measurement is taken, but the measurement is thought to be more representative of the site than a single, non-composited sample. Composite sampling that is representative of the sites is usually the most cost-effective sampling methodology. It avoids the costs of multiple measurements, allowing more sites to be sampled and increasing statistical sample size.

Index period sampling, in which measurements are made during the same period each year (e.g., midsummer), is intended to control short-term variability.

10.3.4 Temporal Variability

All aquatic assemblages go through annual cycles of composition and abundance changes. In addition, short-lived species also exhibit short-term temporal variability. Index period sampling, in which measurements are made during the same period each year (e.g., midsummer), is intended to control short-term variability. Index period sampling is effective if assemblage composition and abundance are relatively stable and predictable among years. If the assemblage is not stable within the index period, it might be necessary to make repeated measurements during a season or year to obtain growing season or annual average estimates of the metrics. Repeated measurements over a season or year are more expensive and reliable than index-period sampling. For cost-effectiveness, assemblages that can be adequately characterized using index period sampling are therefore preferable to those which require repeated sampling, unless the information from the repeated sampling is more valuable.

In view of major seasonal changes in lakes, it is possible to have more than one index period. Warm temperate and subtropical lakes, in particular, might require two or more index periods because biological activity remains high year-round. Multiple index periods must be analyzed separately. Therefore, there will also be separate reference expectations and biocriteria for each season represented by an index period. Two index periods require double the sampling effort of a single index period but provide greater information on biological variability throughout the year.

10.3.5 Classification

Each lake class requires a separate reference characterization (hence, separate reference sites) and separate biocriteria. For better statistical validity, each class should have a minimum of 5 or 10 reference sites (preferably up to 30 sites). Excessive proliferation of lake classes results in an unwieldy and expensive biocriteria program.

10.3.6 Status and Trends

Estimating status and trends of lakes as a resource requires a different sampling design from that proposed here. Unbiased estimation of status requires random selection of sampling units (lakes) within sampling strata (lake classes). One approach is to assign all lakes to the classes and then randomly select a sample

of lakes from each class (list-frame sampling). An alternative approach is to use a grid and sample lakes nearest the grid points, as is being done in EMAP (USEPA 1991e).

Trends can be assessed in single lakes or in a region. Several years of sampling are required for trend assessment. EPA 841-R-93-003 (USEPA 1993d) outlines trend analysis methods for lakes.

10.4 PROGRAM RESOURCES

A successful bioassessment and biocriteria program depends on (1) a clear definition of goals, (2) the active use of biomonitoring data in decision making, and (3) the allocation of adequate resources to ensure a high-quality program.

The implementation of a bioassessment and biocriteria program requires proper management and the appropriate combination of resources and expertise. Agencies already having well-developed programs usually have experienced and well-trained biologists, appropriately equipped facilities, and properly maintained sampling gear. Areas just beginning a bioassessment and/or biocriteria program need to evaluate their existing biological expertise, facilities, and equipment and expand accordingly. A cost-effective way to accomplish this is to coordinate efforts and share data with adjacent states or tribes, especially when lake or reservoir systems cross political boundaries.

10.4.1 Program Elements

Monitoring agencies can and should enhance their programs through cooperation with other agencies. For example, they should seek coordination with staff from state fishery, land management, geology, agriculture, and natural resource agencies. If federally employed aquatic biologists are stationed in a state or if the state has substantial federal lands, cooperative bioassessments and biocriteria development programs could be initiated. Scientists at universities should also be included in the planning and monitoring phases of the program—their students make excellent field assistants and future ecologists and natural resource managers. The selected team of specialists from this above pool of talent can also provide the "expert consensus" referred to earlier in defining reference conditions and developing biocriteria.

A cost-effective way to develop a bioassessment and biocriteria program is to coordinate efforts and share date with adjacent states or tribes, especially when lake or reservoir systems cross political boundaries.

10.4.2 Personnel and Resources

Several trained and experienced biologists and natural resource specialists should be available to provide thorough evaluations, support various activities, and manage quality. They should have training and experience commensurate with the needs of the program. At least one staff member should be familiar with establishing a quality assurance framework.

A biocriteria and biomonitoring program has several required elements, as well as optional elements, that determine the costs and resources of the program. Program elements include:

- *Quality assurance and quality control (e.g., standard operating procedures, training).*
- *Delineated reference conditions with annual monitoring of selected sites.*
- *Multiple assemblage biosurvey.*
- *Habitat assessment.*
- *Status and trends monitoring of a representative sample of lakes (optional).*
- *Computer hardware and software (database management, data analysis) and staff training.*
- *Documentation of program and study plans, periodic updates of analyses, and periodic review of reference conditions and biocriteria.*

Laboratory and field facilities and services should be in place and operationally consistent with the designed purposes of the program so that high-quality environmental data can be generated and processed in an efficient and cost-effective manner (USEPA 1992b). Adequate taxonomic

references and scientific literature should support data processing and interpretation.

Quality management is an important planning aspect that focuses attention on establishing and improving quality in all aspects of the biocriteria development process. Quality management requires that all personnel involved in a biocriteria project (from senior management to field and laboratory technicians) be aware of and responsive to data needs and expectations.

Appendix A

Glossary of Terms

***a posteriori* classification:** a classification made based upon the results of experimentation.

***a priori* classification:** a classification made prior to experimentation.

alkali lakes: also referred to as "soda lakes;" characterized by high pH (≥=pH 10) and a high concentration of salts.

antidegradation statement: statement that protects existing designated uses and prevents high-quality waterbodies from deteriorating below the water quality necessary to maintain existing or anticipated designated beneficial uses.

aquatic assemblage: an association of interacting populations of organisms in a given waterbody, for example, fish assemblage or a benthic macroinvertebrate assemblage.

aquatic community: an association of interacting assemblages in a given waterbody, the biotic component of an ecosystem.

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.

assemblage structure: the make-up or composition of the taxonomic grouping such as fish, algae, or macroinvertebrates relating primarily to the kinds and number of organisms in the group.

beneficial uses: desirable uses that water quality should support. Examples are drinking water supply, primary contact recreation (such as swimming), and aquatic life support.

Best Management Practice (BMP): an engineered structure or management activity, or combination of these, that eliminates or reduces an adverse environmental effect of a pollutant.

biological assessment: an evaluation of the biological condition of a waterbody that uses biological surveys and other direct measurements of resident biota in surface waters.

biological criteria: numeric values or narrative expressions that describe the reference biological condition of aquatic communities inhabiting waters that have been given a designated aquatic life use.

biological indicators: plant or animal species or communities with a narrow range of ecological tolerance that may be selected for emphasis and monitored because their presence and

relative abundance serve as a barometer of ecological conditions within a management unit.

biological integrity: the condition of the aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by an evaluation of multiple attributes of the aquatic biota. Three critical components of biological integrity are that the biota is (1) the product of the evolutionary process for that locality, or site, (2) inclusive of a broad range of biological and ecological characteristics such as taxonomic richness and composition, trophic structure, and (3) is found in the study biogeographic region.

biological monitoring: the use of a biological entity as a detector and its response as a measure to determine environmental conditions. Toxicity tests and biological surveys are common biological monitoring methods.

biological survey (biosurvey): the process of collecting, processing, and analyzing representative portions of a resident aquatic assemblage to determine the assemblage structure and function.

biota: plants, animals and other living resources of a region.

bisection scoring: used when metric value distribution is based upon data from unimpaired reference sites. The 25th percentile becomes the minimum value for the highest score: the difference between the 25th percentile and 0 is divided into two equal parts.

canonical correlation analysis (CC): a linear multivariate ordination procedure using linear canonical equations with multiple dependent and independent variables.

canonical correspondence analysis: a non-linear multivariate ordination procedure.

Carlson's Trophic State Index (TSI): a numerical index for estimating lake trophic state on a scale of 0 to 100 with each increase of 10 in the index representing a doubling of algal biomass.

coefficient of variation: standard deviation (from the mean) expressed as a percentage of the mean.

community component: any portion of a biological community. The community component may pertain to the taxonomic group (fish, invertebrates, algae), the taxonomic category (phylum, order, family, genus, species, stock), the feeding strategy (herbivore, omnivore, predator), or the organizational level (individual, population, assemblage) of a biological entity within the aquatic community.

designated use classifications: classification of a waterbody or segment based on the purposes (beneficial uses) for which the waterbody may be used as specified in water quality standards.

diatoms: any of a number of related microscopic algae, one-celled or in colonies, whose walls consist of two parts or valves and contain silica.

discriminant analysis: a type of multivariate analysis used to distinguish between two groups.

ecological or environmental indicators: measurable features of an ecosystem that singularly or in combination with other features provide managerially useful evidence of water resource or ecosystem quality, or reliable evidence of trends in quality. Indicators can be biological, physical, or chemical measurements, and can sometimes have elements of more than one discipline: for instance, concentrations of chemicals in fish tissue.

ecological integrity: the condition of the biotic (aquatic community) and abiotic components (water chemistry and habitat) of unimpaired waterbodies as measured by assemblage structure and function, water chemistry, and habitat measures.

ecological properties: biotic and habitat attributes of a waterbody.

ecoregions: a relatively homogeneous area defined by similarity of climate, landform, soil, potential natural vegetation, hydrology, or other ecologically relevant variable.

epifauna: benthic animals living on the sediment or among rocks and other structures.

epilimnion: the upper waters (above the metalimnion) of a thermally stratified lake.

expert consensus: a method used to establish reference condition when no candidate sites are available based on the collective experience and expertise of regional biologists.

flowage lakes: areas of a river system which are sufficiently deep, slow moving and wide to have lacustrine characteristics. Unlike reservoirs, they typically have wide inflow and outflow regions.

forebay zone: same as the lacustrine zone of a reservoir.

frustule: the hard shell of a diatom.

gradient analyses: a suite of statistical techniques including principal component analysis, canonical correlation's analysis, and canonical correspondence analysis used to examine the relationships between biotic and environmental factors.

habitat: a place where the physical and biological elements of ecosystems provide a suitable environment including the food, cover, and space resources needed for plant and animal livelihood.

hypoxic: waters that have a very low oxygen level.

infauna: animals that live within the sediments, often in holes they have dug.

inflow zone: area where a river enters a reservoir.

interquartile coefficient: the ratio of the interquartile range of a metric to its scope for detection.

lacustrine zone: area of a reservoir which is most lake-like: current velocities are much slower than for riverine or transitional zones. Little sediment deposition normally occurs since most sediment load has been deposited in the riverine or transitional zones; may thermally stratify; primary productivity predominates.

lake: a body of fresh or salt water of considerable size, whose open-water and deep-bottom zones (no light penetration to bottom) are large compared to the shallow-water (shoreline) zone, which has light penetration to its bottom.

limnology: the study of the functional relationships and productivity of freshwater biotic communities as they are affected by the dynamics of physical, chemical and biotic environmental parameters.

littoral zone: the area of a lake near the shore from the region of the highest seasonal water level to the deepest point at which attached submerged macrophytes occur.

log linear models: statistical modeling techniques for dealing with categorical data.

marl lakes: lakes in which solid calcium carbonate precipitates during periods of high photosynthesis forming a characteristic marl bench in the euphotic zone.

metalimnion: the stratum of steep thermal gradient that separates the epilimnion from the hypolimnion in a thermally stratified lake.

morphoedaphic index (MEI): the ratio of dissolved solids (measured as total dissolved solids, alkalinity, or conductivity) to mean lake depth; MEI has been used to predict the total fish production, phytoplankton standing crop, and total phosphorus concentration of lakes not subject to cultural eutrophication.

multiple metric or multimetric approaches: analysis techniques using several measurable characteristics of a biological assemblage.

multiple use: when a water body has more than one beneficial use designation.

multivariate community analysis: statistical methods (e.g., ordination or discriminant analysis) for analyzing physical and biological community data using multiple variables.

ombrotrophic bog: an acidic wetland which receives all of its nutrients from atmospheric deposition.

ordination analysis: a set of techniques in which sampling units are arranged in relation to one or more coordinate axes such that their relative positions to the axes and to each other provide maximum information about their ecological similarities.

oxycline depth: depth at which dissolved oxygen levels fall below a threshold value.

paleolimnology: the study of the environmental history of inland waters, based primarily on analysis of biological, chemical, and physical characteristics of sediment cores.

pelagic zone: the area of open water beyond the littoral zone.

predictive models: statistical models that can be used to predict biological response based on ecological (habitat) variables.

Principal Components Analysis (PCA): a linear multivariate ordination technique that determines a reduced set of coordinate axes.

principal axes: new variables created by ordination analysis that account for variation in the data.

profundal zone: the sediments beyond the littoriprofundal zone.

reference site: a site on a waterbody which represents the best attainable physical habitat, water chemistry, and biological parameters for specific environmental conditions.

reference condition: The chemical, physical, or biological quality or condition exhibited at either a single site or an aggregation of sites that represent the least impared or reasonably attainable condition at the least impared reference sites.

regression: any of a number of statistical techniques in which the relationship of one (or more) variable(s) is (are) estimated as a function of another variable or variables.

reservoir: a lake created for human use often as the result of impoundment of a river system; classified as lake type 73 by Hutchinson (1957).

risk assessment: a scientific process that includes hazard identification, receptor characterization and endpoint selection, stress-response assessment, and risk characterization.

riverine zone: the relatively narrow and well-mixed area of a reservoir immediately downstream of the river inflow where current velocities decrease and significant sediment transport still occurs.

robust: insensitive to assumption violations, i.e., holds even when the probability model is incorrect.

scope for detection: the range from 0 to the lower quartile for metrics that have high values under unimpaired conditions (e.g., EPT index) or that range from the upper quartile to 100 for metrics that have low values under unimpaired conditions (e.g., percent Chironomidae).

spatial variability: variation in a biological parameter due to different ecological conditions among sites.

temporal variability: variation in a biological parameter due to temporal fluctuations in ecological condition such as changing water chemistry or sunlight, e.g., diurnal and seasonal variations.

Total Maximum Daily Load (TMDL): The total allowable pollutant load to a receiving water such that any additional loading will produce a violation of water-quality standards.

transitional zone: area of a reservoir between the riverine and lacustrine zones; current velocities are intermediate, significant sedimentation occurs, light penetration increases and primary productivity increases.

trisection scoring: used when metric value distribution is based upon data from reference and impaired sites (population distribution). The range of values from the 95th percentile to 0 is divided into thirds with the top third receiving the highest score, the middle third receiving the intermediate score, and the bottom third receiving the lowest score.

trophic state index: any numerical index for estimating trophic state of a lake.

unimodal response: a response in which a species has [a single] peak abundance at [an] optimal value [or range] of an environmental variable and its abundance is lower at higher or lower values of [that] environmental variable.

univariate tests: statistical tests for comparing two or more groups; techniques include t-test, analysis of variance, sign test, Wilcoxon rank test, and the Mann-Whitney U-test.

water quality standards: provisions of state or federal law which consist of a designated use or uses for the waters of the United States, water quality criteria for such waters based upon such uses. Water quality standards are to protect public health or welfare, enhance the quality of the water and serve the purposes of the Clean Water Act (40 CFR 131.3) (USEPA 1983) a law or regulation that consists of the beneficial designated use or uses of a waterbody, the numerical and narrative water-quality criteria that are necessary to protect the use or uses of that particular waterbody, and an antidegradation statement (ITFM 1994).

Appendix B

Comparison of Existing Lakes Protocols

Two research programs—the U.S. Geological Survey's North American Water Quality Assessment (NAWQA) and the USEPA Office of Research and Development's Environmental Monitoring and Assessment Program (EMAP)—incorporate biological monitoring components similar to those of USEPA's Biological Monitoring Programs. The following discussion compares these programs, as well as a program implemented by the Tennessee Valley Authority (TVA) and USEPA's Clean Lakes Program.

NAWQA, EMAP, TVA, and Clean Lakes Program

NAWQA and EMAP are scientific research programs with objectives quite different from those of state-level biological assessment programs (USEPA 1990a, Meador et al. 1993). NAWQA and EMAP differ in the waterbodies they address: NAWQA is designed for streams and rivers and EMAP is designed for all environments. including streams and lakes (EMAP-Surface Waters). The two programs are set apart by the statistical and experimental designs used to accomplish their program.

The NAWQA and EMAP programs make assessments of the environment and trends in environmental quality at the national or regional scales. In these programs, USGS or USEPA scientists and technicians are directly responsible for collecting data from the field and analyzing results. NAWQA and EMAP also use different statistical designs to assess national trends. The NAWQA approach is to sample repeatedly at equal time intervals at the same stations (Gurtz 1994) and to apply nonparametric statistical algorithms to the data, adjusted for seasonal variation (Meador et al. 1993). The EMAP approach is to overlay a region or landscape with a grid and then to monitor a randomly selected subsample of the grid (USEPA 1991d). Both methods are useful to evaluate trends in environmental quality.

The TVA program, which monitors ecological condition of reservoirs,also covers a large percentage of the country. The assessment is based on sediment quality, dissolved oxygen content (DO), chlorophyll *a* concentration, a benthic macroinvertebrate index, and a Reservoir Fish Assemblage Index (RFAI) (TVA 1994).

USEPA's Clean Lakes Program provides support to water quality agencies for lake assessment, protection, and restoration. The program requires assessments of dissolved oxygen (DO), chlorophyll *a*, Secchi disk transparency, macrophytes, sediment chemi-

cal characterization, and tributary streamflow measurements. A complete description of the Clean Lakes Program recommended sampling methodology is found in the Clean Lakes Program Guidance Manual (USEPA 1980a) .

State Mandates

USEPA's Biological Monitoring and Biological Criteria Programs are targeted to state and tribal water quality and environmental management agencies. Monitoring data is collected at the state or local watershed level by state agency personnel or by suitably trained organizations to identify problems and report on control effectiveness. The Biological Monitoring Program accomplishes its mission through training and the translation of research techniques to rapid, efficient and defensible protocols these techniques can be used on an operational basis by state personnel and others to develop biological criteria that can be used in management decision making.

The broad water quality goals of the Clean Water Act are translated by each of the states into water quality standards designed to protect beneficial uses (USEPA 1996a). The USEPA biological monitoring program therefore looks to individual states and tribes rather than federal agencies for the assessment of those beneficial uses, as they have been designated in various administrative codes. The protocols presented in this document are designed to assist states and tribes in making their beneficial use assessments in biological terms. The protocols and measurement techniques will also help states fulfill their obligation to report to Congress biennially on the attainment of designated beneficial uses of their surface waters (USEPA 1994f).

Monitoring Biotic Integrity

Although the statistical designs used by EMAP, NAWQA, and USEPA's biological monitoring programs differ, the monitoring components are generally complementary. These programs assess conditions of lakes, rather than producing comprehensive inventories of biological resources. Some but not all major biological assemblages—invertebrates, fish, plants, birds, microorganisms—are sampled, identified, and counted, using a standardized technique, during a well defined season of the year.

In this kind of monitoring, the focus is on the full array of species captured by the sampling methods to provide cost-effective sampling and standardization. The methods described in this document do not target rare species. Issues of seasonal variation and bias in the sampling protocols are standardized or "indexed" to simplify statistical analysis (Sokal and Rolfe 1969). These monitoring programs produce sensitive and robust estimates of the biological integrity of the aquatic system, as well as the impacts that anthropogenic activities might have on the environment in terms of degrading the aquatic life designated use (Fausch et al. 1984, Karr 1991, USEPA 1989b).

It is desirable for all agencies to use similar water quality monitoring protocols so that data can be shared and compared. The national water quality monitoring council,(NWQMC) which succeeded the Intergovernmental Task Force on Monitoring (ITFM) is the forum for that kind of information exchange, and several of the recommendations of the task force address the issue of data comparability (ITFM 1992). However, in the case of biological monitoring, species characteristics differ regionally, and sampling techniques are not expected to be the same across ecoregions. In USEPA's Biological Monitoring Program and Biological Criteria Program, sampling techniques and index period must be identical for the reference condition and the test locations, and the use of common techniques across agencies within a region would significantly improve the efficiency and power of biological monitoring activities. Table B-1 compares lake habitat and biological monitoring among EMAP, TVA, and the USEPA Biological Monitoring Program. NAWQA has not published specific protocols for lakes and reservoirs.

Table B-1. Comparison of lakes protocols with EMAP, TVA Reservoirs, and Clean Lakes.

	USEPA Lakes Biomonitoring/ Biocriteria Program	EMAP	TVA Reservoirs	Clean Lakes Program
I. Habitat Assessment	Single qualitative estimate to 10 stations, vegetation type, canopy layer, understory, ground cover, shoreline substrate, bank features, human influence, bottom substrate, fish cover, water quality	10 stations, observations are made 10m from shore; plot dimensions are 15m x 15m, vegetation type, canopy layer, understory, ground cover, shoreline substrate, bank features, human influence, bottom substrate macrophyte cover, fish cover, littoral microhabitat, water quality.	Water quality.	Water quality.
II. Benthic Invertebrates	Preferred sampling in the sublittoral zone (profundal zone is an alternative) (Tier 2) Sampling gear depends on substrate: Rocks or gravel - dome sampler: Sand - Peterson or Van Veen. Mud - Ponar. Clay - Peterson or Van Veen. Lake-wide composites of 2 to 3 casts at 3 to 10 stations. 100 individuals at "lowest taxon"; alternatives are more than 100 individuals or identification to family (100 individuals).	Sublittoral zone. K-B cores. 150 individuals. No. 60 mesh used to sort organisms larger than 250μm.	Line of sight transect across width of reservoir; 10 samples collected at equal intervals along transect. Ponar sampler for mud substrates. Peterson sampler for rocky substrate. Organisms identified to lowest level, generally species or genus. Organisms reported as no/m^2.	
III. Fish	Electrofishing; gill nets, seines, fykes, and minnow traps. Identify, measure length and weight of fish. Inspect all fish for external anomalies (supplemental).	Electrofishing, gill nets, seines, fykes, and minnow traps. Identify, measure length and weight of fish. Inspect all fish for external anomalies.	15 electrofishing runs (300m each) in each location of reservoir (inflow, transition, forebay); all habitats (e.g., bluff, rip-rap, mud) sampled in approximate proportion to their occurrence in the sampling location. Identify, measure length and weight of fish. Young of the year fish counted separately from adults. Inspect all fish for external diseases, parasites, anomalies.	Preproject monitoring (phase 1) requires: General discussion of fish populations and ecological relationships. Standard fish flesh analyses for organic and heavy metal contamination if there is significant public consumption. No specific protocols recommended for phase 2 and 3 monitoring.
IV. Zooplankton	Vertical tow with 10μm mesh net (Tier 3). Identify 100 organisms to family or species; measure *Daphnia*.	Vertical tow with dual (bongo) net; 48- and 220μm mesh. Identify to species; measure.	No protocols recommended.	No protocols recommended.

Table B-1. (continued)

	EPA Lakes Biomonitoring/ Biocriteria Program	EMAP	TVA Reservoirs	Clean Lakes Program
V. Sediment Diatoms	Sediment grab for recent diatoms (Tier 2). Sediment cores for paleolimnology (supplemental). Identify to species.	Single sediment core at index site with modified K-B corer.	No protocols recommended.	No protocols recommended.
VI. Birds	No protocols recommended.	Special team of ornithologists visit lake. May-July. Canoe shoreline transect; record birds seen or heard for 5 minutes at 200m intervals.	No protocols recommended.	No protocols recommended.
VII. Phytoplankton (including Chlorophyll)	Chlorophyll *a* (Tiers 1-3) Seasonal composite from 6 to 12 spaced samples (Tier 3).	Chlorophyll *a* only (1.5m).	Chlorophyll *a*: Mean of euphotic zone concentration. Growing season mean of 6 to 10 sampling periods.	Chlorophyll *a*: Depth integrated from the top 6ft. of the water column preferably from the deepest part of the lake. Monthly sample September-April, biweekly sample May-August.
VIII. periphyton	Sample 3 to 5 substrates, composite 300-500 diatom frustules identified to species (supplemental).	No protocols recommended.	No protocols recommended.	No protocols recommended.
IX. Submerged Macrophytes	Estimate cover, identify dominant species (Tier 1). Tiers 2-3: 5 to 10 samples. Sample with rake. Identify to species, weigh each (net weight).	Estimate percent cover (habitat component).	Macrophyte coverage determined by color aerial photography. Boat surveys to identify dominant species.	Document species composition, distribution, and depth during the growing season. Include community types and abundance along with species list.

Appendix D

Biological Assemblages

D.1 ALGAE

Algae dominate the primary production of most lake ecosystems, occurring as free-floating phytoplankton or attached periphyton. Phytoplankton are the base of most lake food webs, and fish production is linked to phytoplankton primary production (e.g., Ryder et al. 1974). Excessive nutrient and organic inputs from human activities in lakes and their watersheds lead to eutrophication, characterized by increases in phytoplankton biomass, macrophyte biomass, nuisance algae blooms, loss of water clarity, and loss of oxygen from bottom waters. From a human perspective, problems may also include loss of aesthetic appeal, decreases in desirable gamefish, loss of accessibility from increased macrophyte production, and increased cost of treating drinking water.

Measurements of algae include estimation of total biomass with water column chlorophyll *a* concentration, and identification and counts of individual species within several subgroups, including periphyton (attached forms), periphytic diatoms, phytoplankton (free-living, floating), phytoplankton diatoms, and sediment diatoms. Diatoms are identifiable by their frustules (valves) of silica and thus do not require preservation and identification of soft parts (Dixit et al. 1992). Analysis of diatoms can be done on frustules preserved in lake sediments, attached assemblages on natural substrates (rocks, macrophytes), or attached assemblages that colonize artificial substrates. Planktonic chrysophyte scales are also preserved in lake sediments, and their spatial distribution is less variable than that of diatom assemblages (Smol et al. 1984).

Algal assemblages respond rapidly (in days) to changes in their environment with concomitant changes in overall abundance, growth rates, and species composition, and therefore do not integrate conditions of the lake. Algal species have characteristic optimal nutrient and trace element requirements, and specific tolerances for cations, salinity, pH, etc. (e.g., Tilman 1982). Changes in physical and chemical water quality (nutrient concentrations, loadings, salinity, temperature, turbidity) can thus lead to a rapidly changed species composition (Charles and Smol 1994, Dixit et al. 1992). These mechanisms form the basis of algal associations, seasonal assemblage succession (Hutchinson 1967), and the well-known responses to nutrient enrichment (Reynolds 1984). Temporal variability is the greatest disadvantage of indicators based on measurements of the algal assemblage, and either repeated sampling or temporally integrated samples are required to obtain a

seasonal or annual assemblage estimate. The algal assemblage seasonal succession cycles are only general and their exact timing and composition are not predictable (Reynolds 1984). In addition, assemblage composition and abundance are influenced by grazing pressure from zooplankton. For example, lakes with abundant, large-bodied zooplankton suspension feeders may have greater water clarity than similar lakes without the grazers (e.g., Edmondson and Litt 1982). Other advantages and disadvantages of using algal assemblages are listed in Table D- 1.

Primary Production

Nutrient enrichment from human activities generally leads to increased biomass of algae in lakes. Measurement of algal and other plant biomass, or a surrogate (e.g., chlorophyll *a*), is a good indicator of eutrophication and is clearly related to biological integrity, meeting the first criterion for successful metrics. Trophic state is an expression of the production of a lake ecosystem. We use Carlson and Simpson's (1996) definition of trophic state based solely on biomass and operationally measured by variables that estimate biomass.

Carlson's Trophic State Index (TSI) (Carlson 1977) is the most widely used index for eutrophication in lake monitoring programs, and is based on epilimnetic chlorophyll *a* concentration, total phosphorus concentration, and Secchi depth. Index values range from below 0 (ultraoligotrophic) to above 100 (hypereutrophic). Each of three measures (chlorophyll, total P, Secchi) is used to calculate an independent index, and the indices can be compared to identify whether algal growth is limited by phosphorus, light, or other nutrients. Chlorophyll is the most accurate and is the preferred indicator of trophic state. Total phosphorus and Secchi depth indices are also used as surrogates in the absence of chlorophyll data, and can be used to identify factors contributing to algal growth when all three are measured (Carlson and Simpson 1996). Measurement of primary productivity is not recommended because it is expensive to measure and frequently difficult to interpret.

Other indices have been developed and might be appropriate for different lake ecoregions in the country. A nitrogen index can be included to identify nitrogen limitation (Carlson 1992, Kratzer and Brezonik 1981). Other trophic state models (e.g., Dillon and Rigler 1974, Larsen and Mercier 1976, Vollenweider 1975) use annual phosphorus loading rates or retention fractions, and rely on measurements of nutrient concentrations rather than the biological response to nutrient loading. See Carlson and Simpson (1996) for a complete discussion of the trophic state concept.

Both algae and macrophytes contribute to a lake's plant biomass, therefore, metrics for both algal and macrophyte biomass are preferred for whole-lake trophic state (Canfield et al. 1983, Carlson and Simpson 1996) (see section 4.2 for macrophyte biomass).

Table D-1. Advantages, disadvantages, and alternatives to using algal assemblages.

Advantages	Disadvantages	Alternatives
Species composition, abundance respond to water quality: - Nutrients (N, P, Si) - pH, alkalinity - Metals - Temperature Field sampling relatively easy. Identification rapid to division and family level. Simple biomass indicators (chlorophyll *a* or dry weight). Historic and prehistoric record in sediment diatoms.	Require taxonomic expertise for species indentification. Strong temporal variability; do not integrate (except sediment diatoms). Quantitative inference of water quality requires large calibration data set.	Water quality measures (N, P). pH, alkalinity measurement. Metals analysis. BOD, COD. ATP

The Tennessee Valley Authority (TVA) uses chlorophyll *a* concentrations for one of its reservoir assessment metrics. The metric is based on the mean growing season water column concentration and a single maximum concentration (Dycus and Meinert 1994, TVA 1995). Reservoirs are considered mesotrophic or oligotrophic, based on natural watershed geochemistry and expected chlorophyll *a* concentrations. Low, moderate, and high mean concentrations of chlorophyll *a* are rated "good"," fair," and "poor," respectively, with differing definition of these three categories for mesotrophic and oligotrophic reservoirs. A very high single-sample maximum (> 30 µg/L) reduces the rating by one class. Thus, a good rating implies chlorophyll concentrations within the range expected and no extreme blooms. In addition, for mesotrophic reservoirs unusually low concentrations (< 3 µg/L) are rated "fair." Also, if this low concentration occurred despite sufficient phosphorus, it was considered an indication of limitations other than nutrients and resulted in a poor rating.

Phytoplankton Species Composition

Many different levels of algal monitoring and assessment exist. Metrics based on indicator taxa can be quite simple, such as qualitative estimates of relative dominance of algal divisions (Table D-2). For example, dominance by diatoms might be rated "good", and dominance by cyanobacteria might be rated "poor," requiring only a rapid, qualitative estimate of the relative abundances of diatoms and cyanobacteria. Indicator genera could also be used, For example, abundant populations of the cyanobacteria *Oscillatoria* or *Anabaena* indicate eutrophication (e.g., Edmondson and Lehman 1981). Certain diatoms and chrysophytes are sensitive to pH and dissolved aluminum (Charles and Whitehead 1986, Smol et al. 1984).

Algal assemblage data, consisting of taxonomic identifications and abundance (relative or absolute) of each taxa, can be analyzed in two ways:(1) by determining assemblage metrics based on species structure, or (2) by multivariate assemblage analysis. Simplified field and laboratory procedures are possible for some (but not all) of the species structure metrics.

Due to the high temporal variability of plankton, several samples during the growing season might be needed for accurate assemblage analysis.

Assemblage metrics, as defined and used in assessment of biological integrity, rely on the comparison of a metric to a reference value. Assemblage metrics possible for use in algal analysis include (Bahls 1993):

- Diversity metrics, such as number of taxa , percent contribution of dominant taxon, and Shannon-Wiener diversity which incorporates both number of taxa and evenness, and is less sensitive to sample size.
- Indicator taxa (e.g., bloom-forming cyanobacteria) respond to acid, eutrophication, metals, organics, salinity, and climate. Responses are reliable to water chemistry and many responses of individual species are unimodal.
- Indices and ratios.
- Pollution tolerance index, based on tolerance groups of Lange-Bertalot (Bahls 1993).

Table D-2. Potential algal metrics.

Metric	Optimal Condition	Impaired Condition
Trophic state.	Similar to reference expectation.	Substantially higher or lower.
No. of Taxa.	High	Reduced number of taxa.
% dominance.	Low	High dominance.
Indicator taxa, ecological categories, and tolerance indices.	Similar assemblage to reference.	Abundance of indicator, taxa or high index values (e.g., cyanobacteria acidophilic taxa, tolerant species, etc.)
Similarity indices.	Similar to reference.	different.

This index is functionally similar to the Hilsenhoff Biotic Index for invertebrates (Hilsenhoff 1987).

- Similarity indices, comparing the similarity in assemblage composition to reference conditions (canonical correspondence, other ordination; Jaccard's similarity, other).
- Ratios of algal divisions (e.g., cyanobacteria:total) or other functional groupings (e.g., motile cells:total).
- Ecological categories.
- Trophic state categories (eutrophic, oligotrophic, acidophilic).
- Inferred water chemistry.
- Requires a large calibration data set to develop predictive model of water chemistry.

Spatial Variability—Phytoplankton can be patchily distributed in a lake, affecting the variability of a sampling program. Most phytoplankton patchiness is the result of water motion and identifiable water masses, such as Langmuir circulation, vertical stratification, and embayments with limited exchange to open water due to morphometry or submerged vegetation. Effects of these can be minimized by taking vertically integrated samples in mid-lake, with a vertical tow, a pump, or a series of bottle samples.

Temporal Variability—The largest single disadvantage of phytoplankton sampling, including biomass and chlorophyll *a* measurement, is temporal variability. The algal assemblage seasonal succession cycles are only general, and their exact timing and composition are not predictable (Reynolds 1984). The variability is best controlled with repeated sampling (typically monthly or weekly) using a minimum of 10 samples to obtain either an annual average or an index period average (e.g., growing season, spring overturn, peak biomass) (Knowlton and Jones 1989).

Sedimented Diatoms

Diatoms and chrysophytes preserved in lake sediments are integrators of lake history and make it possible to infer changes in other biotic assemblages (Charles et al. 1994, Dixit et al. 1992). Environmental variables, such as alkalinity, aluminum, dissolved organic carbon, salinity, nickel. conductivity, calcium, total nitrogen, total phosphorus, Secchi transparency, and trophic state have been inferred using diatom-based predictive models (Charles et al. 1994, Dixit et al. 1992, Fritz 1990).

The diatom fossil record can aid in establishing reference conditions. See Appendix C for methods. Surface sediments represent recent or current lake conditions and usually integrate the assemblage over 1 or more years (Dixit et al. 1992). Presettlement conditions may be characterized by sediment cores of 0.5 to 1.0m depths (Charles et al. 1994). Dating sediment cores is possible using pollen or radioactivity of 210Pb (radon decay product).

Periphyton

Periphyton Spatial Variability—Periphyton abundance and species composition might be variable around the periphery of a lake owing to differences in water quality, local variation of runoff from the shore, differences in substrate, and other factors. Periphyton may be scraped from natural substrates, or artificial substrates may be deployed for periphyton colonization (Kentucky DEP 1993, Bahls 1993, Florida DEP 1996, Oklahoma CC 1993). A composite sample from several substrates at several sites should remove most of the effects of local spatial variability.

Temporal Variability—Like phytoplankton, periphyton are subject to changing water chemistry and seasonal succession. Sampling during an index period in a time of relative stability might remove most of the confounding effects of time.

Response of Metrics—Although periphyton have been used successfully in streams (e.g., Bahls 1993, Patrick 1949), their application as lake indicators is relatively new. Metrics of periphytic diatoms have shown promise for bioassessment, based on investigation of undisturbed reference lakes in Montana (Gerritsen and Bowman 1994), but actual response to disturbance or pollution is as yet unknown. Periphyton are considered an experimental assemblage for lake assessment because of limited information on response to stressors.

D.2 SUBMERGED MACROPHYTES

Aquatic plants respond to nutrients, light, toxic contaminants, salt, and management. A lack of macrophytes might indicate water quality problems due to herbicides, salinization, or excessive turbidity. Submerged and floating macrophytes respond to nutrients in the sediment (Barko et al. 1992), and an overabundance of submerged or floating leaved plants can be an indicator of excess nutrients. Exotic species (e.g., Eurasian water milfoil) often become dominant and cause weed problems under eutrophic conditions. In addition, submerged macrophytes are sensitive to shading by turbidity and by dense periphyton growth. Many species are sensitive to phytotoxins, such as copper and herbicides.

Submerged macrophytes are extensively managed. Exotic species frequently dominate eutrophic lakes, and control attempts include harvesting, herbicides, and grass carp. Natural macrophytes are managed where they are thought to interfere with recreation.

Extreme eutrophication in shallow lakes may have alternate stable states: one dominated by macrophytes, the other by phytoplankton (Scheffer et al. 1992). Management of such lakes to promote the macrophyte dominated state includes removal of planktivorous fish and introduction of macrophytes and piscivorous gamefish (Hosper et al. 1992).

Macrophytes respond more slowly to environmental changes than do phytoplankton or zooplankton and might be better integrators of overall environmental conditions (Table D-3). This would allow a single sampling event per year, during the time of maximum abundance of macrophytes. Both floating leaved and emergent plants are easily assessed from aerial photographs, which permit estimates of total area covered and percent cover (density) within stands. For the purposes of lake assessment, emergent vegetation (i.e., semi-terrestrial) is lake habitat, but floating and submerged vegetation are lake biota.

Macrophyte Indicators

Extent and Percent Cover—Extent and percent cover of rooted vegetation are easily obtained from rapid surveys or remote sensing (aerial or satellite imaging). These methods have been used successfully to monitor the status and trends of submerged vegetation in estuaries (e.g., Orth and Moore 1983). Extent of both floating-leaved and emergent vegetation can be estimated from aerial photos or from shorezone surveys. Wetlands can also be estimated from maps developed by the National Wetlands Inventory (NWI), although these would not indicate the extent of littoral emergent vegetation in most lakes. When compared to expected or reference values, the extent and percent cover of macrophytes and emergents provide an assessment of the overall integrity of the lake system. Loss of emergents and wetlands on a lake margin indicates lost wildlife habitat and possibly increased nutrient and sediment input. Nuisance weed problems might indicate eutrophication, and loss of native macrophytes (compared to reference) might indicate excess turbidity or toxic contamination.

Technical Issues

Spatial Variability—With suitable substrate and sufficient light, macrophytes colonize the littoral areas in lakes and reservoirs. Spatial variability of cover and extent within these areas can be a result of one or more of the factors listed below:

- Substrate type—bedrock would be colonized by periphyton instead of macrophytes.
- Topography.
- Disease.
- Insect infestation.
- Local sources of nutrients and turbidity.

Vegetation functional measurements such as net growth, primary productivity, etc., are time consuming and require repeated monitoring at different times in the growing season. It is not clear that the information gained from functional measurements is any better for assessment and management purposes than remote, wide-scale measurements.

Assemblage Metrics—Identification of taxa and relative abundance counts or biomass estimates of each allow calculation of similar assemblage metrics described for the algae assemblages (Table D-4).

To minimize effects of variability, several sites are sampled in a lake and combined into a composite sample.

Temporal Variability—Aquatic macrophyte assemblages on the whole are usually at maximum cover and extent in midsummer. Temporal variability is avoided by sampling the macrophyte assemblage at approximately the same time every year. Interannual variability of macrophyte cover can be high (Scheffer et al. 1992); if so, total vegetated area may not be an effective metric.

Research Needs—It is generally accepted that macrophytes respond to nutrients by expanding their extent and cover. Research is needed to determine which species respond to contaminants such as acid, metals, organics, and salinity. Macrophytes might respond to individual contaminants or only a combination of contaminants. They might respond to contaminants only at extreme levels or conditions.

D.3 BENTHIC MACROINVERTEBRATES

Benthic invertebrate assemblages in lakes correspond to particular habitat types and can be classified according to the three basic habitats of lake bottom: littoral, sublittoral, and profundal. The littoral habitat of lakes usually supports larger and more diverse populations of benthic invertebrates than do the sublittoral and profundal habitats (Moore 1981, Wiederholm 1984). The vegetation and substrate heterogeneity of the littoral habitat provide an abundance of microhabitats occupied by a varied fauna, which in turn enhances invertebrate production. The littoral habitat is also highly variable due to seasonal influences, land use patterns, riparian variation, and direct climatic effects producing high-energy areas. The epifauna species composition, number of individuals, areal extent, and growth form vary with the species composition of the macrophyte beds, making it difficult to determine the benthic status accurately.

The sublittoral habitat, below the area of dense macrophyte beds, but above typical thermoclines, lacks the heterogeneity of the littoral habitat; However, it is also less subject to littoral habitat variables and influences. The sublittoral habitat is rarely exposed to severe hypoxia but might also lack the sensitivity to toxic effects that is found in the profundal habitat. The sublittoral habitat supports diverse infaunal populations, and standardized sampling is easy to implement because a constant depth and

Table D-3. Advantages, disadvantages, and alternatives to using macrophyte assemblages.

Advantages	Disadvantages	Alternatives
Respond to: - Nutrients - Metals - Herbicides - Turbidity - Water level change Structural component; littoral habitat for fauna. Sampling relatively easy (aerial photography or transects); simple abundance metrics. Integrators of environmental conditions. Endpoints of concern (weeds, wetlands, SAV loss).	Subject to management (planted, removed, poisoned). Not important in some regions.	TSI Secchi Nutrient analysis. Metals analysis. Herbicide analysis.

substrate can be selected for sampling. Therefore, the sublittoral habitat is the preferred habitat for surveying the benthic assemblage in most regions.

The profundal habitat, in the hypolimnion of stratified lakes, is more homogeneous due to a lack of habitat and food heterogeneity, and hypoxia and anoxia in moderately to highly productive lakes are common. The profundal habitat is usually dominated by three main groups of benthic organisms including chironomid larvae, oligochaete worms, and phantom midge larvae (*Chaoborus*) (Wiederholm 1984). Many species of chironomids and tubificid oligochaetes are tolerant to low dissolved oxygen, such that these become the dominant profundal invertebrates in lakes with hypoxic hypolimnia. As hypoxia becomes more severe tubificids can become dominant over chironomids (Hergenrader and Lessig 1980). In cases of prolonged anoxia, the profundal assemblage might disappear entirely. If hypoxia is rare in reference lakes of the region, and if toxic sediments are suspected to occur in some lakes, then the profundal habitat might be preferred for the region.

Benthic macroinvertebrates are moderately long-lived and are in constant contact with lake sediments. Contamination and toxicity of sediments will therefore affect those benthic organisms which are sensitive to them (Wiederholm 1984). Acidification of lakes is accompanied by shifts in the composition of benthic assemblages to dominance by species tolerant of acidic conditions (Perry and Troelstrup 1988, Schindler et al. 1989). Effects of rapid sedimentation are less well-known but appear to cause shifts toward lower abundances and oligotrophic species assemblages as well as more motile species (Masters 1992, Wiederholm 1984).

Benthic macroinvertebrates are present year-round and are often abundant, yet not very motile. However, the benthos integrate environmental conditions at the sampling point (Table D-5). To date, TVA, EMAP, and several states (Florida, Oklahoma, North Dakota) have surveyed benthos as part of lake bioassessment in the United States. Developmental work by TVA, USEPA, and several states is likely to refine metrics based on macroinvertebrates.

Invertebrate Indicators

Primary emphasis in the past has been placed on chironomids and oligochaetes as indicators of lake trophic status. Several indices and classification systems have been developed for lake trophic state using chironomid and oligochaete assemblages as indicators (e.g., Naumann 1932). The trophic indices, most of which were developed for lakes of northern Europe, rely on relative abundances of chironomid species, the ratio of tolerant to intolerant tubificid oligochaetes, or the ratio of oligochaetes to chironomids (reviewed in Wiederholm 1980). Ratios are unstable metrics because numerator and denominator are independent (Barbour et al. 1992); proportions or percentage metrics work better.

TVA is using benthic macroinvertebrate composition as one of five assessment indicators in reservoirs (Dycus and Meinert 1992, Dycus and

Table D-4. Potential macrophyte metrics.

Metric	Optimal Condition	Impaired Condition
% cover or biomass in available habitat colonized.	Similar to reference.	Substantially more or less than reference.
% cover, biomass in vegetated areas.	Similar to reference.	Substantially more or less than reference.
No. of taxa.	High	Low
% cover, biomass of dominant species.	Low	High
No. of exotic species.	Zero	≥ 1
% cover, biomass of exotics.	Zero	High

Meinert 1993, Dycus and Meinert 1994). TVA benthic composition metrics evaluate richness, composition, abundance, and indicator taxa. The condition of macroinvertebrate assemblages in TVA reservoirs is strongly associated with hypoxia in the reservoirs (after Dycus and Meinert 1992). The EMAP surface waters pilot project is also using benthic macroinvertebrates for assessing the biological condition of lakes and has found that number of taxa among benthic macroinvertebrates corresponds to level of disturbance in a watershed (USEPA 1993a).

Lake benthic metrics that are responsive to stresses, are in general, similar to stream invertebrate metrics (Table D-6). Metrics used successfully by TVA in assessing reservoirs include (TVA 1994, TVA 1995):

- Number of taxa.
- Number of long-lived taxa (*Corbicula*, *Hexagenia*, mussel, snails).
- Number of EPT taxa (*Ephemeroptera*, *Plecoptera*, *Trichoptera*).
- Proportion as *Tubificidae*.
- Proportion as dominant taxon.
- Total abundance excluding *Chironomidae* and *Tubificidae*.
- Percentage of samples on a transect with no organisms present.

Invertebrate metrics demonstrated to respond to stresses in Florida lakes include (FDEP 1994, Gerritsen and White 1997):

- Number of taxa.
- Shannon-Wiener diversity.
- Percent oligochaetes.
- Number of ETO taxa (*Ephemeroptera*, *Trichoptera*, *Odonata*).
- A tolerance index similar to HBI.

Biological assessment using benthic macroinvertebrates must focus on a subset of assemblages (defined by habitat and season) to avoid costly sampling of all assemblages. Assemblage composition is affected by substrate, macrophytes, depth, and season. The optimal assemblage for reasons of cost, variability, and interpretation appears to be the sublittoral assemblage of epifauna and infauna. The littoral assemblage is highly variable and costly to sample, and the profundal assemblage might be uniformly impacted by hypoxia in many regions of the country. Hypoxia might be natural in deep, mesotrophic lakes or in warm water lakes. If hypoxia is an expected profundal condition, sublittoral benthos is the preferred assemblage. If hypoxia is rare or not expected in the reference condition, profundal benthic sampling might be preferred.

Table D-5. Advantages, disadvantages, and alternatives to using macroinvertebrate assemblages.

Advantages	Disadvantages	Alternatives
Respond to: - DO - Sediment metals - Other sediment toxins - Organic enrichment - Fish Integrators of environmental conditions. Low mobility. Moderate temporal variability. Trophic link to fish, birds.	High spatial variability due to habitat dependence. Littoral habitat sampling difficult. Metrics not well developed or tested in lakes. Laboratory identification and count can be time-consuming, requires expertise.	DO Sediment TOC. Toxicity bioassays. Fish assemblage.

Technical Issues

Spatial Variability—To account for spatial variability within the sampling area of a lake, at least three grabs must be taken. The grabs can be combined into a composite sample to save money, but valuable information is lost. For example, data on spatial variability is lost, but more importantly effects of one sample with a very large density of a single taxon will be more significant in a composite sample than in the average of individual samples. In large lakes or lakes with heterogenous bottom substrate, five or more sites might need to be sampled. Selection of the epifauna or infauna for sampling will depend on the major substrate type present and the overall objectives of the biosurvey. For example, with sediment problem the benthic infauna would be the appropriate part of the assemblage to sample. If the major substrate type present is hard substrate or vegetation, the epifauna should be sampled.

Toxic or contaminated sediments are more likely to be a stress on profundal invertebrates because sediments accumulate in the deep, depositional areas and infaunal oligochaetes might be more sensitive to toxicity than are other invertebrates. However, the sublittoral habitat has certain advantages for sampling macrobenthos because it is subject to hypoxia less frequently than the profundal habitat and because the sublittoral area typically has greater number of taxa, including some mayflies and caddisflies than the profundal area.

Temporal Variability

The issue of seasonality needs further investigation to determine the most effective index period for sampling or the sampling frequency. Sampling period can be either during the most stressful period or during a time after recruitment when the populations have stabilized. The selected period should be of the least consequence to the identification and sampling process, especially if the sampling is designed for volunteer monitoring groups. For example, samples taken right after recruitment will have early instars that are difficult to identify. If more than one period is designated, the appropriate sampling frequency needs to be established.

Sampling Strategies

The sampling area should focus on the most predominant substrate available and the metrics

Table D-6. Potential benthic metrics.

Metric	Response to stress
No. of taxa.	Reduced
Shannon-Weiner diversity.	Reduced
Mean no. of individuals per taxon.	Variable
% contribution of dominant taxon.	Elevated
% intolerant species.	Reduced
% oligochaetes.	Elevated under organic enrichment.
ETO taxa (*ephemeroptera, trichoptera, odonates*).	Reduced under enrichment or DO stress.
% non-insects.	Reduced
Crustacean + mollusc taxa.	Reduced under acid stress.
% crustaceans and molluscs.	Reduced under acid stress.
Tolerance indices (e.g., HBI [Hilsenoff 1987]; Hulbert's Lake Condition Index [LCI]).	Reduced
% suspension feeders.	Reduced
% shredders.	Reduced under enrichment (not useful in very large lakes).
Abundance (exclude *Chironomidae* and *Tubificidae*).	Reduced
No. of samples with no organisms present.	Increased

should be developed independent of microhabitat variation. The type of sampling gear will depend on the substrate being sampled as each substrate has its own optimal sampling gear. Standardized sampling techniques for each gear type should be implemented to allow for the comparison of data. Processing of samples should be standardized by using a standard net size of 595 μm (No. 30 mesh).

The objective is to adequately characterize the sampling unit which is a single lake, embayment, or lake basin. Heterogeneity within a sampling unit (lake) is not of interest in bioassessment. Samples from several sites are combined into a single composite for analysis and characterization of the lake. To get a representative sample of benthic invertebrates, it is necessary to sample at several locations, such as, three to five areas of the sublittoral zone around the lake. Sampling at each site might also consist of several grab, which can be composited to save money.

Research Needs—Six recommendations for further study were identified during the development of this document by the Benthic Workgroup:

1. Metric development and calibration must allow for regional modifications. The need for regional modification of metrics must be clear so that states do not discount the program as "not working" if the metrics being used are not suitable for their region.

2. Sampling methodology must be based on regional characteristics and must be appropriate to the needs of states. Regional adaptations will be based on substrate, habitat, lake type, and other environmental characteristics. Design strategies should also include ways to evaluate the design, and identify specific problems/ characteristics so that states can easily identify whether or not a specific design is working in their region. One suggestion was that a questionnaire accompany the guidance to specifically identify whether the sampling methods were found to be suitable for the region after use for a predetermined period of time.

3. The appropriate number of replicate samples needs to be investigated in order to tighten confidence intervals and to resolve the best returns on the data for the investment. A determination must also be made as to whether multiple sampling efforts should be conducted on the same day or on separate days.

4. Investigate seasonality so that the best index period(s) is selected for sampling. Ideally,sampling should occur during the period that will least affect the field identifications and yield the most valuable information. More than one index period might be needed to address specific objectives. Cost-effective strategies will focus on reduced frequency of sampling.

5. Investigate the applicability of vertical stratification of biomass as in estuarine sediments, and develop a surrogate infaunal trophic index for lakes that might have universal application.

6. Investigate the occurrence and causes of morphological deformities of benthic organisms in response to stressors. This type of metric would provide information on individual health or sublethal effects.

7. Evaluate potential for field identification as cost saving measure.

D.4 ZOOPLANKTON

Lake zooplankton consist primarily of crustaceans, rotifers, and, to a lesser extent, semi-planktonic insect larvae of the genus *Chaoborus*. Many zooplankton species found in north temperate lakes are cosmopolitan or wide-ranging in their distribution (Hutchinson 1967). There is a strong positive relationship between the number of crustacean zooplankton species and lake surface area (Dodson 1992, Fryer 1985), and weaker positive relationships between number of species and lake productivity, and the number of neighboring lakes (Dodson 1992).

More than any other assemblage, zooplankton structure and function are controlled externally by both higher and lower trophic levels (fish predators and algal food) and internally by planktonic predators (Lewis 1979, Zaret 1980, Carpenter et al. 1987) (Table D-7). Zooplankton composition and abundance are variable in time

with numbers changing one to three orders of magnitude within weeks. The complexity of open water zooplankton dynamics is in part due to trophic interactions taking place in a three-dimensional environment of reduced structure (Gerritsen 1980).

The trophic cascade can be modified by nutrient enrichment and internal interactions (Carpenter et al. 1987) and can in turn affect physical characteristics such as light penetration and temperature (Mazumder et al. 1990).

Zooplankton Indicators

Zooplankton indicators that have been investigated rely on measurement of plankton size structure, and trophic categories (Stemberger and Lazorchak 1994) (Table D-8). From the ecological interactions listed above, zooplankton body size is a potential indicator of the presence or absence of planktivorous forage fish, and of the absence or presence, respectively, of large piscivores. Use of zooplankton body size as an indicator (Mills and Schiavone 1982, Mills et al. 1987, O'Gorman et al. 1991) showed that mean zooplankton body size can predict populations of yellow perch and migration of alewives. Furthermore, dominance by large, visible *Daphnia* species (e.g., *D. pulex, D. galeata*) indicates the presence of large piscivores, circumneutral pH, and the absence of blue-green algal blooms (Edmondson and Litt 1982, Mills et al. 1987).

Several zooplankton species, especially some of the larger predators and *Daphnids*, are sensitive to acidification, and acidic lakes have fewer zooplankton taxa than circumneutral lakes (Baker and Christensen 1991). Large *Daphnia* (> 1 mm) are used as an indicator of trophic balance in operational biomanipulation in Europe (Hosper et al. 1992, Hosper and Meijer 1993), and lakes with large *Daphnia* have lower chlorophyll concentrations than comparable lakes without (Mazumder 1994).

The EMAP Surface Waters program is testing selected zooplankton metrics in New England lakes. EMAP zooplankton sampling consists of a single vertical tow at the deepest point of a lake, using a dual (bongo) net, with a fine (48μm) net and a coarse (202μm) net (USEPA 1994a, USEPA 1994b).

Technical Issues

Spatial Variability—Zooplankton are subject to many of the same water movements that affect phytoplankton. In addition, many species perform diurnal vertical migration. Integrated sampling of the mid-lake water column with a vertical or oblique tow is usually sufficient for relative abundances of zooplankton species. To avoid possible effects of vertical migration, samples should not be taken near dawn or dusk.

Table D-7. Advantages, disadvantages, and alternatives to using zooplankton assemblages.

Advantages	Disadvantages	Alternatives
Respond to: - Fish - Phytoplankton - Thermal loading - Acidity - Pesticides Field sampling and counting relatively easy (but does require taxon. expertise). Trophic link to fish. Sedimentary record for some groups.	Response to human stressors and impacts not well documented. Interpretation difficult: respond to both higher and lower trophic levels. Do not integrate well (high temporal variability).	Fish assemblage. Trophic state (Secchi depth, chlorophyll, phosphorus). Algae

Index Period—Zooplankton assemblages are not stable in time undergoing seasonal succession. To the extent that assemblages are seasonally predictable, they can be sampled within an index period. Mid-summer or mid-winter are relatively stable periods. Midsummer is preferred to coincide with other assemblages.

Research Needs—Although preliminary results from EMAP are encouraging, the responsiveness and reliability of many zooplankton-based metrics are not yet well known. Response of zooplankton metrics to stressors, needs to be tested in different regions of the country. Seasonal variability and predictability of zooplankton assemblages needs to be analyzed to determine optimal index periods and the minimum number of samples required to characterize a lake.

D.5 FISH

Fish populations are powerful structuring forces on other lake assemblages through feeding interactions (trophic cascades). Abundant populations of piscivorous fish reduce planktivorous forage fish species, releasing predatory zooplankton from predation, resulting in dominance by large-bodied, suspension-feeding zooplankton (e.g., Brooks and Dodson 1965, O'Brien 1979). The large suspension-feeding zooplankton can in turn reduce phytoplankton abundance, increasing water clarity and altering the thermal structure of the lake (Mazumder et al. 1990). The trophic cascade also influences, and is influenced by, nutrient dynamics (Carpenter et al. 1987).

It is well known that fish production is tied to lake primary production (e.g., Oglesby 1977, Ryder et al. 1974). In fact, oligotrophic lakes are often fertilized by fishery agencies to enhance sport fish production. In addition, there are regional, geographic differences in fish abundance that are not explained by trophic state (Nürnberg 1996). Moderate to severe eutrophication reduces and might eliminate desirable sport fish due to loss of habitat, poor water quality, and food web simplification (NRC 1992). Fish are highly dependent on habitat for spawning and for refuge. Some species (e.g., yellow perch, most salmonids) spawn in streams; others require clean rock or gravel habitat in the lake. Submerged vegetation provides cover for both forage fish and piscivores, and recolonization of littoral areas by macrophytes increases sportfish abundance, as well as improving water quality.

More than any other assemblage, fish are subject to management which can confound assessment efforts (Table D-9). Exotic piscivorous sport fish (e.g., striped bass, Pacific salmon) are widespread in lakes throughout the United States, and many of these populations are maintained by regular stocking. Stocking to maintain a population results in an artificially large population—especially juveniles—with resultant trophic cascade effects on zooplankton and phytoplankton. This problem is especially pronounced in "put and take" fisheries, where large numbers of hatchery-reared adults are released for a fishing season and decimate invertebrate assemblages during the season. In general, if exotic piscivorous species reproduce naturally, biological integrity is less likely to be affected.

Table D-8. Potential zooplankton metrics.

Metric	Response to stress
% large *Daphnia* (> 1 mm).	Low
No. of taxa.	Reduced under contamination or stress.
% dominance.	High
Size structure (% of large animals or % of small animals).	Dominated by small species (e.g., rotifers).
Trophic structure metrics - No. of trophic links - Complexity measures - % large predators - No. of predator species	Simplified trophic structure.

Fish Indicators

Assemblage Composition and Abundance—Measurements of fish assemblage composition and relative abundance can be incorporated into several metrics, including the Index of Biotic Integrity (IBI), an index of several assemblage-level metrics and their variations, and multivariate assemblage analysis. Field measurements for these are the relative abundances of species in the habitat.

Index of Biotic Integrity (IBI)—The Index of Biotic Integrity (IBI) incorporates attributes of fish assemblages to evaluate human effects on a stream and its watershed (Karr 1991, Karr et al. 1986). Those attributes cover the range of ecological levels from the individual through population, community, and ecosystem. IBI consists of 6 to 12 measures, or metrics, in 4 broad categories: species composition, trophic composition, fish abundance, and condition (Karr 1991). A site is assigned scores for the resemblance of each metric to the reference (unimpacted or least impacted) condition expected for that area. Total scores of all metrics result in an overall score for the site.

As with other multimetric indices, component metrics of IBI require adaptation and calibration to the geographic regions in which they will be applied, thus incorporating biogeographic variation of assemblages and systems into the assessment (Karr 1991). This may include deletion or replacement of selected IBI metrics and is done with the development of a reference site data base. Local adaptations of IBI for streams have been developed for several regions of the United States (Karr 1991, Leonard and Orth 1986, Miller et al. 1988, Steedman 1988).

Although lakes and reservoirs differ in physical attributes from rivers and streams (the former being more homogenous), the valued attributes, or biological integrity, of fish assemblages apply equally. These attributes include species composition, trophic composition, abundance, and condition. Differences between lake and stream habitats lie in the expectations for the attributes and will be reflected in reference site data. An index used by TVA on its reservoirs is based on 12 metrics and is called the Reservoir Fish Assemblage Index, or RFAI (Jennings et al. 1995, Hickman and McDonough 1996) (Table D-10). The status of this index is discussed under Research Needs.

The major problem in applying IBI to lakes is obtaining representative samples of fish assemblages in lakes. Quantitative sampling in lakes is not as reliable as that in streams because of lake morphology, bottom types, and gear efficiency. Modification of IBI for lakes may include use of relative abundances based on subsamples from constant-effort sampling. Sampling gear and protocols for different habitats of lakes will need to be standardized.

Table D-9. Advantages, disadvantages, and alternatives to using fish assemblages.

Advantages	Disadvantages	Alternatives
Respond to: - DO - Pesticides - Metals - Organic enrichment - Eutrophication - Acidification - Thermal loading Tolerances to stress known. Integrators of environmental conditions. Easy taxonomic ID; expertise widespread. Universal endpoint.	Filed sampling is time consuming and expensive, with high spatial variance and gear problems. Intensively managed. - Stocking - Angling impact of sampling The only index which has been developed (RFAI), has only been tested regionally.	DO Trophic state. Toxicity bioassays. Contaminant analysis. pH, alkalinity measurement.

Qualitative Screening—Widespread familiarity with the condition of sport and forage fish in natural resource agencies permits qualitative screening assessment using expert knowledge of local and state fisheries experts (USEPA 1989b). The intent is to serve as a screening tool and to maximize the use of existing knowledge of fish assemblages with a questionnaire polling state fish biologists and university ichthyologists believed knowledgeable about the fish assemblages in lakes of concern. Unlike field surveys, questionnaires can provide information about tainting or fish tissue contamination and historical trends and conditions. Disadvantages of questionnaires include inaccuracy caused by hasty responses, a desire to report conditions as better or worse than they are, and insufficient knowledge.

Contaminants in Fish Tissue—Contaminant concentrations in fish tissue have been monitored to assess the extent of environmental contamination and to estimate risks to human health from consuming fish. Contaminant concentration is an excellent indicator of health risk, but it is not an indicator of biological integrity.

Pathology—Pathological abnormalities (lesions, tumors, growth anomalies) of fish are monitored as overall indicators of environmental degradation, including effects of severe eutrophication, sediment contamination, and acidification. Significant rates of pathology typically occur only in the most severely polluted habitats and in populations of nonmigratory, bottom-feeding fish. Pathology can be incorporated into multimetric indices, such as IBI (Dionne and Karr 1992).

Technical Issues

The major problem in developing fish indices for lakes is obtaining representative samples of fish assemblages in lakes. Quantitative sampling in lakes is not as reliable as in streams because of lake morphology, bottom types, and variable gear efficiency. Modification of IBI for lakes can include use of relative abundances based on subsamples from constant-effort sampling. Sampling gear and protocols for different lake habitats will need to be standardized.

Spatial and Temporal Variability—Fish are highly mobile and respond rapidly to gradients in physical habitat and water chemistry. They actively avoid harmful conditions. Physical and chemical parameters that affect fish spatial distribution include:

Table D-10. Fish assemblage metrics under investigation by TVA. After Dycus and Meinert (1994) and Hickman and McDonough (1996).

Metric	Optimal Condition	Impaired Condition
Species Richness and Composition		
- No. of taxa.	High	Reduced
- No. of Lepomis sunfish species.	High	Reduced
- No. of sucker species.	High	Reduced
- No. of intolerant species.	High	Reduced
- % tolerant individuals.	Low	High
- % dominance by one species.	Low	High
Trophic Composition		
- No. of piscivore species.	High	Reduced
- % omnivores.	Low	High
- % invertivores.	High	Low
Reproduction Composition		
- No. of lithophilic spawning species.	High	Low
Abundance		
- Total individuals.	Similar to reference.	Reduced
Fish Health		
- % individuals with anomalies.	Low	Increased

- Habitat.
- Cover.
- Dissolved oxygen.
- pH.
- Temperature.
- Turbidity.
- Light.

Many fish seek specific habitats for activities such as feeding, resting, and spawning. Their movement between habitats is dependent on time of day and season. Although fish populations are relatively stable compared to smaller, shorter-lived plankton and benthos, fish mobility and behavior make fish difficult to sample.

Index Period—Sampling during the spring coincides with optimal biological conditions and may show recovery from environmental stress periods. However, to avoid spring spawning, sampling is usually conducted in late summer and early fall. Seasonal changes in the relative abundances of the fish assemblage occur primarily during reproductive periods and (for some species) the spring and fall migratory periods. If fish sampling is required during this period, then changes in relative abundance will be important. Mid to late summer is often a time of oxygen stress and should show the greatest effects from environmental stress.

Sampling Gear—Obtaining both qualitative and quantitative data on fish populations is limited by gear selectivity and the fish mobility (USEPA 1992b). All sampling gear is selective. The habitat or portion of habitat sampled and efficiency of gear for a particular species in one area does not necessarily apply to different species nor to the same species in another area. Temporal and spatial changes in relative abundance of a species can be assessed under a given set of conditions if those species are readily collected with a particular kind of gear.

Electrofishing is the technique used most often by agencies that monitor fish assemblages. The EMAP Surface Water Northeast Lake Pilot Survey found electrofishing the most effective single-gear technique (USEPA 1994a, USEPA 1994b). The RFAI for TVA reservoirs includes electrofishing as a collection technique (Hickman and McDonough 1996). Other considerations with respect to electrofishing are:

- Many agencies already have the equipment available.
- Electrofishing is easy to use and produces quick results.
- It is depth- and species-selective and does not effectively sample catfish or any fish in deeper water. It can be difficult to get close to some fish (e.g., northern pike).
- It can be difficult to get equipment into remote areas.

Seining, an active sampling technique, can be used in the littoral areas (straight seines). Haul seines and trawls are used in deeper open water areas. Seining or trawling is not effective in areas with bottom obstructions that can tear or foul the net. Although the results are expressed as number of fish captured per unit effort, quantitative seining is very difficult. This method is more useful in determining the variety of fish rather than the number of fish inhabiting the water.

Athough gill nets are a passive technique with several disadvantages, they might be the most appropriate gear type for sampling deep sublittoral habitats. Gill nets are size-selective, depending on mesh size and do not obtain representative samples of the total population. They are most effective on lake herring, trout, lake whitefish, yellow perch, walleyes, and northern pike (USEPA 1992b). There is a high mortality rate of fish caught in gill nets and occasional mortality of nontarget species such as turtles, muskrats, beavers, and diving waterfowl. Trap and fyke nets are effective in shallow areas. Like gill nets, they are also passive and do not obtain a representative sample of the total population. Meador et al. (1993) and Weaver (1993) recommended a multi-gear approach that takes advantage of differences in gear selectivity and efficiency to achieve a more accurate representation of the fish assemblage structure. TVA uses shoreline electrofishing for the shallow littoral zone and experimental gill nets for the sublittoral/limnetic zone (Hickman and McDonough 1996).

Research Needs—TVA has been actively developing assessment tools for its reservoirs for several years. The move to a multimetric approach for reservoir fish began in 1990. Successive steps in this development process have brought continued improvement to the RFAI. Potential improvements in the fish indices include using a simple random sampling design rather than a fixed station design to enhance statistical validity with little increase in variability. Use of the index in reservoirs or other river systems is necessary to test its performance under a wider range of conditions than is available in the Tennessee River. Correlation with known human-induced impacts remains a critical need before general acceptance of the fish index as a reliable method to address reservoir environmental quality.

A related issue is the effect of game fish management on IBI or other fish assemblage metric scores. Nearly all lakes are stocked or have been stocked in the past, and these practices can affect the biological assemblages in a lake. Stocking lakes with large piscivores is also used in biomanipulation to improve water clarity of eutrophic lakes (e.g., Hosper et al. 1992).

Appendix E

Statistical Analysis Methods for Biological Assessment

INTRODUCTION

A central premise of biological assessment is comparison of the biological resources of a waterbody to an expected reference condition (Figure E-1). Impairment of the waterbody is judged by its departure from the expected condition. This approach presumes that the purpose of management is to prevent, identify, and subsequently repair anthropogenic damage to natural resources. Biological assessment of waterbodies is predicated on our ability to define, measure, and compare an assessment endpoint between similar systems. This guidance outlines analytic methodologies to perform two tasks shown in Figure E-1:

- Characterization of the biological expectation.
- Determining whether a site deviates from that expectation.

All of the methods considered here use the same general approach: sites are assessed by comparing the assemblage of organisms found at a site to an expectation derived from observations of many relatively undisturbed reference sites (Figure E-1) The expectations are modified by classifying the reference sites to account for natural variability, and each assessment site is classified using non-biological (physical, chemical, geographic) information. Biological variables are tested for response to stressors by comparison of reference unimpaired sites and known impaired sites. A set of "rules" are developed from this information, which are then used to determine if the biota of a site deviate from the expectation, indicating that the site is impaired.

Several analytic methods have been developed to assess the condition of water resources from biological data, beginning with the saprobien system in the early 20th century to present-day development of biological markers. This appendix outlines three methods for analyzing and assessing waterbody condition from assemblage and community-level biological information:

1. Multimetric assessment using an index that is the sum of several metrics. This is the basis of the Index of Biotic Integrity (IBI) (Karr et al. 1986), the Invertebrate Community Index (ICI) (Ohio EPA 1990); the Rapid Bioassessment Protocol (USEPA 1989b); and state indices developed from these (e.g., Southerland and Stribling 1995, Barbour et al. 1996a, Barbour et al. 1996b).
2. Multimetric assessment using an index that is developed from a multivariate discriminant model to discriminate reference from impaired sites. This is the basis

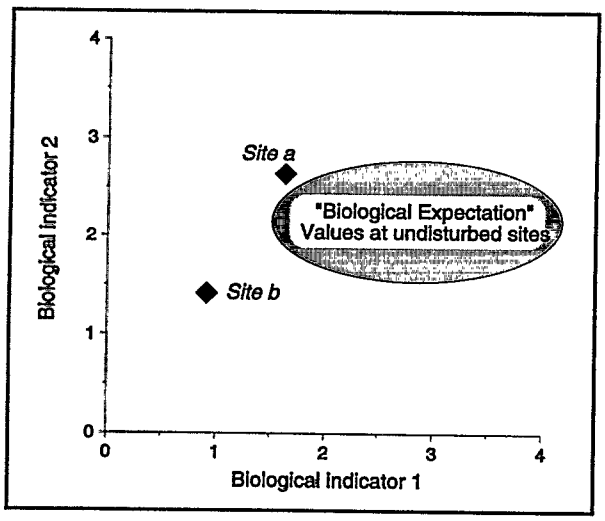


Figure E-1 Graphical representation of bioassessment. Assessment sites a and b are compared to an ideal biological expectation. Site a is near to its expectation; Site b deviates from it and is considered to be impaired.

of stream bioassessment in Maine (Davies et al. 1993), and of the estuarine invertebrate indices developed by the EMAP-Estuaries program (USEPA 1993e, USEPA 1994h,USEPA 1994i, Engle et al.1994).

3. Assessment using multivariate ordination of species abundances. This methodology has been used widely in assessment of streams in Britain (e.g., Wright et al. 1984); assessment of marine macroinvertebrates in the North Sea (e.g., Warwick and Clarke 1991); and in assessment of benthic macroinvertebrates in the Great Lakes (e.g., Reynoldson et al. 1995).

Many other methods are possible, as well as permutations of the three methods above, all of which are beyond the scope of this document. The three approaches were selected because:

- They use community and assemblage data.
- The methods are not restricted to any one assemblage. The examples all use benthic macroinvertebrates and fish (freshwater and estuarine), but any other assemblage could also be used, such as phytoplankton, zooplankton or macrophytes.
- The methods are general, and have been used by many agencies in many areas. The examples used to illustrate the methods have also been carried out over wide geographic areas with many sites, demonstrating the generality of the methods.
- The examples used to illustrate the methods are concise. Methods were fully documented, and have been carried to completion, that is, assessment of biological impairment and non-impairment.

The optimal analysis methodology should also be cost-effective and easy to communicate to managers and the public. Both the multimetric index and the discriminant model index (approaches 1 and 2) are easy to apply in a continuing operational monitoring program because data from an individual site are entered into a formula, and the site's deviation from reference conditions can be known immediately (Gerritsen 1995). The ordination approach (3) requires reanalysis of a the entire reference data set for each new batch of monitoring sites. The multimetric index (approach 1) is the easiest to explain to managers and the public because it does not rely on specialized concepts such as multivariate statistics. The ordination approach (3) may be most cost-effective if the biological survey is a single event—a large number of sites are surveyed once, and there is no plan to continue monitoring or to survey new sites.

Characterization of Reference Conditions

Reference conditions establish the basis for comparison and for detecting impairment of waterbodies. They should be applicable to an individual waterbody, such as a stream or lake and also to similar waterbodies on a regional scale (USEPA 1996a).

Classification Tools

The objective of classification is to group similar waterbodies together, so that reference conditions will reflect reasonable expectations for assessing waterbodies. There are two fundamental approaches to classifications: *a priori* or rule based, where known rules are applied to classifying objects; and *a posteriori*, or data-based, where rules for classifying objects are derived from data obtained from the objects (waterbodies) themselves (Conquest et al. 1994).

For example, a rule-based classification may divide mountain and lowland streams by elevation or stream gradient. An *a posteriori* classification would examine data from all streams, and determine if there is a basis for separating them into two or more classes (not necessarily including elevation or gradient). The *a posteriori* approach requires a relatively large sample of reference sites to derive the classes and rules, with both biological and physical-chemical data from each site.

The basic assumption of classification is that physical habitat and water quality largely determine the composition of biological communities in waterbodies. Therefore, if waterbodies are classified adequately, reference biological community types should correspond to the classification. Classification is often an iterative process of refining the classification scheme as new data are obtained, until a satisfactory classification emerges that accounts for variation in the reference site biological data.

Several statistical tools can assist in site classification, but there is no set procedure. If *a priori* classification is based on well-developed prior knowledge, then graphical analysis of biological data, followed by any necessary modifications and tests of the resultant classification, may be sufficient.

If a rule-based classification is not self-evident, then it may be necessary to develop an alternative classification from the data using one or more analytical classification approaches. These methods include several cluster analysis methods, and several approaches to ordination analysis, including principal components analysis (PCA), correspondence analysis (CA) and its variants, and non-metric multidimensional scaling (NMDS).

In statistical terminology, each site is a sample unit (SU) (Ludwig and Reynolds 1988). Ideally, sample units should be independent, which is generally achievable in small streams and in many lakes, where each waterbody can be a separate sample unit. Large lakes and reservoirs, large rivers, and estuaries may include several sample units within the same waterbody. For large and complex lakes and estuaries, it may be necessary to define a site as a contiguous basin or embayment. Any portion of the waterbody that is partially isolated from the rest by bottom topography or water motion should be considered a separate site and sampled accordingly. This also applies to the three zones of large reservoirs (riverine, transition, and forebay) and to salinity zones of estuaries (e.g., fresh, mesohaline polyhaline), which have different biological communities and dynamics even though they are not hydrologically isolated (Thornton 1990b). Thus, large waterbodies (including large reservoirs) may comprise several sites or SUs. Sites (SUs) are considered independent and are kept separate in analysis; no "average" is estimated for a multiple-site waterbody. Multiple sites are not strictly independent and will need to be considered carefully in reference condition characterization and in metric response evaluation.

Large rivers may be more problematic in that sites on a river are serially linked by water flow. Sites are defined as river reaches of some minimum length that exhibit some (but not complete) independence. Sample units (reaches) may be defined by length (e.g., a set length or a multiple of stream widths), as the reach between major tributaries, or as segments downstream of major impacts and discharges (e.g., urban areas).

Graphical Analysis

A key graphical display is box-and-whisker plots (Figure E-2). These show population attributes of the data: central tendency, spread, and outliers. In the display used here, the central point is the median value of the variable; the box shows the 25th and 75th percentiles (interquartile range); and the whiskers show values within the inner fences (Figure E-2). Points beyond the fences may be considered outliers or extreme values. Box-and-whisker plots are simple, straightforward, powerful, and the interquartile ranges are used to evaluate whether there is a real difference between two areas and whether a metric is a good candidate for use in assessment. Graphing the data should always be a first step in data analysis.

Statistical methods used by biologists are frequently tests of whether two or more populations have different means using t-tests, analysis of variance, or various nonparametric methods. However, the fundamental problem of biological assessment is not to determine whether two populations (or samples) have a different mean, but to determine whether an

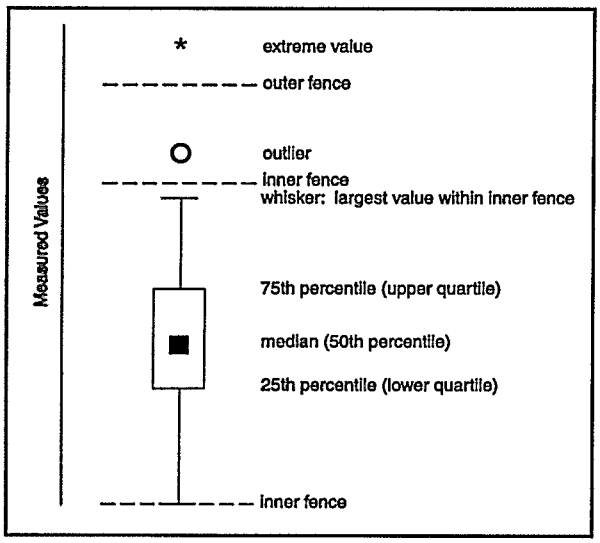


Figure E-2. Box and whisker diagram (after Tukey 1977). The box is the interquartile range (25th - 75th percentile). Inner fences are the quartiles ± 1.5 × interquartile range; outer fence is 3 × interquartile range. Ends of whiskers are the most extreme observations within the inner fences.

individual site is a member of the least-impaired reference population (Figure E-1). If it is not, then a second question is how far it has deviated from that reference. Therefore, biological assessment requires the entire distribution of a metric, which is effectively displayed with a box-and-whisker plot.

Ordination Methods

The purpose of ordination analysis is to reduce the complexity of many variables (for example, abundance of 200 species from 50 sites) into fewer variables, such that the sites and the species are ordered on the new variables (Figure E- 3). The new variables are called the principal axes of the analysis; the first axis accounts for the most variation in the original data, the second accounts for somewhat less variation, and so on. Typically, only the first two to four axes of the analysis are presented because higher axes contribute little to the variance explained and because one cannot present or conceptualize more than three axes simultaneously.

Principal Components Analysis—One of the most commonly used ordinations is principal components analysis (PCA). In PCA, the new variables (principal axes) are linear combinations of the original data; that is, the relationship between each principal axis and the abundance of each species can be expressed as a straight line, as in simple linear regression (Jongman et al. 1987). Thus, PCA is a multivariate extension of linear regression (Figure E-3), making the assumption that a variable will have a maximum value at one end of a principal axis and minimum value at the other. Because the principal axes can be seen as environmental gradients to which the species respond, ordination is also called gradient analysis (Jongman et al. 1987).

The procedure of PCA is an eigenanalysis of the correlation matrix among variables in the original data matrix. The variables may be species abundance, calculated assemblage metrics, or environmental (chemical and habitat) variables. Eigenanalysis results in as many eigenvalues as there are rows (or columns) in the correlation matrix, and each eigenvalue and corresponding eigenvector describes an axis of the ordination. The eigenvalue of an axis is the variance accounted for by that axis. Often, only the first two or three axes explain significantly more variance in the original data than a random axis. Rules for determining the number of significant axes are explained in Jackson (1993b). Details of formulas and calculations for PCA, as well as variations of PCA, are in Ludwig and Reynolds (1988).

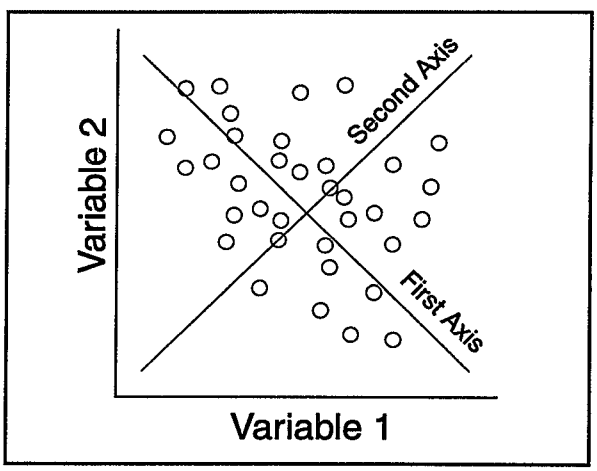


Figure E-3. Ordination. The relationship of species 1 and species 2 can be described by translating and rotating the axes, so that most of the variance is on the first axis. In this 2-dimensional example, the observations have been reduced to a single dimension, the first axis, which is a linear combination of species 1 and species 2.

Because PCA is linear, and assumes multivariate normal distributions, data transformations are often necessary. Species abundance data usually have many zeros in the data matrix, and no transformation will normalize them. PCA is not useful for species abundance data, although it can be made to work well for data that are normal or can be transformed to a normal distribution (e.g., environmental variables, assemblage attributes such as number of taxa , etc.).

Correspondence Analysis Family—A problem with linear ordinations such as PCA is that species do not always respond linearly to gradients; in fact, a unimodal response to environmental gradients is much more common (Jongman et al. 1987). A unimodal response is one in which a species has peak abundances at certain optimal values of an environmental variable (for example, pH or nutrient concentration) and abundances are lower at both higher and lower values of the environmental variable. There are many examples of environmental optima for aquatic organisms; optima are supported by uptake kinetics, and they form the basis for resource-based competition and seasonal succession (e.g., Tilman 1982).

Multivariate ordination based on unimodal responses to environmental gradients is called correspondence analysis. As in PCA, correspondence analysis also seeks new variables to explain the species abundances on fewer axes and is frequently "detrended" to eliminate a mathematical artifact from its calculation (Jongman et al. 1987).

Ordination can also be done to develop associations between the species abundances and measured environmental variables. In this case, both species abundance and the environmental variables are related to the principal axes and the whole procedure can be regarded as a multivariate, multiple regression. The linear form is called canonical correlation (CC); the unimodal form is called canonical correspondence analysis (CCA). Because it assumes unimodal responses, CCA is thought to be a realistic and robust multivariate ordination (ter Braak 1986, Palmer 1993).

In CCA, each species, site, and environmental variable has a score on each of the principal axes. Results of CCA are presented graphically by plotting the scores on two of the axes (usually the first two) (Figure E-4). Plotting site scores with environmental variable scores shows the relationship between the sites and the environmental variables and can also show clustering of sites.

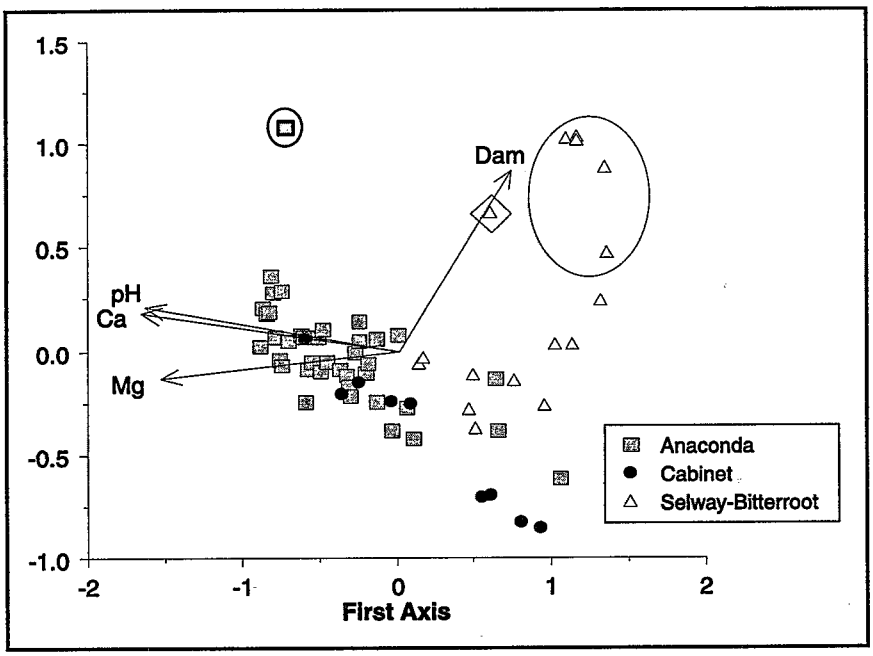


Figure E-4. Canonical corresponence analysis of periphytic diatom assemblages from Rocky Mountain lakes. Site scores (points) and environmental variables (arrows) on the first two axes. Points within ovals are lakes with dams at their outlet; single point inside diamond is dammed by a natural glacial moraine.

Nonmetric Multidimensional Scaling—Nonmetric multidimensional scaling (NMDS) is increasing in use in ecological application because it offers several advantages over other ordination methods. Because the ordination works on a matrix of distance ranks, it is distribution-free and hence unaffected by non-normality and nonlinearity in the data (Ludwig and Reynolds 1988). It is robust and produces interpretable

ordinations from different ecological data sets. The disadvantages of NMDS are that it is iterative and subject to local minima (SYSTAT 1992) and that no canonical form has yet been developed. It is possible, however, to estimate correlations of environmental (explanatory) variables with the axes of NMDS.

Like cluster analysis, NMDS uses a distance metric among sample units (sites), and results can be sensitive to the choice of the distance metric (Jackson 1993a). Bray-Curtis distance and the relative distance metrics (relative Euclidean distance and chord distance) tend to work best (Kenkel and Orloci 1986, Ludwig and Reynolds 1988).

The objective of NMDS is to obtain a "best fit" between the dissimilarity measures and the distances calculated in ordination space. The dissimilarities have as many dimensions as there are sites, but the ordination reduces these to a smaller number, usually 2 or 3. The procedure is to rank the distances in the similarity matrix from smallest to largest, then to calculate an initial starting ordination (termed the initial configuration) directly from the dissimilarity matrix. Intersite distances are calculated from the initial configuration, ranked, and compared to the ranked dissimilarities. A best solution is sought iteratively, changing the configuration so that the two rankings (dissimilarities and configuration) become more similar. Goodness of fit of the configuration to the dissimilarities is measured by Kruskal's stress coefficient (Ludwig and Reynolds 1988) or Guttman's coefficient of alienation (SYSTAT 1992). Iterations stop when stress or alienation reaches a minimum value.

NMDS is available on many commercial statistical software packages. Distance measures used by ecologists, especially Bray-Curtis distance and chord distance, are not usually available in these packages and must be calculated separately. Relative Euclidian distance is also only rarely available; however, if an input matrix of percent abundances of species is used, then Euclidean distance will yield relative Euclidean distance.

Results from NMDS are a final configuration, consisting of coordinates for each site in the 2 or 3 dimensional ordination. As in other ordinations, points close to each other in the ordination space (Figures E-3 and E-4) represent sites with similar species composition.

Classification Analysis—Classification, or the placement of objects into categories, is an innate human activity. A wide variety of formal classification procedures have been developed (see Gauch 1982 for a review). Only two will be discussed here, cluster analysis and two-way indicator species analysis (TWINSPAN).

Cluster Analysis—Cluster analysis is known as an agglomerative classification, that is, it successively builds clusters until all objects have been joined in a single cluster. Cluster analysis begins with a matrix of intersite dissimilarities. The smallest dissimilarity in the matrix is selected and those two sites are joined in a cluster. The algorithm then calculates the dissimilarities between the new cluster and all other sites or clusters. Again, the smallest dissimilarity is selected, the two objects are joined, and the process repeats itself until all objects are joined. Results can be shown in a dendrogram (Figure E-5), where the bars connecting clusters represent the dissimilarity between them. Final clusters are identified by choosing a cutoff dissimilarity value. The cutoff dissimilarity value clearly affects the number of clusters (Figure E-5): it may range from one to the number of sites. The number of clusters should be small, and should explain as much variance of the biological data as possible.

Classification with cluster analysis is not as straightforward and objective as is implied by a dendrogram produced by a mathematical algorithm. First, several algorithms may be used for recalculating dissimilarities among agglomerated clusters of sites, and each algorithm may produce different results. A favored algorithm for ecological data is the unweighted pair-group method (UPGMA) (Ludwig and Reynolds 1988, Reynoldson et al. 1995). Second, the dissimilarity measure affects results. As in NMDS analysis, relative dissimilarity measures (relative Euclidean, chord distance) and Bray-Curtis distance work best for species-abundance data (Ludwig and Reynolds 1988). Finally, as noted above, selection of a distance cut point for defining clusters is subjective (Figure E-5).

Two-way Indicator Species Analysis (TWINSPAN)—TWINSPAN was developed by Hill (1979), and is a divisive technique. Instead of

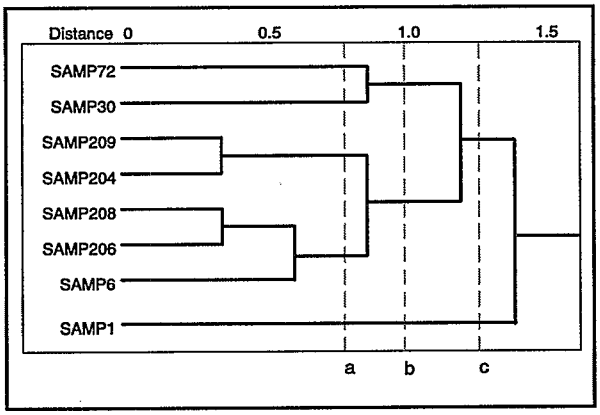


Figure E-5. Dendrogram from cluster analysis. Cutpoints a, b, c are at distances 0.75, 1.0, 1.25, respectively, and result in 5, 4, and 2 clusters, respectively.

building up clusters from individual sites, divisive methods start with the entire data set and divide it into two. The division process is repeated until a specified number of clusters are obtained (Gauch 1982). TWINSPAN first ordinates the data, then divides the sample into two clusters near the middle of the first ordination axis. Ordination is by reciprocal averaging, which is a variation of correspondence analysis. New ordinations are repeated on each daughter cluster, and the daughters are in turn divided on their first ordination axis. TWINSPAN is only available in specialized software packages.

Discriminant Model—The objective of a discriminant model is to predict community type, or community composition, from non-biological data. Development of such a model requires a data set with both biological and non-biological data, and testing of the model requires a second, similar data set. Discriminant analysis is best illustrated with a simple example (e.g., Ludwig and Reynolds 1988, Johnson and Wichern 1992). Suppose that abundances of two species are examined in riffle and pool sites of streams (Figure E-6) and we wish to develop a model that will discriminate between riffle and pool sites, using only the biological data. As shown in the figure, pool sites tend to have greater abundances of both species. Using either species alone to form the rule would lead to frequent errors. Discriminant analysis finds a best fit straight line to separate the groups; the heavy line of Figure 6 is the border and the hatched line perpendicular to it is the discriminant function. Sites with positive scores are more likely to be pools, and sites with negative scores are more likely to be riffles.

Discriminant function analysis involves computation of a pooled variance-covariance matrix of the groups, and solving for the coefficients of the discriminant function. Formulae and computations are shown in Ludwig and Reynolds (1988), Johnson and Wichern (1992), Pielou (1977), and other multivariate statistics textbooks. Discriminant analysis also allows calculation of multivariate distance (Mahalanobis D^2) between groups, and an F-test for group differences. A limitation of discriminant function analysis is that it is linear; strong nonlinearity of the data will reduce its power to separate groups.

Rule-Based Classification: Characterization of Reference Conditions

The objective of reference characterization is to describe (characterize) each of the reference classes in terms of biological indicators and other descriptive variables. The first step is to support or reject the *a priori* classification, followed by modifying it to arrive at a parsimonious and robust classification; that is, one with the fewest classes that explains the most variance in the reference data set.

There is no single "best" classification nor are resources generally available to determine all possible differences between all waterbodies in a region. The key to classification is practicality within the region or state in which it will be applied; local conditions determine the classes. Classification will depend on regional experts familiar with the range of conditions in a region as well as biological similarities and differences among waterbodies. Ultimately, classification can be used to develop a predictive model of those chemical and physical characteristics that affect the values of the biological metrics and indices in reference sites.

A useful classification scheme is hierarchical, beginning at the highest (regional) level and stratifying as far down as necessary (Conquest et al. 1994). The procedure is to classify waterbodies according to region and then to increase the stratification in the classification hierarchy to a reasonable point for the given region. Although several classification levels are possible, in practice, only one, or at most two,

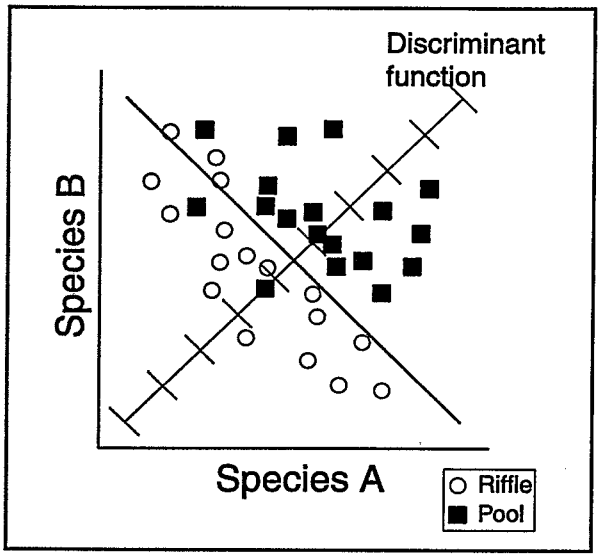


Figure E-6. Illustration of discriminant function analysis. Neither species A nor species B can be used alone to distinguish riffle from pool sites. Discriminant analysis estimates a linear border between the two site classes (heavy line), and a discriminant function (graduated line). The discriminant function is a linear combination of the input variables (species A and species B), and yields a probability that a site belongs to the riffle or pool class.

relevant levels would typically be used. Classification should avoid a proliferation of classes that do not contribute to assessment. One or two relevant levels of the hierarchy will yield the best classification scheme. Potential hierarchical classifications for streams, lakes, and estuaries, respectively, are given in Gerritsen (1995),USEPA (1996a), and USEPA (1997a).

Confirmation of *a priori* Classification

Univariate Tests—Univariate tests of classifications include all the standard statistical tests for comparing two or more groups: t-test, analysis of variance, sign test, Wilcoxon rank test, and Mann-Whitney U-test (USEPA 1996b, Ludwig and Reynolds 1988). These methods are used to test for significant differences between groups (classes) to confirm or reject the classes. They are univariate, with a single dependent (response) variable. Biological variables (metrics) may require transformation to meet assumptions of t-tests and ANOVA, or non-parametric tests (e.g., rank tests, Mann-Whitney) may be used. See USEPA (1996b) for discussions on the use of these and other univariate tests for biocriteria. Failure to confirm the classification for any single response variable does not mean that it will fail for other response variables. Because assessment is based on multiple variables (metrics or species composition), multivariate tests might be more convenient than a succession of individual tests.

Discriminant Analysis—Discriminant analysis can be used as a form of multivariate, one-way analysis of variance that tests differences between a set of groups based on several response variables. It is used as a test of classifications (Conquest et al. 1994), provided that the assumptions of linearity and normality are met. Many statistical software packages provide discriminant analysis.

Gradients—On occasion, environmental gradients might not allow formation of discrete site classes. For example, the number of zooplankton taxa in lakes is usually related to lake size (e.g., Dodson 1992). Similarly, fish and invertebrate number of taxa in streams is typically related to stream size (order, discharge or watershed area) (e.g., Ohio EPA 1987, DeShon 1995)

Ordination—The *a priori* classification may also be confirmed with one of the ordination methods. Sites are plotted in ordination space using different symbols for the *a priori* classes. If classes overlap completely in ordination space, then there is no apparent difference in their species composition (or other variables used in the ordination), and it may be appropriate to aggregate the coinciding classes. Species or variable scores can be plotted in ordination space to determine which contribute most to separation among classes. Correlation coefficients of environmental variables with the site scores will show if there are environmental gradients that are associated with the ordination and with the site classes. Examples and detailed methods for ordinations are given in Jongman et al. (1987) and Ludwig and Reynolds (1988).

A Posteriori Classification

This method of classification determines classes from the structure of the data, rather than from pre-existing knowledge or hypotheses. Because

the principal goal of classification in biocriteria programs is to account for biological variation, the biological data (typically species composition data) are used for classifying. As with *a priori* methods, only data from reference sites are used to develop the classification (e.g., Moss et al. 1987, Wright et al. 1984, Reynoldson et al. 1995).

Test sites must also be assigned to appropriate classes, so that they can be compared to reference sites. Because anthropogenic degradation affects the biota of the waterbodies, assigning test sites to classes using their biological data may lead to incorrect classification (Figure E- 7). Therefore, the classification also requires a method to assign test sites to classes, using non-biological measures that are not affected by anthropogenic degradation. Following an *a posteriori* classification, this is typically a discriminant function model that is constructed from the reference data set (Norris 1995).

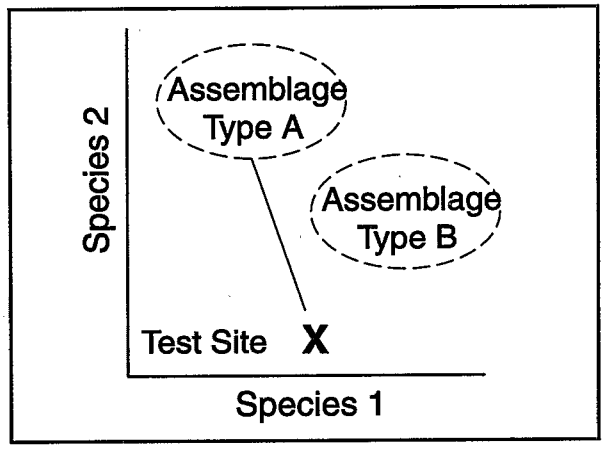


Figure E-7. Misclassification of test sites. A test site (X) that was originally in assemblage Type A has been degraded (arrow). If biological data are used to classify the test site, then it would be classified as Type B because it is now more similar to Type B. If, on the other hand, non-biological measures that are not affected by degradation are used to classify the test site, then it would be correctly identified as Type A and the degree of biological degradation could be assessed.

Identifying Classes

Classification is a subjective activity even when it is done with seemingly objective quantitative methods. The subjectivity is due in part to the information that will be used to decide if objects are similar or not, and in part to the methods and their variations that will be used to classify the objects. For example, we may say that Miami is similar to Havana. We may also say that, during the Cold War era, Havana was similar to Moscow. Does it then follow that Miami is similar to Moscow (SYSTAT 1992)? This example illustrates that the variables used to determine similarity (climate, economic system) profoundly affect the resultant classification.

There are several different quantitative methods to classify objects, each of which may result in different classification. Furthermore, each classification method requires subjective decisions on the similarity measure to use in the classification and on the number of classes to identify. Thus, classification remains subjective, even when done with seemingly quantitative algorithms. Classifications developed from biological data should make sense in the physical and chemical context of the habitats. *A posteriori* classification is developed from the biological data set. Species abundance data are examined, and groups of sites are identified that are similar to each other. Usually, this is done with a similarity (or dissimilarity) measure and a form of cluster analysis. Subjective decisions are required to select the classification methodology, the similarity measure, and the number of groups to identify.

As was stated above, the general objective of classification is parsimony of classes (few classes) to obtain a large partitioning of variance among the classes. Too few classes results in large variability within each class, and too many classes results in trivial differences among classes.

Assigning Test Sites to Classes

After reference site classes have been determined, using cluster analysis or some other *a posteriori* classification a model is developed to enable test sites to be assigned to one of the reference classes. This is typically a discriminant model developed from non-biological data of the reference sites. Data for the discriminant model should be measurements that are not affected by anthropogenic degradation, such as stream gradient, sinuosity, natural water chemistry, lake depth, watershed soil type, etc. (Norris 1995). The output of a discriminant model is a discriminant function that assigns sites to one of the classes. It is developed from reference site data, and should be tested with an independent reference site data set.

Multimetric Index Method

The indices currently used are variations of the Index of Biotic Integrity (IBI) for fish assemblages in streams, developed by Karr and his coworkers (e.g., Karr 1981, Karr et al. 1986). The concept was extended to benthic invertebrate assemblages (Ohio EPA 1987, USEPA 1989b, Barbour et al. 1992, Kerans and Karr 1994).

Each index is the sum of several (up to 12) standardized component metric scores. Metric scores are usually on an ordinal scale of 1 to 5 (Karr et al. 1986), or 0 to 6 (USEPA 1989b) or as a percentage of the reference metric value (Maxted et al. 1994). Component metrics consist of measures such as total number of taxa, percent abundance of the dominant taxon, number of species and percent abundance of intolerant groups, and percent abundance of functional feeding groups such as planktivorous fish or invertebrate shredders.

Metric Variability

Metrics that are too highly variable within the reference sites are unlikely to be effective for assessment. Relative variability is often measured with the coefficient of variability, defined as the standard deviation divided by the mean (expressed as percent):

$$CV = \frac{s}{\bar{x}} \times 100$$

The CV is a measure of how large the variability is compared to the mean. Ideally the CV should be small, which can be achieved with a small variance or with a large mean value. However, some metrics might have low values under reference conditions (e.g., number of exotic species), and CV will always be large for such metrics. For example, if a sample of 10 reference sites, each with 10 taxa, includes a single site with a single exotic species, then the CV of the number of exotic species is over 300 percent. Furthermore, the multimetric approach calls for comparison of metric values to a percentile of the reference population values and is thus a distribution-free approach. Because the CV is the ratio of the sample standard deviation to the mean, it might not adequately express variability for non-normal distributions.

An alternative measure to the CV is the "interquartile coefficient," which is based on quartiles of the reference distribution and the expected change of the metric rather than its parameters (Gerritsen and Bowman 1994). In operational bioassessment, metric values below the lower quartile of reference conditions are typically judged as not meeting reference expectations (e.g., Ohio EPA 1990). The range from 0 to the lower quartile can be termed a "scope for detection." For those metrics with low values under reference conditions and high values under impaired conditions, the scope for detection is the range from the 75th percentile to the maximum possible value (e.g., 100 percent) (Figure E- 8).

The larger the scope for detection, compared to the interquartile range, the easier it will be to detect deviation from the reference condition. The "interquartile coefficient" is thus defined here as the ratio of the interquartile range to the scope for detection:

$$C_{IQ} = \frac{IQ}{D_s}$$

where IQ = interquartile range

$$D_s = \begin{cases} \text{25th percentile (for metrics that decrease with impairment); or} \\ \text{maximum possible value - 75th percentile (for metrics that increase with impairment} \end{cases}$$

The interquartile coefficient is analogous to the CV and is used similarly, but it is bidirectional and is calculated from percentiles in the same way that assessment uses percentiles. In general, an interquartile coefficient greater than 1 indicates excessive variability of a metric.

Metric Response

Response of metrics to stresses is evaluated by comparison of reference sites to test sites. The simplest comparison is using box-and-whisker plots of the metric distribution in reference and test sites (Figure E-8) or by univariate tests of metrics in reference and test sites. Alternatively, it may be possible to develop an empirical model of metric response to stressors.

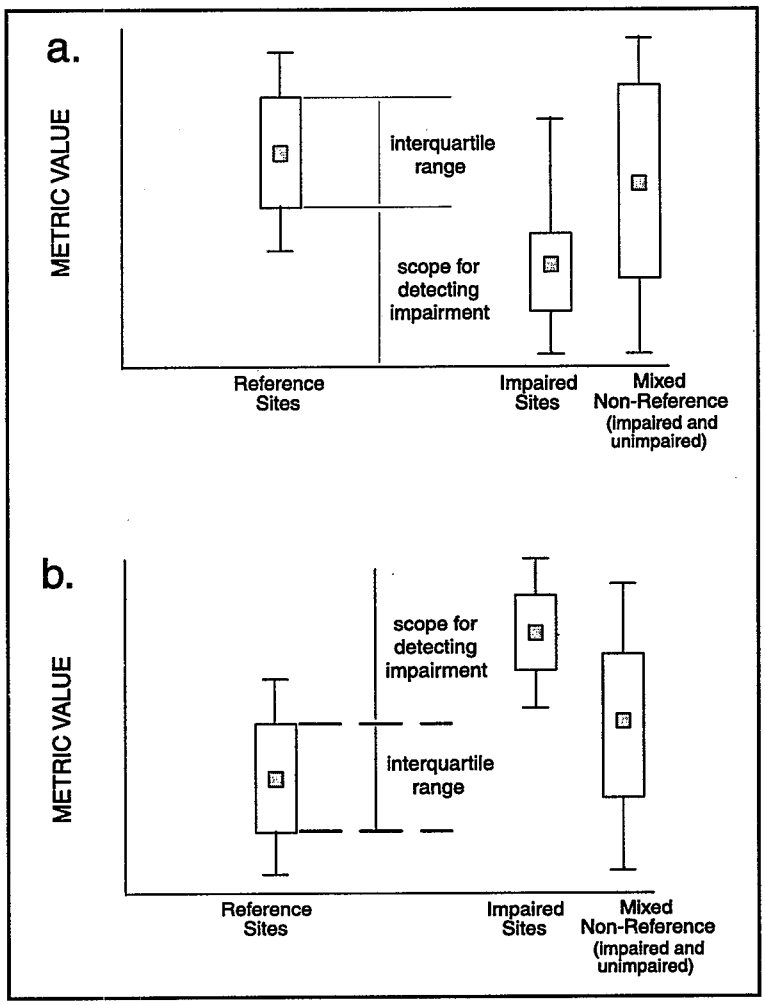


Figure E- 8. Assessing candidate metrics that have (a) high values under reference conditions, and (b) low values under reference conditions.

Several approaches are available including multiple regression, canonical correlation, canonical correspondence analysis, and log-linear models (Ludwig and Reynolds 1988, Jongman et al. 1987).

Metrics are judged responsive if there are significant differences in central tendency or in variance between reference and test sites (Figure E- 8). If the test sites are known to be affected by anthropogenic pollution or disturbance, then mean or median values of responsive metrics should be substantially different between reference and test sites (Figure E-8). If the test sites simply do not meet reference criteria (i.e., they might be a mix of impaired and unimpaired sites, or sites with different stressors), then the variance in the test sites should be larger than that in the reference sites (Figure E-8). If possible, it is advisable to separate test sites according to the stressors or types of impairment (e.g., habitat degradation, toxic substances, organic enrichment) so that response to each stressor can be determined.

When selecting metrics, it is important to visually examine the distribution of metrics in reference sites and in impacted sites. Metrics are selected for inclusion based on their responsiveness, typically by visual examination of box and whisker plots (e.g., Fig. E-8) or scatterplots (Barbour et al. 1996a, Fore et al. 1996). If there is no overlap of the data points, or if the overlap is restricted only to the whiskers of the box plots, then the metric responds strongly to the impairment. A strong response here implies that at least 75% of affected sites have no overlap with at least 50% of the reference sites. A minimum response strength might be defined as no overlap of the median of one site type with the quartile of the other; implying that at least 50% of affected sites are below the 25th percentile of reference sites.

Many biologists may be tempted to use statistical significance tests to select metrics, but slavish reliance on significance tests does not contribute to biological understanding (Yoccoz 1991) and may weaken a multimetric index. If sample size is small (say, n = 6 in both reference and impact sites), then significance tests (at $\alpha = 0.05$) will have low power and responsive metrics may be rejected. On the other hand, if sample size is large (say, n = 30 in both site categories), then it would be possible to detect a statistically significant difference that is biologically meaningless. In this case, metrics that do not contribute to meaningful assessment could be selected, simply because statistical significance was detected. A better measure is the expected frequency with which a metric will fall below a threshold to register impairment. Frequency

can be estimated with a box and whisker plot, but not with a significance test. For example, if the median of impaired sites is below the quartile of reference sites (Figure E-4), then we estimate that impaired test sites will be below the reference quartile in at least 50% of all observations.

Metrics that are responsive to known or unknown stresses are retained for index development. Finally, responsive metrics are evaluated for redundancy, where redundancy means a tight correlation (r>0.9) and a linear relationship. A metric that is linearly correlated with another might not contribute new information to the assessment. Pairs of metrics with correlation coefficients greater than 0.9 should be examined carefully to determine whether they are linear and if both metrics are necessary. Often, strongly correlated metrics are calculated from the same raw data, or their method of calculation ensures correlation. For example, Shannon-Wiener diversity and percent abundance of the dominant taxon are linearly correlated in any data set. A scatterplot of the strongly (>0.9) correlated metrics should be examined; if there is an apparent nonlinear or curved relationship, then both should be retained. If all the points fall very close to a straight line, then one of the metrics can be safely eliminated.

Multimetric Index Development

Multimetric indices are typically developed by summing the metrics that proved responsive to disturbance. The first step is to standardize the different numerical scales of metrics (e.g., number of taxa; % of individuals that are predators) into unitless scores (e.g., Karr et al. 1986, Gerritsen 1995). The scores may be ordinal, or they may be a percentage of a reference value. Ordinal scores are more commonly used, and correspond to categories such as "impaired" and "unimpaired." The index is the sum (or mean) of the metric scores, and is likewise compared to index values at reference sites. Index values at reference sites are then used to establish biocriteria. Socio-political decisions must then determine the numerical values of biocriteria corresponding to aquatic life use categories.

Metric Scoring

Several methods may be used for scoring metrics, all of which are based on the metric distribution in reference sites. Metrics may be given ordinal scores (most often 1, 3, or 5); corresponding to impaired, intermediate, or unimpaired biota, respectively, or may be given a score which is the metric's percentage of the reference value (Figure E-9).

All of these require comparison to some measure of the reference value distribution: an upper percentile, a lower percentile, or a central tendency (Figure E-9). Although a central tendency of the reference sites (e.g., the mean value) may be intuitively attractive as a basis of comparison, there are two important reasons for using percentiles instead:

- An assessment methodology must be able to take into account natural variability of ecological systems. We know that aquatic biota may differ from riffle to riffle in the same unimpacted stream. Central tendency does not take into account the natural variability, and scoring criteria based on central tendency

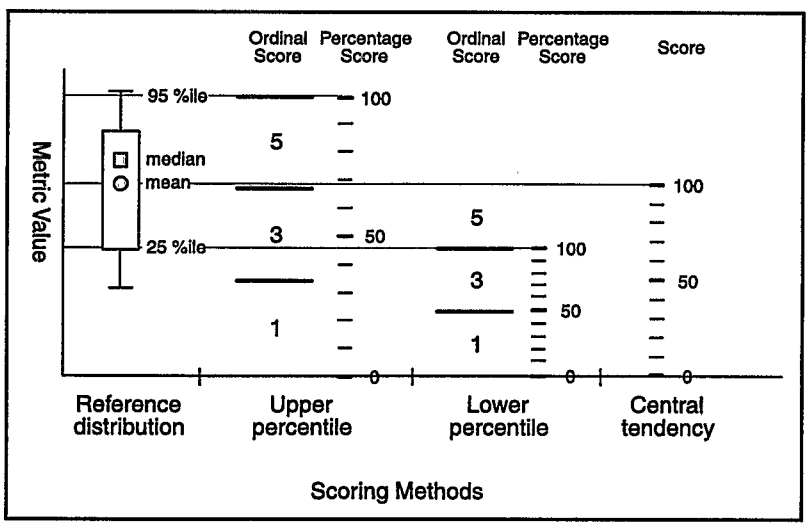


Figure E-9. Illustration of alternative scoring methods, using an upper percentile, a lower percentile, or a central tendency. Most common score breakdowns (5-3-1 ordinal, or percentage) are shown for each, but other ordinal scores have also been used (e.g., 6-4-2-0).

may result in lowered scores for many sites that are within the expected variability of natural, undisturbed sites.

- A second disadvantage of central tendency measures occurs when reference sites, upon which the reference condition is empirically based, are known to be affected to some degree by human activities. Reference sites are often selected to be the least anthropogenically affected in a region, but may still be subject to regional and wide-ranging impacts (e.g., USEPA 1996a). Examples include estuaries and large rivers which receive inputs from their entire watersheds, and small streams and lakes in extensively altered agricultural ecoregions (e.g., the Cornbelt Plains). Use of central tendency then reflects the general (and unquantifiable) degradation of the region and will not result in reference conditions that represent the biological potential.

Two approaches are used to develop metric expectations and scoring criteria (Simon and Lyons 1995). The first approach uses defined reference sites that meet criteria for representative reference sites. Data from the reference sites are used to define expectations and develop metric scoring criteria (Simon and Lyons 1995). The principal scoring criterion (between meeting and not meeting reference expectations) is typically based on a lower percentile of the reference distribution; for example, the 25th percentile (Ohio EPA 1990, Barbour et al. 1996a, Barbour et al. 1996b). In this method, values above the 25th percentile are considered unimpaired (similar to reference conditions) and values below the 25th percentile are considered impaired to some degree. The range from 0 to the 25th percentile is bisected, with values in the top half receiving a score of 3 and those in the bottom half receiving a score of 1 (FigureE-9). This approach also lends itself to scores using percent of reference value (Figure E-9).

The second approach does not include definition of reference criteria, but uses information from the entire range of sites, from the most to the least affected by anthropogenic pollution and disturbance. A large and representative survey data set is required to develop the reference criteria. Reference expectations and scoring criteria are based on the best values observed for each metric, even if the best values do not occur in the least affected sites (Simon and Lyons 1995). The most common scoring method is trisection (Karr et al. 1986) using the 95th percentile of the metric distribution. Metric values from 0 (or the lowest possible value) to the 95th percentile are trisected; values in the top one-third receive a 5, values in the middle third receive a 3, and values in the bottom third receive a 1 (most impaired).

Choice of scoring method should be based on the approach used for defining reference sites, rather than on the method that will produce the most conservative or most liberal scoring. If reference sites are representative of relatively unimpaired conditions, then the lower percentile cutoff and bisection is preferred. If reference sites are not definable, then scoring criteria based on the "best" values are the only alternative.

To account for covariables such as size, the data are plotted, a locally weighted estimate is made of the appropriate percentile (95th or 25th), and the range below it is trisected or bisected (Figure E-10).

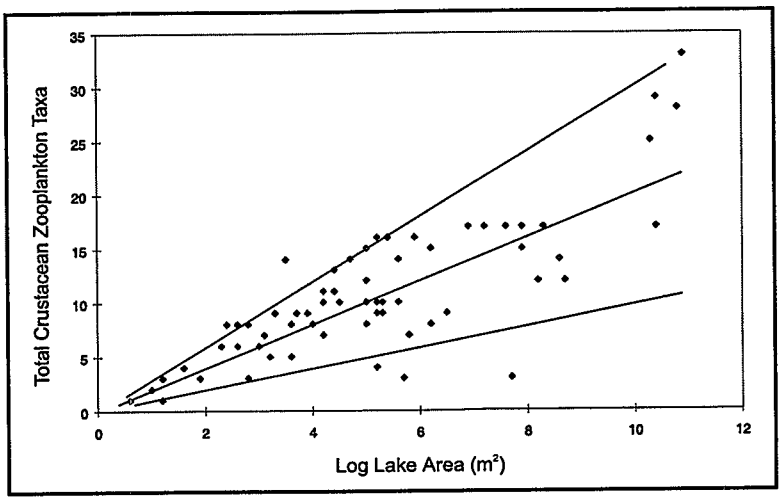


Figure E-10. Total crustacean zooplankton taxa in North American lakes (redrawn after Dodson 1992).

Additive Index

The index is the sum of the scores of the selected metrics that prove responsive to disturbance. Criteria for index values are also generated from the reference sites, just as with

individual metrics. A perfect index score is unlikely in the reference sites, therefore, a reference expectation is developed for the total score. Because the index is the sum of several metrics, the Central Limit Theorem predicts that it will have a lower coefficient of variation than individual metrics, and can be approximated better by a normal distribution than can individual metrics (Fore et al. 1994). Because of these properties, multimetric indices can usually distinguish 3 to 5 statistically significant gradations of impairment, based on comparison of a single sample to the reference distribution (Fore et al. 1994, Gerritsen 1995).

Discriminant Model Index

Discriminant analysis may be used to develop a model that will divide, or discriminate, observations among two or more predetermined classes. Output of discriminant analysis is a function that is a linear combination of the input variables, and that obtains the maximum separation (discrimination) among the defined classes. The model may then be used to determine class membership of new observations. Thus, given a set of unaffected reference sites, and a set of degraded sites (due to toxicity, low DO, or habitat degradation), a discriminant function model can identify variables that will discriminate reference from degraded sites.

Developing biocriteria with a discriminant model requires a training data set to develop the discriminant model, and a confirmation data set to test the model. The training and confirmation data may be from the same biosurvey, randomly divided into two, or they may be two consecutive years of survey data, etc. All sites in each data set are identified by degradation class (e.g., reference vs impaired) or by designated aquatic life use class. To avoid circularity, identification of reference and impaired, or of designated use classes, should be made from non-biological information such as riparian zone modification; known discharges, known contamination, toxicity, nonpoint sources, impervious surface in the watershed, land use practices, etc.

One or more discriminant function models are developed from the training set, to predict class membership from biological data. After development, the model is applied to the confirmation data set to determine its performance: The test determines how well the model can assign sites to classes, using independent data that were not used to develop the model. More information on discriminant analysis is in any textbook on multivariate statistics (e.g., Ludwig and Reynolds 1988, Jongman et al. 1987, Johnson and Wichern 1992).

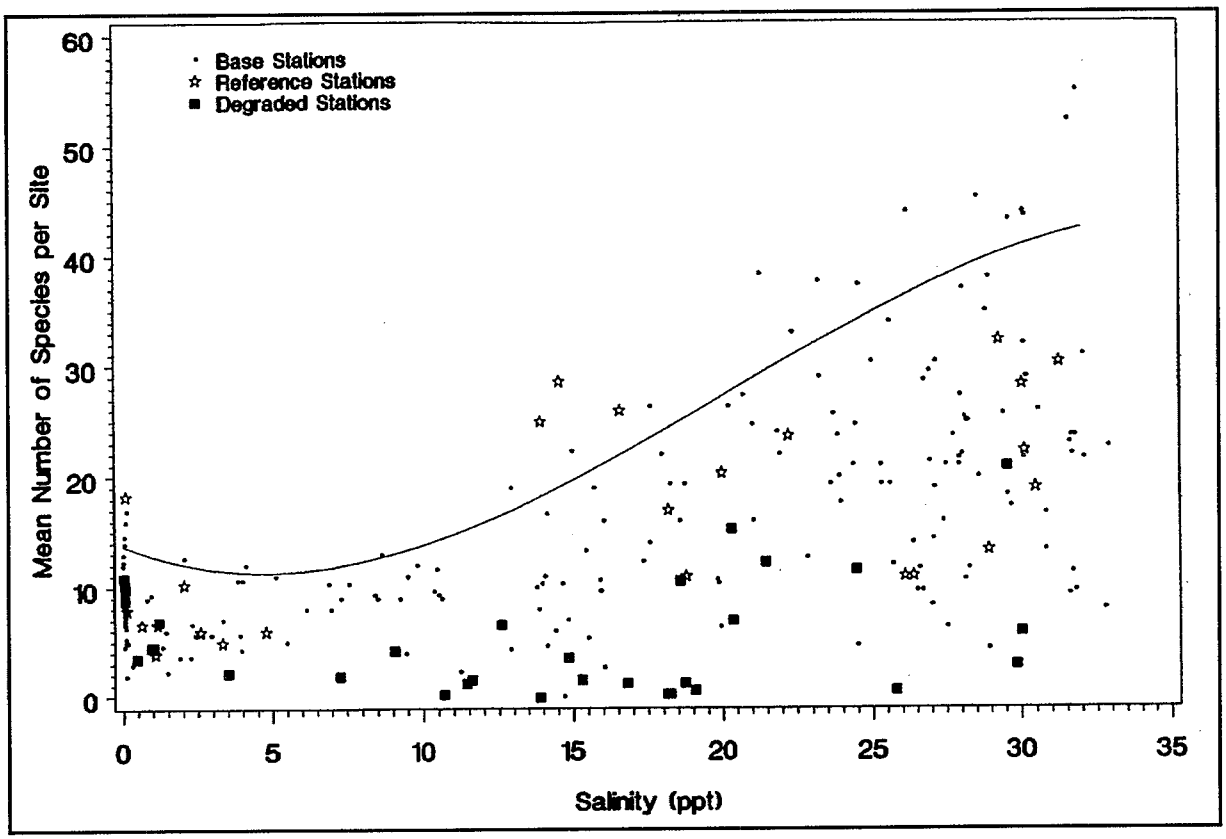
Classification of Estuaries with a Discriminant Model Index

A straightforward *a priori* classification for estuaries was that used by EMAP-Estuaries: first, regionalization into the biogeographic provinces used by NOAA, the U.S. Fish and Wildlife Service and USEPA; second, stratification of estuaries by physical characteristics of shape and size; and third, measurement of physical covariates that affect assemblage composition, principally salinity, depth, and sediment attributes (for benthos) (USEPA 1993e).

Estuaries were classified as large estuaries, large tidal rivers, and small estuaries. Data collected in 1990 showed that large estuaries had the greatest number of taxa, and tidal rivers the fewest taxa. On that basis, the three estuary classes were retained for further analysis.

It has long been known that estuarine faunal diversity is highest at the seaward end of estuaries, in full-salinity seawater. The lowest number of taxa are found in brackish waters that are too saline for freshwater organisms, and too fresh for marine-adapted organisms. To characterize reference conditions in EMAP estuaries, it was therefore necessary to predict the number of taxa that could be found at any given salinity. Figure E-11 shows effect of the covariate, salinity, on the number of taxa captured in benthic grabs. In the example shown, the reference expectation for EMAP estuaries in the Virginian Province was a third order polynomial regression of a running average 90th percentile of the data shown, given by the line in the figure.

Of five possible covariates considered in EMAP-estuaries, only salinity was deemed to have a strong enough effect on benthic macroinvertebrates to justify adjusting reference expectations for it (USEPA 1993e, USEPA 1994h). The observed number of taxa per site was corrected by dividing by the expected number of taxa for the site, obtained from the regression model, to yield percent of expected number of taxa (USEPA 1993e, Engle et al. 1994, USEPA 1994h).



Index

From the set of sampled sites in the EMAP data set, reference and degraded sites were identified based on predetermined criteria. Criteria for reference sites were:

- Summertime bottom DO never less than 1 ppm.
- No contaminants observed in the sediment (1980).
- No toxicity observed in the sediment.

Reference sites had to meet all three criteria. Reference site salinities ranged from < 5 ppt to > 18 ppt.

Sites were rated as degraded if they met either of the two following criteria:

- One or more hypoxic events with bottom DO < 0.3 ppm.
- The concentration of at least one sediment contaminant exceeding the ER-M value, and *Ampelisca* bioassay indicated toxicity (< 75% survival and significantly different from control) (USEPA 1993e).

Stepwise discriminant analysis was used to determine which metrics could best discriminate between reference and degraded sites. Because number of taxa was deemed an important indicator by itself, it was "forced" into the discriminant model. The eventual discriminant model had five variables:

- % expected number of taxa.
- No. of amphipods.
- % of abundance as bivalves.
- Mean weight per polychaete.
- No. of capitellid polychaetes.

The model correctly classified 89% of degraded sites, and 86% of reference sites, using the learning data set to test its performance. Discriminant scores were normalized to a range of 0 to 10, for ease in communication of index scores (USEPA 1993e).

The original discriminant model was developed from the 1990 EMAP-Virginian Province sampling effort. The model was subsequently tested with the EMAP-Virginian Province data set collected in 1991, an independent test (USEPA 1994h). The 1990 model failed to discriminate the 1991 data correctly, so a new discriminant model was developed using both 1990 and 1991 data sets. The revised model used the following 3 variables:

- Mean abundance of opportunistic species.
- Biomass/abundance ratio for all species.
- Mean number of infaunal species per grab.

The revised model was subsequently tested with another independent data set, the 1992 Virginian Province data. The revised model correctly discriminated the 1992 reference and degraded sites, and the model was not further modified (USEPA 1994i). The model correctly identified 83% of reference sites and 100% of degraded sites.

Designated Aquatic Life Use Classes

An alternative to the above methodology is to develop biocriteria directly for administrative aquatic life use classes (Davies et al. 1993). In this approach, data from a set of sites (the training set) are assigned to predetermined aquatic life use classes. The classes are determined by regulation and might be (for example): (a) pristine; (b) altered habitats, but native species maintained; (c) discharges and vegetation permitted, native communities altered, but fishable-swimmable goals met; or (d) nonattainment. Experts assign sites to one of the four classes based on the narrative descriptions of the aquatic life use classes (above) and biological data from the training set sites (Davies et al. 1993).

One or more discriminant models to predict class membership are developed from the training set. The purpose of the discriminant analysis here is not to test the classification (the classification is administrative rather than scientific), but to assign test sites to one of the classes.

An example of this approach is the biocriteria adopted by Maine for streams (Davies et al. 1993). Stream biologists assigned a training set of streams to four life use classes. A two-stage discriminant modeling process was used to develop discriminant models for assigning test streams to use classes. The first stage was a model to predict membership in each of the four classes, expressed as a probability for each. The second stage was a set of three discriminant models that predict two-way class membership (i.e., nonattainment (NA) versus A or B or C; NA or C versus A or B; and NA or C or B versus A). A selection procedure was used to select predictive variables for the models, and the second-stage models were constrained to exclude predictive variables used in the first-stage model. This approach is detailed by Davies et al. (1993).

Multivariate Ordination Model

The third approach that has been successfully used for development of biocriteria uses multivariate ordination to determine if test sites are different from reference sites. The comparisons are usually made graphically, in ordination space (c.f., Figures E-1, E-3, and E-9); such that if a site is outside of the area on an ordination diagram defined by reference sites, it is judged to be degraded.

Classification of reference sites is often *a posteriori*, using one of several clustering methods on the biological data. Following definition of biological clusters (reference classes), a discriminant model is developed using physical-chemical data to allow classification of test sites (Moss et al. 1987, Wright et al. 1984, Reynoldson et al. 1995, Norris 1995).

Cluster analysis must be done with great caution because there are many similarity measures and many clustering algorithms, many of which may produce different, and often unintelligible, results (Ludwig and Reynolds 1988, Jackson 1993a). In general, the best results for bioassessment purposes have been achieved with UPGMA, and with TWINSPAN, a divisive technique (Reynoldson et al. 1995, Moss et al. 1987, Gauch 1982). The most successful similarity measures have been Bray-Curtis similarity, chord distance, and relative Euclidean distance (Kenkel and Orloci 1986, Ludwig and Reynolds 1988).

Ordination analysis is often done after determination of clusters, to see whether the identified clusters also separate in ordination space. The clusters are now treated the same as an *a priori* classification. Ordination methods most often used at this stage include correspondence analysis and non-metric multidimensional scaling (e.g., Moss et al. 1987, Reynoldson et al. 1995).

Test sites are then assigned to one of the reference classes with the discriminant model (using physical-chemical data), and test sites are compared to their respective reference population in ordination space.

The ordination approach is illustrated with the classification of benthic macroinvertebrate assemblages from Great Lakes reference sites (Reynoldson et al. 1995). Cluster analysis of benthic macroinvertebrates from 96 reference sites revealed five groups of sites. Cluster analysis used the Bray-Curtis distance measure, and the clustering algorithm was unweighted pair group mean averages (UPGMA) (Reynoldson et al. 1995). The sites were subsequently visually depicted with ordination by NMDS, which showed each cluster occupying a unique area in ordination space.

Following classification, a discriminant model was developed to identify class membership of sites using physical and chemical (i.e., non-biological) data. The model was developed with the reference sites as the calibration data set, for subsequent use with test sites to identify the class of benthic assemblage the test site should belong to. Sites were selected for uniform characteristics (< 2 km from shore, < 30 m depth, fine-grained sediment), and the criterion for reference sites was large distance (> 10 km) from known discharges (Reynoldson et al. 1995). First, explanatory variables for input into the discriminant function analysis were identified with correlation analysis of physical-chemical variables. Variables that were significantly correlated with any of the three ordination axes were used for the discriminant analysis. Of 25 variables examined, 18 were strongly correlated with the ordination and were input into the stepwise discriminant analysis. Of these, nine produced the best model to predict class membership. Biological assemblage group membership was correctly predicted by the discriminant model for 87% of the sites, ranging from 64%-100% for each of the five assemblage groups (Reynoldson et al. 1995).

Biological integrity of test sites was assessed by first assigning a test site to one of the five assemblage groups, based on the discriminant model applied to the site's physical-chemical data. The biological assemblage structure of the test site was then compared to assemblage structure of the reference sites of that group, by plotting the positions of reference sites and the test site in ordination space, and determining if the test site was within the region defined by the reference sites (e.g., Figure E-12). The approach was used for an assessment of benthic sites in Collingwood Harbour, Ontario, which had been contaminated with metals. Benthic macroinvertebrate assemblages were different from reference sites within two boat slips, and the authors concluded that sediment remediation was justified in the boat slips. Outer reaches of the harbor exceeded Ontario sediment metals criteria but benthic assemblages in the outer harbor were similar to reference sites of their respective classes. Because there were no discernible biological differences, the authors concluded that sediment remediation could not be justified in the outer harbor (Reynoldson et al. 1995).

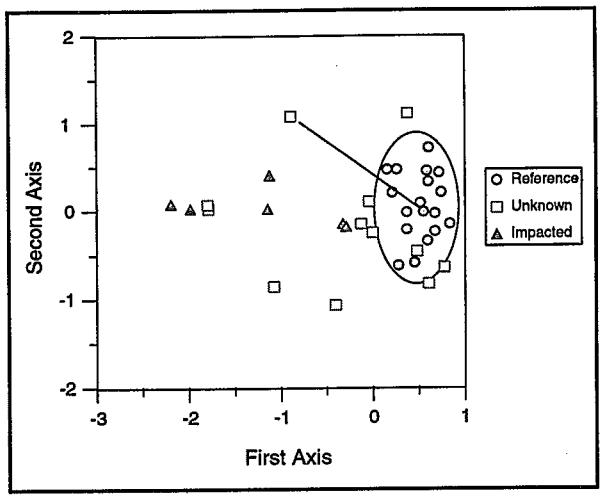


Figure E-12. Assessment by ordination. Solid circles are reference sites, known impacted sites (triangles) deviate from the reference group, primarily on the first axis. Impairment may be judged by whether a site is outside the region bounding reference sites (ellipse), or by the distance between a site and the reference centroid (arrow).

Appendix F

Executive Summaries of State Pilot Studies

The following includes executive summaries from three states (Maine, Vermont and Wisconsin) that performed and completed pilot studies based upon the Lake and Reservoir Bioassessment and Biocriteria Technical Guidance Document. These pilot studies are distinct from the case studies incorporated throughout this document and, thus, are presented separately. The purpose of each pilot study was not to evaluate individual state sampling field methods but rather to test the utility of the data analysis and biocriteria development guidance in this manual. Each state adapted its study to incorporate its available data. All three states reported favorably on the usefulness of this guidance manual and indicated the various stages of their own biocriteria development. For more information regarding these studies, please contact the state's project leader as listed in each executive summary.

Geographic Analysis and Categorization of Maine Lakes: A Trial of the Draft Lake Bioassessment and Biocriteria Technical Guidance

Maine's test of the Draft Lake Bioassessment and Biocriteria Technical Guidance focused on lake classification, selection of reference sites and metric evaluation from extant data. Dataset limitations prevented the creation of a lake biocriteria program for the entire state; instead, various aspects of the guidance were 'spot tested.'

Classification of Lakes

Although Maine lakes are quite diverse, it was decided that delineating 10 or fewer lake classes would be most practical. The classification effort focused on 451 lakes which had complete datasets, utilizing cluster analysis on a combination of morphometric and chemical variables including surface area, flushing rate, maximum depth, mean depth, drainage area, elevation, color, alkalinity and specific conductance. The ecoregion approach suggested in the guidance resulted in 3 'modified' ecoregions (based on Omernik 1987) each having two lake classes. Clustering was primarily based on surface area and depth variables.

In this portion of the test, the biggest constraints were related to the difficulties of doing statistical analysis with available software. Three other concerns were evident: (1) how to handle outlier lakes (i.e., very large lakes), (2)

how to incorporate additional lakes as data becomes available, and (3) the selection bias represented by monitored lakes in the datasets.

Selection of Reference Lakes
Each of the lakes in the dataset had a development ranking assigned from 1990 Census Data. Rankings from a subset of these lakes were found to be similar to rankings derived from the examination of United States Geological Service (USGS) topographic maps. Lakes having low development rankings were screened by professionals to eliminate lakes impacted by activities unrelated to population. Box and whisker plots were used to compare trophic status of reference lakes to non-reference lakes in each lake class. These comparisons often showed little separation between reference and non-reference lakes. This may be partially explained by the disproportionate number of reference and non-reference lakes. The technique used to choose reference lakes may need refinement but appears to be compatible with the guidance emphasis on using readily available information.

Biological Parameters
Maine examined metrics from Tier 2B level biological data (phytoplankton and zooplankton) to evaluate their potential utility. Some metrics were suggested in the guidance, others had basis in current scientific literature and a few were suggested by the nature of the dataset. Forty-six phytoplankton metrics were examined; the thirteen metrics showing potential utility were reduced to four after the elimination of redundant metrics: total cell volume, % volume *Cyanophyta*, % volume chrysophytes and the ratio of volume of *Cyanophyta* to desmids. Seven out of nineteen zooplankton metrics showed potential utility in screening for trophic increases. Metrics were eliminated with inherent redundancy and scoring was developed for two: total abundance and the ratio of cladocerans to copepods. Cumulative distribution plots for reference sites and non-reference sites were utilized to determine scoring levels for the two metrics. A multimetric index was not developed due to the low number of lakes and the overlap of the Tier 2B biological data.

The guidance provided a reasonable framework for the development of the biological metrics. However, the literature suggests that rotifers respond to trophic changes as well as crustacean zooplankton, a point previously overlooked in the guidance. Potential phytoplankton metrics should not only be based on count data, but also total cross-sectional area and/or volume measures. The variation in phytoplankton cell size is so great that cell numbers do not approximate standing crop as well as volume or area. Although a multimetric index was not developed, it appears that reasonable guidance is provided in the draft to accomplish this.

General Comments and Applicability within Maine's Lake Management Strategy
There are some concerns that this experience will be extrapolated to the rest of the country. For example, Maine has a large number of lakes and most are of glacial origin. Maine has only 3 lakes receiving point source discharges; Maine's lake management focus continues to be on trophic status and NPS pollution control. The test of this document was from this perspective, and other states may have other or additional management priorities. Another concern is whether or not the gain is worth the additional cost of obtaining biological (phytoplankton and zooplankton) data. One aspect that may be beyond the scope of the guidance document is the lack of reference lakes for atmospheric deposition impacts (in particular, Hg accumulation). If one assumes that atmospheric deposition is somewhat uniform over regions, it becomes nearly impossible to select unimpacted reference lakes and strong reliance must be placed on the term "minimally impaired" as used in the guidance.

Overall, this test has been considered a success despite some limitations of the dataset, and Maine has shifted from 'test' mode to 'development' mode. Biological samples have been collected from 100 potential reference lakes (1996) which are currently being analyzed and similar samples from 100 lakes of unknown status (1997) which will be analyzed and used as a test dataset in the future. The development of biocriteria for Maine's lakes will be an ongoing process over the next few years as time and funding permits. It is anticipated that the results will be useful to concerned citizens at the local level as well as biologists at the state level.

Project Contacts:
Linda Bacon and Roy Bouchard
Maine Department of Environmental Protection
Division of Environmental Assessment
Bureau of Land and Water Quality
phone (207) 287-3901
fax (207) 287-7191

Biocriteria Development for Vermont Lakes-Pilot and Field Phases

Project Summary, April, 1998

The development of test biocriteria by the Vermont Department of Environmental Conservation (VTDEC) was conducted in order to evaluate methods presented in this Lake and Reservoir Bioassessment and Biocriteria Technical Guidance Document. Methods for conducting each phase of the project were taken directly from the document, and a comprehensive Pilot Phase final report is available. The Pilot Phase of this project was completed during 1995, using existing information contained in the Vermont Lakes and Ponds Database. Following this effort, a field program was initiated to develop reference level biocriteria beginning in 1996. Results and lessons learned from the Pilot Phase and results from the Field Phase (1996-1997) are presented in this summary. Movement towards implementation of fully developed lake biocriteria for Vermont is discussed.

Field Phase

Taking the lessons learned from the Pilot Phase, a comprehensive bioassessment project was designed using this Guidance. To avoid the difficulties of classification, a regional approach to definition of lake biological reference conditions was adopted by planning assessments of both Vermont and New Hampshire lakes. To date, this cooperative Field Phase has evaluated 29 lakes. Ten additional lakes are scheduled for assessment during 1998. Data results from 1996 and 1997 are available, and an overview of analyses conducted with these data to date is presented below. The reader should note that trial criteria presented below are provisional, and should be considered in development. The present study lake set contains 23 candidate reference lakes, and six known impaired or test lakes. The test lakes are either culturally eutrophied, anthropogenically sedimented, or have perpetually anoxic hypolimnia. The geographic distribution of 1996-1997 study lakes is presented in Figure F-1.

Bioassessment Assemblages, Metrics, and Methods

Following recommendations from the Pilot Phase, the trophic state, phytoplankton, benthic macroinvertebrate, and macrophyte assemblages were selected for assessment. Trophic state parameters (Secchi disk transparency, chlorophyll *a*, algal biovolume) were collected bi-weekly at a central location in the lake. Phytoplankton were enumerated from a season-wide, whole lake composite, consisting of composited, bi-weekly, depth-integrated samples of the photic zone acquired from a fixed station network. Discrete bi-weekly composites were retained in archive for future analysis if necessary. Profundal and sublittoral benthic macroinvertebrate samples were collected as triplicate composites using an Ekman dredge, from a fixed station network on each lake. Triplicate composited samples of benthic macroinvertebrates from rocky-cobbled, littoral-mud, and macrophyte bed habitats were collected using a sweep net. A timed collection period of 20 minutes total per composite sample per habitat was employed to ensure quantitative data comparability among lakes. The entire littoral zone was surveyed for macrophytes, whereby species were identified and abundances classified using the Braun-Blanquet

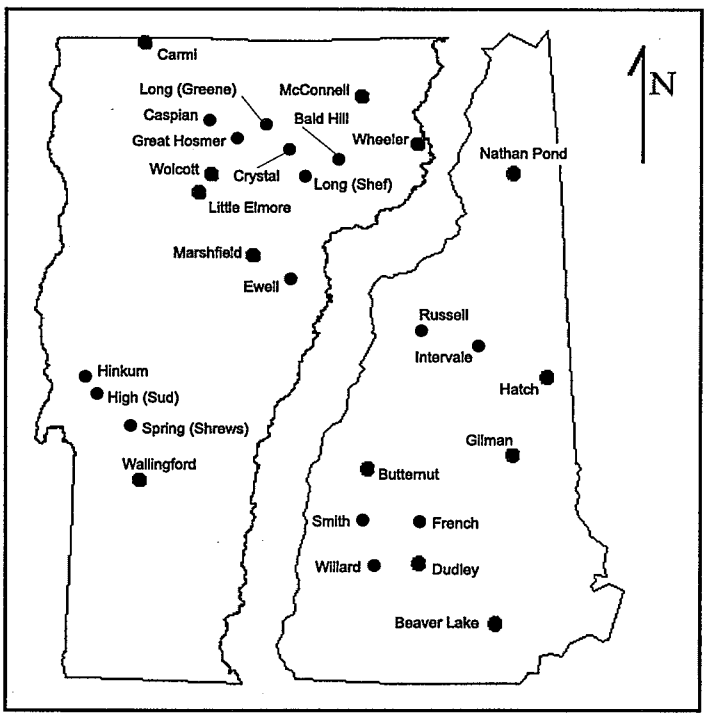


Figure F-1. Location of 1996-1997 study lakes in the Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project

scale. Benthic macroinvertebrates and macrophytes were assessed during the mid-late summer index period (approximately August 1-August 31). Habitat quality was assessed at the time of the macrophyte survey. A quality assurance program was employed to ensure the precision, accuracy, comparability and representativeness of data collected. Table F-1 presents selected metrics under evaluation for the Field Phase of this bioassessment project.

Table F-1. Selected tier two metrics evaluated for 1996-1997 Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project study lakes.

Trophic State and Physico-chemical:
Alkalinity
Conductivity
Dissolved oxygen
Algal biovolume - Sweet TSI
Chlorophyll *a* - Carlson TSI
Secchi transparency - Carlson TSI
Benthic Macroinvertebrate:
No. of taxa
% dominants
Shannon-Weiner index of diversity
% intolerant species
COTE index (Coleoptera, Odonata, Ephemeroptera, Trichoptera)
% intolerant chironomids
No. of Crustacea - Mollusca taxa
Functionality (ie. shredder, scrapers...)
Macrophytes:
% cover - littoral zone
% cover - littoral zone, nuisance species
No. of species
Relative species dominance
No. of rare species
No. of *Potamogeton spp.*
No. of *Utricularia spp.*
% occurrence by structural morphology
Phytoplankton:
Total density
Total biovolume
Shannon-Weiner diversity
% *Anabena spp., Aphanizomenon spp., Anacystis spp.*
% *cyanobacteria* (density and biovolume)
% diatoms (density and biovolume)
% chlorophytes (density and biovolume)
% euglenophytes (density and biovolume)
% phyrrophytes (density and biovolume)
% cryptophytes (density and biovolume)

Preliminary Lake Classification

In selecting candidate reference and test lakes, the same classification metrics were used as for the Pilot Phase. An a-priori classification was adopted using alkalinity as a classifying variable. A cutoff of approximately 15mg/l was used to classify lakes as poorly buffered (15 mg/l as $CaCO_3$), or well-buffered (> 15mg/l as $CaCO_3$). Existing lake assessment data suggests that these two lake classes correspond to tannic and clear water lakes (one exception being the clear, but lower-alkalinity Hatch Pond, NH). Thus for ease of presentation, low-alkalinity, poorly buffered lakes are called tannic, while higher-alkalinity, well buffered lakes are called clear. Table F-2 provides the range of physico-chemical attributes for each of these classes.

This proposed classification was validated with phyto-plankton data using canonical correspondence analysis (Figure F-2). The position of clear and tannic lakes is well separated along the second axis, as are the relative positions of the algal orders. Test lakes with increased blue-green algae in the community separate along the first axis. This ordination suggests that there is variability in the phytoplankton assemblage biometrics which can be explained by the proposed classification.

Criteria Development, Phytoplankton

All phytoplankton data were examined using Tukey box plots to identify metrics which discriminate between reference and test lakes. Metrics thus 'appearing' discriminatory were tested by calculating interquartile coefficients. To avoid 'double counting' of impairments, the selected metrics were examined for covariance. A high degree of covariance was noted between the percent composition of cyanobacteria and percent composition of *Aphanizomenon flos-aquae, Anabaena flos-aquae,* and *Anacystis marina* ($r=0.85$, $p<0.05$). The latter metric was retained as it more accurately defines occurrences of undesirable blue-green algal blooms in test lakes.

A total of five metrics were selected for each lake class to construct a phytoplankton index. Trial criteria were developed by scoring the metric ranges using the 'bi-section' method presented in this Guidance. Trial criteria are presented in Table F-3. Lakes were scored using these criteria, and the distribution of scores is presented in Figure F-3. For tannic lakes, both test lakes met reference criteria. None of the clear test lakes met reference criteria.

Table F-2. Ranges of selected attributes of candidate reference lakes falling into two classes evaluated in conjunction with the Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project.

	Tannic lakes	Clear lakes
Size (ac)	20-96	20-789
Depth (m)	3-17	7-43
Alkalinity	6-14	9.6-100*
Mean Secchi disk transparency	1.7-11.5	2.5-7.8

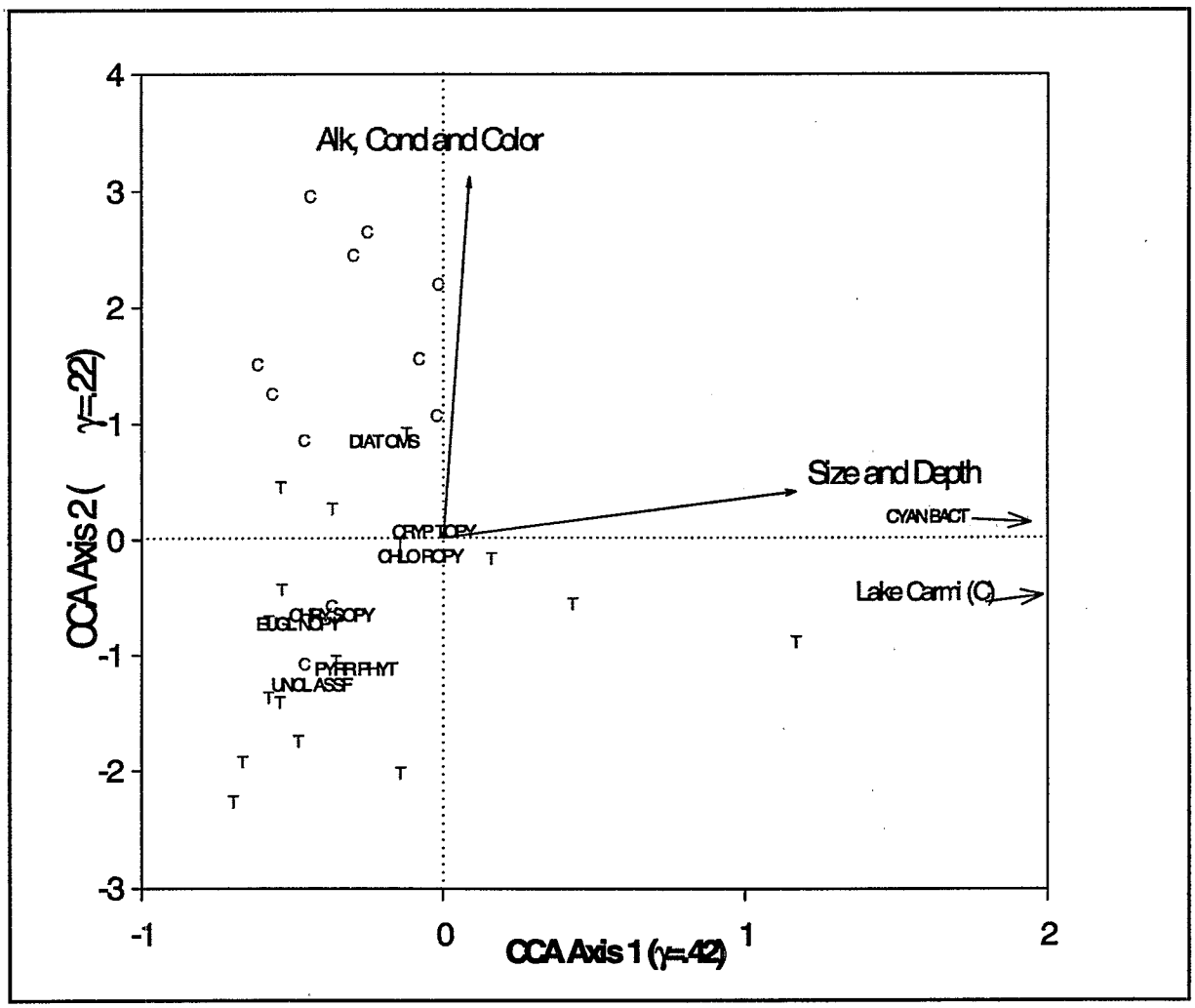


Figure F-2. Canonical correspondence ordination triplot for unclassified reference and test lakes evaluated in conjunction with the Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project. Lakes are denoted as clear (C) or tannic (T). Eigenvalues (Y) are provided for each axis. For simplification, physical variables are grouped and presented by their relative position to the ordination axes. Relative percent composition by algal orders are scaled by a factor of 2 for ease of interpretation.

Table F-3. Trial phytoplankton assemblage biocriteria for Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project study lakes.

Tannic lake metrics	Interquartile coefficient	*Score attributed:* 1	3	5
Total density	0.40	>980	862-980	<862
Total biovolume	0.89	>361K	287K-361K	<287K
% cryptophytes	0.66	>47	28-47	<28
% diatoms-biovolume	0.99	<11	11-13.4	>13.4
*% APHA-ANFA-ANMA**	0.56	>10	1-10	<1
Clear lake metrics		*Score attributed:* 1	3	5
Total density	0.15	>780	620-780	<620
Total biovolume	0.09	>580K	389K-580K	<389K
% cryptophytes	0.76	>9	9-19	<19
% diatoms-biovolume	0.67	<42	42-63.5	>63.5
*% APHA-ANFA-ANMA**	0.06	>5	13-5	<3

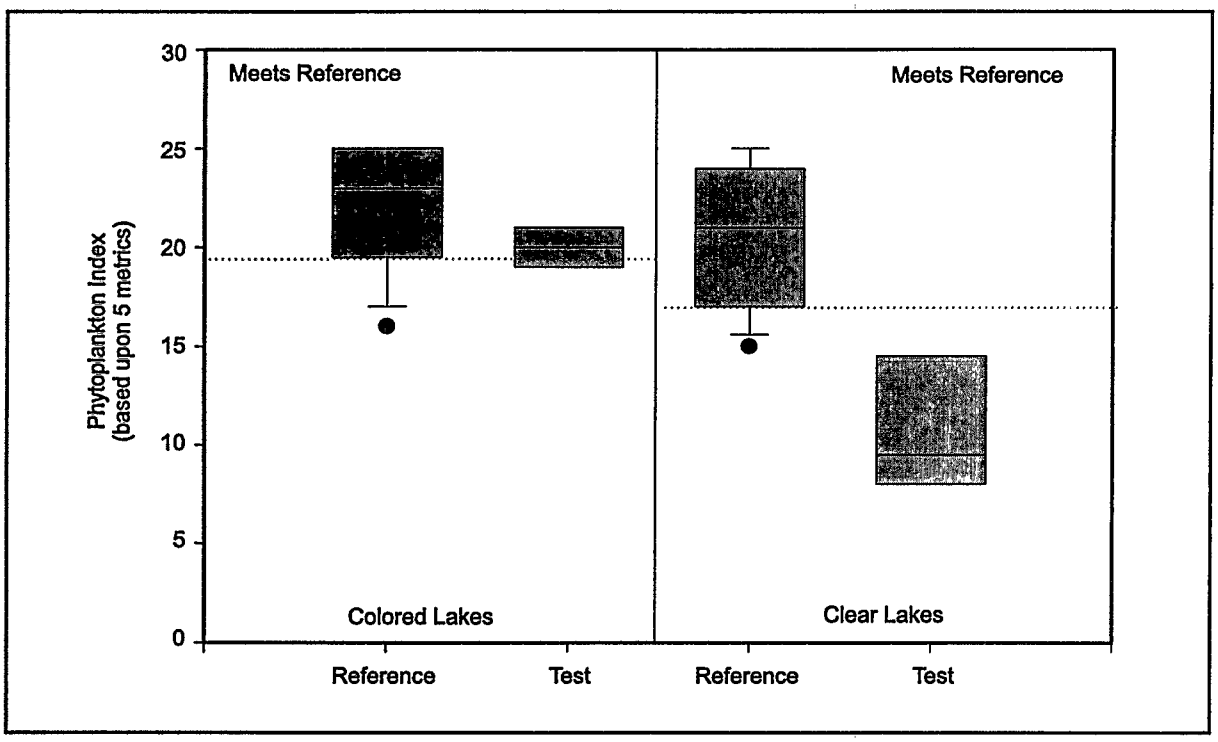


Figure F-3. Distribution of phytoplankton index scores for 29 classified study lakes. The dotted line corresponds to the lower quartile of the reference distribution. Lakes which score above this value are considered meeting reference conditions.

Criteria Development, Littoral Zone Macrophytes

The methods by which macrophyte data were collected do not permit calculation of a Shannon-Weiner index of diversity, yet this is an important measure describing the macrophyte assemblage. To provide an alternate measure, a *relative species dominance* metric is proposed where impairment is indicated by metric values increasing above reference. This metric is calculated as:

$$\frac{\text{\% cover - littoral zone}}{\text{No. of species}}$$

An alternate way of assessing macrophyte communities is to determine the relative contribution by different structural groups. Analogous to relative percent composition by algal divisions in the phytoplankton, or by function in the macroinvertebrates, relative percent occurrence by structural grouping can affect other biological assemblages and vary with impairments to lake water quality. To evaluate this for the study lakes, seven structural groupings were proposed (Table F- 4).

Interquartile coefficients for macrophyte metrics were calculated, and many metrics were found to be insensitive. The most discriminating metrics were nevertheless retained for trial criteria development in the interest of assessing test lakes against a reference condition. The criteria presented in Table F-5 are at best draft, and should be considered in development pending the acquisition of additional data. Reference and test lake scoring is presented in Figure F-4.

Table F-4. Proposed structural macrophyte groupings for use in bioassessment of Vermont and New Hampshire lakes.

Proposed structural group	Representative example species
Emergent erect	*Pontederia spp.*
Emergent pronate	*Sparganium minimum*
Floating leaved	*Brasenia schreberi*
Submerged narrow-leafed	*Najas spp.*
Submerged broad-leafed	*Potamogeton amplifolius*
Submerged whorled	*Ceratophyllum spp.*
Submerged mat-like	*Eriocaulon spp.*

Table F-5. Trial macrophyte assemblage biocriteria for Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project study lakes.

Metrics	Tannic Lakes				Clear Lakes			
	IC	1	3	5	IC	1	3	5
Percent cover-littoral zone	0.55	>26	21-26	<21	1.19	>37	22-37	<22
Relative species dominance	0.28	>1.5	1.2-1.5	<1.2	1.87	>1.4	1-1.4	<1
% occurrence floating leaved	1.60	>25	20.1-25	<20.1	0.37	>18	15-18	<15
% occurrence submerged narrow leaved	1.05	<6	6-10	>10	3.76	>35	32-35	>32
% occurrence submerged whorled	2.06	>12	6-12	<6	0.52	>12	7-12	<7

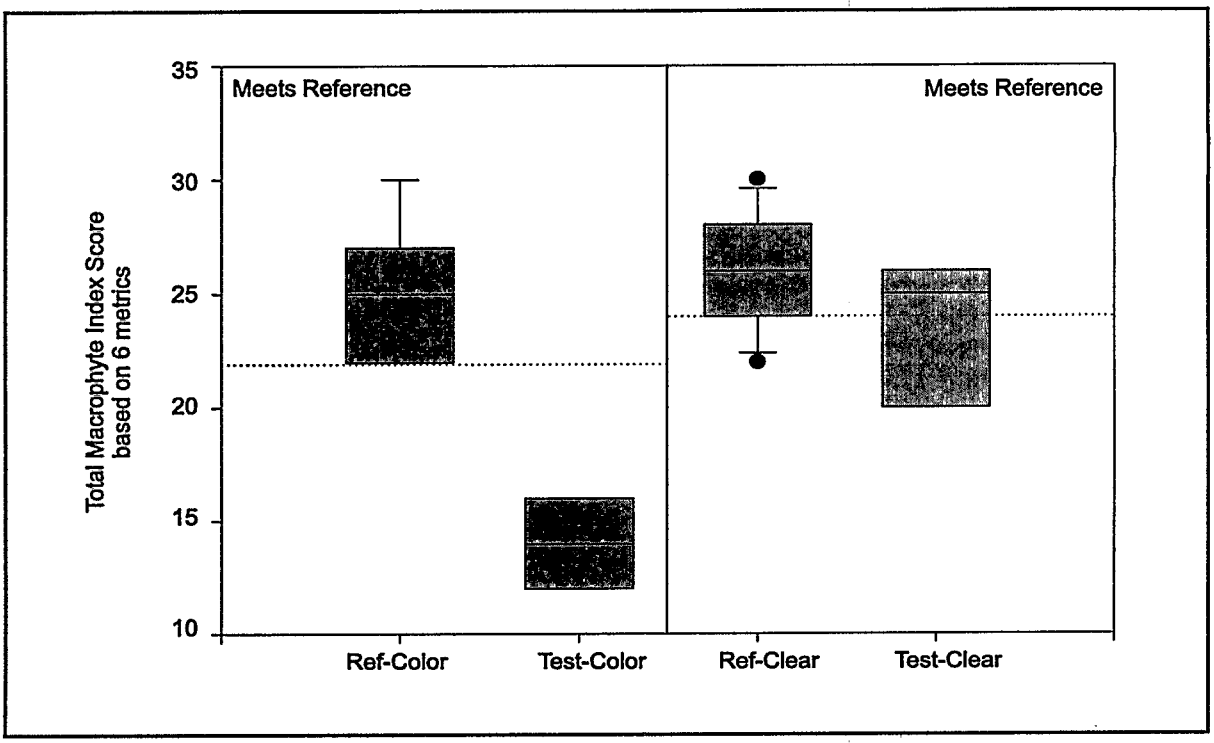


Figure F-4. Distribution of macrophyte index scores for 29 classified lakes. The dotted line corresponds to the lower quartile of the reference distribution. Lakes which score above this value are considered meeting reference conditions.

Criteria Development, Trophic State Indices
Trophic state indices were calculated for Secchi disk transparency and chlorophyll *a*, using Carlson's algorithms, and for total algal biovolume using Sweet's algorithm. This latter index is calculated as:

$$(\text{Log}_2 (B+1)) \times 5$$

where B=0.001 x algal biovolume in um^3/ml.

Trophic state indices are useful in regions where there exists a wide range of trophic conditions. In this study lake set, even lakes which are considered impaired by eutrophication score only at the low range of eutrophy, with no lake exceeding 62 on the unitless trophic state index scale. Therefore, the calculated scoring ranges for criteria developed from this study set are extremely narrow. Interquartile coefficients and draft trophic state criteria are presented in Table F-6.

The trophic state index calculated from algal biovolume is presented in this section to assess whether this metric has greater discrimination than its untransformed analogue, algal biovolume (presented in the phytoplankton section above). Comparison of the interquartile coefficients for algal biovolume, and for the calculated algal biovolume trophic state index suggests that this metric is most discriminating in the phytoplankton assemblage biocriteria, (interquartile coefficient of 0.89 and 0.09 vs. 0.27 and 1.18 for tannic and clear lakes respectively). Accordingly, algal biovolume should best be retained in the phytoplankton assemblage criteria in refinements of these trophic state criteria. The distributions of reference and test lakes, scored by trophic state indices, are shown in Figure F-5.

Results, Benthic Macroinvertebrates
As of this writing, taxonomic data have been validated only for the 1996 study lakes. A wide variety of macroinvertebrate metrics have been calculated from these data, and it appears that the macroinvertebrates could provide highly discriminating metrics (Table F-7). It is anticipated that a lake macroinvertebrate index will have metrics from each of the five habitats

Table F-6. Interquartile coefficients (IC) and trial biocriteria scoring for trophic state indices - Bioassessment and Paleolimnology of Vermont and New Hampshire Lakes Project study lakes.

	Biocriteria scoring			
Tannic lake metrics	IC	1	3	5
TSI Chlorophyll-a	0.35	>52	49-52	<49
TSI algal biovolume	0.27	>52	49-52	<49
TSI Secchi disk transparency	1.82	>52	50-52	<50
Clear lake metrics				
TSI Chlorophyll-a	1.50	>48	44-48	<44
TSI algal biovolume	1.18	<46	42-46	<42
TSI Secchi disk transparency	0.93	>40	39-40	<39

evaluated. Based upon review of the 1996 taxonomy, and observations from the 1997 samples currently in taxonomy, it is believed that profundal zone samples may not provide useful data. Indeed, profundal zone samples are comprised almost entirely of *Chironomidae, Chaoboridae*, and *Oligochaeta*, and variation in overall density is dependent on hypolimnetic oxygen conditions. If these observations hold true, profundal zone macroinvertebrate assessments will likely be dropped from Vermont's bioassessment protocols.

Moving Toward Implementation

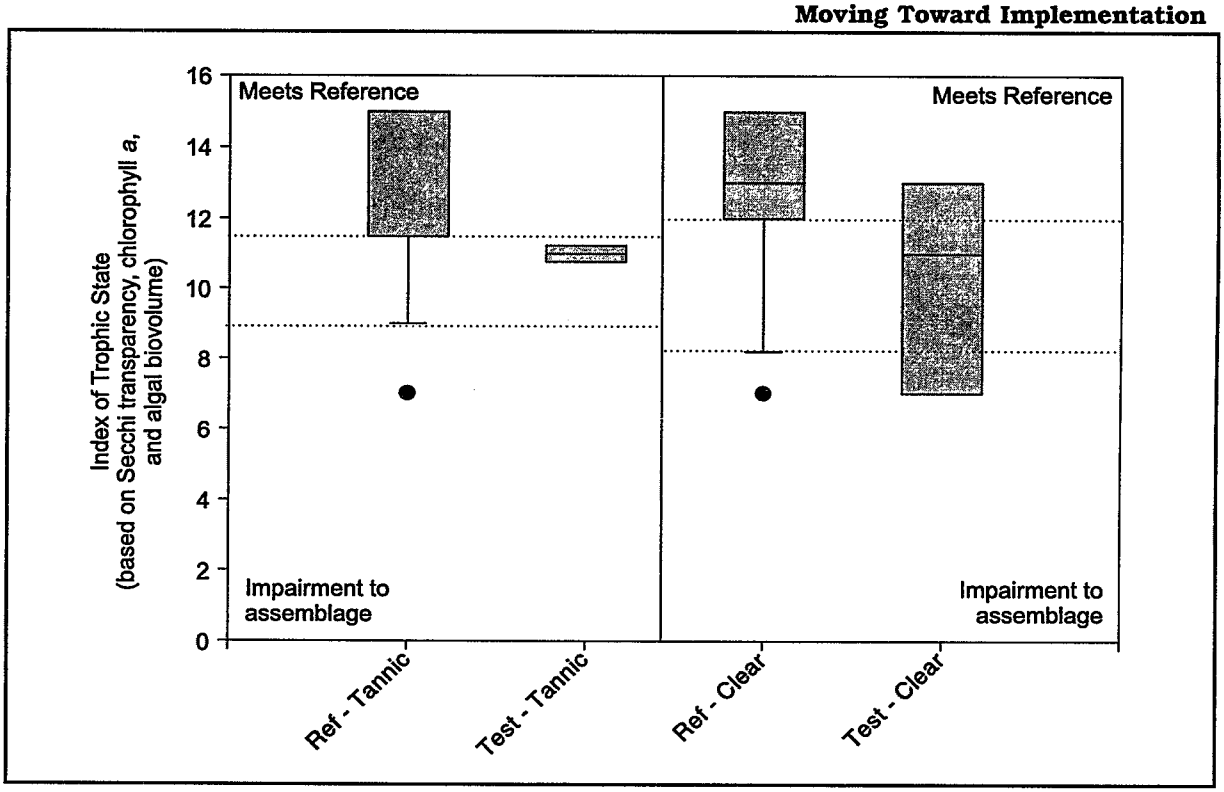


Figure F-5. Distribution of trophic state index scores for 29 classified lakes. The dotted lines correspond to the lower quartile and lower 5th percentile of the reference distribution. Lakes which score above the lower quartile value are considered meeting reference conditions. Lakes which score below the 5th percentile are considered impaired.

Table F-7. Potentially robust macroinvertabrate metrics for five lakes habitats evaluated in conjunction with the Paleolimnology and Bioassessment of Vermont and New Hampshire Lakes Project.

Habitat	Macroinvertebrate metric
Profundal	*VT-BI*; % other; % intolerant chironomids*
Sublittoral	*% collector-filterer, % predator, % shredder-detritivore, % shredder-herbivore*
Littoral rocky-cobbled	*VT-BI, Shannon-Wiener diversity, % collector-filterer, % predator, % shredder-detritivore, % shredder-herbivore, % coleopterans, % trichopterans, % oligochaetes*
Littoral macrophyte-beds	*VT-BI, Shannon-Wiener diversity, % collector-gatherer, % predator, % coleopterans, % trichopterans, % oligochaetes*
Littoral fine-muds	*VT-BI, Shannon-Wiener diversity, % collector-filterer, % predator, % coleopterans, % trichopterans, % oligochaetes*

**VT-BI is the Vermont stream biotic index*

Vermont's efforts toward developing useful biocriteria are by no means complete. The Field Phase outlined above was designed to provide the States of Vermont and New Hampshire with the baseline experience and information needed to move forward with a long-term sustainable bioassessment program. The trial criteria presented herein should be reevaluated and further refined in conjunction with that longer term program. The present Field Phase provides a large volume of useful data, but it may not be sustainable over the long-term. Using results from the 1996 to 1998 study lakes, the need for robust data to develop (and assess compliance with) criteria will be balanced with the fixed personnel and operating expenses of a small State agency.

Presently, the classification of study lakes is provisional. The reference condition for poorly-buffered (tannic) lakes is well characterized, though a group of poorly buffered and clear lakes might need to be characterized separately. Also there exists the need to assess the reference condition for a variety of well buffered (clear) lake types. Progress on this will be made during the 1998 field season.

This cooperative Vermont/New Hampshire initiative carries with it a paleolimnological component designed specifically to determine the historical condition of candidate reference lakes. Application of paleolimnological models to the sediments of selected candidate reference lakes will ensure that the underlying biological information used to develop criteria is indeed of reference quality.

Vermont has already seen the benefits of biological assessment as a tool for evaluating lakes. Data from this Field Phase have been used to refine and update Aquatic Life Use Support in Vermont's 305(b) inventory for every Vermont study lake bioassessed to date. While numeric criteria are not yet ready for inclusion into Vermont's Water Quality Standards, it is anticipated that subsequent revisions to Standards will contain lake biological criteria.

Readers of this Guidance are encouraged to communicate with the Project Contact directly for information regarding Vermont's bioassessment program.

Project Contact:
Neil Kamman, Aquatic Biologist
Vermont Department of Environmental Conservation
103 S. Main St., 10N
Waterbury, VT 05671-0108
phone (802) 241-3777
fax (802) 241-3287
neilk@dec.anr.state.vt.us
http://www.anr.state.vt.us/water1.htm

An Evaluation of the Draft Lake and Reservoir Bioassessment and Biocriteria Technical Guidance Document Using Wisconsin Lakes

Wisconsin's test of the Draft Lake Bioassessment and Biocriteria Technical Guidance focused on the development of a multimetric index for Wisconsin lakes. Such an index would provide a more accurate determination of use impairment for the biennial 305(b) water quality assessment, and changes as a result of watershed best management practices under the nonpoint priority watershed program. This index would also allow more informed permitting decisions, as well as proactive management by rapidly detecting emerging pollutant threats to lakes.

Background

Although bioassessment could detect changes from a broad range of anthropogenic sources, this study only involved lakes that have been impaired by eutrophication. While it would be beneficial to include other pollutants, e.g., acid precipitation and mercury, there was not sufficient information in our data set from lakes impacted by such pollutants. All of the lakes in our data set have experienced some degree of impairment from anthropogenic sources. As such, the reference lakes would be classified as "least impaired." Better reference sites could be selected if they had been chosen prior to data collection.

This study examines the trophic variables: total phosphorus, chlorophyll, Secchi depth and the biotic communities of phytoplankton, zooplankton, sedimented diatoms, and macrophytes. Analysis involved a comparison of index development using Tier 1A (single visit) Tier 1B (multiple visits) as well as Tier 2A and 2B. Macroinvertebrates and fish were not used in the analysis.

Data used for this analysis was largely collected under the Long Term Lakes program of the Wisconsin Department of Natural Resources. Collection began in 1986 and continues through the present. Samples are collected five times annually: late winter, June, July, August, and during fall turnover. Parameters that are analyzed from these collections include Secchi depth, chlorophyll, total phosphorus, phytoplankton, and zooplankton. In addition, the macrophyte assemblage was surveyed occasionally during this time period. Not all of the samples are available for this analysis. Trophic variables (Secchi, chlorophyll, and P) were available for the years 1986-1994 while phytoplankton from 1986 were used and zooplankton from 1986 and 1988 were used. As part of another project, sediment samples were collected from the main basin of most of these lakes in 1991 for diatom analysis. Seven reservoirs of varying trophic characteristics were sampled in 1994 for trophic variables, phytoplankton, and macrophytes.

Classification

The lakes sampled covered three ecoregions, but not all regions contain reference quality lakes. Therefore, lakes were not separated into three ecoregions. In fact, all but two of the reference lakes were found in the northern lakes and forests ecoregion. The reference lakes were chosen based upon low levels of development in their watersheds. All of the lakes had some development on the shoreline but most were summer homes and the density was relatively low. A criteria that was not used in the selection of the reference lakes was their known trophic status. It was felt that lakes that may have naturally had higher nutrient levels but low development should be included in the reference lakes. Since all of the lakes had some degree of disturbance in their watersheds the reference condition was calculated using the *tri-section* method.

The reference lakes, test lakes, and reservoirs exhibited a wide range of morphological conditions and watershed size (Table F-8). These data were used to determine how robust various metrics were in assessing unknown lakes. When possible, metrics were constructed based upon multiple visits as well as single visits during an index period. August was chosen as the index period.

Metric Development

At least 4 metrics were developed for each biological entity (Table F-9). Trophic variables of the lakes were described using Carlson's Trophic Status Index (TSI), modified for Wisconsin lakes. Trophic status for chlorophyll, phosphorus, and Secchi depth was described using the equations:

$$WTSI_{SD} = 60 - (32.2 \text{ Log SD})$$

$$WTSI_{TP} = 60 - (33.2(0.96 - .054 \log TP)$$

$$WTSI_{Chl} = 60 - (33.2(0.76 - 0.52 \log Chl)$$

Reference metrics are unusable if they are excessively variable. Variability of the metrics was measured by determining the ratio of the interquartile range to the scope of detection. The scope of detection was defined as the distance from the lower quartile to the minimum value possible when reference values were higher than the test cases. When the reference values were lower than the test cases, the scope of detection was defined as the distance from the upper quartile to the maximum value possible. An interquartile coefficient greater than one generally indicates the metric is too variable to detect impairments.

Some of the metrics had an interquartile coefficient greater than one, but each of the biological units had at least one metric with a coefficient less than one (Table F-9). For metrics to be useful there must also be good separation between the reference and test lakes. Each biological unit also had at least one metric that fulfilled this condition. The list of metrics that were judged to be robust and useful are listed in Table F-10.

Discussion

This study has formulated a draft bioassessment index for Wisconsin lakes using biocriteria metrics. Different metrics were not formulated for individual lake classes or for separate ecoregions largely because of the lack of sufficient reference lakes. Instead, the lakes for all the regions were combined. The Wisconsin Lake Index was tested on 13 lakes using all of the metrics and tested independently for Tiers 1A, 1B, 2A, and 2B. A comparison was made of each lake's classification at the four different levels. All the lakes received the same classification ("departing from reference conditions") under Tier 1A and 1B. Tier 2 appeared to be more descriptive of lake condition than Tier 1. In fact, two lakes classified as "departing from reference conditions" under Tier 1, were categorized as "impaired" using Tier 2 sampling techniques.

This analysis has allowed us to make some recommendations concerning which metrics are useful for developing a Wisconsin Lakes Index. Since all the lakes received the same score under Tier 1A and 1B it is suggested that only a single visit during the index period (August) is necessary if lakes are to be classified using Tier 1 only. In addition, we suggest that the macrophyte metrics be expanded under Tier 1. All of the macrophyte metrics, with the exception of density, can be determined with little or no extra effort under the suggested Tier 1 metrics in the draft document. Density should be included as a Tier 2 metric for macrophytes.

Although there was not complete agreement between Tier 2A and 2B they were similar enough to suggest that only one sampling trip is sufficient. In addition, it is suggested that the zooplankton metrics be eliminated until it is better understood how these metrics relate to the lake's impairment.

The diatom metrics tested were not as useful as expected. The only metric that proved robust enough to use was the percentage of *Stephanodiscus*. Although this metric was useful there was a great deal of variability across the test lakes. In the 13 lakes where the index was tested, the diatom metric tended to indicate that the lake was less impaired compared with most of the other metrics. This metric may not be as robust as some others. We recommend its usage but with reservations. The diatom metrics likely would be more useful if TSI values were calculated using the entire diatom assemblage but this would entail considerably more work, including detailed taxonomic knowledge.

A summary of the recommendations are:

Tier 1 Only one sampling trip (during August). Metrics: trophic state variables, macrophyte metrics except density.

Tier 2 Only one sampling trip (during August). Metrics: trophic state variables, macrophytes, phytoplankton, diatoms.

It is evident from this analysis that lake assessment using biocriteria is a more robust technique than using the traditional indices: phosphorus, Secchi depth, and chlorophyll by themselves. The additional information from the biota gives a much more accurate picture of a lake's health, especially its biological integrity.

Improvements could be made in developing a Wisconsin Lake Index if better reference conditions were used to define the metrics. Since most of the reference lakes used in this study had some lakeshore development they were not ideal choices. Another ongoing study has identified sufficient reference lakes in each of the major ecoregions in the state, and will develop metrics for sedimented diatoms.

Table F-8. Morphological data for the study lakes and reservoirs.

Lake Name	County	MORPHOLOGY							
		lake area (hectares)	max. depth (m)	mean depth (m)	lake vol. (m^3)	watershed (hectares)	*watershed lake area*	hydrology description	S.D.F.
REFERENCE LAKES									
Bear Paw Lake	Oconto	19.8	6.1			414	20.9	seepage	1.88
Eau Claire Lake (Lower)	Douglas	324.6	12.5	6.7	21763552	2813	8.7	drainage	1.73
Eau Claire Lake (Upper)	Bayfield	403.1	28.0	8.8	35627887	2046	5.1	drainage	2.30
Escanaba Lake	Vilas	118.6	7.9	4.3	5059742	518	4.4	drainage	2.10
Franklin Lake	Oneida	65.2	7.6			259	4.0	seepage	1.46
Keyes Lake	Florence	81.7	23.5			570	7.0	drainage	1.61
Lac Courte Oreilles	Sawyer	2039.2	27.4	10.4	211327493	17042	8.4	drainage	2.55
Lost Lake	Florence	37.2	13.7	5.2	1929165	52	1.4	seepage	1.09
Patten Lake	Florence	103.2	15.8	5.5	5661681	2098	20.3	drainage	1.78
Round Lake	Chippewa	87.4	5.5	3.0	2664321	298	3.4	seepage	1.28
Silver Lake	Barron	136.4	27.7	11.6	15795967	1935	14.2	seepage	2.25
TEST LAKES									
Amnicon Lake	Douglas	172.4	9.4	3.0	5254632	1251	7.3	drainage	2.70
Bass Lake	St. Croix	168.8	10.7					seepage	
Big Cedar Lake	Washington	377.2	32.0	10.4	39086570			drainage	2.25
Big Green Lake	Green Lake	2972.8	71.9	31.7	942360318	958	0.3	drainage	1.39
Big Long Lake	Manitowoc	48.6	11.6			518	10.7	seepage	2.22
Big McKenzie	Burnett	479.6	21.6	5.8	27771841	1709	3.6	drainage	1.47
Browns Lake	Racine	160.3	15.2	2.4	3907670			drainage	1.82
Butternut Lake	Price	407.1	9.8	4.3	17372357	11139	27.4	drainage	2.52
Cedar Lake	Polk/St. Croix	448.0	8.5			8614	19.2	drainage	1.42
Clark Lake	Door	351.3	7.6	2.1	7494635	4817	13.7	drainage	1.53
Crystal Lake	Sheboygan	61.5	18.6	6.1	3749785			seepage	1.87
Fish Lake	Dane	87.4	18.9					seepage	1.26
Fox Lake	Dodge	1062.3	5.8	2.1	22665227	15022	14.1	drainage	2.19
Friess Lake	Washington	47.3	14.6	8.2	3896569			drainage	1.51
Kentuck Lake	Vilas	387.3	12.2	4.0	15345746	777	2.0	drainage	1.30
Lac La Belle	Waukesha	471.1	13.7	3.4	15793500			drainage	1.69
Long Lake	Chippewa	425.7	30.8	6.1	25952456	1746	4.1	drainage	3.08
Long Lake	Fond du Lac	168.8	14.3	6.7	11315962			drainage	1.76
Mason Lake	Adams	346.0	2.7	2.1	7382388	9453	27.3	drainage	1.92
Minocqua Lake	Oneida	550.4	18.3	7.0	38583309	20720	37.6	drainage	3.68
Nagawicka Lake	Waukesha	371.1	27.4	11.0	40719699			drainage	1.98
Pelican Lake	Oneida	1450.8	11.9			2590	1.8	drainage	1.91
Pewaukee Lake	Waukesha	1008.9	13.7	4.6	46126050			drainage	1.94
Pike Lake	Marathon	83.0	10.4	4.0	3287229	829	10.0	drainage	1.32
Pike Lake	Washington	211.2	13.7					drainage	1.19
Ripley Lake	Jefferson	169.2	13.4	5.5	9280717			seepage	1.40
Rock Lake	Jefferson	554.8	17.1	4.9	27057656			drainage	1.43
Rollingstone Lake	Langlade	271.9	3.7			2512	9.2	drainage	1.32
Sand Lake	Rusk/Chippewa	106.0	30.5	8.8	9371994	251	2.4	seepage	1.98
School Section Lake	Waupaca	15.8	11.6	8.5	1346962			drainage	1.76
Shell Lake	Washburn	1044.1	11.0	7.0	73194807	4159	4.0	seepage	1.43
Silver Lake	Waupaca	27.5	5.2	2.1	587137			seepage	1.05
Squaw Lake	St. Croix	52.2	9.8	4.0	2068549	259	5.0	seepage	2.91
Thunder Lake	Oneida	742.6	2.7			2590	3.5	drainage	1.80
White Clay Lake	Shawano	94.7	14.0	4.3	4040886	1036	10.9	drainage	1.56
Whitewater Lake	Walworth	259.0	11.6					drainage	2.80
Wilson Lake	Iron	65.6	6.4	4.3	2797537	207	3.2	drainage	1.89
RESERVOIRS									
Big Eau Pleine	Marathon	2764.0	14.0	4.9	134794885	8536	3.1	drainage	5.48
Brule	Florence	120.2	19.5	6.1	7326882	271949	2262.6	drainage	2.51
Caldron Falls	Marinette	412.0	12.2	4.6	18835266	124796	302.9	drainage	4.70
Dutch Hollow Lake	Sauk	85.0	12.2			1313	15.5	drainage	2.31
Gile	Iron	1369.5	7.6			18130	13.2	drainage	3.19
Minong	Washburn	632.9	6.4	2.7	17362489	60507	95.6	drainage	4.48
Rainbow	Oneida	823.5	8.5			194249	235.9	drainage	3.53
Redstone Lake	Sauk	247.7	11.0	4.3	10568472	7677	31.0	drainage	4.69
St. Croix	Douglas	774.2	8.5	2.1	16517554	32437	41.9	drainage	3.38
Willow	Oneida	2551.9	9.1	3.0	77783359	84693	33.2	drainage	7.29

Table F-9. Summary of degree of separation and the interquartile coefficient for all of the metrics.

Metric	Separation	Interquartile Coefficient
Trophic Variables		
$WTSI_{Chl}$	yes	0.08
$WTSI_{SD}$	yes	0.05
$WTSI_{TP}$	yes	0.06
Phytoplankton		
Richness	no	0.8
Ana, Aph, Micro	no	4.2
% Blue-green density	yes	1.0
% Blue-green biomass	yes	0.4
Zooplankton		
Richness	yes	0.8
Daphnia size	yes	1.0
Herbivore/Predator	yes	0.6
Copepod/Cladocera	no	
Large Predator	no	
Chydorus	no	0.3
Diatoms		
Richness	maybe	0.15
Diversity	no	0.5
% Planktonic taxa	no	2.3
% Aulacoseira spp.	no	0.3
% Stephanodiscus spp.	yes	0.02
% Cyclotella	yes	7.3
Macrophytes		
Richness	no	0.9
% Coverage of littoral zone	no	1.5
Max depth of growth	no	0.8
% Exotic taxa	yes	1.0
% Sensitive taxa	yes	1.0

Table F-10. Metrics that possess good separation between reference and test lakes as well as an interquartile coefficient less than or equal to 1.0.

Trophic Variables	$WTSI_{Chl}$
	$WTSI_{SD}$
	$WTSI_{TP}$
Phytoplankton	% blue-green density
	% blue-green biomass
Zooplankton	Herbivore/predator
	Daphnia size
	No. of taxa
Diatoms	No. of taxa
	% *Stephanodiscus*
	% *Cyclotella*
Macrophytes	% exotic species
	% sensitive species

Project Contacts:
Paul Garrison and James Johnson
Wisconsin Department of Natural Resources
Bureau of Research
1350 Femrite Drive
Monona, WI 53716

Appendix G

Literature Cited

Alcaraz, M., G.A. Paffenhofer, and J.R. Strickler. 1980. Catching the algae: A first account of visual observations on filter-feeding calanoids. Pages 241-248 In *ASLO Special Symposium 3: The evolution and ecology of zooplankton communities*. University Press of New England, Hanover, NH.

APHA. 1992. *Standard methods for the examination of water and wastewater*. 16th ed. American Public Health Association, Washington, DC.

ASTM. 1990. *Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing*. ASTM Designation E 1391 - 90. American Society of Testing and Materials, West Conshohocken, PA.

——. 1998 (draft). *Standard test methods for measuring the toxicity of sediment-associated contaminants with fresh water invertebrates*. ASTM Designation E 1706 - 95b. American Society of Testing and Materials, West Conshohocken, PA.

Bahls, L. 1993. *Periphyton bioassessment methods for Montana streams*. Montana Water Quality Bureau, Department of Health and Environmental Science, Helena, MT.

Baker, J.P. and S.W. Christensen. 1991. Effects of acidification on biological communities in aquatic ecosystems. Pages 83-106 In D.F. Charles, ed. *Acidic deposition and aquatic ecosystems: Regional case studies*. Springer-Verlag, New York, NY.

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(3):386-391.

Barbour, M.T., J. Gerritsen, G.O. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M.L. Bastian. 1996a. 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, and J.S. White. 1996b. *Development of the Stream Condition Index (SCI) for Florida*. Prepared for Florida Department of Environmental Protection, Tallahassee, FL.

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. *Environmental Toxicology and Chemistry* 11:437-449.

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, eds. *Biological Assessment and Criteria: Tools for water resource planning and decision making.* Lewis Publishers, Boca Raton, FL.

Barbour, M.T. and S. Thornley. 1990. Quality assurance issues related to field assessment of aquatic ecosystems. Pages 167-183 In D. Hart, ed. *Proceedings of the Third Annual Ecological Quality Assurance Workshop.* Burlington, Ontario.

Barko, J.W., H.L. Eakin, and D.G. McFarland. 1992. *Interrelationships between macrophyte growth and sediment nutrient availability.* Abstracts, NALMS 12th Annual International Symposium. North American Lake Management Society, Madison, WI.

Battarbee, R.W. 1973. A new method for the estimation of absolute microfossil numbers, with reference especially to diatoms. *Limnology and Oceanography* 18:647-653.

Birks, H.J.B., S. Juggins, and J.M. Line. 1990. Lake surface-water chemistry reconstructions from paleolimnological data. Pages 301-313 In B. J. Mason, ed. *The Surface Waters Acidification Programme.* Cambridge University Press, Cambridge, MA.

Blalock, H.M., Jr. 1979. *Social Statistics.* Revised 2nd edition. McGraw-Hill Book Company, New York, NY.

Brooks, J.L. and S.I. Dodson. 1965. Predation, body size and composition of plankton. *Science* 150:28-35.

Camburn, K.E., J.C. Kingston, and D.F. Charles. 1984-1986. *PIRLA Diatom Iconograph.* PIRLA unpublished report series, Report no. 3, Indiana University, Bloomington, IN.

Canfield, D.E., Jr. 1981. *Chemical and trophic state characteristics of Florida lakes in relation to regional geology.* Final report to Florida Cooperative Fish and Wildlife Unit, University of Florida, Gainesville, FL.

Canfield, D.E., Jr., K.A. Langeland, M.J. Maceina, W.T. Haller, J.V. Shireman, and J.R. Jones. 1983. Trophic state classification of lakes with aquatic macrophytes. *Canadian Journal of Fisheries and Aquatic Sciences* 40:1713-1718.

Caper, J., G. Power, and F.R. Shivers, Jr. 1983. *Chesapeake waters, pollution, public health and public opinion.* Tidewater Publishers, Centreville, MD.

Carlson, R.E. 1977. A trophic state index for lakes. *Limnology and Oceanography* 22:361-369.

Carlson, R.E. 1992. Expanding the trophic state concept to identify non-nutrient limited lakes and reservoirs. Pages 59-71 In *Enhancing the State's Lake Management Programs: Proceedings of a conference, Chicago, 1991.* North American Lake Management Society, Madison, WI.

Carlson, R.E. and J. Simpson. 1996. *Volunteer Lake Monitoring in the Upper Mid-west: Programs, Techniques, and Technical Recommendations.* North American Lake Management Society, Madison, WI.

Carpenter, S.R., J.F. Kitchell, J.R. Hodgson, P.A. Cochran, J.A. Elser, M.M. Elser, D.M. Lodge, D. Kretchmer, X. He, and C.N. von Ende. 1987. Regulation of lake primary productivity by food web structure. *Ecology* 68:1863-1876.

Charles, D.F. and J.P. Smol. 1994. Long-term chemical changes in lakes: Quantitative inferences using biotic remains in the sediment record. *Advances in Chemistry* 237:1-51.

Charles, D.F., J.P. Smol, and D.R. Engstrom. 1994. Paleolimnological approaches to biological monitoring. Pages 233-293 In L. L. Loeb, and A. Spacie, eds. *Biological monitoring of aquatic systems.* CRC Press, Boca Raton, FL.

Charles, D.F. and D.R. Whitehead. 1986. The PIRLA project: Paleoecological investigation of recent lake acidification. *Hydrobiologia* 143:13-20.

Cochran, W.G. 1963. *Sampling Techniques.* John Wiley and Sons, Inc., New York, NY.

Conover, W.J. 1980. *Practical Nonparametric Statistics. 2nd edition.* John Wiley and Sons, Inc., New York, NY.

Conquest, L.L., S.C. Ralph, and R.J. Naiman. 1994. Implementation of large-scale stream

monitoring efforts: Sampling design and data analysis issues. Pages 69-90 In L. Loeb and A. Spacie, eds. *Biological Monitoring of Aquatic Systems*, Lewis Publishers, Boca Raton, FL.

Courtemanch, D.L., S.P. Davies, and E.B. Laverty. 1989. Incorporation of biological information in water quality planning. *Environmental Management* 13:35-41.

Crowder, A. and D.S. Painter. 1991. Submerged macrophytes in Lake Ontario: Current knowledge, importance, threats to stability, and needed studies. *Canadian Journal of Fisheries and Aquatic Sciences* 48:1539-1545.

Cumming, B.F., J.P. Smol, J.C. Kingston, D.F. Charles, H.J.B. Birks, K.E. Camburn, S.S. Dixit, A.J. Uutala, and A.R. Selle. 1992. How much acidification has occurred in Adirondack regionlakes (New York, USA) since preindustrial times? *Canadian Journal of Fisheries and Aquatic Sciences* 49:128-141.

Davic, R.D. and J.E. DeShon. 1989. The Ohio Lake Condition Index: A new multiparameter approach to lake classification. *Lake Research Management* 5:1-8.

Davies, S.P., L.T Somides, D.L. Courtemanch, and F. Drummond. 1993. *Maine biological monitoring and biocriteria development program.* Maine Department of Environmental Protection, Bureau of Water Quality Control, Division of Environmental Evaluation and Lake Studies, Augusta, ME.

DeBernardi, R. 1984. Methods for the estimation of zooplankton abundance. Pages 59-86 In J.A. Downing and F.H. Rigler, eds. *A manual on methods for the assessment of secondary productivity in fresh waters. 2nd edition.* Blackwell Scientific Publications, Oxford, England.

DeShon, J.E. 1995. Development and application of the Invertebrate Community Index (ICI). Pages 217-244 In W.S. Davis and T.P. Simon, eds. *Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making.* CRC Press, Boca Raton, FL.

Dillon, P.J. and F.H. Rigler. 1974. The phosphorus-chlorophyll relationship in lakes. *Limnology and Oceanography* 19:767-773.

Dillon, P.J., W.A. Scheider. R.A. Reid, and D.S. Jeffries. 1994. Lakeshore capacity study: Part l-Test of effects of shoreline development on the trophic status of lakes. *Lake Reservation Management* 8(2):121-129.

Dionne, M. and J.R. Karr. 1992. Ecological monitoring of fish assemblages in Tennessee River reservoirs. Pages 259-281 In D.H. McKenzie, D.E. Hyatt, and V.J. McDonald, eds. *Ecological indicators.* Elsevier, New York, NY.

Dixit, S.S., B.F. Cumming, H.J.B. Birks, J.P. Smol, J.C. Kingston, A.J. Uutala, D.F. Charles, and K.E. Camburn. 1993. Diatom assemblages from Adirondack lakes (New York, USA) and the development of inference models for retrospective environmental assessment. *Journal of Palaeolimnology* 8:27-47.

Dixit, S.S., J.P. Smol, J.C. Kingston, and D.F. Charles. 1992. Diatoms: Powerful indicators of environmental change. *Environmental Science Technology* 26:23-33.

Dodson, S.I. 1992. Predicting crustacean zooplankton richness. *Limnology and Oceanography* 37:848-856.

Donigian, A.S., Jr. and W.C. Huber. 1991. *Modeling of nonpoint source water quality in urbanand non-urban areas.* EPA 68-03-3513(#29). U.S. Environmental Protection Agency, Athens, GA.

Dycus, D.L. and D.L. Meinert. 1992. *Reservoir vital signs monitoring - 1991: Summary of vital signs and use impairment monitoring on Tennessee Valley Reservoirs.* TVA/WR—92/8. Water Resources Division, Tennessee Valley Authority, Chattanooga, TN.

——. 1993. Monitoring and evaluation of aquatic resource health and use suitability in Tennessee Valley Authority reservoirs. Draft. *Water Management.* Tennessee Valley Authority, Chattanooga, TN.

——. 1994. Tennessee Valley reservoir and stream quality 1993 summary of vital signs and use suitability monitoring - Volume 1. *Water Mangement.* Tennessee Valley Authority, Chattanooga, TN.

Eakins, J.D. and R.I. Morrison. 1978. A new procedure for the determination of ^{210}Pb in lake and marine sediments. *International Journal of*

Radio-Isotopes 29:531-536.

Edmondson, W.T. and J.T. Lehman. 1981. The effect of changes in the nutrient income on the condition of Lake Washington. *Limnology and Oceanography* 26:1-29.

Edmondson, W.T. and A.H. Litt. 1982. Daphnia in Lake Washington. *Limnology and Oceanography* 27:272-293.

Engle, V.D., J.K. Summers, and G.R. Gaston. 1994. A benthic index of environmental condition of Gulf of Mexico estuaries. *Estuaries* 17:372-384.

Environment Canada. 1994. *Development of methodologies for whole-sediment toxicity tests with benthic invertebrates: a. Optimum age for initiation and duration of Hyallela azteca survival and growth test, b. Formulation of artificial sediment(s) for use in reference toxicity tests with benthic invertebrates.* Environment Canada Method Development and Application Division Technology Development Directorate. TS - 27.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine Freshwater Research* 42:555-67.

Fausch, D.D., J.R. Karr, and P.R. Yant. 1984. Regional application of an index of biotic integrity based on stream fish communities. *Transactions of the American Fisheries Society* 113:39-55.

FDEP. 1994. *Lake bioassessments for the determination of nonpoint source impairment in Florida.* Draft. Florida Department of Environmental Protection, Biology Section, Division of Administrative and Technical Services, Tallahassee, FL.

——. 1996. *Standard operating procedures for biological assessment.* Florida Department of Environmental Protection, Biology Section, Tallahassee, FL.

Fore, L.S., J.R. Karr, and L.L. Conquest. 1994. Statistical properties of an index of biotic 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. 1975. Biological integrity of water - An historical approach. Pages 127-140 In R.K. Ballentine and L.J. Guarrai, eds. *The Integrity of Water*, U.S. Environmental Protection Agency, Washington, DC.

Fritz, S.C. 1990. Twentieth-century salinity and water-level fluctuations in Devils Lake, North Dakota: Test of a diatom-based transfer function. *Limnology and Oceanography* 35:1771-1780.

Fryer, G. 1985. Crustacean diversity in relation to the size of waterbodies: Some facts and problems. *Freshwater Biology* 15:347-361.

Gauch, H.G. 1982. *Multivariate analysis in community ecology.* Cambridge University Press, Cambridge, MA.

Gerritsen, J. 1980. Adaptive responses to encounter problems. Pages 52-62 In W.C. Kerfoot, ed. *The evolution and ecology of zooplankton communities.* University Press of New England, Hanover, NH.

Gerritsen, J. 1995. Additive biological indices for resource management. *Journal of the North American Benthological Society* 14:451-457.

Gerritsen, J., and M.L. Bowman. 1994. *Periphytic diatom assemblages on high elevation Rocky Mountain lakes: Characterization of reference conditions.* Prepared for the Montana Department of Health and Environmental Quality, Helena, MT and the U.S. Environmental Protection Agency, Washington, D.C.

Gerritsen, J., and J.S. White. 1997. *Development of a biological index for Florida Lakes.* Prepared for Florida Department of Environmental Protection, Tallahassee, FL.

Glew, J.R. 1988. A portable extruding device for close interval sectioning of unconsolidated core samples. *Journal of Palaeolimnology* 1:235-239.

——. 1989. A new trigger mechanism for sediment samplers. *Journal of Palaeolimnology* 2:241-243.

Gray, J.S. 1989. Effects of environmental

stress on species fish assemblages. *Biological Journal of the Linnean Society* 37:19-32.

Green, G.R. and B.A. Beckman. 1992. The composite index of coincident indicators and alternative coincident indexes. *Survey of Current Business.* U.S. Department of Commerce, Bureau of Economic Analysis, Washington, DC.

Gurtz, M.E. 1994. Design of biological components of the National Water-Quality Assessment (NAWQA) Program. Pages 323-354 (Chapter 15) In S.L. Loeb and A. Spacie, eds. *Biological Monitoring of Aquatic Systems.* Lewis Publishers, Boca Raton, FL.

Hall, R.I. and J.P. Smol. 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. *Freshwater Biology* 27:417-434.

Hecky, R. E. and H. J. Kling. 1981. The phytoplankton and protozooplankton of the euphotic zone of Lake Tanganyika: Species composition, biomass, chlorophyll content, and spatio-temporal distribution. *Limnology and Oceanography* 26:532-547.

Heiskary, S.A. 1989. Integrating ecoregion concepts into state lake management programs. Pages 89-100 In *Enhancing the States' Lake Management Programs Proceedings.* North American Lake Management Society, Madison, WI.

Heiskary, S.A. and C.B. Wilson. 1989. The regional nature of lake water quality across Minnesota: An analysis for improving resource management. *Journal of the Minnesota Academy of Science* 55:71-77.

Heiskary, S.A., C.B. Wilson, and D.P. Larsen. 1987. Analysis of regional patterns in lake water quality: Using ecoregions for lake management in Minnesota. *Lake Reservation Management* 3:337-344.

Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources.* Elsevier, Amsterdam.

Hergenrader, G.L. and D.C. Lessig. 1980. Eutrophication of the salt valley reservoirs, 1968-1973, III. The macroinvertebrate community: Its development, composition, and change in response to eutrophication. *Hydrobiologia* 75:7-25.

Hickman, G.D. and T.A. McDonough. 1996. Assessing the reservoir fish assemblage index: A potential measure of reservoir quality. *American Fisheries Society Symposium* 16:85-97.

Hill, K.R. 1986. Classification of Iowa lakes and their fish standing stock. *Lake Reservation Management* 2:105-109.

Hill, M.O. 1979. TWINSPAN-A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of individuals and attributes. Cornell University, Ithaca, NY.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomology* 20:31-39.

Hosper, S.H. and M.L. Meijer. 1993. Biomanipulation, will it work for your lake? A simple test for the assessment of chances for clear water, following drastic fish-stock reduction in shallow, eutrophic lakes. *Ecological Engineering* 2:63-72.

Hosper, S.H., M.L. Meijer, and P.A. Walker, eds. 1992. Handleiding Actief Biologisch Beheer. RIZA/OVB (Netherlands Institute for Inland Water Management and Wastewater Treatment/ Organization for Improvement of Inland Fisheries) Lelystad, The Netherlands (in Dutch).

Hughes, R.M., S.A. Heiskary, W.J. Matthews, and C.O. Yoder. 1994. Use of ecoregions in biological monitoring. Pages 125-151 In S.L. Loeb and A. Spacie, eds. *Biological monitoring of aquatic systems.* Lewis Publishers, Boca Raton, FL.

Hughes, R.M. and D.P. Larsen. 1988. Ecoregions: An approach to surface water protection. *Journal of Water Pollution Control Federation* 60:486-493.

Hurlbert, S.H. 1971. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* 52:577-586.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monography* 54(2).

Hutchinson, G.E. 1957. *A treatise on limnology. Vol. 1. Geography, physics and chemistry.* John

Wiley and Sons, New York, NY.

Hutchinson, G.E. 1965. *The ecological theater and the evolutionary play.* Yale University Press, New Haven, CT.

Hutchinson, G.E. 1967. *A treatise on limnology. Vol. 2. Introduction to lake biology and the limnoplankton.* John Wiley and Sons, New York, NY.

Hynes, H.B.N. 1994. Historical perspective and future direction of biological monitoring of aquatic systems. Chapter 2 In S.L. Loeb and A. Spacie, eds. *Biological monitoring of aquatic systems.* CRC Press, Boca Raton, FL.

ITFM. 1992. *Ambient water-quality monitoring in the United States. First year review, evaluation, and recommendations.* ITFM. Interagency Advisory Committee on Water Data. Intergovernmental Task Force on Monitoring Water Quality, Water Information Coordination Program, Washington, DC.

Intergovernmental Task Force on Monitoring Water Quality (ITFM). 1994. *Water Quality Monitoring in the United States (Region V): 1993 Report of the Intergovernmental Task Force on Monitoring Water Quality.* Intergovernmental Task Force on Monitoring Water Quality, Washington, DC.

Jackson, D.A. 1993a. Multivariate analysis of benthic invertebrate communities: The implications of choosing particular data standardizations, measures of association, and ordination methods. *Hydrobiologia* 268: 9-26

Jackson, D.A. 1993b. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology* 74: 2204-2214.

Jennings, M.J., L.S. Fore, and J.R. Karr. 1995. Biological monitoring of fish assemblages in Tennessee Valley reservoirs. *Regulated Rivers: Research & Management* 11:263-274.

Johnson, R.A. and D.W. Wichern. 1992. *Applied Multivariate Statistical Analysis.* John Wiley & Sons, New York, NY.

Johnson, R.J. 1989. Water quality standards for lakes. Pages 123-128 In G.H. Flock, ed. *Water quality standards for the 21st Century.*

U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Jongman, R.H.G., C.J.F. ter Braak, and O.F.R. van Tongeren. 1987. *Data analysis and community and landscape ecology.* Pudoc. Wageningen, the Netherlands.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21-27.

——. 1991. Biological integrity: A long-neglected aspect of water resource management. *Ecological Application* 1:66-84.

Karr, J.R. and D.R. Dudley. 1981. Ecological perspective on water quality goals. *Environmental Management* 5:55-68.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. *Assessing biological integrity in running waters: A method and its rationale.* Illinois Natural History Survey, Special Publication 5.

Kenkel, N.C. and Orloci, L. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: Some new results. *Ecology* 67: 919-928.

Kennedy, R.H., K.W. Thornton, and D. Ford. 1985. Characterization of the reservoir ecosystem. In D. Gunnison, ed. *Microbial processes in reservoirs.* Dr. W. Junk Publishers, Boston, MA.

Kennedy, R.H. and W.W. Walker. 1990. Reservoir nutrient dynamics. Pages 109-132 In K.W. Thornton, B.L. Kimmel and F.E. Payne, eds. *Reservoir limnology: Ecological perspective.* John Wiley and Sons, New York, NY.

KDEP. 1993. *Methods for assessing biological integrity of surface waters.* Kentucky Department of Environmental Protection, Division of Water, Frankfort, KY.

Kerans, B.L., and J.R. Karr. 1994. A Benthic Index of Biotic Integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Application* 4:768-785.

Kingston, J.C. 1986. Diatom analysis. In D.F. Charles and D.R. Whitehead, eds. *Paleoecological investigation of recent lake acidification: Methods and project descriptions.* EA-4906, Res. Proj. 2174-10. Electric Power Research Institute, Palo Alto, CA.

Knowlton, M.F. and J.R. Jones. 1989. Comparison of surface and depth-integrated composite samples for estimating algal biomass and phosphorus values and notes on the vertical distribution of algae and photosynthetic bacteria in midwestern lakes. *Archiv für Hydrobiologie. Supplement* 83:175-96.

Kratzer, C.R. and P.L. Brezonik. 1981. A Carlson-type trophic state index for nitrogen in Florida lakes. *Water Research Bulletin* 17:713-715.

Lahiri, K. and G.H. Moore. 1991. *Leading economic indicators: New approaches and forecasting records.* Cambridge University Press, Cambridge, MA.

Larsen, D.P. and H. Mercier. 1976. Phosphorus retention capacity of lakes. *Journal of Fisheries Research Board of Canada* 33:742-750.

Leach, J.H. and R.C. Herron. 1992. A review of lake habitat classification. Chapter 3 In W.D.N. Busch and P.G. Sly, eds. *The development of an aquatic habitat classification system for lakes.* CRC Press, Boca Raton, FL.

Leonard, P.M. and D.J. Orth. 1986. Application and testing of an index of biotic integrity in small, coolwater streams. *Transaction of American Fisheries Society* 115:401-414.

Lewis, W.M., Jr. 1979. *Zooplankton community analysis.* Springer-Verlag, New York, NY.

Line, J.M., C.J.F. ter Braak, and H.J.B. Birks. 1994. WACALIB version 3.3-a computer program to reconstruct environmental variables from fossil assemblages by weighted averaging and to derive sample-specific errors of prediction. *Journal of Palaeolimnology* 10:147- 152.

Ludwig, J.A. and J.F. Reynolds. 1988. *Statistical ecology.* John Wiley and Sons, New York, NY.

Martin, R.G. and R.D. Arneson. 1978. Comparative limnology of a deep-discharge reservoir and a surface-discharge lake on the Madison River, Montana. *Freshwater Biology* 8:33-42.

Masters, A.E. 1992. *Reservoir vital signs monitoring, 1991 - Benthic macroinvertebrate community results.* TVA/WR-92/3. Tennessee Valley Authority, Water Resources Division, Chattanooga, TN.

Maughan, J.T. 1993. *Ecological assessment of hazardous waste sites.* Van Nostrand Reinhold, New York, NY.

Maxted, J.R., E.L. Dickey, and G.M. Mitchell. 1994. *Habitat quality of a Delaware nontidal stream.* Delaware Department of Natural Resources and Environmental Control, Dover, DE.

Mayewski, P.A., M. Wumkes, J. Klinck, M.S. Twickler, J.S. Putscher, K.C. Taylor, A.J. Gow, E.D. Waddington, R.B. Alley, J.E. Gibb, P.M. Grootes, D.A. Meese, M. Ram, M. Whalen, and A.T. Wilson. 1994. Record drilling depth struck in Greenland. *Eos* 75:113-124.

Mazumder, A. 1994. Patterns of algal biomass in dominant odd- vs. even-link lake ecosystems. *Ecology* 75(4):1141-1149.

Mazumder, A., W.D. Taylor, D.J. McQueen, and D.R.S. Lean. 1990. Effect of fish and plankton on lake temperature and mixing depth. *Science* 247(1):312-315.

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. U.S. Geological Survey, Washington, DC.

Meinert, D.L., S.R. Butkus, and T.A. McDonough. 1992. *Chickamauga Reservoir Embayment Study - 1990.* TVA/WR-92/28. Tennessee Valley Authority, Water Resources Division, Chattanooga, TN.

Miller, D.L., P.M. Leonard, R.M. Hughes, J.R. Karr, P.B. Moyle, L.H. Schrader, B.A. Thompson, R.A. Daniels, K.D. Fausch, G.A. Fitzhugh, J.R. Gammon, D.B. Halliwell, P.L. Angermeier, and D.J. Orth. 1988. Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13:12-20.

Mills, E.L., D.M. Green, and A. Schiavone, Jr. 1987. Use of zooplankton size to assess the community structure of fish populations in freshwater lakes. *North American Journal of Fisheries Management* 7:369-378.

Mills, E.L. and A. Schiavone, Jr. 1982. Evalua-

tion of fish communities through assessment of zooplankton populations and measures of lake productivity. *North American Journal of Fisheries Management* 2:14-27.

Moore, J.W. 1981. Factors influencing the species composition, distribution and abundance of benthic invertebrates in the profundal zone of a eutrophic northern lake. *Hydrobiologia* 85:505-510.

Moss, D., M.T. Furse, J.F. Wright, and P.D. Armitage. 1987. The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 41-52.

NALMS. 1992. *Developing eutrophication standards for lakes and reservoirs.* North American Lake Management Society, Alachua, FL.

Naumann, E. 1932. Grundzuge der regionalen limnologie. *Binnengewasser* 11.

Norris, R.H. 1995. Biological monitoring: The dilemma of data analysis. *Journal of North American Benthological Society* 14:440-450.

NRC. 1992. *Restoration of aquatic ecosystems.* National Research Council. National Academy Press, Washington, DC.

Nürnberg, G.K. 1996. Trophic state of clear and colored, soft- and hardwater lakes with special consideration of nutrients, anoxia, phytoplankton and fish. *Lake and Reservation Management* 12(4):432-447.

O'Brien, W.J. 1979. The predator-prey interaction of planktivorous fish and zooplankton. *American Science* 67:572-581.

——. 1990. Perspectives on fish in reservoir ecosystems. Pages 209-226 In K.W. Thornton, B.L. Kimmel, and F.E. Payneln, eds. *Reservoir limnology: Ecological perspectives.* John Wiley and Sons, New York, NY.

O'Gorman, R., E.L. Mills, and J.S. DeGisi. 1991. Use of zooplankton to assess the movement and distribution of alewife (*Alosa pseudoharengus*) in south-central Lake Ontario in spring. *Canadian Journal of Fisheries Aquatic Sciences* 48:2250-2257.

Oglesby, R.T. 1977. Relationships of fish yield to lake phytoplankton standing crop, production, and morphoedaphic factors. *Journal of Fisheries Research Board of Canada* 34(12):2271-2279.

Ohio EPA. 1987. *Biological criteria for the protection of aquatic life. Vols. 1-3.* Ohio Environmental Protection Agency, Monitoring and Assessment Program, Surface Water Section, Division of Water Quality, Columbus, OH.

——. 1990. *The use of biocriteria in the Ohio EPA surface water monitoring and assessment program.* Ohio Environmental Protection Agency, Ecological Assessment Section, Division of Water Quality Planning and Assessments, Columbus, OH.

OCC. 1993. *Development of rapid bioassessment protocols for Oklahoma utilizing characteristics of the diatom community.* Oklahoma Conservation Commission, Oklahoma City, OK.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geography* 77(1):118- 125.

Oreskes, N., K. Schrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641-646.

Orth, R.J. and J.W. Moore. 1983. Chesapeake Bay: An unprecedented decline in submerged aquatic vegetation. *Science* 222:51-53.

Palmer, M.W. 1993. Putting things in even better order: The advantages of canonical correspondence analysis. *Ecology* 74(8):2215-2230.

Patalas, K. 1971. Crustacean plankton and the eutrophication of St. Lawrence Great Lakes. *Journal of Fisheries Research Board of Canada* 29:1451-1462.

Patrick, R. 1949. A proposed biological measure of stream conditions, based on a survey of the Conestoga Basin, Lancaster County, Pennsylvania. *Proceedings of the Philadelphia Academy of Natural Sciences* 101:277-341.

Perry, J.A. and N.H. Troelstrup. 1988. Whole ecosystem manipulation: A productive avenue for test system research? *Environmental Toxicology and Chemistry* 7:941 -951.

Peterman, R.M. 1990. The importance of reporting statistical power: The forest decline andacidic deposition example. *Ecology* 71(5):2024-2027.

Peters, J.A. 1988. Quality control infusion into stationary source sampling. Chapter 22 In L.H. Keith, ed. *In Principles of environmental sampling.* American Chemical Society, Washington, DC.

Peters, R.H. 1991. *A critique for ecology.* Cambridge University Press, Cambridge, MA.

Pielou, E.C. 1977. *Mathematical ecology. 2nd edition.* John Wiley and Sons, Inc., New York, NY.

Porter, K.G. 1977. The plant-animal interface in freshwater ecosystems. *American Science* 65:159-170.

PSEP. 1995. *Recommended guidelines for conducting laboratory bioassays on Puget Sound sediments.* Prepared for U.S. Environmental Protection Agency Region 10, Office of Puget Sound, Seattle, WA and Puget Sound Water Quality Authority, Olympia, WA.

Ranasinghe, J.A., S.B. Weisberg, J. Gerritsen, and D.M. Dauer. 1994. *Assessment of Chesapeake Bay benthic macroinvertebrate resource condition in relation to water quality and watershed stressors.* Prepared by Versar, Inc. for Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD.

Reynolds, C.S. 1984. *The ecology of freshwater phytoplankton.* Cambridge University Press, Cambridge, MA.

Reynoldson, T.B., R.C. Bailey, K.E. Day, and R.H. Norris. 1995. Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology* 20:198-219.

Rohm, C.M., J.M. Omernik, C.W. Kiilsgaard. 1995. Regional patterns of total phosphorus in lakes of the Northeastern United States. *Lake and Reservoir Management* 11(1):1-14.

Roth, N.E., M.T. Southerland, J.C. Chaillou, J.H. Vølstad, S.B. Weisberg, H.T. Wilson, D.G. Heimbuch, J.C. Seibel. 1997. *Maryland biological stream survey: Ecological status of non-tidal streams in six basins sampled in 1995.* CBWP-MANTA-EA-97-2. Prepared for Maryland Department of Natural Resources, Chesapeake Bay and Watershed Programs, Annapolis, MD.

Ryder, R.A., S.R. Kerr, K.H. Loftus, and H.A. Regier. 1974. The morphoedaphic index, a fish yield estimator: Review and evaluation. *Journal of Fisheries Research Board of Canada* 31:663-688.

Scheffer, M., M.R. de Redelijkheid, and F. Noppert. 1992. Distribution and dynamics of submerged vegetation in a chain of shallow eutrophic lakes. *Aquatic Botany* 42:199-216.

Schindler, D.W. 1988. Effects of acid rain on freshwater ecosystems. *Science* 239:149-158.

Schindler, D.W., S.E.M. Kasian, and R.H. Hesslein. 1989. Biological impoverishment in lakes of the midwestern and northeastern United States from acid rain. *Environmental Science and Technology* 23:573-580.

Schueler, T.R. 1987. *Controlling urban runoff: A practical manual for planning and designing urban BMPs.* Document No. 87703. Metropolitan Washington Council of Governments, Washington, DC.

Schuster, E.G. and H.R. Zuuring. 1986. Quantifying the unquantifiable: Or have you stopped abusing measurement scale? *Journal of Forestry* 1986:25-30.

Scott, E.M., G.D. Hickman, and A.M. Brown. 1992. *Reservoir vital signs monitoring, 1991 - Fish community results.* TVA/WR-92/5. Tennessee Valley Authority, Water Resources Division, Chattanooga, TN.

Scott, L.C., J.A. Ranasinghe, A.T. Shaughnessy, J. Gerritsen, T.A. Tornatore, R. Newport. 1991. *Long-term benthic monitoring and assessment program for the Maryland portion of Chesapeake Bay: Level 1, Comprehensive report. Vol. 1.* Versar, Inc., Columbia, MD.

Simon, T.P. and J. Lyons. 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. Pages 245-262 In W.S. Davis and T.P. Simon, eds. *Biological assessment and criteria. Tools for water resource planning and decision making.*

CRC Press, Boca Raton, FL.

Smith, V.H. 1983. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* 221:669-671.

Smol, J.P. 1983. Paleophycology of a high Arctic lake near Cape Herschel, Ellesmere Island. *Canadian Journal of Botany* 61:2195-2204.

Smol, J.P. 1986. Chrysophyte analysis, Section 7. In D.F. Charles and D.R. Whitehead, eds. *Paleoecological investigation of recent lake acidification: Methods and project description.* EA-4906, Research Project 2174-10. Electric Power Research Institute, Palo Alto, CA.

Smol, J.P., D.F. Charles, and D.R. Whitehead. 1984. Mallomonadacean microfossils provide evidence of recent lake acidification. *Nature* 307(5952):628-630.

Smol, J.P., B.F. Cumming, and J.C. Kingston. 1990. QA in paleolimnology. In D.R. Hart, ed. *QA guidelines for biology in aquatic environmental protection.* Prepared for Environment Canada, National Water Research Institute, Burlington, Ont.

Sokal, R.R. and F.J. Rohlf. 1969. *Biometry.* W.H. Freeman & Co., San Francisco, CA.

Southerland, M.T., and J.B. Stribling. 1995. Status of biological criteria development and implementation. Pages 81-96 In W.S. Davis and T.P. Simon, eds. *Biological assessment and criteria: Tools for water resource planning and decision making.* CRC Press, Boca Raton, FL.

Southwood, T.R.E. 1977. Habitat, the templet for ecological strategies? *Journal of Animal Ecology* 46:337-65.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical Methods. 7th edition.* The Iowa State University Press, Ames, IA.

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.

Steinhart, C.E., L.4. Schierow, and W.C. Sonzogni. 1982. An environmental quality index for the Great Lakes. *Water Resources Bulletin* 18:1025-1031.

Stemberger, R.S. and J.M. Lazorchak. 1994. Zooplankton assemblage responses to disturbance gradients. *Canadian Journal of Fisheries and Aquatic Sciences* 51:2535-2547.

Stribling, J.B., J. Lathrop-Davis, M.T. Barbour, J.S. White, and E.W. Leppo. 1995. *Evaluation of environmental indicators for the wetlands of Montana: The multimetric approach using benthic macroinvertebrates.* Prepared for the State of Montana, Department of Health and Environmental Sciences, Helena, Montana.

Suter, G.W., II. 1993. *Ecological risk assessment.* Lewis Publishers, Boca Raton, FL.

SYSTAT, Inc. 1992. SYSTAT: Statistics, Version 5.2 edition. SYSTAT, Inc., Evanston, IL.

ter Braak, C.J.F. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179.

——. 1988. CANOCO - A FORTRAN program for canonical community ordination. Agricultural Mathematics Group, Wageningen, The Netherlands.

ter Braak, C.J.F. and S. Juggins. 1993. Weighted averaging partial least squares regression(WA-PLS): An improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269/270:485-502.

Thompson, S.K. 1992. *Sampling.* John Wiley and Sons, New York, NY.

Thornton, K.W. 1990a. Perspectives on reservoir limnology. Pages 1-14 In K.W. Thornton, B.L. Kimmel and F.E. Payne, eds. *Reservoir limnology: Ecological perspectives.* John Wiley and Sons, New York, NY.

——. 1990b. Sedimentary process. Pages 43-69 In K.W. Thornton, B.L. Kimmel, and F.E. Payne, eds. *Reservoir Limnology: Ecological Perspectives.* John Wiley and Sons Inc., New York, NY.

Tilman, D. 1982. *Resource competition and community structure.* Princeton University Press, Princeton, NJ.

Trebitz, A.S., S.A. Nichols, S.R. Carpenter, and R.C. Lathrop. 1993. Patterns of vegetation change in Lake Wingra following a Myriophyllum spicatum decline. *Aquatic Botany* 46:325-340.

Tukey, J. W. 1977. *Exploratory data analysis.* Addison-Wesley, Reading, MA.

TVA. 1994. Tennessee Valley reservoir and stream quality - 1993: Summary of vital signs and use suitability monitoring. Vol. I. *Water Management.* Tennessee Valley Authority, Water Resources Division, Chattanooga, TN.

——. 1995. Aquatic ecological health determinations for TVA reservoirs - 1994: An informal summary of 1994 vital signs monitoring results and ecological health determination methods. *Water Management.* Tennessee Valley Authority, Water Resources Division, Chattanooga, TN.

Uhlmann, D. 1971. Influence of dilution, sinking and grazing rate on phytoplankton populations of hyperfertilized ponds and microecosystems. *Mitt. Internat. Verein. Limonl.* 19:100-124.

Underwood, A.J. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4(1):3-15.

USEPA. 1980a. *Clean lakes program guidance manual.* EPA 440-5-81-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1980b. *National accomplishments in pollution control 1970-1980: Some case histories.* U.S. Environmental Protection Agency, Office of Planning and Management, Program Evaluation Division, Washington, DC.

——. 1983. *Water Quality Standards Handbook.* U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

——. 1984. *Technical support manual: Waterbody surveys and assessments for conducting use attainability analyses. Vol. 111: Lake Systems.* U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

——. 1985a. *Clean lakes program: A review of the first decade.* EPA 440/5-85-033. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

——. 1985b. *Methods for measuring the acute toxicity of effluents to freshwater and marine organisms.* EPA 600/4-85/013. U.S. Environmental Protection Agency, Cincinnati, OH.

——. 1985c. *Technical support document for water quality-based toxics control.* EPA 440/4-85/03. U.S. Environmental Protection Agency, Washington, DC.

——. 1987a. *Handbook of methods for acid deposition studies.* EPA 600/4-87/026. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

——. 1987b. *Report of the national workshop on in-stream biological monitoring and criteria.* In-stream Biol. Criteria Comm. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Environmental Research Laboratory, Region 10, Corvallis, OR.

——. 1987c. *Surface water monitoring: A framework for change.* U.S. Environmental Protection Agency, Office of Water, Office of Policy, Planning and Evaluation, Washington, DC.

——. 1989a. *Overview of selected EPA regulations and guidance affecting POTW management.* EPA 440/69-89/008. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1989b. *Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish.* EPA/440/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1990a. *Biological criteria: National program guidance for surface waters.* EPA 440/5-90-004. U.S. Environmental Protection Agency, Washington, DC.

——. 1990b. *Lake and reservoir restoration guidance manual. 2nd edition.* EPA 440/4-90-006. Prepared by North America Lake Management Society for U.S. Environmental Protection Agency, Washington, DC.

——. 1990c. *Macroinvertebrate field and laboratory methods for evaluating the*

biologicalintegrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Washington, DC.

——. 1990d. *Near coastal program plan for 1990: Estuaries.* EPA 600/4-90/033. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Lab, Narragansett, RI.

——. 1991a. *Biological criteria: Guide to technical literature.* EPA 440/5-91-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1991b. *Biological criteria: Research and regulation. Proceedings of a symposium.* EPA 440/5-91-005. U.S. Environmental Protection Agency, Office of Water, Health and Ecological Criteria Division, Washington, DC.

——. 1991c. *Biological criteria: State development and implementation efforts.* EPA 440/5-91-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1991d. *Design Report for EMAP, the Environmental Monitoring and Assessment Program.* EPA 600-3-91-053. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

——. 1991e. *The Environmental Monitoring and Assessment Program (EMAP) - Surface Waters monitoring research plan, fiscal year 1991.* EPA 600/3-91/022. U.S. Environmental Protection Agency, Corvallis, OR.

——. 1991f. *Technical support document for water quality-based toxics control.* EPA 505/2-90-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1992a. *Draft ecoregions/subecoregions of Florida.* U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Corvallis, OR.

——. 1992b. *Fish field and laboratory methods for evaluating the biological integrity of surface waters.* EPA 600/R-92/111. U.S. Environmental Protection Agency, Cincinnati, OH.

——. 1992c. *Framework for ecological risk assessment.* EPA 630-R-92-001. U.S. Environmental Protection Agency, Washington, DC.

——. 1992d. *Monitoring Guidance of the National Estuary Program.* EPA 842-B-92-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1992e. *Procedures for initiating narrative biological criteria.* EPA 822-B-92-002. U.S.Environmental Protection Agency, Office of Science and Technology, Washington, DC.

——. 1992f. *Sediment classification methods compendium.* EPA 823-4-92-006. U.S. Environmental Protection Agency, Office of Water. Washington, DC.

——. 1993a. *EMAP-Surface waters 1991 pilot report. Review draft.* EPA 620/R-93/003. U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR.

——. 1993b (draft final). *EPA requirements for quality assurance project plans for environmental data operations.* EPA QA/R-5. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC.

——. 1993c. *Guidelines for preparation of the 1994 state water quality assessments (305(b) reports).* EPA 841-B-93-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1993d. *Statistical methods for the analysis of lake water quality trends.* EPA 841-R-93-003. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

——. 1993e. *Virginian Province demonstration report. EMAP-Estuaries: 1990.* EPA 620/R-93/006. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

——. 1994a. *EMAP Surface waters field operations manual for lakes.* U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, NV.

——. 1994b. *Environmental Monitoring and Assessment Program: Integrated quality assurance project plan for the Surface Waters Resource Group, 1994 activities, Rev. 2.00.* EPA 600/X-91/080. U.S. Environmental Protection Agency, Las Vegas, NV.

——. 1994c (draft). *EPA program guidance on*

the TMDL concept. U.S. Environmental Protection Agency, Washington, DC.

——. 1994d. *Guidance for the Data Quality Objectives Process.* EPA 600/R-96/055. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

——. 1994e (draft). *Guidance on implementation of biological criteria.* U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

——. 1994f. *National water quality inventory: 1992 report to Congress.* PA 841-R-94-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1994g. *Periphytic diatom assemblages of high elevation Rocky Mountain lakes: Characterization of reference conditions.* U.S. Environmental Protection Agency, Washington, DC.

——. 1994h. *Statistical summary: EMAP-Estuaries Virginian Province - 1991.* EPA 620/R-94/005. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI.

——. 1994i. *Statistical summary: EMAP - Estuaries Virginian Province - 1992.* EPA 620/R-94/019. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI.

——. 1994j. *USEPA procedures for assessing the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates.* EPA 600/R-94/024. U.S. Environmental Protection Agency, Environmental Research Laboratory, Duluth, MN.

——. 1994k (draft). *The watershed protection approach 1993/94 supplement to 1992 annual report.* U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1995. *Generic quality assurance project plan guidance for programs using community level biological assessment in wadable streams and rivers.* EPA 841-B-95-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1996a. *Biological criteria: Technical guidance for streams and small rivers.* EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

——. 1996b. *Biological criteria: Technical guidance for survey design and statistical evaluation of biosurvey data.* Prepared for U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

——. 1996c. *Guidance for data quality assessment, practical methods for data analysis, EPA QA/G-9, QA96 Version.* EPA 600/R-96/084. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

——. 1996d. Proposed guidelines for ecological risk assessment. U.S. Environmental Protection Agency. *Federal Register* 61(175):47552-47631.

——. 1996e. *Biological assessment methods, biocriteria, and biological indicators: Bibliography of selected technical, policy, and regulatory literature.* EPA 230-B-96-001. U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Washington, DC.

——. 1996f. *Summary of state biological assessment programs for streams and rivers.* EPA 230-R-96-007. U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Washington, DC.

——. 1997a (review draft). *Estuarine/near coastal marine waters bioassessment and biocriteria technical guidance.* Prepared for U.S. Environmental Protection Agency, Office of Water, Washington DC.

——. 1997b. *Field and laboratory methods for macroinvertebrate and habitat assessment of low gradient nontidal streams.* Mid-Atlantic Coastal Streams Workgroup. U.S. Environmental Protection Agency, Environmental Services Division, Region 3, Wheeling, WV.

——. 1997c (draft). *Lake regions of Florida.* U.S. Environmental Protection Agency, Office of Research and Development, Western Ecology Division, Corvallis, OR.

——. 1997d (draft). *Revision to rapid bioassessment protocols for use in streams and rivers: Periphyton, benthic macroinvertebrates, and fish.* EPA 841-D-97-002. U.S. Environmental Protection Agency, Office of Water,

Washington, DC.

Vighi, M. and G. Chiaudani. 1985. A simple method to estimate lake phosphorus concentrations resulting from natural, background, loadings. *Water Research* 19:987-991.

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* 15:392-299.

Vollenweider, R.A. 1968. *Scientific fundamentals of eutrophication of lakes and flowing waters with special reference to phosphorus and nitrogen.* OECD/DAS/CSI/68.27. OECD, Paris.

——. 1975. Input-output models with special reference to the phosphorus loading concept in limnology. *Schweizerische Zeitschrift fur Hydrologie* 37:53-83.

Wallace, J.B., J.W. Grubaugh, and M.R. Whiles. 1996. Biotic indices and stream ecosystem processes: Results from an experimental study. *Ecological Applications* 6(1):140-151.

Warwick, R.M., and K.R. Clarke. 1991. A comparison of some methods for analyzing changes in benthic community structure. *Journal of Marine Biological Association* 71:225-244.

Watson, N.H.F. 1974. Zooplankton of the St. Lawrence Great Lakes: Species composition, distribution, and abundance. *Journal of the Fisheries Research Board of Canada* 31:783-794.

Weaver, M.J., J.J. Magnuson, and M.K. Clayton. 1993. Analyses for differentiating littoral fish assemblages with catch data from multiple sampling gears. *Transactions of the American Fisheries Society* 122:1111-1119.

Weber, S.P., B. Shaw, and S.A. Nichols. 1995. *The aquatic macrophyte communities of eight northern Wisconsin flowages.* Technical Report, College of Natural Resources, University of Wisconsin, Stevens Point, WI.

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

Wetzel, R.G. 1990. Reservoir Ecosystems: Conclusions and speculations. Pages 227-238 In K.W. Thornton, B.L. Kimmel, and F.E. Payne, eds. *Reservoir limnology: Ecological perspectives.* John Wiley and Sons, New York, NY.

Wiederholm, T. 1980. Use of benthos in lake monitoring. *Journal of the Water Pollution Control Federation* 52:537-547.

——. 1984. Responses of aquatic insects to environmental pollution. Pages 508-557 In V.H. Resh and D.M. Rosenberg, eds. *The ecology of aquatic insects.* Praeger, New York, NY.

Wonnacott, T.H. and R.J. Wonnacott. 1969. *Introductory Statistics. 2nd Edition.* John Wiley and Sons, Inc., New York, NY.

Wright, J.C. 1967. Effect of impoundments on productivity, water chemistry, and heat budgets of rivers. Pages 188-199 In *Reservoir Fishery Resources.* American Fisheries Society. Washington, DC.

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.

Yoccoz, 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. and E.T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109-144 In W.S. Davis and T.P. Simon, eds. *Biological Assessment and Criteria: Tools for water resource planning and decision making.* Lewis Publishers, Boca Raton, FL.

Zaret, T.M. 1980. *Predation and freshwater communities.* Yale University Press, New Haven, CT.