5. Using Biological Data As Indicators of Water Quality

Contents

5. Using Biological Data As Indicators of Water Quality

A primary objective of the Clean Water Act is to restore and maintain the chemical, physical, and biological integrity of the Nation's waters (Section 101(a)). In 1991, EPA issued a policy statement regarding the "Use of Biological Assessments and Criteria in the Water Quality Program" (U.S. EPA 1991a). This policy states in part:

> *To help restore and maintain the biological integrity of the Nation's waters, it is the policy of the Environmental Protection Agency that biological surveys shall be fully integrated with toxicity and chemical-specific assessment methods in State water quality programs. EPA recognizes that biological surveys should be used together with whole-effluent and ambient toxicity testing, and chemical-specific analyses to assess attainment/non-attainment of designated aquatic life uses in State water quality standards. EPA also recognizes that each of these three methods can provide a valid assessment of designated aquatic life use impairment.*

The framework described in this chapter is intended to help states and other jurisdictions make better decisions when using biological assessment data and other data to determine impairments and list waterbodies. This framework includes a discussion of the key elements a state's methodology should contain for using biological assessment data and provides information on different methodologies and approaches that can be used to support different water quality determinations.

The references section of this chapter lists key EPA guidance documents that provide technical information to develop and implement effective bioassessment programs for assessing attainment of water quality standards and identification of impaired waters. All of these documents are available through EPA's website: http://www.epa.gov/ost/biocriteria/index.html.

Throughout this chapter, the various key elements and approaches for using biological data are rated from Level 1 to Level 4, to reflect the rigor and level of quality each provides. Level 4 is the highest rigor and quality and provides a relatively high level of certainty in an assessment. Level 1 data describe much less rigorous approaches that are still valuable but present a relatively high degree of uncertainty in the assessments and decisionmaking based on those assessments.

5.1 How Does the State Use Biological Data in the Context of WQS?

Biological assessments, or bioassessments, are an evaluation of the biological condition of a waterbody using biological surveys and other direct measurements of the resident living organisms. Biological assessment data are important for measuring the attainment of WQS for the protection of aquatic life. Biological assessment data can provide a clear picture of whether a waterbody is meeting its designated aquatic life use(s) and can validate whether existing water quality criteria for toxic chemicals, whole-effluent toxicity, physical characteristics, and habitat quality are adequately protecting the designated aquatic life use(s). Biological assessments reflect the total cumulative impact of all stressors over a period of time on a waterbody on the biological community. As such, they are a unique waterbody response measurement, providing information about a waterbody that no other measurement can. For this reason, a state should use biological assessment data as a core indicator for making aquatic life use determinations, as long as the state can provide documentation of the adequate quality and rigor of the key elements of the state's bioassessment program.

Biological data of different varieties can be used by states in assessing the status of waterbodies and in making listing determinations. This section, and the following three sections, primarily focus on the use of bioassessment data. Other types of biological data are discussed briefly in Section 5.3.7.

5.1.1 Using Biological Data in the Context of State Water Quality Standards

As with all assessment methodologies, the better developed the methodology, the better the tool will be in its application. Bioassessment data can be used by states to develop biological criteria, or biocriteria for their WQS. Biocriteria are numeric values or narrative descriptions that are established to protect the biological condition of the aquatic life inhabiting waterbodies of a given designated aquatic life use. Biocriteria can be formally adopted into a state's WQS and used as waterbody response criteria in a regulatory fashion similar to other water quality criteria. To date only a few states have taken this approach. More commonly, biocriteria are developed by states as quantitative endpoints to interpret their narrative biological quality standards. Most states have some form of a narrative biological condition standard formally adopted into their WQS.

Bioassessment data can also be used to establish or refine the aquatic life designated use(s) for a waterbody. By doing so, a state can develop biologically based aquatic life uses that may be more appropriately protective of the biological integrity of the waterbody than simple broad aquatic life use categories (e.g., cold water/warm water), or other uses unrelated to the natural biological quality and variability that a waterbody may be designated for. To make improvements, bioassessment data can be used to refine or tier aquatic life designated uses and to quantitatively define the level of biological condition associated with each tier. With tiered aquatic life uses, a state can set numeric biocriteria that clearly define the upper and lower bounds of biological conditions expected within each aquatic life tier. When approached in this fashion, a state will have aquatic life WQS that clearly and precisely define what the management objective is for a given waterbody and the numeric benchmarks above and below which the objective is or is not achieved.

The more biological assessment data are used to refine the aquatic life uses, to develop biocriteria to protect those uses, and to assess attainment of waterbodies against those standards, the more precise and reliable the assessments will be. Once bioassessment data are used both in designating the aquatic life use and in assessing use attainment, the more confidence a state can have in its decisions regarding waterbody status, the need for listings or the capability to de-list.establish For these reasons, states should use biological assessment data as a core indicator for revising and improving their aquatic life designated uses and criteria as long as the state can provide documentation of the adequate quality and rigor of the key elements of the state's bioassessment program.

A highly developed biological assessment program (like the Level 4 program described below), coupled with biologically based aquatic life uses and numeric biocriteria to protect those uses, constitutes the most effective combination for assessing and managing aquatic life resources, and should be the goal of all states. The more developed the bioassessment and criteria program, the more reliable and appropriate listing decisions will be, and the better and more effective the management efforts can be to restore those waterbodies.

Some states, such as Maine, Ohio, Vermont, Florida, Maryland, Kentucky, and Oregon, have already constructed biological assessment and standards programs for streams and small rivers incorporating tiered aquatic life uses derived from their bioassessment data, and are protecting those uses through numeric or narrative biocriteria. Most other states are developing programs and are at different levels of implementation and at different levels of the quality continuum described below. This guidance attempts to address all programs regardless of the stage of development, as EPA believes defensible assessment and listing decisions can be made at any stage of development or level of quality. However, the lower the level of development of the biological assessment program, the more restricted the assessment and listing decisions will be. The guidance and information described in this chapter attempts to provide recommendations for the full range of assessment and standards programs for streams and small rivers, but should also be applicable to other waterbody types as programs for those waterbodies are developed by states in the future.

Quality levels for aquatic life designated uses within WQS programs

A clear description of how biological data are used to interpret applicable WQS is an important element of a state's assessment methodology. States use a variety of approaches for integrating biological data in their WQS programs. The most common approaches are described below.

- **Level 1:** Minimal WQS program for aquatic life use protection. Possibly only one, or very few, aquatic life uses that apply to all waterbodies that are not defined or interpreted with biological assessment information. No numeric biocriteria and possibly only a generalized narrative biocriterion without defined implementation procedures or translator mechanisms. The majority of the key bioassessment elements (see Section 5.2) are also Level 1.

- **Level 2:** Basic WQS program for aquatic life use protection. A state has aquatic life designated uses with different categories or subcategories related to recreational fisheries, cold water/warm water fisheries, species of concern, or other descriptions. Biological assessment approaches have not yet been developed that define attainment of the aquatic life uses. No tiering of aquatic life uses using the bioassessment data. A biocriteria index has been developed for interpreting the bioassessment data, and the index is used to interpret the state's narrative biocriterion that applies to all waterbodies and all uses, but the index has not been related to the designated aquatic life uses. Most of the key bioassessment elements are at Level 2 or above.

- **Level 3:** Intermediate WQS program for protecting aquatic life use. A state, territory, or authorized tribe has developed bioassessment protocols and has derived a biocriteria index for interpreting the bioassessment data and implementing their narrative criteria. Numeric biocriteria, however, are not yet adopted, but the state has adopted well-defined biologically based designated uses for their waterbodies and also has specific biological descriptions or methods used to define the uses. Rather than applying one narrative biological criteria to all uses, the state may have tiered narrative biocriteria for each use that are measured by bioassessment data. The tiered narrative biocriteria are adopted into their WQS, and well-described implementation procedures or translator mechanisms that define quantitative thresholds are described either in the WQS or in other implementing regulations or policies and procedures documents such as the continuing planning process or consolidated assessment and listing methodology.. Most of the bioassessment key elements are at Level 3 or above.

- **Level 4:** Advanced WQS program for protecting aquatic life use. A state, territory, or authorized tribe has tiered aquatic life uses developed using bioassessment data that reflect a continuum of biological conditions based on regional reference conditions and natural biological integrity. Numeric biocriteria are adopted in the WQS for each tiered aquatic life use, and well-described implementation procedures or translator mechanisms that define quantitative thresholds are included either in the WQS or in other implementing regulations or policies and procedures documents such as the state, territory, or authorized tribe's continuous planning process or consolidated assessment and listing methodology.. A Level 4 program includes a monitoring and assessment program that accurately and precisely assesses the quality of biological conditions in any given waterbody and compares it against the aquatic life use tiers and the biocriteria thresholds. Most of the key bioassessment program elements are at Level 4 or better.

EPA recommends that states use biological assessments to refine, or tier, their aquatic life uses. A tiered approach to classification should articulate appropriate ecological expectations for state waters (e.g., reference conditions) and specify goals for individual waterbodies (e.g., tiered, designated aquatic life uses). Appropriate water quality criteria may then be adopted into state standards to protect the specific designated uses. The water quality criteria and any needed implementation procedures should provide for quantifiable measurement of each specified use. This approach will better protect high-quality waters, provide for more accurate evaluation of effectiveness of controls and best management practices, and enhance public confidence and participation in the WQS-setting and waterbody listing process.

The states of Maine, Vermont, and Ohio have well-described use classification systems in their standards. Currently, most states are using or preparing to use Level 2 or 3 (i.e., they have adopted narrative biocriteria and either have well-developed bioassessment procedures in place or are in the process of validating procedures and decision thresholds). Many states have rigorous bioassessment programs that can serve as a basis for implementing or adopting numeric biocriteria in their water quality standards. As of 1996, all but three states had either developed, or were in the process of developing bioassessment approaches for streams. Thirty states used bioassessment to interpret aquatic life use attainment and 28 states had narrative biocriteria (U.S. EPA 1996a). Today only a few states have numeric biological criteria in their standards (e.g., Maine, Ohio, and Florida for streams; Delaware for estuaries). EPA is updating the status of state and tribal bioassessment programs. Preliminary indications are that state and tribal program growth and sophistication have continued beyond 1996 levels.

5.1.2 Using Bioassessment Data To Determine Impairments in a Level 1 Program

For states with a Level 1 aquatic life standards program and a Level 1 bioassessment program, using bioassessment data to determine and list impairments may be tenuous, with relatively low precision. Level 1 bioassessments can detect severe impairments but have less power to distinguish degrees of impairment or degrees of biological recovery, and therefore tend to provide all-or-nothing results. Because of this, assessments and listing based on Level 1 programs require a significant amount of best professional judgment and scientific interpretation. Highly altered waterbodies may be reasonably listed as impaired, but determining recovery and restoration of the waterbodies, without reference conditions or tiered biological quality thresholds derived from higher level bioassessment data, may be more difficult. Level 1 bioassessment data may best be used to screen waterbodies for further study (i.e., those waterbodies with severe impairments revealed using Level 1 bioassessments should be listed as priority waterbodies for more intensive bioassessments, conducted at Level 2 or above, before actually being listed). In addition, Level 1 bioassessment programs generate data that generally should not be used to conclude a waterbody is attaining WQS, where other aquatic life data show exceedances of 304(a) aquatic life criteria, etc.

5.1.3 Using Bioassessment Data To Determine Impairments in a Level 2 Program

In a state with a Level 2 aquatic life use protection program, assessment and listing decisions using bioassessment data may involve some scientific and regulatory judgment using thresholds for attainment that are derived from the bioassessment data as interpretations of the narrative aquatic life standard. For a state with a Level 2 aquatic life WQS program and having only one general aquatic life use for all or most of its waterbodies, and also having a Level 2 or higher bioassessment program, decisions should be made as to where along a continuum of biological condition the state would identify a threshold level that is considered acceptable, and therefore in attainment of the standard. A quartile approach may be useful for determining attainment thresholds from the bioassessment data (see Section 5.3).

Under such an approach, once quantitative thresholds are established from the bioassessment data, impairment occurs when a bioassessment of a waterbody shows a statistically significant departure of biological condition from the threshold. (Oregon, Kentucky, and North Carolina now use this approach.)

For a state with a Level 2 aquatic life standards program, it may be necessary to document the procedures and rationale for interpreting the narrative standard and the statistical derivation of the decision thresholds that were derived from the bioassessment data. For additional guidance on this, see Section 5.3.

5.1.4 Using Bioassessment Data To Determine Impairments in a Level 3 Program

When a state has not yet adopted numeric biocriteria, but has adopted well-defined biologically based designated uses for its waterbodies with specific biological descriptions or methods that define the biologically based uses, a state may conclude that impairment occurs when a biological assessment shows that the waterbody is not achieving the biologically based designated use in accordance with the state's methods and procedures. Usually these states will have a well-defined quantitative threshold in their regulations or are implementing procedures that define the upper and lower bounds of biological condition acceptable for each aquatic life use tier derived from their bioassessment data. (Vermont and Maine currently use this approach.)

5.1.5 Using Bioassessment Data To Determine Impairments in a Level 4 Program

If a state has adopted numeric biocriteria in its WQS for all waterbodies that numerically define the range of biological condition for each of their aquatic life use tiers, the state may conclude that impairment occurs when the biological condition of a waterbody with a given designated aquatic life use is significantly less than the numeric biocriterion defining the lower end biological condition threshold for that use. With a Level 4 aquatic life standards program, determining waterbody impairments is more definitive because of the quantitative precision the numeric biological thresholds provide. As such, a state should have a high level of confidence in the listing decisions it makes. Likewise, adequate restoration of a waterbody, as a result of a TMDL or other management action, is readily determined when bioassessments in the restored waterbody show that the biological condition has improved to above the lower numeric biological threshold for the particular aquatic life use. In this case, delistings of restored waterbodies should become a straightforward process. With tiered aquatic life uses and numeric biocriteria thresholds defining the highest and lowest acceptable biological conditions within the aquatic life use tier, it is also feasible to track biological degradation in waterbodies. As a waterbody's biological condition is tracked and the waterbody begins to exhibit a condition that is approaching the lower level for the aquatic life use tier, the waterbody may be listed as "threatened" so that necessary actions can be taken to prevent the waterbody from deteriorating any further. (Ohio is one state using this approach.)

5.2 How does the State Use the Key Elements of a State Biological Assessment Approach to Assess and Document Data Quality, Including the Use of Other Data?

The rigor and quality of the biological data are integral to a biological assessment program. Depending on how they are derived, not all biological data are necessarily of equal value for assessing WQS attainment/impairment. The following sections outline the key elements that should be included in a state bioassessment program to ensure that good-quality data are the basis from which reliable attainment/impairment decisions are made as well as decisions regarding sampling and monitoring design (5.2.1), classification of waterbodies (5.2.2), choice of reference conditions (5.2.3), choice of indicator assemblages (5.2.4), choice of field and laboratory protocols (5.2.5), and precision of the biological methods (5..6). Table 5-1 provides a summary of the key elements for the four levels of rigor in conducting bioassessments.

Table 5-1. Defining levels of rigor for key components of a biological assessment

	Level 1 information	Level 2 information	Level 3 information	Level 4 information
Temporal Coverage	No index period is identified and sampling can be scattered throughout the year. This approach is not recommended because it does not help to establish a reliable benchmark reflecting the natural cycles of spawning, recruitment, migration, and mortality.	A seasonal period is identified for convenience in sampling or to match existing programs. Sampling outside the index period may be done, but usually is reserved for emergency response monitoring.	A well-documented seasonal index period(s) is identified or coverage is comprehensive (periodic sampling occurs throughout the year). Index periods are selected based on known ecology to minimize natural variability, maximize gear efficiency, and maximize the information gained on the assemblage (U.S. EPA 1999a). Reference conditions are calibrated for the index period(s).	A well-documented seasonal index period(s) identified; multiple sampling during index period likely; reference conditions calibrated for the index periods.
Natural Classification	No classification of ecosystems. This approach is not recommended, because natural variability is not partitioned to improve the benchmarks for assessment.	Minimal classification limited to individual watersheds or basins. This approach may not recognize stream continuum principles where headwaters differ in function from mainstem. In estuaries and lakes, classification may apply only to portions or embayments.	Classification recognizes geographical or other similar organization. This approach usually is based on landscape features and supplemented with instream or other waterbody characteristics.	Classification based on a combination of landscape features and physical habitat structure of waterbody type. This approach provides the best classification scheme for assessment.
Reference Conditions	No reference condition formally established. Presence and absence of key taxa may constitute the basis for assessment. Professional opinion may be used to support assessment of attainment. This approach may be difficult to defend, especially in listing determinations, than those relying on more formal scientific evidence.	Reference conditions preestablished by professional biologist and based on known ecology of area. A site-specific control or paired watershed approach may be selected for assessment. Regional sites generally are not used at this level.	Reference conditions may be site-specific, but normally are based on watershed-scale assessments. Regional reference sites have likely been developed for the relevant waterbody type and are the basis for assessment and monitoring.	Regional reference conditions are established for each waterbody class and consist of sites and/or other specified means of establishing regional expectations for assessing and monitoring each waterbody.

Table 5-1. Defining levels of rigor for key components of a biological assessment (continued)

	Level 1 information	Level 2 information	Level 3 information	Level 4 information
Indicator Assemblages	Visual observation of biota; poor taxonomic resolution.	One assemblage (usually of an invertebrate); adequate but consistent taxonomic resolution.	Single assemblage collected and analyzed; high data quality and higher taxonomic resolution.	Two or more assemblages collected and analyzed; taxonomic resolution to the lowest practical taxon (mostly genus/species).
Field and Lab Protocols	Documentation of methods is cursory, and the methods usually are not written as SOPs. Methods may be highly variable, relying primarily on best professional judgment.	Methods generally are well documented, but QA/QC may be minimal. Training of biologists may be oriented to new or inexperienced staff only.	Methods are well documented and SOPs are updated periodically. An effective QA/QC program is in place. Training is provided periodically throughout the year for all staff to raise skill levels and enhance interaction and consistency.	Same as level 3, but methods cover multiple assemblages.
Precision of Assessments	Precision of method is low or not measured. Replicate data for estimating precision are not normally available. Capability of indicator to distinguish between human and natural influences is unknown.	Precision of method is moderate. Method is better documented to enable more consistent sampling and higher precision. Capability of indicator to distinguish between human and natural influences has been determined based on studies conducted in other states or regions.	Precision is moderately high, maintained through rigorous methods, training, and periodic refinements or improvements to the implementation of the methods. Capability of indicator to distinguish between human and natural influences has been documented within the state, but is generally based on impaired and reference sites without a gradient of stressors/human influence.	Normally precision is highest, reflective of the most rigorous methods development and QA/QC, with good repeatability in assessments and a high level of confidence in analytical results. Capability of indicator(s) to distinguish between human and natural influences is quite high and based on a gradient of stressors/human influence, which may also include impaired and reference sites.

Table 5-1. Defining levels of rigor for key components of a biological assessment (continued)

	Level 1 information	Level 2 information	Level 3 information	Level 4 information
Thresholds	No formal index or community-based endpoint. Assessment may be based only on presence or absence of targeted or key species. (Some citizen monitoring groups use this level.) Attainment thresholds not specified. .	A biological index or endpoint is established for specific waterbodies, but is likely not calibrated to waterbody classes or statewide application. Index is probably relevant only to a single assemblage. Watershed monitoring can be used where regional reference conditions have not been established. Attainment thresholds may be based on dividing the total possible index or model score into equal parts (e.g., quarters, thirds).	A biological index, or model, has been developed and calibrated for use throughout the state or region for the various classes of a given waterbody type. The index is probably relevant only to a single assemblage, but may or may not be applicable to several states or tribes. Several states conduct assessments using Level 3 information (e.g., Florida, Arizona). Attainment thresholds in these states are based on discriminant model or distribution of candidate reference sites.	Biological index(es), or model(s) for multiassemblages is (are) developed and calibrated for use throughout the state or region. Integrated assessments using the multiple assemblages are possible, thus improving both the assessment and diagnostic aspects of the process. (Ohio and Idaho are examples of states using this approach.) Attainment thresholds in these states are the same as Level 3, except power analysis is used to determine the number of assessment categories.

5.2.1 Index Period or Other Temporal Conditions During Which a State Collects Biological Data

As part of its monitoring program design, a state should clearly identify waterbodies of interest by including them in what is called a target population. In the biological assessment program, such identification typically is done by waterbody and ecoregion type, along with selection of an index period. Because it may not be possible to adequately monitor each waterbody or waterbody type, most monitoring programs collect data from a representative sample of waterbodies in a target population (e.g., EMAP, MBSS). If the monitoring program takes a well-designed sample survey approach or a comprehensive, nonrandom approach (as Ohio does), the state may obtain statistically valid inferences about the condition of the target population.

A state should also document the index period (time of year and duration) when it will sample the condition of the biological community, or specify that it will sample year-round. EPA recommends establishing index periods for a particular season, time of the day, or other window of opportunity when signals are determined to be strong and reliable. Further, EPA recommends that only results from similar index periods be compared when assessing WQS attainment/impairment.

The sampling does not need to occur during the more severe or worst-case conditions. However, understanding the dynamics of how an ecosystem functions at different times enables an investigator to better interpret data from prescribed index periods. The use of an index period also allows a better concentration of sampling during a period when reference conditions have been characterized. A specified index period is used in most state bioassessment programs, although the level of specificity varies.

5.2.2 Natural Classification of Waterbodies

The state should clearly document how it determines the natural variability of its biological data. Classification is useful in evaluating natural variability and distinguishing it from variability resulting from human-induced changes. Classification of waterbodies may be based on waterbody type (e.g., rivers, streams, lakes, wetlands, estuaries), watershed drainage size, ecological regions, elevation, temperature, and other physical features of the landscape and/or waterbody. The number of classifications the state can analyze may be limited by the number of samples taken and the availability of candidate reference sites within each class. EPA recommends classifying more specifically than simply by waterbody type, because it is highly unlikely that the biological condition of any given waterbody type is uniform throughout the entire state. States should list the classification approach(es) used, if any, for all waterbody types monitored.

Ecoregions have been used successfully as primary classification schemes (e.g., in Ohio; see Yoder and Rankin 1995), or as aggregates of ecoregions (for example, in Florida, see Barbour et al. 1996; in Wyoming, see Gerritsen et al. 2000). Ecoregions are areas of relative ecosystem homogeneity (or similar quality) defined by similarity of land form, soil, vegetation, hydrology, and general land use. For example, streams of a given ecoregion are more similar to one another than they are to streams in another ecoregion. In coastal marine areas, large-scale biogeographic provinces are similar in concept to ecoregions. These provinces are characterized based on latitude, climate, and similarities in land form (Holland 1990). For wetlands, classifying by hydrogeomorphic type has been used by many states in evaluating natural variability among wetland types. For a discussion of various methods of classifying wetlands see *Methods for evaluating wetland condition: wetlands classification* (U.S. EPA 2002a).

Ecoregions are not the only classifications of freshwater ecosystems; Hawkins et al. (2000a) point out that the amount of biotic variation related to landscape features is not large, and augmenting classifications based on local habitat features accounts for substantially more variation than the larger-scale environmental features. Some states have used other landscape factors such as elevation and rainfall to classify their waterbodies (Spindler 1996).

5.2.3 Reference Conditions

Reference conditions should be defined to assess a waterbody's ecological health and establish water quality goals. Reference conditions serve as the benchmark of biological integrity against which a waterbody's conditions are compared. The assessment and listing methodology should describe how the state developed its reference conditions and whether they are based on assessment of reference sites or were developed through other means. The assessment and listing methodology may incorporate by reference the state's biological assessment methods and indicate which of the four levels of rigor best characterizes those methods.

State assessment methodologies should clearly document how reference sites are selected and used. A reference condition can be derived from reference sites, an empirical model of expectations that may include knowledge of historical conditions, or a model extrapolated from ecological principles. Normally, actual sites that represent best attainable conditions of a waterbody are used. Generally, EPA recommends the use of a regional reference condition based on an aggregate of sites that allows for broader application in state water resource programs than individual, site-specific conditions (U.S. EPA 1996a).

Where reference sites are not available (e.g., for large ecosystems such as rivers, estuaries, near-shore coastal areas, and in significantly altered systems such as urban centers and cropland areas), a disturbance gradient might be constructed to extrapolate to an appropriate reference condition (Karr and Chu 1999). This approach requires some knowledge of both stressor gradients and biological condition gradients.

Abiotic factors also may be used in selecting candidate reference sites. Use of these factors helps to avoid circularity in defining biological characteristics that become the basis of reference conditions. Candidate reference sites then are evaluated to determine the degree of human modification that has occurred. Factors considered may include human population density and distribution, road density, and the presence of mining, logging, agriculture, urbanization, grazing, or other land uses. This information can be gleaned from GIS data layers, maps, and/or evaluations by resource managers. Candidate sites should be eliminated if they have undergone extensive human modification, especially to riparian zones. Candidate sites can be selected by from probabilistic sampling (*a posteriori* determination) or from targeted sites (*a priori* selection).

Abiotic selection criteria can range from a few chemical criteria to a whole range of factors as discussed above. The rigor of the criteria also varies from very conservative, which may restrict the number of candidate reference sites selected, to very liberal, which may increase the number. Although EPA prefers a conservative approach, states may take different approaches based on their knowledge of the reference sites. EPA suggests using a conservative approach when greater uncertainty exists as to whether the candidate sites are likely to represent the highest quality waters. State methodologies should include documentation of these decisions.

It is very important that the state or tribe verify in the field the current conditions of candidate reference sites. A candidate site should be eliminated if conditions preclude its ability to serve as a reference for high-quality water. A reference site may be natural, minimally impaired (somewhat natural), or best available (altered system).

In summary, when reference sites are used to establish reference conditions, the state should document how (by what criteria) it selects reference sites and how it uses them to define regional reference conditions (e.g., by combining sites in a regional reference condition, or through other approaches such as a paired watershed or upstream/downstream design).

5.2.4 Indicator Assemblages

State assessment and listing methodologies should document both the assemblages used as indicators and the level of taxonomy used to assess them. Biological indicators can be separated into four principal assemblages that are used for assessing WQS attainment/impairment decisions: benthic macroinvertebrates, fish, algae, and aquatic macrophytes. Research is under way on birds and amphibians as candidate assemblages for wetlands, marshes, and headwater and ephemeral streams, as well as other waterbody types (U.S. EPA 2001- MAIA).

Although a single assemblage may be sufficient to make a WQS attainment determination, EPA recommends the use of more than one to enhance confidence in the assessment finding. Each assemblage serves a different function in the aquatic community, has differing habitat ranges and preferences, and may be susceptible to stress in varying manners and degrees. Several states routinely collect and analyze more than one assemblage in their water quality assessments, although different agencies within a state may collect the data.

Benthic macroinvertebrates

The benthic macroinvertebrate assemblage inhabits the sediment or bottom substrates of waterbodies and responds to a wide array of stressors in different ways. Often it is possible to determine the type of stress that has affected a macroinvertebrate community (U.S. EPA 1990a, 1999a). Because many macroinvertebrates have life cycles of a year or more and are relatively immobile, macroinvertebrate community structure generally is a function of past conditions in the specific waterbody. The benthic assemblage is the most common assemblage used in bioassessments for state water quality programs (U.S. EPA 1996b).

Taxonomy: Genus/species taxonomic identification provides the most representative information on ecological relationships and best resolution in sensitivity to impairment (U.S. EPA 1999a). In the Northwest, it is standard practice in bioassessments for all macroinvertebrates in the subsample to be identified to the lowest possible taxonomic level, generally genus or species (Hayslip 1993). However, in some geographical regions of the United States, family-level identification is used more commonly and may be sufficient for assessments (Hawkins and Norris 2000, Bailey et al. 2001). The scientific determination of level of taxonomy should include a knowledge of adaptive radiation within the fauna (i.e., estimates of the number of genera and/or species per family). For example, the higher the ratio of genera to families, the less likely a family-level identification approach will be adequate. Naturally depauperate systems, such as low-gradient streams or oligotrophic lakes, may warrant family-level indices. In lakes and estuaries, biomass measurements are done on taxonomic groupings (e.g., family or genus) as part of bioassessments.

Whatever level of taxonomic rigor is chosen, the state should clearly document it in its assessment and listing methodology. A macroinvertebrate "voucher collection" for each major basin, ecoregion, site class, or other appropriate study unit is recommended highly. Such a collection contains a representative of each taxon and serves as a basin record and reference for checking identifications as well as a providing a data quality check. A senior aquatic taxonomist should check the specimens entered into the collection for accurate identification and, if necessary, send them out to recognized experts for verification. Ideally, the voucher collection should be housed in a museum or university. The state's protocols for establishing and maintaining such a collection also need to be described (or referenced) in its assessment and listing methodology.

Fish

A bioassessment conducted using a fish assemblage requires that all fish species (and size classes), not just game fish, be collected. Fish are good indicators of long-term effects and broad habitat conditions because they are relatively long-lived and mobile (Karr et al. 1986). The fish assemblage also integrates various features of environmental quality, such as food and habitat availability. The physical degradation of streams can cause changes in the food web and the composition and distribution of habitats (Lonzarich 1994). The objective of the fish assemblage portion of any protocol is to collect a representative sample using methods designed to (a) collect all except rare species in the assemblage and (b) provide a measure of the relative abundance of species in the assemblage. Fish assemblages in streams are used more commonly in bioassessments conducted in the eastern and midwestern United States than elsewhere, although some programs in other regions are investigating their utility (U.S. EPA 1996b). Fish assemblages in the streams and rivers of the western United States have been the subject of fewer studies because of their more depauperate nature. Also, fish diversity is low naturally in headwaters and other small streams, as well as in intermittent streams, making fish less viable indicators than other assemblages. Fish are considered important indicators in larger waterbody types (e.g., lakes, estuaries); however, here too, fish assemblages have been used less often than other methods in water quality assessments because of mobility and sampling difficulties in these waterbodies (U.S. EPA 1998, 2000a).

Taxonomy: All fish species should be identified to species level either in the field or the laboratory, depending upon the expertise of the field crew. As with benthic macroinvertebrates, it is important to retain voucher specimens (ideally in a museum or university), and EPA recommends that a taxonomic expert verify and make determinations on any problematic taxa. Additional information on species of interest may be obtained by recording total length and weight. In addition, fish may be examined for external anomalies.

Periphyton or phytoplankton

Algae are primary producers and responsive indicators of environmental change. The periphyton assemblage serves as a good biological indicator in streams and shallow areas because of its naturally high number of species and rapid response to exposure and recovery. Most algal taxa can be identified to species level by experienced biologists, and the tolerance or sensitivity to specific changes in environmental conditions is known for many species (Rott 1991, Dixit et al. 1992). Because periphyton is attached to the substrate, this assemblage integrates physical and chemical disturbances to a stream reach. However, few state environmental agencies have developed protocols for the periphyton assemblage in streams. Recently, Idaho proposed a method to use diatoms in assessing the biointegrity of large Idaho rivers (IDEQ 1999). Phytoplankton is a common assemblage used in lake (U.S. EPA 1998) and estuary (U.S. EPA 2000a) assessments.

Taxonomy: In general, EPA recommends identifying algae to the species level in rivers and wadeable streams because (1) differences among assemblages that may occur at the species level will be better characterized and (2) large differences exist in ecological preferences among algal species within the same genus. However, substantial information can be gained just by identifying algae to the genus level. Although valuable ecological information may be lost, the costs of genus-level analyses are less, especially if inexperienced analysts are involved (U.S. EPA 1999a, Chapter 6).

Identifying diatom genera in assemblages can provide valuable characterizations of biotic integrity and environmental conditions, and may be a good approach when implementing a new program and only an inexperienced analyst is available. As the analyst gains more experience counting, the taxonomic level of the analyses should improve. Eventually, the cost of counting and identifying algae to the species level becomes not much greater than the cost of analysis to the genus level (U.S. EPA 1999a, Chapter 6).

For assessing lakes, EPA recommends sampling the phytoplankton assemblage and counting and identifying cells to the order or genus level. Simplified field and laboratory procedures are possible for measurements based on higher taxonomic levels such as division or order. Identification to the species level is considered supplemental at this time, because it is not clear that the information gained represents a substantial improvement over higher levels of taxonomy (U.S. EPA 1998).

Aquatic macrophytes

Aquatic macrophytes include vascular plants (grasses and forbs) and may be emergent or submergent. Vascular aquatic macrophytes are a vital resource because of their value as extensive primary producers and habitat for fish and waterfowl (U.S. EPA 2000a). As an ecological indicator, this assemblage is most important in estuaries (U.S. EPA 2000a) and wetlands (U.S. EPA 2002). Excessive nutrient loadings lead to prolific phytoplankton and epiphytic macroalgal growth on grasses that outcompete the macrophytes (U.S. EPA 2000a).

Taxonomy: Macrophytes are identified to species level or categorized as emergent, submergent, or floating leaf for purposes of assessment. The taxonomy serves a basis for areal coverage or standing crop biomass analyses. Because submerged aquatic vegetation distributions are specific to a given habitat parameter (i.e. salinity, depth, etc.), they most commonly occur in monotypic stands with some mixed beds (e.g., *Zostera and Ruppia*). In these cases the taxonomy analysis will be less revealing than using abiotic parameters as early warning measures (e.g., light attenuation coefficient, total suspended solids, chorophyll *a*).

Whatever assemblage(s) are used, states should document the rationale for choosing them and include in their assessment and listing methodologies the value and purpose of the assemblage(s) in attainment and listing decisions. If not documented elsewhere, the consolidated assessment and listing method is the appropriate place to document these decisions. The scientific credibility of every assessment depends on how the assemblage is selected.

5.2.5 Field and Laboratory Protocols for Indicator Assemblages

Standardization of laboratory and field methods should be done to establish the validity and reliability of biological data. Whatever assemblage is chosen, the methods for sample collection and laboratory analysis should be documented fully. EPA has published a generic quality assurance (QA) project plan guide for programs using community-level biological assessment in wadeable streams and rivers (see U.S. EPA 1995). The development of standard operating procedures (SOPs) for field and laboratory methods should include an effective quality assurance program with quality control (QC) checks. To minimize bias, reduce error, and maintain a high level of data integrity, the SOPs and QA/QC plan should identify the specific procedures for all aspects of the biological program.

Information on data quality objectives and quality assurance/quality control procedures usually is documented in a separate quality assurance project plan and standard operating procedures document, which can be referenced in the state's general assessment and listing methodology. This information should be available for other parties to use as a reference in developing compatible monitoring projects.

Considerations for macroinvertebrate assemblage sampling and laboratory analysis

Habitat type: The three basic macroinvertebrate habitat types available to sample are (a) artificial substrate, (b) multihabitat, and (c) single habitat. Each type offers sampling advantages and disadvantages. Some choices are more appropriate in some regions of the country than in others. State assessment methodologies should describe which habitat they are sampling and why they have chosen it.

Each of the three habitat types is commonly used throughout the United States to sample aquatic organisms. However, at a minimum the following considerations should be met when selecting which one to sample: (1) adherence to strict quality control procedures to provide consistency and avoid sampling error, (2) reliance in choosing a single habitat type based on its availability and dominance as a productive organism habitat (e.g., cobble in streams, kelp beds in coastal areas, or mud in estuaries), (3) preference for a multihabitat approach in systems with diverse habitat, and (4) use of artificial substrates, which leads to sampling habitat that is natural for the system(s) under study (e.g., rock baskets in cobble streams or lakes, or multiplate Hester Dendy substrates to represent woody debris in streams). A state's assessment methodology should describe which habitat type it is sampling and why it was chosen.

Gear/number of samples: In streams, macroinvertebrate samples usually are taken with either a Surber sampler, Hess sampler, D frame net, or artificial sampler. State methodologies need to specify the gear type to be used. In addition, they need to document the specific characteristics of that gear (e.g., the standard mesh size for nets, if applicable) and the number of samples taken from the habitat type.

For riffle sampling, EPA's Rapid Bioassessment Protocols (RBPs) recommend sampling a minimum of 2 or 3 m^2 of stream bottom (U.S. EPA 1999a). The RBPs recommend compositing (combining) riffle samples into a single sample representative of the stream reach; however, replicates are taken at a proportion of the sites (usually 10% of the sites) to measure sampling precision (U.S. EPA 1999a). Others (Kerans et al. 1992) recommend taking replicate samples at all sites (i.e., taking more than one sample from a stream reach and keeping it separate for taxonomic identification and enumeration). Three to five replicates are commonly used at each site in many research studies (Resh and McElvary 1993), though scientific debate continues on the appropriate number of samples per site/reach. The same approach (i.e., compositing samples with replicates for precision estimates) is recommended for lakes (U.S. EPA 1998) and estuaries (U.S. EPA 2000a) (however, the gear for infaunal sampling consists of grab samplers (e.g., Ponar). Again, state assessment methodologies should document (or reference) their sampling approach.

Subsampling: Bioassessment programs designed to support assessing WQS attainment/impairment decisions rely on timely and cost-effective laboratory processing of benthos samples. Alternatively, analysts sometimes use a predetermined fraction of the field sample for identification and enumeration, called "subsampling," the goal of which is to provide an unbiased representation of a larger sample (Barbour and Gerritsen 1996). Crucial to the reduction of costs and time associated with processing benthic samples, the subsampling procedures developed by Hilsenhoff (1987) and modified by Plafkin et al. (U.S. EPA 1989) have been implemented in many state programs. As an improvement to the mechanics of the technique, Caton (1991) designed a sorting tray and method that allow for rapid isolation of organisms and easy removal of all organisms and debris as well as the elimination of any investigator bias in the process. In Rocky Mountain streams of Wyoming, a 200-organism subsample was found to be optimal in terms of information return for the investment (Gerritsen et al. 1996). Most agencies in the Northwest use either a 300- or 500-organism subsample. However, proportional subsampling may be a viable alternative to fixed-count subsampling, and has been advocated as more accurate in some cases (Courtemanch 1996, Cuffney et al. 2000).

Whatever procedure and number of organisms are subsampled for identification, the state's assessment methodology should clearly document (or reference) the approach used. Precision estimates are important to help interpret results from subsampling efforts. An approach whose precision is considered low indicates lower confidence in the interpretation of data than one whose precision is considered high. For instance, subsampling 100 organisms, as opposed to 300 or 500, will provide less information about taxa richness because the probability of capture is less. However, knowing with precision how taxa richness is estimated from only 100 organisms may, in limited circumstances, still allow an agency to adequately assess the condition of a site. EPA recommends that states test the level of subsampling and establish precision measurements for application to their subsampling levels.

Considerations for fish assemblage sampling and laboratory analysis

Reach length or sampling area: The most recent revision to the Rapid Bioassessment Protocols (U.S. EPA 1999a) describes two acceptable methods for site or reach selection. The first is a fixed distance method such as that used by Ohio EPA (150–200 meters) and Massachusetts DEP (100 meters). The second is a proportional distance method such as that used by the EPA Office of Research and Development's EMAP program (40 times the stream width). In lakes and estuaries, fish sampling is to occur in the littoral zone along the shoreline, or in the pelagic areas for a specified distance or time (U.S. EPA 1998, 2000a).

Field methods: The RBPs recommend electrofishing as a standard sampling technique for use in streams and small areas (U.S. EPA 1993a). Single-pass removal through electrofishing is sufficient to obtain a representation for biological assessments (Bauer and Burton 1993). However, in some cases electrofishing may not be allowed in order to accommodate the presence of endangered species, or may not be practical for other reasons. In these cases, other methods such as snorkeling or seines are used. Snorkeling may miss some smaller, nongame species of fish and therefore is less useful for assemblage-level analysis. Sampling gear used in large waterbodies, such as rivers, lakes, and estuaries, consists of seines, gill nets, and trawls. The method selected should be documented in the assessment methodology.

Considerations for periphyton and phytoplankton assemblage sampling and laboratory analysis

Field methods: The two major categories for periphyton sampling differ as to the type of substrate sampled (natural versus artificial). For an accurate assessment of the assemblage, samples should be collected during periods of stable instream flow.

For natural substrates, samples may be collected from either all available microhabitat types or from a single habitat type. The procedures for sampling from all available microhabitats have been adapted from the Kentucky and Montana protocols (Kentucky DEP 1993, Bahls 1993) and are reported in the latest version of the RBPs. An alternative to compositing several microhabitats is to select a single habitat type that sufficiently characterizes the study reach. The most accurate way to decrease sample variability is to collect from only one type of habitat within a reach and to composite many samples within that habitat (Rosen 1995). If multiple habitats are sampled, the samples should be kept separate, by habitat, for analysis.

Periphyton also can be sampled by collecting from artificial substrates that are placed in aquatic habitats and colonized over a period of time. This procedure is especially useful in larger (nonwadeable) streams, rivers with no riffle areas, wetlands, and lake environments. Kentucky (Kentucky DEP 1993), Florida (Florida DEP 1996), and Oklahoma (Oklahoma CC 1993) have used this technique successfully. Either surface (floating) or benthic (bottom) periphytometers are used and fitted with glass slides, glass rods, clay tiles, plexiglass plates, or similar substrates that occur in the study area. The minimum requirements for periphyton investigations are as described in the Rapid Bioassessment Protocols (U.S. EPA 1999a) for streams. The minimum requirements for phytoplankton investigations are as described in the Lakes Bioassessment and Biocriteria Document (U.S. EPA 1998) and the Estuarine Bioassessment and Biocriteria Document (U.S., EPA 2000a).

Phytoplankton standing stock is estimated by chlorophyll *a* measurements. One approach is to collect three replicate samples at each station at one-half the Secchi depth using a Kemmerer or Van Dorn sampler (U.S. EPA 2000a). Another approach is to collect a depth-integrated sample through the entire photic portion of the water column. The same techniques for phytoplankton collections are applicable to lakes and reservoirs (U.S. EPA 1998) and estuaries and coastal marine waters (U.S. EPA 2000a).

Laboratory analysis: Generally, two types of algae can be identified for assessment: soft algae (nondiatoms) and diatoms. Some states identify the diatoms only. For data on diatom abundance, EPA recommends counting a minimum of 300 to 500 valves or frustules and recording taxa and number counted on bench sheets. Chlorophyll *a* also is analyzed in conjunction with taxonomic identification. Chlorophyll *a* is analyzed fluorometrically or spectrophotometrically following disruption of cells (by grinding) and extraction with acetone (APHA 1992). Once again, documentation of the methods selected by the state and adequate QA/QC procedures to ensure that high quality data are available for making WQS attainment decisions are important.

Considerations for macrophyte assemblage sampling and laboratory analyses

Field methods: For large waterbodies (i.e., large rivers, lakes or reservoirs, wetlands, estuaries or coastal marine areas), areal coverage and distribution of submerged aquatic macrophytes may be estimated from aerial photographs, if available, and ground-truthed at the site (U.S. EPA 2000a). The dominant taxa may be field-identified from vegetation samples collected in shallow waters. Detailed macrophyte monitoring and assessment procedures are included in U.S. EPA (1992), Ferguson and Wood (1994), and Orth et al. (1993). Macrophyte surveys in streams and wetlands usually require site visits to identify the diversity of species and delineate the areal coverage and standing crop biomass.

Laboratory analysis: Most identifications of macrophytes are done in the field. However, voucher collections and samples for biomass determinations are returned to the laboratory.

5.2.6 Precision of Biological Methods

State methodologies should document the capability of selected biological indicators to distinguish between human and natural influences. The value of a biological index lies in its capacity to be used reliably as a signal of environmental degradation. The capability of the indicator to discern differences among sites along a known gradient of disturbance should be examined critically.

The discriminatory capability of the indicator or index is determined by observing its response to environmental stress. The preferred way to do this testing is by establishing a gradient of stress based on nonbiological factors such as contaminant concentrations, physical habitat quality, or land uses (Karr and Chu 1999). Alternatively, binomial discriminatory capability can be determined by comparing biological differences between high-quality reference sites and stressed sites (U.S. EPA 1999b). Engle (2000) and McCormick and Peck (2000) address discriminatory capability for estuarine and freshwater systems, respectively. The document Evaluation Guidelines for Ecological Indicators (U.S. EPA 2000b) and the revised Rapid Bioassessment Protocols (U.S. EPA 1999a) also address this issue.

Whatever assemblage or combination of assemblages is used, the state's assessment and listing methodology should document its value and purpose in making WQS attainment and listing decisions. Fundamental requirements for a biological assessment include understanding the performance of the method (e.g., bias and precision) as well as the effects of natural variability on the method's ability to detect a gradient of environmental impairment. Biological assessments are most useful when the sample is representative of the site examined and the assemblage measured, the data are an accurate reflection of that sample, and the methods distinguish natural and measurement variability (i.e., "noise") from a true environmental effect (i.e., "signal").

Method precision indicates the level of confidence in a site characterization, partly through the liklihood that the assessment could be replicated. Precision in a bioassessment requires consideration of variability resulting from both human and natural sources. Therefore, each step in the sampling and analysis process, including sampling precision, laboratory sorting precision, and taxonomic identification precision (ITFM 1995), should be addressed.

Bias also is an important consideration. Certain sampling gear or procedures, for example, are biased in terms of the types of biota they collect or the types of environmental conditions in which they are most efficient. It is important to understand such sources of bias and how they may interact with natural sources of variation (e.g., flow, season, geomorphology) to influence site characterization. Quality assurance programs encourage the continued documentation of variability to ensure the ability to detect long-term trends. An ongoing quality assurance program also is useful for periodically reevaluating the performance of the indicator and the adequacy of reference conditions.

Two fundamental characteristics of a biological assessment are that samples are representative of the site or assemblage of interest, and that the analytical data accurately reflect the sample. Measurement of precision in these two requirements determines the level of confidence in the assessment. Precision is measured to identify errors and allow inferences to be made about the repeatability of an assessment. Once the precision of a method is known, the likelihood of replicating an assessment can be estimated and the level of confidence in an assessment can be characterized. More specific guidance on documenting measurement error, as well as temporal and spatial variability, is provided below.

Estimating and documenting measurement error

The process of collecting and analyzing biological data has inherent sources of variability that can obscure the discriminatory ability of an indicator. It is important to estimate effects of these sources of variability to ensure that monitoring objectives are addressed satisfactorily and so that data quality and comparability can be documented (Diamond et al. 1996, MDCB 1999). A major source of variability in biological assessments is measurement error. Measurement error is the degree to which one accurately characterizes the sampling unit or site and includes two general components: (1) natural spatial and temporal variability within the sample unit and (2) human or method errors. Natural spatial and temporal variability may lead to differences in precision or bias in an indicator that can result in inaccurate characterization of a site. Human or method errors include inconsistencies in sampling effort across sites, inappropriate use of sampling gear, inaccuracies in laboratory sorting and processing, and misidentified organisms. All of these errors can also result in mischaracterization of a site.

Human and methodological errors are controlled by using standardized and comparable methods, proper training of personnel, and quality assurance procedures (U.S. EPA 1995). Quality assurance procedures include examination of replicate field samples at some subset of the sample units (e.g., 10% of the sites) and reexamination of a proportion of samples by an independent taxonomist. For programs in which multiple field sampling crews are used, it is important to document variability in results caused by personnel. Side-by-side sampling by different field crews is done to document the magnitude of variability as a source of measurement error. Adequate training and similar experience shared across crews helps ensure that this source of error is minimized.

Documenting temporal variability among and within field seasons

It is unlikely in a monitoring program that data can be collected simultaneously from a large number of sites. Instead, sampling may be conducted over several days, weeks, or months. In many cases, indicators are implemented only during a particular season, time of day, or other window of opportunity when their signals are determined to be strong, stable, and reliable, or when stressor influences are expected to be greatest. This optimal time frame, or index period, can reduce sources of error in site characterization resulting from temporal variability (U.S. EPA 1999a). However, because an index period can span several weeks or months, it may be prudent to estimate and document variability within a field season or index period. This process is best accomplished by analyzing multiple samples, collected over time, from reference sites.

Although resource constraints often limit assessments to single index periods, it is useful to understand seasonal effects on an indicator, particularly in cases involving unexpected monitoring demands, such as spills, emergencies, and time-critical decisionmaking. Understanding the seasonal variability and expectations for biological data, using candidate reference sites, could allow data to be used for studies outside the primary index period or for other programmatic needs.

Documenting temporal variability across years

Indicator responses may change over time, even when environmental conditions remain relatively stable. Changes may be due to weather, succession, population cycles, or other natural interannual variations. Available estimates of variability across years should be examined to ensure that the indicator reflects true trends in ecological condition for characteristics that are relevant to the assessment question. To determine interannual stability of an indicator, EPA recommends that monitoring be conducted for several years at stable reference sites with minimal influence of stressors/pollutants.

Documenting spatial variability

Indicator responses to various environmental conditions must be consistent across a site class to enable reliable assessments. Locations within the reporting unit that are known to share similar ecological conditions should exhibit similar indicator results. If spatial variability occurs because of natural regional differences in physiography or habitat (e.g., elevation), it may be necessary to adjust indicator expectations and/or stratify the reporting area into more homogeneous subunits.

Use of a regional reference condition, based on an aggregate of high-quality sites, will account for “natural” spatial variability. This information is then used to determine the discriminatory capability of the indicator. Partitioning the natural variability on a spatial scale (i.e., site classification) ensures that biological response to various stressors will be similar within the site class.

5.2.7 Use of Other Types of Biological Data

Additional types of biological data may be available to, or generated by, a state for determining the status of waterbodies and for making decisions regarding impairments and the need for listings. For example, if a state shares a waterbody with another state, it must consider existing and readily available data from the state that shares the waterbody. Additional data may include fish population data (fisheries surveys, population modeling, impingement/entrainment data), endangered species data, migration data, spawning data, etc. Using these other types of biological data, states may decide to list waterbodies as impaired and initiate TMDLs to manage the impairments. When doing so, states should clearly justify the assessment and impairment decision by documenting how the biological data illustrate a violation of their water quality standards, either the designated uses, the criteria or the antidegradation policy. In many cases, additional types of biological data are interpreted as indicating violations of narrative standards or designated uses. For example, New York State listed 152 miles of the Hudson River, from the Troy Dam south, because of thermal changes from power plants leading to fish mortality occurring during power plant cooling water intake. Based on 24 years of fisheries studies, data indicated that tens to hundreds of millions of eggs, larvae, and juvenile fish of several species were killed per year by large volume, once-through cooling water users. The cumulative impact of multiple once-through cooling facilities substantially reduces the young-of-the-year population for the entire river. Data indicated that this reduction was 25%-79% for spottail shiner, 27%-63% for striped bass, 52%-60% for American shad, 44%-53% for Atlantic tomcod perch, and 33% for bay anchovy. All perennial fresh waters in New York’s WQS (including the Hudson River) have a narrative standard that states these waters shall be suitable for fish propagation and survival.

Additionally, data indicating the presence of introduced, exotic, or invasive species may be used to make a use impairment decision. This is up to the state’s discretion in determining whether a particular species is predominating the waterbody to the extent of impairing aesthetic or recreational uses. However, for making aquatic life use impairment decisions, the approaches outlined above will consider such species’ presence in the calculation of metrics and the associate index. If the biological assessment is sufficiently robust, the impact of introduced, exotic, or invasive species will be shown by the data. The state should also be aware of any threatened and/or endangered species that may reside in or near the waterbody of concern, and may judge the use to be impaired if water quality does not support the species of concern.

5.3 How Does the State Analyze Biological Data to Determine WQS Attainment?

An important step in a bioassessment program is to analyze the data to make WQS attainment decisions and identify any impairment. The establishment of decision thresholds as benchmarks in the water quality standards of the state, or in other implementing regulations or policy or procedures documents, is key to the data analysis. This section describes the overarching strategy for analysis of biological data (5.3.1), the multimetric approach to analyzing data (5.3.2), the combination of metrics and multiple discriminant analysis (5.3.3), and a modeling approach using observed/ expected taxa (5.3.4).

State bioassessment programs should incorporate at least two key elements for analyzing bioassessment data to develop thresholds or decision criteria. These elements are index development and threshold selection. Index development can include single or multiple metrics, discriminant models, or other predictive models of the aquatic community. Thresholds are the "criteria" above which the waterbody is considered to be in attainment. The index should be developed and then verified on independent data sets. Then the attainment threshold should be established and documented. Selecting this threshold, or criterion, is perhaps the most critical element in reporting and documenting attainment status. States typically establish this threshold, and then add other thresholds to distinguish among higher (e.g., outstanding natural resource waters, excellent warmwater habitat, or excellent/good habitat) and lower assessment categories (e.g., limited resource waters, fair/poor/very poor). All thresholds, and the rationales for their selection, should be documented either in the applicable WQS, or in other implementing regulations or policies and procedures documents such as the state, territory, or authorized tribe's continuous planning process or consolidated assessment and listing methodology. More detailed descriptions of the various analytical approaches taken by states appear in Table 5-2, along with the level of information they provide. For estuaries, a different approach is used including reference thresholds for biological effects of contaminants (Long et al. 1995), sediment toxicity, and bottom dissolved oxygen (Schimmel et al. 1994).

5.3.1 Analysis of the Biological Data

Numerous methods are available for analyzing biological indicator data to assess WQS attainment status, including both univariate and multivariate analysis techniques. Bioassessment programs functioning at Levels 3 and 4 (see Table 5-1) have focused on three primary approaches. Sections 5.3.2, 5.3.3, and 5.3.4 go into more detail on each of the three approaches for states and tribes with programs of lower level of rigor to refine and enhance their existing programs.

States do not need to develop their own data analysis methods. Use of existing tools is acceptable and encouraged. Each state does need to document the specific tool it will be using (e.g., a specific multimetric index) and how it will apply this tool. Each state should document the level of information on the indicator index used (whether multimetric or discriminant/predictive model).

EPA recommends that each state establish its analytical threshold based on index values from a statistical distribution of candidate reference sites, or a discriminant model from a range of aquatic life conditions that includes reference conditions. Estimates of variance, such as a standard deviation, as well as power analysis (Fore et al. 1996) can assist in determining how many assessment levels an index may represent.

Regardless of approach, the primary purpose of an analytical threshold is to establish levels of biological quality that can be used for determining WQS attainment status in the aquatic system of interest. States need to carefully document their rationale for selecting thresholds, including thresholds that define gradations in quality or attainment status such as "good/fair/poor" or "full/partial attainment/nonattainment." The threshold should allow for relatively

Table 5-2. Description of component biological variables and predicted direction of variable response to increased perturbation

Variable	Description	Direction
Generalized core variable *Richness (assemblage)*		
1. Taxa richness	Measures the overall variety of the assemblage. Measure of biodiversity.	Decrease
2. Specific family/order richness	Number of taxa in various taxonomic families or orders that are ecologically informative. Examples are number of mayflies (Ephemeroptera) or number of darter species.	Decrease
3. Threatened and endangered species	Usually rare taxa where habitat and ecological viability at risk of depletion and ultimate extinction.	Decrease
4. Rare species	Taxa with low numbers of individuals in population.	Decrease
Composition (assemblage)		
5. Expected biota (observed/expected)	A modeled prediction of taxonomic composition in undisturbed waterbodies within natural site classes. Endpoint approaches 1 for attainment of natural or expected condition.	Decrease
6. Relative abundance	Percent composition of taxonomic groups to total number of individuals in sample, or composed within a particular taxonomic hierarchy. Examples are percent green sunfish (increases with perturbation) and percent stoneflies (Plecoptera), which decrease with perturbation.	Decrease (Increase for certain groups)
7. Compositional redundancy	Measures the change in dominance or redundancy of relative abundance as stressors increase. An example is the increase in percent dominance of one taxon as others are diminished. Evenness and diversity indexes include redundancy components	Increase
8. Keystone taxa	Targeted populations considered crucial to maintenance of assemblage or community. Example is presence of brook trout.	Decrease
9. Alien species	Taxa that are not indigenous to a particular area.	Decrease
Function (population, assemblage, or system)		
10. Reproductive success (population)	Measures some aspect of spawning and nursery success; may be representation of a variety of larval stages. Examples: young-of-year, juvenile index.	Decrease
11. Trophic structure (system)	Measures capacity of ecosystem to support primary and secondary producer/consumers. May comprise several metrics.	Decrease or Increase
12. Guilds (assemblage)	How organisms earn their living. May be trophic, habitat, feeding, or reproductive associations of organisms. May comprise several taxa.	Decrease or Increase
13. Long-lived taxa guild (assemblage)	Support of multi-year life cycle guild indicates extended good water/habitat quality. Examples are reproducing populations of trout and common abundance of stonefly nymphs.	Decrease

Table 5-2. Description of component biological variables and predicted direction of variable response to increased perturbation continued

Variable	Description	Direction
Response to stress (individual, population or assemblage)		
14. Anomalies, disease, deformities, aberrances	Sublethal effects from disease and/or toxicants: e.g., lesions, tumors, or eroded fins in fish; deformed chironomid head capsules; anomalies in striae patterns or frustule shape of diatoms.	Increase
15. Changes in regional species distribution	Range restrictions or expansions of individual species; this is typically depicted on maps comparing where a species was collected historically against current collection locales. Sensitive species typically experience range reductions; invasive aliens and tolerant species expand their ranges.	Increase
16. Other specific response signatures (individual, population or assemblage)	Any measure that has capability of diagnosing stressors. Examples include % aberrant diatoms linked to heavy metal contamination; and measures of individual health (lesions, tumors linked to toxicity).	Characteristic response
Fish-specific (variations) *Richness*		
1. Native taxa richness	Number of different native species, measure of biodiversity.	Decrease
Composition		
2. Morphological composition	Used mostly in the fish assemblage to measure affinity for water column or bottom substrate. An example is % round-bodied sucker.	Increase (?)
3. Habitat preference composition	Measures integrity of habitat to support variety of taxa. An example is number of headwater species.	Decrease
4. Genetic diversity	Genetic variation occurs even when phenotypes appear identical. The use of molecular techniques (e.g., gel electrophoresis to distinguish allozymes) are used to assess genetic diversity.	Decrease
5. Salmonid guilds	Population metrics that characterize various life stages of top carnivores.	Decrease
6. Temperature preference richness (temperature guilds)	Usually related to cold-water forms and those that are stenothermic.	Decrease
Function		
7. Specialized spawners (spawning guilds)	Excludes strategies tolerant to siltation. Measure of ability of stream reach to support a variety of reproductive strategies; affected by toxics, turbidity, sedimentation.	Decrease
8. Specialized feeders (feeding guilds)	Excluding omnivores; measure of trophic/food web complexity of fish assemblage.	Decrease
9. Biomass	Composite weight (biomass). Measure of relative productivity.	Decrease
10. Abundance	Number of individuals (abundance).	Decrease
11. Migration	Daily migrations are typically for feeding and/or predatory avoidance; most seasonal migrations are for reproduction.	Decrease

Table 5-2. Description of component biological variables and predicted direction of variable response to increased perturbation continued

Variable	Description	Direction
12. Anadromous spawning	Fish that spend most of their lives in salt water migrating to fresh water to spawn (e.g., salmon, striped bass, shad).	Decrease
13. Top carnivores	Measure of ability of food chain to support top level; affected by toxics, turbidity.	Decrease
Response to stress		
14. Morbidity	The rate of diseased or affected organisms in a specific location.	Increase
15. Tissue contamination	Measurement of pollutant(s) concentration in living organisms.	Increase
Macroinvertebrate-specific (variations) *Composition*		
1. Diversity indexes	Integrates richness and evenness in mathematical algorithm. An example is the Shannon-Wiener Diversity Index.	Decrease
2. % Dominant taxon	A specific measure of compositional redundancy found in several macroinvertebrate multimetric indices.	Increase
Function		
3. Habit representation (flow/habitat guilds)	Most preferred habit measure is % clingers, which include the insects having a fixed retreat or adaptations for attachment to surfaces in flowing water. Excludes molluscs and other non-insect taxa.	Decrease
4. Voltinism (life cycle guilds)	Measure of long-lived macroinvertebrates (univoltine, life cycles of 1 or more years) or short life cycles (multivoltine, several per year).	Increase or Decrease
Algal-specific (variations) *Composition*		
1. Diversity indexes	As described for macroinvertebrates	Decrease
2. Community similarity	Integrates richness and relative abundance for comparing the composition among sites. Adapted from Whitaker and Fairbanks. Need reference site composition or modeled composition of reference conditions, similar to O/E measure.	Decrease
Function		
3. Autecological affinity (chemical guilds)	Measures the ecological preferences of diatoms and is useful along a stressor gradient. Examples are acidobiontic, alkaliphlic, etc.	Increase or Decrease
4. Biomass	Measures indication of nutrient problem and potential for nuisance algal growth.	Increase

Source: Revised from EPA 1999a and 2000b.

straightforward decisions when biological data are compared against the thresholds to facilitate water quality management decisions. State decisions applying the threshold also need to be documented.

5.3.2 The Multimetric Approach

The most common method of data analysis is use of a multimetric index, which combines several biological variables into a single, unitless index. These variables, or metrics, are characteristics of the biota that change in some predictable way with increased human influence (Barbour et al. 1995). Use of multiple metrics to assess biological conditions maximizes the information available regarding the functions and processes of aquatic communities. For a metric to be of value, it should be (1) ecologically relevant both to the biological assemblage or community under study and to the specified program objectives, and (2) it must be sensitive to stressors (Barbour et al.1995). All metrics that fit these two criteria are potential metrics for consideration. Further analysis of this "universe" will likely eliminate some metrics because of insufficient data or because the range in data is not sufficient to distinguish between natural variability and anthropogenic effects. The analysis should identify the candidate metrics that warrant further consideration (i.e., those that are most informative).

The selected metrics can be used independently or together, depending upon the state's specific program design. A pioneer in the use of multimetric indices for bioassessment, Ohio EPA has developed indices for fish and macroinvertebrate assemblages of its streams and rivers (Yoder and Rankin 1995).

In multimetric analyses, several metrics are calculated and scored from low to high in a common scoring system. Scoring is needed because some metrics respond in different directions to anthropogenic stressors. For example, the abundance of tolerant organisms (density) increases as conditions degrade, whereas the number of intolerant taxa (richness) decreases as conditions degrade. Once the metrics have been scored using a common scale, the scores of all metrics are summed or averaged for a final index score. A multimetric index originally developed for fish assemblages in Midwestern streams (Karr 1981, Karr et al. 1986) has been adapted to streams and rivers throughout the United States and tested in lakes, reservoirs, and estuaries. Because modifications in the index may be appropriate for different regions and among waterbody types, a process for calibrating an index for ecological specificity is required. That process involves two primary steps: (1) selecting candidate metrics and testing for those that should become core metrics, and (2) developing an index by transforming metric values to unitless scores and aggregating as a multimetric index. Examples of generic metrics that are used in various water resource programs are described in Table 5-2. The response of these metrics along a biological gradient provides a means to assess condition to different levels of impairment.

Selection of metrics

Examples of ecologically relevant attributes include components of diversity, identity, composition, function, invasion by exotics, and rare and endangered species. Potential measures relevant to the ecology of the waterbody within the region or state should be evaluated.

Representative metrics from each of four primary categories should be selected: (1) *richness*, which measures for diversity or variety of the assemblage; (2) *composition*, which measures for identity and dominance; (3) *tolerance*, which measures for sensitivity to perturbation; and (4) *trophic measures*, which provide information on feeding strategies and guilds. Karr and Chu (1999) suggest that measures of individual organism health (i.e., anomalies or deformities) be used to supplement other metrics. Karr has expanded this concept to include metrics that are reflective of landscape-level attributes, thus providing a more comprehensive, multimetric approach to ecological assessment (Karr and Chu 1999).

Core metrics should be selected following initial candidate metric screening to identify those that discriminate between "good" and "poor" quality ecological conditions. Metrics that are responsive to specific pollutants or stressors, where the response is well characterized, are most useful as diagnostic tools. Core metrics should be selected to represent diverse aspects of structure, composition, individual health, or processes of the aquatic biota. Together they form the foundation for a sound, integrated analysis of the biotic condition to judge attainment of biological criteria or designated aquatic life uses. The ability of a metric to discriminate between reference conditions and stressed conditions (determined by abiotic, or nonbiological, judgement criteria) is crucial to selecting core metrics. Multiple metrics should be selected to provide a strong and predictable relationship with biological conditions.

Combining metrics into an index

Two basic approaches are used to develop metric expectations and scoring criteria as a basis for index development (Simon and Lyons 1995). The approaches are to use data from reference sites (i.e., composited reference condition) or data from sites representing a range of conditions (i.e., a disturbance gradient). If reference sites are used, there should be a sufficient number of reference sites and samples available to define reference conditions. If data from sites representing a range of conditions (disturbance gradient) are used, they should reflect the entire range of abiotic influence, from minimal human influence to degradation. In either case, a regional reference condition should be developed for each site class (typically termed a bioregion).

Metrics vary in their scale; they may be integers, percentages, or dimensionless numbers. Prior to developing an integrated index for assessing biological condition, it a state should standardize core metrics via a transformation to unitless scores. Recent research has shown that transforming metric values into unitless scores is best done on a numerical scale from 100 to 0 (Hughes et al. 1998, U.S. EPA 1999a). Under such an approach, the data from all sites for each metric, including reference sites, are truncated to the 95th percentile to prevent outliers and extreme values from adversely influencing scoring criteria. (Note: For those metrics that tend to increase in value as the disturbance gradient increases, the 5th percentile is used.) The range from the 95th percentile to the minimum possible value is then subdivided from 100 to 0, with 100 being the maximum score. Finally, the summation of all metric scores is averaged to provide a 100-point scale for the index.

An index provides a means of integrating information from a composite of biological metrics. Aggregation of metrics simplifies management and decisionmaking so that a single-index value is used to determine whether action is needed. The common elements in the development of any analytical assessment tool are use of (1) an initial data set to develop (calibrate) the index and (2) a confirmation data set to test (validate) the index. The initial and confirmation data may be from the same set of biological data, randomly divided, or they may be from two consecutive years of biological data used separately. All sites in each data set are identified by degradation class (e.g., reference versus stressed). To avoid circularity, identification of reference and stressed classes should be made based on nonbiological (abiotic) information, such as the quality of the riparian zone and other habitat features, the presence of known discharges and nonpoint sources, the extent of impervious surface in the watershed, and the extent of land use practices, among other indicators.

Analytical threshold

The population of reference sites normally is used to determine the threshold that separates acceptable from unacceptable biological condition. Reference can also be used to refine aquatic life designated uses by clearly defining the level of biological condition associated with each use as discussed earlier in Section 5.1.1. A population statistic, such as the 25th percentile (Yoder and Rankin 1995, DeShon 1995, Barbour et al. 1996) or 10th percentile (Roth et al. 1997) of the reference sites is a commonly used threshold for multimetric indices. A 25th or 10th percentile is used to recognize that conditions at candidate reference sites are variable, and those at the lower end of the reference scale have a certain level of uncertainty in their quality. This recognition does not mean that 25% of the candidate reference sites are impaired, but that these sites may need closer scrutiny or investigation to assess their condition. The greater the uncertainty in accurately selecting true reference sites, the higher the threshold percentile should be. In addition, precision estimates of the bioassessment methods provide a range of values in which a site condition may not be assessed confidently as either acceptable or unacceptable. In such case, more investigation may be warranted.

5.3.3 Combining Metrics and Multiple Discriminant Analysis

A variety of approaches can be used to combine metrics for an attainment determination. Maine DEP employs a hierarchical decisionmaking technique, which is an example of a discriminant model that uses a variety of biological metrics. It begins with statistical models (linear discriminant analysis) to make an initial prediction of the classification of an unknown sample by comparing it with characteristics of each class identified in the baseline database (Davies et al. 1993). The output of the primary statistical model is a list of probabilities of membership for each of four groups designated as classes A (the highest aquatic life use), B, C, and nonattainment (NA) of Class C. All sites are given an *a priori* aquatic life use of A, B, or C based on waterbody uses and administrative decisions. Stream biologists from Maine DEP assigned a training set of streams to form aquatic life use classes and tested the argument with water chemistry data (see Davies et al. 1993 for description of how ALUS classes were established). Subsequent models are two-way discriminant models to distinguish between a given class and any higher classes as one group, and any lower classes as a second group. The model uses 31 quantitative measures of community structure, including the Hilsenhoff Biotic Index, Generic Species Richness, EPT, and EP values. Monitored test sites are then assigned to one of the four classes based on the probability of that result, and uncertainty is expressed for intermediate sites. The classification can be the basis for management action if a site does not meet its designated use (A, B, or C) or the basis for reclassification to a higher class if the site has improved.

Analytical threshold

The Maine DEP discriminant models predict the membership of a site in one of Maine's aquatic life use classes A, B, or C, or nonattainment (NA). Assignment to a single class must be based on a probability predicted by the submodel of 0.6 or greater. If the model indicates a site is actually in a lower biological class than its designated legislative class, then the site is not attaining its aquatic life use (e.g., the site is listed as Class B, but the discriminant analysis assigns the biota to Class C). If the model fails to assign a class by the required probability, best professional judgment is used.

5.3.4 Modeling Approach Using Observed/Expected Taxa

Another approach, which is used in Oregon and extensively by the U.S. Forest Service, is based on an empirical (statistical) discriminant function model that predicts the aquatic macroinvertebrate fauna that would be expected to occur at a site in the absence of environmental stress (Simpson et al. 1996). A comparison of the invertebrates predicted to occur at the test sites with those actually collected provides a measure of biological impairment at the tested sites. The predicted taxa list also provides a "target" description of the invertebrate community to measure the success of restoration measures. The type of taxa predicted by the model also may provide clues as to the type of impact a sampled site is experiencing. This information can be used to facilitate further investigations and design control/restoration measures. The models are based on a stepwise progression of multivariate and univariate analyses and have been developed for several regions and various habitat types found in lotic systems. Each model is tailored to specific regions (or states) to provide the most accurate predictions for the seasonal and habitat sampled. (See Hawkins et al. 2000b for a more complete description of how this is done.) This approach is being evaluated by EPA. States using this observed/expected approach will need to describe in their methodologies how their model was built and tested for waterbodies.

Analytical threshold

Oregon combines metrics and multivariate models to assess biological condition. In deciding to list or delist impaired waters, Oregon considers aquatic communities (primarily macroinvertebrates) to be impaired if they are found to be at 60% or less of the expected reference community for both multimetric scores and multivariate model scores. Streams with either multimetric scores or multivariate scores between 61% and 75% of expected reference communities are considered to be "streams of concern." Streams with greater than 75% of expected reference communities using either multimetric or multivariate models are considered unimpaired.

5.3.5 Determining Water Quality Standards Attainment

As stated in section 5.1, biological assessment data are important for measuring the attainment of water quality standards for the protection of aquatic life. Biological assessments reflect the total cumulative impact of all stressors over a period of time on a waterbody using the biological community as an indicator. In order for States to best use biological assessment data when determining water quality standards attainment, States should either define their Aquatic Life Uses in their WQS in terms of the expected biological condition for that class and type of waterbody, adopt numeric biocriteria in their WQS, or evaluate the bioassessment data pursuant to well-described implementation procedures or translator mechanisms that define quantitative thresholds that are described either in the WQS or alternatively in other implementing regulations or policies and procedures documents such as the state, territory, or authorized tribe's continuous planning process or consolidated assessment and listing methodology. Each of these approaches should have adequate documentation in the assessment and listing methodology of how the data will be used when addressing all the key elements of a State biological assessment program. Additionally, this documentation should include caveats relating to the known quality and rigor of the data which has been documented earlier in Table 5.1 and Section 5.2.

Although biological data and biological standards can be used to identify water quality impairments, biological data alone, does not usually identify the causes of impairments. Identification of the causes of biological impairments usually requires evaluation of the biological data and other information on watershed conditions. The state, territory, or authorized tribe's assessmesnta nd listing methodology should describe how biological data will be used to determine the cause of an impairment and whether a use is impaired by a pollutant, if this has not already been established in the WQS or other implementing policy or procedure document. For guidance on procedures for identifying causes of biological impairment, see the Stressor Identification guidance document (U.S. EPA 2001).

5.4 References

American Public Health Association (APHA). 1992. Standard methods for the examination of water and wastewater. 18th ed. American Public Health Association, American Water Works Association, and Water Pollution Control Federation. Washington, DC.

Bahls L. 1993. Periphyton bioassessment methods for Montana streams. Water Quality Bureau, Department of Health and Environmental Science, Helena, MT.

Bailey RC, Norris RH, Reynoldson TB. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. J N Am Benthol Soc 20(2):280-286.

Barbour MT, Gerritsen J. 1996. Subsampling of benthic samples: A defense of the fixed organism method. J N Am Benthol Soc 15:386-392.

Barbour MT, Gerritsen J, Griffith GE, Frydenborg R, McCarron E, White JS, Bastian ML. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. J N Am Benthol Soc 15:185-211.

Barbour MT, Stribling JB, Karr JR. 1995. The multimetric approach for establishing biocriteria and measuring biological condition. In: Davis W, Simon T, eds. Biological assessment and criteria: Tools for water resource planning and decisionmaking. Ann Arbor, MI: Lewis Publishers, pp. 63-76.

Bauer SB, Burton TA. 1993. Monitoring protocols to evaluate water quality effects of grazing management on western rangeland streams. U.S. Environmental Protection Agency, Region 10. Seattle, WA. EPA 910-R-93-017.

Caton LW. 1991. Improved subsampling methods for the EPA "rapid bioassessment" benthic protocols. Bull N Am Benthol Soc 8(3):317–319.

Courtemanch DL. 1996. Commentary on the subsampling procedures used for rapid bioassessments. J N Am Benthol Soc 15:381–385.

Cuffney TF, Moulton SR, Carter JL, Short TM. 2000. Abstract. Fixed-count and proportional benthic invertebrate subsampling methods: A comparison of efficacy and cost. Bull N Am Benthol Soc 17(1):144.

Davies SP, Tsomides L, Courtemanch DL, Drummond F. 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.

DeShon JE. 1995. Development and application of the invertebrate community index (ICI). In: Davis WS, Simon TP, eds. Biological assessment and criteria: Tools for water resource planning and decision making. Boca Raton, FL: Lewis Publishers, pp. 217–243.

Diamond JM, Barbour MT, Stribling JB. 1996. Characterizing and comparing bioassessment methods and their results: A perspective. J N Am Benthol Soc 15:713-727.

Dixit SS, Smol JP, Kingston JC, Charles DF. 1992. Diatoms: Powerful indicators of environmental change. Environ Sci Technol 26:23–33.

Engle VD. 2000. Application of the indicator evaluation guidelines to an index of benthic condition for Gulf of Mexico estuaries. Chapter 3. In: Jackson L, Kutz J, Fisher W, eds. Evaluation guidelines for ecological indicators. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. EPA 620-R-99-005.

Ferguson RL, Wood LL. 1994. Rooted vascular aquatic beds in the Albemarle-Pamlico estuarine system. National Marine Fisheries Service, Beaufort, NC. Project No. 94-02.

Florida Department of Environmental Protection (FL DEP). 1996. Standard operating procedures for biological assessment. Florida Department of Environmental Protection, Biology Section. July 1996.

Fore LS, Karr JR, Wisseman RW. 1996. Assessing invertebrate responses to human activities: Evaluating alternative approaches. J N Am Benthol Soc 15(2):212-231.

Gerritsen J, Barbour MT, King K. 2000. Apples, oranges, and ecoregions: On determining pattern in aquatic assemblages. J N Am Benthol Soc 19:487-496.

Gerritsen J, White J, Barbour MT. 1996. Variability of Wyoming stream habitat assessment and biological sampling. Prepared for Wyoming Department of Environmental Quality, Sheridan, WY.

Hawkins CP, Norris RH, Gerritsen J, Hughes RM, Jackson SK, Johnson RK, Stevenson RJ. 2000a. Evaluation of the use of landscape classifications for the prediction of freshwater biota: Synthesis and recommendations. J N Am Benthol Soc 19(3):541-556.

Hawkins CP, Norris RH, Hogue JN, Feminella JW. 2000b. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol Appl 10:1456-1477.

Hawkins CP, Norris RH. 2000. Effects of taxonomic resolution and use of subsets of the fauna on the performance of RIVPACS-type models. In: Wright JF, Sutcliffe DW, Furse MT, eds. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK. pp. 217-228.

Hayslip GA. 1993. EPA Region 10 in-stream biological monitoring handbook (for wadable streams in the Pacific Northwest). Region 10, U.S. Environmental Protection Agency, Environmental Services Division, Seattle, WA. EPA 910-9-92-013.

Hilsenhoff WL. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol 20:31-39.

Holland AF, ed. 1990. Near coastal program plan for 1990: Estuaries. Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, RI. EPA 600-4-90-033.

Hughes RM, Kaufmann PR, Herlihy AT, Kincaid TM, Reynolds L, Larsen DP. 1998. A process for developing and evaluating indices of fish assemblage integrity. Can J Fish Aquat Sci 55:1618-1631.

Idaho Department of Environmental Quality (IDEQ). 1999. Draft V.2, Idaho rivers ecological assessment framework. IDEQ River Bioassessment Team.

Intergovernmental Task Force on Monitoring Water Quality (ITFM). 1995. The strategy for improving water-quality monitoring in the United States: Final report of the Intergovernmental Task Force on Monitoring Water Quality. U.S. Geological Survey, Reston, VA.

Karr JR. 1981. Assessment of biotic integrity using fish communities. Fisheries 66:21-27.

Karr JR, Chu EW. 1999. Restoring life in running waters: Better biological monitoring. Washington, DC: Island Press.

Karr JR, Fausch KD, Angermeier PL, Yant PR, Schlosser IJ. 1986. Assessment of biological integrity in running waters: A method and its rationale. Special Publication 5. Illinois Natural History Survey, Champaign, IL.

Kentucky Department of Environmental Protection. 1993. Methods for assessing biological integrity of surface waters. Division of Water, Kentucky Department of Environmental Protection, Frankfort, KY.

Kerans BL, Karr JR, Ahlstedt SA. 1992. Aquatic invertebrate assemblages: Spatial and temporal differences among sampling protocols. J N Am Benthol Soc 11:377-390.

Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ Manage 19:81-97.

Lonzarich D. 1994. Stream fish communities in Washington: Patterns and processes. PhD dissertation, University of Washington.

McCormick FH, Peck DV. 2000. Application of the indicator evaluation guidelines to a multimetric indicator of ecological condition based on stream fish assemblage. Chapter 4. In: Jackson L, Kurtz J, Fisher W, eds. Evaluation guidelines for ecological indicators. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. p. 107. EPA 620-R-99-005.

Methods and Data Comparability Board (MDCB). 1999. Towards a definition of a performance-based approach to laboratory methods. Version 5.2. In: http://wwwdwimdn.er.usgs.gov/ pmethods.

Oklahoma Conservation Commission (OCC). 1993. Development of rapid bioassessment protocols for Oklahoma utilizing characteristics of the diatom community. Oklahoma Conservation Commission, Oklahoma City, OK.

Orth RJ, Nowak JF, Anderson GF, Whiting JR. 1993. Distribution of submerged aquatic vegetation in the Chesapeake Bay and its tributaries and Chincoteague Bay—1992. Prepared by Virginia Institute of Marine Science, Gloucester Point, Virginia for the U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis, MD.

Resh VH, McElvary EP. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg DM, Resh VH, eds. Freshwater biomonitoring and benthic macroinvertebrates. New York: Chapman and Hall, pp. 159-194.

Rosen BH. 1995. Use of periphyton in the development of biocriteria. In: Davis WS, Simon TP, eds. Biological assessment and criteria. Tools for water resource planning and decision making. Boca Raton, FL: Lewis Publishers, pp. 209-215.

Roth NE, Southerland MT, Chaillou JC, Volstad JH, Weisberg SB, Wilson HT, Heimbuch DG, Seibel JC. 1997. Maryland biological stream survey: Ecological status of non-tidal streams in six basins sampled in 1995. Report no. CBWP-MANTA-EA-97-2. Maryland Department of Natural Resources, Annapolis, MD.

Rott E. 1991. Methodological aspects and perspectives in the use of periphyton for monitoring and protecting rivers. In: Whitton BA, Rott E, Friedrich G, eds. Use of algae for monitoring rivers. Institut fur Botanik, University of Innsbruck, Austria.

Schimmel SC, Melzian BD, Campbell DE, Stubel CJ, Benyi SJ, Rosen JS, Buffum HW. 1994. Statistical summary: EMAP-Estuaries Virginian Province—1991. Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, RI. EPA 620-R-94-005.

Simon TP, Lyons J. 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. In: Davis WS, Simon TP, eds. Biological assessment and criteria. Tools for water resource planning and decision making. Boca Raton, FL: Lewis Publishers, pp. 245-262.

Simpson J, Norris R, Barmuta L, Blackman P. 1996. Australian River assessment system: National river health program predictive model manual. http.//ausrivas.canberra.au.

Spindler P. 1996. Using ecoregions for explaining macroinvertebrate community distribution among reference sites in Arizona. 1992. Arizona Department of Environmental Quality, Hydrologic Support and Assessment Section, Flagstaff, AZ.

U.S. Environmental Protection Agency (U.S. EPA). 1989. Rapid bioassessment protocols for use in streams and rivers. Benthic macroinvertebrates and fish. Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughes RM. Office of Water Regulations and Standards, Washington, DC. EPA 440-4-89-001.

U.S. EPA. 1990a. Biological criteria: National program guidance for surface waters. Office of Water Regulations and Standards, Washington, DC. EPA 440-5-90-004.

U.S. EPA. 1991a. Policy on the use of biological assessments and criteria in the water quality program. Attachment A of Memorandum from Tudor Davies, Director. Office of Science and Technology to Water Management Division Directors, Regions I–X.

U.S. EPA. 1992. Framework for ecological risk assessment. Washington, DC. EPA 630-R-92-001.

U.S. EPA. 1993a. EPA region 10 in-stream biological monitoring handbook (for wadable streams in the Pacific Northwest). Gretchen Hayslip. Region 10, Environmental Services Division, Seattle, WA. EPA 910-9-92-013.

U.S. EPA. 1995. Generic quality assurance project plan guidance for programs using community-level biological assessment in wadable streams and rivers. Office of Water, Washington, DC. EPA 841-B-95-004.

U.S. EPA. 1996a. Biological criteria: Technical guidance for streams and small rivers. Gibson G, Barbour M, Stribling J, Gerritsen J, Karr J. Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC. EPA 822-B-96-001.

U.S. EPA. 1996b. Summary of state biological assessment programs for streams and rivers. Davis WS, Snyder BD, Stribling JB, Stoughton C. Office of Planning, Policy, and Evaluation, Washington, DC. EPA 230-R-96-007.

U.S. EPA. 1998. Lake and reservoir bioassessment and biocriteria Gerritsen J, Carlson R, Charles DL, Dycus D, Faulkner C, Gibson GR, Kennedy RH, Markowitz SA. Technical guidance document. Office of Water, Washington, DC. EPA 841-B-98-007.

U.S. EPA. 1999a. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd ed. Barbour MT, Gerritsen J, Snyder BD, Stribling JB. Office of Water, Washington, DC. EPA 841-B-99-002.

U.S. EPA. 1999b. Quantifying physical habitat in wadeable streams. Kaufmann PR, Levine P, Robison EG, Seeliger C, Peck DV. Office of Research and Development, Washington, DC. EPA/620/R-99/003.

U.S. EPA. 2000a. Estuarine and coastal marine waters: Bioassessment and biocriteria technical guidance. Gibson GR, Bowman ML, Gerritsen J, Snyder BD. Office of Water, Washington, DC. EPA 822-B-00-024.

U.S. EPA. 2000b. Evaluation guidelines for ecological indicators. Jackson LE, Kurtz JC, Fisher WS. Office of Research and Development, Research Triangle Park, NC. EPA 620-R-99-005.

U.S. EPA. 2001. Stressor identification guidance document. Office of Water, Washington, DC. EPA 822-B-00-025.

U.S. EPA. 2002a. Methods for evaluating wetland condition: Wetlands classification. Office of Water; Washington, DC. EPA 822-R-02-017.

U.S. EPA. 2002b. Methods for evaluating wetland condition: Using vegetation to assess environmental conditions in wetlands. Office of Water; Washington, DC. EPA 822-R-02-020.

Yoder CO, Rankin ET. 1995. Biological criteria program development and implementation in Ohio. In: Davis WS, Simon TP, eds. Biological assessment and criteria: Tools for water resource planning and decision making. Boca Raton, FL: Lewis Publishers, pp. 109-144.