9 BIOLOGICAL DATA ANALYSIS

States are faced with the challenge of not only developing tools that are both appropriate and cost-effective (Barbour 1997), but also the ability to translate scientific data for making sound management decisions regarding the water resource. The approach to analysis of biological (and other ecological) data should be straightforward to facilitate a translation for management application. This is not meant to reduce the rigor of data analysis but to ensure its place in making crucial decisions regarding the protection, mitigation, and management of the nation's aquatic resources. In fact, biological monitoring should combine biological insight with statistical power (Karr 1987). Karr and Chu (1999) state that a knowledge of regional biology and natural history (not a search for statistical relationships and significance) should drive both sampling design and analytical protocol.

A framework for bioassessment can be either an *a priori* or *a posteriori* approach to classifying sites and establishing reference condition. To provide a broad comparison of the 2 approaches, it is assumed that candidate reference sites are available from a wide distribution of streams. In the first stage, data collection is conducted at a range of reference sites (and non-reference or test sites) regardless of the approach. The differentiation of site classes into more homogeneous groups or classes may be based initially on *a priori* physicochemical or biogeographical attributes, or solely on *a posteriori* analysis of biology (Stage 2 as illustrated in Figure 9-1). Analysts who use multimetric indices tend to use *a priori* classification; and analysts who use one of the multivariate approaches tend to use *a posteriori*, multivariate classification. However, there is no reason *a priori* classification could not be used with multivariate assessments, and vice-versa.

Two data analysis strategies have been debated in scientific circles (Norris 1995, Gerritsen 1995) over the past few years — the multimetric approach as implemented by most water resource agencies in the United States (Davis et al. 1996), and a multivariate approach advocated by several water resource agencies in Europe and Australia (Wright et al. 1993, Norris and Georges 1993). The contrast and similarity of these 2 approaches are illustrated by Figure 9-1 in a 5-stage generic process of bioassessment development. While there are many forms of multivariate analyses, the 2 most common multivariate approaches are the Benthic Assessment of Sediment (BEAST) used in parts of Canada, the River Invertebrate Prediction and Classification System (RIVPACS) used in parts of England and its derivation, the Australian River Assessment System (AusRivAS) used in Australia.

The development of the reference condition from the range of reference sites (Figure 9-1, Stage 4), is formulated by a suite of biological metrics in the multimetric approach whereas the species composition data are the basis for models used in the multivariate approach. However, both multivariate techniques differ in their probability models. Once the reference condition is established, which serves as a benchmark for assessment, the final stage becomes the basis for the assessment and monitoring program. In this fifth and final stage (Figure 9-1), the multimetric approach uses established percentiles of the population distribution of the reference sites for the metrics to discriminate between impaired and minimally impaired conditions. Where a dose/response relationship can be established from sites having a gradient of conditions (reference sites unknown), an upper percentile of the metric is used to partition metric values into condition ranges. The BEAST multivariate technique uses a probability model based on taxa ordination space

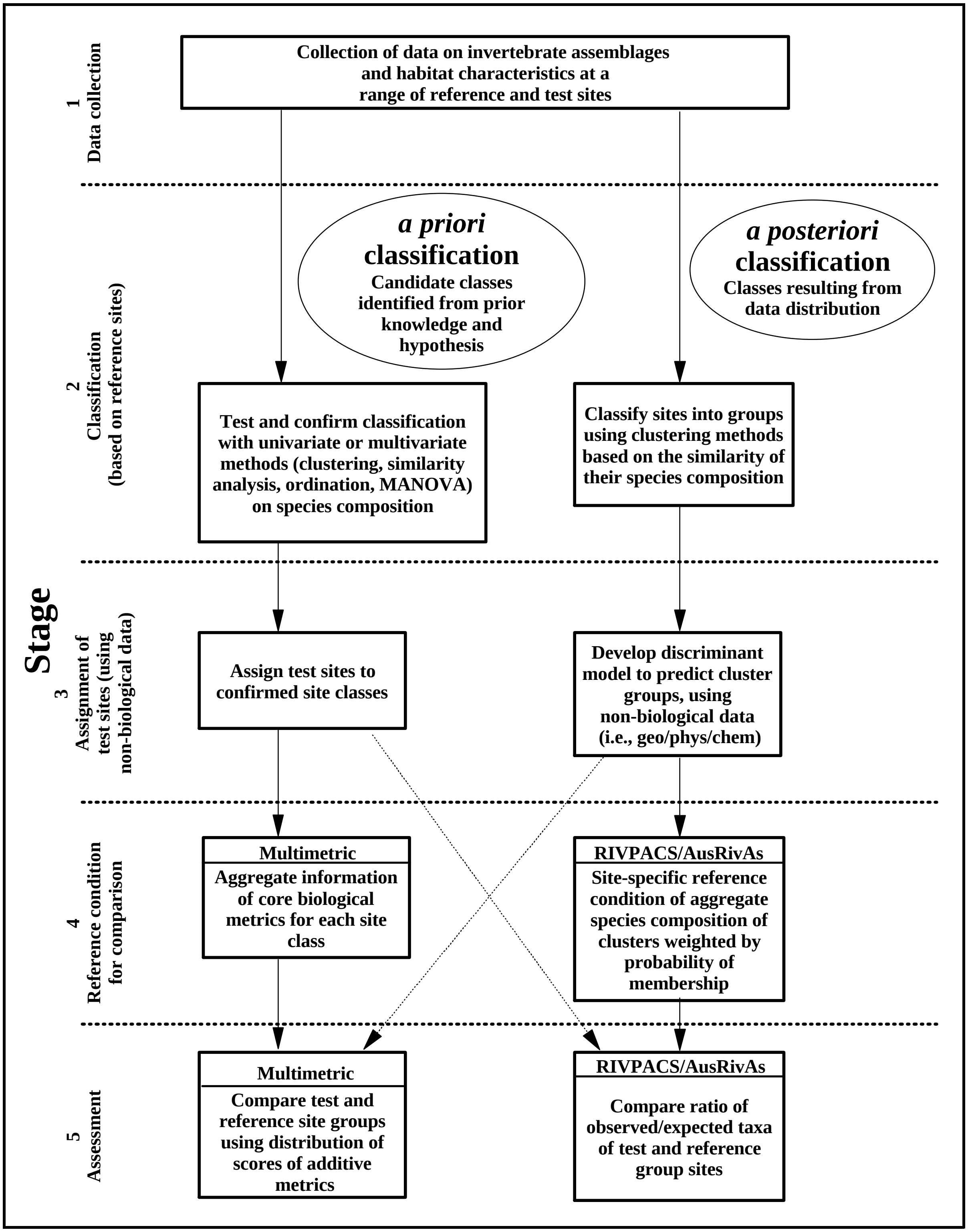

Figure 9-1. Comparison of the developmental process for the multimetric and multivariate approaches to biological data analysis (patterned after ideas based on Reynoldson, Rosenberg, and Resh, unpublished data).

and the “best fit” of the test site(s) to the probability ellipses constructed around the reference site classes (Reynoldson et al. 1995). The AusRivAS/RIVPACS model calculates the probability of expected taxa occurrence from the weighted reference site groups.

The bioassessment program in Maine is an example of a state that uses a multivariate analysis in the form of discriminant function models and applies these models to a variety of metrics. Decisions are made with regard to attainment (or non-attainment) of designated aquatic life uses. The approach used by Maine is based on characteristics of both the multivariate and multimetric approach. In this chapter, only the multimetric approach to biological data analysis is discussed in detail. Discussion of multivariate approaches is restricted to the overview of the discriminant function model used by Maine and the AusRivAS/RIVPACS technique.

9.1 THE MULTIMETRIC APPROACH

Performing data analysis for the Rapid Bioassessment Protocols (RBPs) or any other multimetric approach typically involves 2 phases: (1) Selection and calibration of the metrics and subsequent aggregation into an index according to homogenous site classes; and (2) assessment of biological condition at sites and judgment of impairment. The first phase is a developmental process and is only necessary as biological programs are being implemented. This process is essentially the characterizing of reference conditions that will form the basis for assessment. It is well-documented (Davis and Simon 1995, Gibson et al. 1996, Barbour et al. 1996b) and is summarized here. Developing the framework for reference conditions (i.e., background or natural conditions) is a process that is applicable to non-biological (i.e., physical and chemical) monitoring as well (Karr 1993, Barbour et al. 1996a).

The actual assessment of biological condition is ongoing and becomes cost-effective once Phase 1 has been completed, and the thresholds for determining attainment or non-attainment (impairment) have been established. The establishment of reference conditions (through actual sites or other means) is crucial to the determination of metric and index thresholds. These thresholds are essential elements in performing the assessment. It is possible that reference conditions (and resultant thresholds) will need to be established on a seasonal basis to accommodate year-round sampling and assessment. If data are available, a dose/response relationship between specific or cumulative stressors and biological condition will provide information on a gradient response, which can be a powerful means of determining impairment thresholds.

The 2 phases in data analysis for the multimetric approach are discussed separately in the following section. The reader is referred to supporting documentation cited throughout for more in-depth discussion of the concepts of multimetric assessment.

9.1.1 Metric Selection, Calibration, And Aggregation Into an Index

The development of biological indicators as part of a bioassessment program and as a framework for biocriteria is an iterative process where the site classification and metric selections are revisited at various stages of the analysis. However, once this process has been completed and the various technical issues have been addressed, continued monitoring becomes cost-effective. The conceptual process for proceeding from measurements to indicators to assessment of condition is illustrated in Figure 9-2 (Paulsen et al. 1991; Barbour et al., 1995; Gibson et al., 1996).

Index development outlined in this section requires a stream classification framework to partition natural variability and in which metrics are evaluated for scientific validity. The core metrics representing various attributes of the targeted aquatic assemblage can be either aggregated into an index or retained as individual measures.

Step 1. Classify the Stream Resource

Classification is the partitioning of natural variability into groups or classes of stream sites that are relatively homogeneous with regard to physical, chemical, and biological attributes.

Site classification provides a framework for organizing and interpreting natural variability among streams; ecoregions are a principal example of a classification framework (Omernik 1995). However, classification variables can be at a coarser or finer scale than ecoregions or subecoregions, such as elevation and drainage area. Elevation was determined to be an important classification variable in montane regions of the country (Barbour et al. 1992, 1994, Spindler 1996). Spindler (1996) found that benthic data adhered more closely to elevation than to ecoregions. Ohio EPA (1987) found that stream size (or drainage area) was a covariate and not a determinant of stream classes. The number of fish species increased with stream size (Figure 9-3).

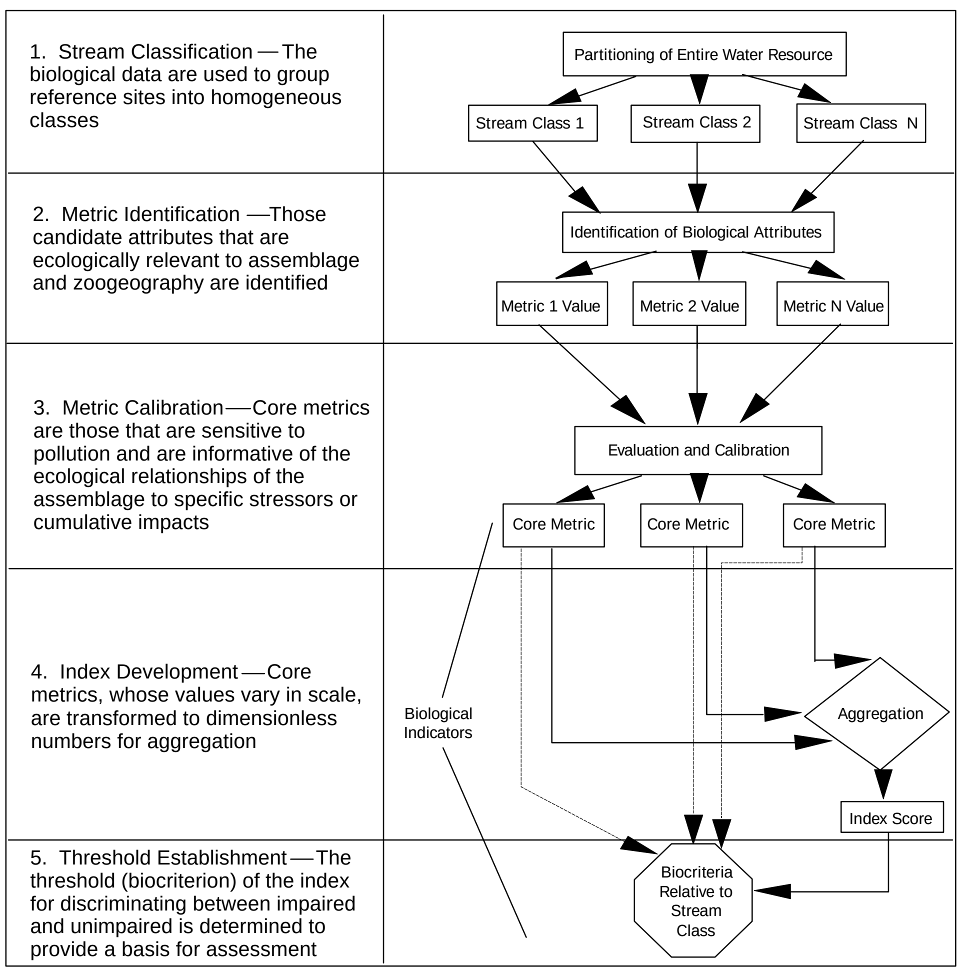

Figure 9-2. Process for developing assessment thresholds (modified from Paulsen et al. [1991] and Barbour et al. [1995]). Dotted lines indicate use of individual metric information to aid in the evaluation of biological condition and cause of impairment.

Classification is best accomplished with reference sites that reflect the most natural and representative condition of the region. Candidate reference sites that are based on minimally degraded physical habitat and water chemistry are used as the basis for stream classification. Quantitative criteria for reference sites aid in a consistent framework for selection. An example of quantitative criteria for identifying reference sites in a statewide study for Maryland (Roth et al., 1997) is presented below (a reference site must meet all 12 criteria):

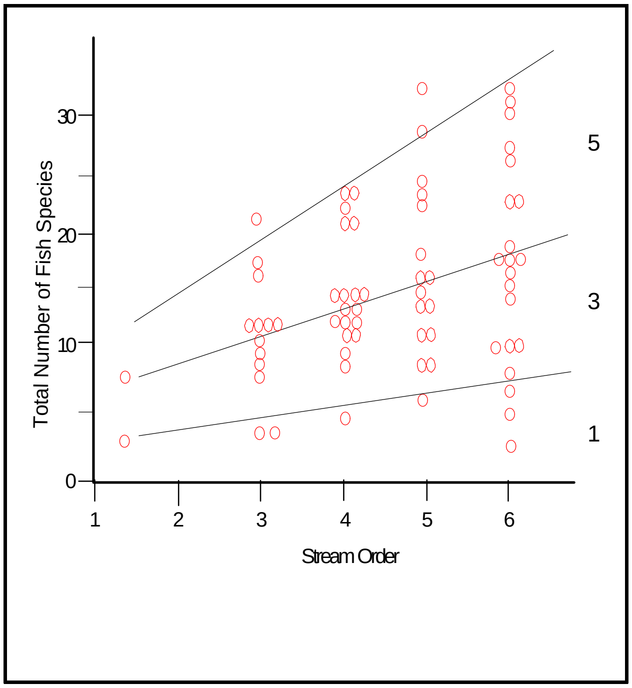

Figure 9-3. Species richness versus stream size (taken from Fausch et al. 1984).

1. pH ≥ 6; if blackwater stream, then pH < 6 and DOC ≥ 8 mg/l
2. ANC ≥ 50 µeq/l
3. DO ≥ 4 ppm
4. nitrate ≤ 300 µeq/l
5. urban land use ≤ 20% of catchment area
6. forest land use ≥ 25% of catchment area
7. remoteness rating: optimal or suboptimal
8. aesthetics rating: optimal or suboptimal
9. instream habitat rating: optimal or suboptimal
10. riparian buffer width ≥ 15 m
11. no channelization
12. no point source discharges

Sites are initially classified according to distinctive geographic, physical, or chemical attributes. Refinement and confirmation of the site classes is accomplished using the biological data (Figure 9-4). Classification is used to determine whether the sampled sites should be placed into specific groups that will minimize variance *within* groups and maximize variance *among* groups. As an example, 3 ecoregionally based delineations (bioregions) were effective at partitioning the variability among reference sites in Florida (Figure 9-5).

Components of Step 1 include:

! Identify classification alternatives. Use physical and chemical parameters that are minimally influenced by human activity to identify classes for testing.

! Identify candidate reference sites that meet the criteria of most "natural" conditions of region.

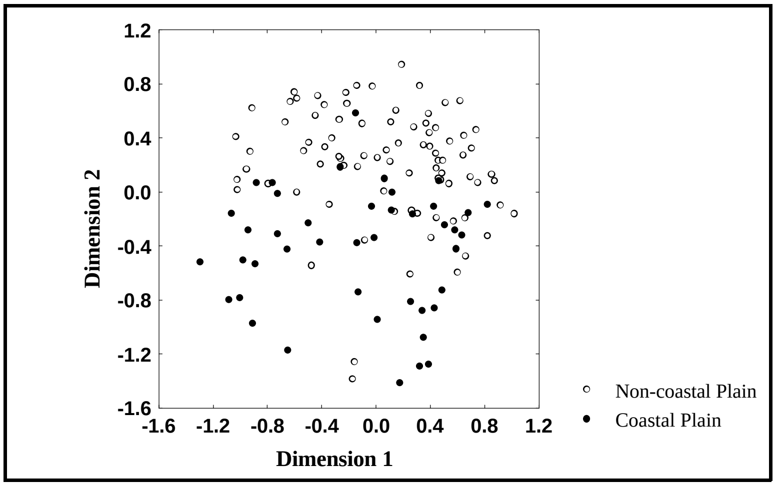

Figure 9-4. Results of mutivariate ordination on benthic macroinvertebrate data from "least impaired" streams from Maryland, using nonmetric multidimensional scaling (NMDS) of Bray-Curtis dissimilarity coefficients.

! Test alternative classification schemes of subecoregion, stream type, elevation, etc., using multiple metric and non-metric biological characteristics including measures such as species composition and EPT taxa (Figure 9-5). Several multivariate classification and ordination methods, and univariate descriptions and tests, can assist in this process (Reckhow and Warren-Hicks 1996, Gerritsen 1995, 1996, Barbour et al. 1996b).

! Evaluate classification alternatives and determine best distinction into groups or classes using biological data. By confirming resource classification based on biological data, site classes are identified that adequately partition variability.

Step 2. Identify Potential Measures For Each Assemblage

> A *metric* is a characteristic of the biota that changes in some predictable way with increased human influence.

Metrics allow the investigator to use meaningful indicator attributes in assessing the status of assemblages and communities in response to perturbation. The definition of a metric is a characteristic of the biota that changes in some predictable way with increased human influence (Barbour et al. 1995). For a metric to be useful, it must have the following technical attributes: (1) ecologically relevant to the biological assemblage or community under study and to the specified program objectives; (2) sensitive to stressors and provides a response that can be discriminated from natural variation. The purpose of using multiple metrics to assess biological condition is to aggregate and convey the information available regarding the elements and processes of aquatic communities.

All metrics that have ecological relevance to the assemblage under study and that respond to the targeted stressors are potential metrics for testing.

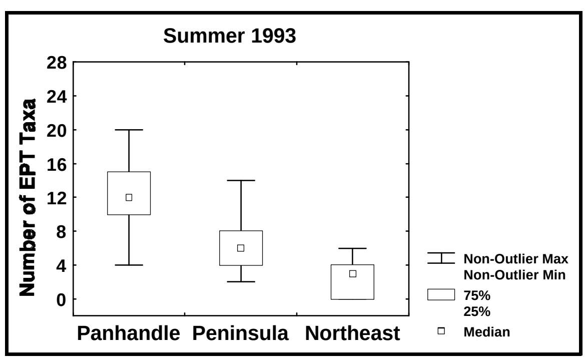

Figure 9-5. An example of a metric that illustrates classification of reference stream sites in Florida into bioregions.

From this "universe" of metrics, some will be eliminated because of insufficient data or because the range of values is not sufficient for discrimination between natural variability and anthropogenic effects. This step is to identify the candidate metrics that are most informative, and therefore, warrant further analysis.

The potential measures that are relevant to the ecology of streams within the region or state should be selected to ensure that various aspects of the elements and processes of the aquatic assemblage are addressed. Representative metrics should be selected from each of 4 primary categories: (1) richness measures for diversity or variety of the assemblage; (2) composition measures for identity and dominance; (3) tolerance measures that represent sensitivity to perturbation; and (4) trophic or habit measures for information on feeding strategies and guilds. Karr and Chu (1999) suggest that measures of individual health 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 et al. 1987). See Table 9-1 for potential metrics that have been useful for periphyton, benthic macroinvertebrates, and fish are summarized in Chapters 6, 7, and 8, respectively.

Components of Step 2 include:

- Review value ranges of potential metrics, and eliminate those that have too many zero values in the population of reference sites to calculate the metric at a large enough proportion of sites.

- Use descriptive statistics (central tendency, range, distribution, outliers) to characterize metric performance within the population of reference sites of each site class.

- Eliminate metrics that have too high variability in the reference site population that they can not discriminate among sites of different condition. The potential for each measure is based on possessing enough information and a specific range of variability to discriminate among site classes and biological condition.

Step 3. Select Robust Measures

Core metrics are those that will discriminate between good and poor quality ecological conditions. It is important to understand the effects of various stressors on the behavior of specific metrics. Metrics that are responsive to specific pollutants or stressors, where the response is well-characterized, are most useful as a diagnostic tool. Core metrics are those that 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.

> The ability of a biological metric to *discriminate* between “known” reference conditions and “known” stressed conditions (defined by physical and chemical characteristics) is crucial in the selection of *core metrics* for future assessments.

Discriminatory ability of biological metrics can be evaluated by comparing the distribution of each metric at a set of reference sites with the distribution of metrics from a set of “known” stressed sites (defined by physical and chemical characteristics) within each site class. If there is minimal or no overlap between the distributions, then the metric can be considered to be a strong discriminator between reference and impaired conditions (Figure 9-6).

As was done with candidate reference sites (see Step 1), criteria are established to identify a population of "known" stressed sites based on physical and chemical measures of degradation. An example set of criteria established for Maryland streams for which failure indicated a stressed site for testing discriminatory power (Roth et al. 1997) is as follows:

- ! pH $\leq$ 5 and ANC $\leq$ 0 µeq/l (except for blackwater streams, DOC $\geq$ 8 mg/l)
- ! DO $\leq$ 2 ppm
- ! nitrate > 500 µM/l and DO < 3 ppm
- ! instream habitat rating poor and urban land use > 50% of catchment area
- ! instream habitat rating poor and bank stability rating poor
- ! instream habitat rating poor and channel alteration rating poor

Table 9-1. Some potential metrics for periphyton, benthic macroinvertebrates, and fish that could be considered for streams. Redundancy can be evaluated during the calibration phase to eliminate overlapping metrics.

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

Step 3 can be separated into 2 elements that correspond to discrimination of core metrics (element 1) and determination of biological/physicochemical associations (element 2). Components of these elements include:

Element 1 *Select core measures that are best for discriminating degraded condition*

! Good (reference) designations of stream sites should be based on land use, physical and chemical quality, and habitat quality.

! Poor (stressed) designations of stream sites for testing impairment discriminations are also based on judgement criteria involving land use, physical and chemical and quality, and habitat quality.

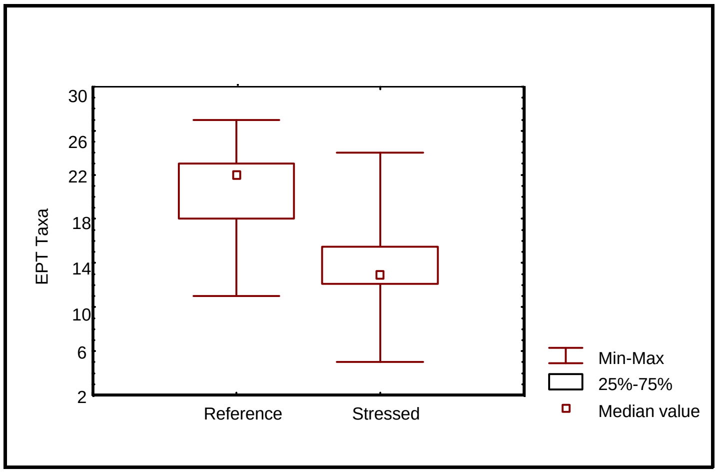

Figure 9-6. Example of discrimination, using the EPT index, between reference and stressed sites in Rocky Mountain streams, Wyoming.

! Determine which biological metrics best discriminate between the reference sites and sites with identified anthropogenic stressors.

! Those metrics having the strongest discriminatory power will provide the most confidence in assessing biological condition of unknown sites.

Element 2 *Determine the associations/linkages between candidate biological and physicochemical measures*

! Plot relationship of metric values against various stressor categories, e.g., chemical concentrations, habitat condition and other measured stressors.

! If desired, multivariate ordination models may be used to elucidate gradients of response of metrics to stressors.

! Monotonic relationships between metrics and stressors allow the use of extreme values (highest or lowest) as reference condition.

! Some metrics may not always be monotonic. For example, total biomass and taxa richness values may exceed the reference at intermediate levels of nutrient enrichment.

! Multiple metrics should be selected to provide a strong and predictable relationship with stream condition.

An *index* provides a means of integrating information from a composite of the various measures of biological attributes.

Step 4. Determine the best aggregation of core measures for indicating status and change in condition

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

Where possible, the scoring criterion for each metric is based on the distribution of values in the population of sites, which include reference streams; for example, the 95th percentile of the data distribution is commonly used (Figure 9-7) to eliminate extreme outliers. From this upper percentile, the range of the metric values can be standardized as a percentage of the 95th percentile value, or other (e.g., trisected or quadrisected), to provide a range of scores. Those values that are closest to the 95th percentile would receive higher scores, and those having a greater deviation from this percentile would have lower scores. For those metrics whose values *increase* in response to perturbation (see Table 7-2 for examples of "reverse" metrics for benthic macroinvertebrates) the 5th percentile is used to remove outliers and to form a basis for scoring.

Alternative methods for scoring metrics, as illustrated in Figure 9-7, are currently in use in various parts of the US for multimetric indexes. A "trisection" of the scoring range has been well-documented (Karr et al. 1986, Ohio EPA 1987, Fore et al. 1996, Barbour et al. 1996b). A "quadrisection" of the range has been found to be useful for benthic assemblages (DeShon 1995, Maxted et al. in press). More recent studies are finding that a standardization of all metrics as percentages of the 95th percentile value yields the most sensitive index, because information of the component metrics is retained (Hughes et al. 1998). Unpublished data from statewide databases for Idaho, Wyoming, Arizona, and West Virginia, are supportive of this third alternative for scoring metrics. Ideally, a composite of all sites representing a gradient of conditions is used. This situation is analogous to a determination of a dose/response relationship and depends on the ability of incorporating both reference and non-reference sites.

Aggregation of metric scores simplifies management and decision making so that a single index value is used to determine whether action is needed. Biological condition of waterbodies is judged based on the summed index value (Karr et al. 1986). If the index value is above a criterion, then the stream is judged as "optimal" or "excellent" in condition. The exact nature of the action needed (e.g., restoration, mitigation, pollution enforcement) is not determined by the index value, but by analyses of the component metrics, in addition to the raw data and integrated with other ecological information. Therefore, the index is not the sole determinant of impairment and diagnostics, but when used in concert with the component information, strengthens the assessment (Barbour et al. 1996a).

Components of Step 4 include:

- Determine scoring criteria for each metric (within each site class) from the appropriate percentile of the data distribution (Figure 9-7). If the metric is associated with a significant covariate such as watershed size, a

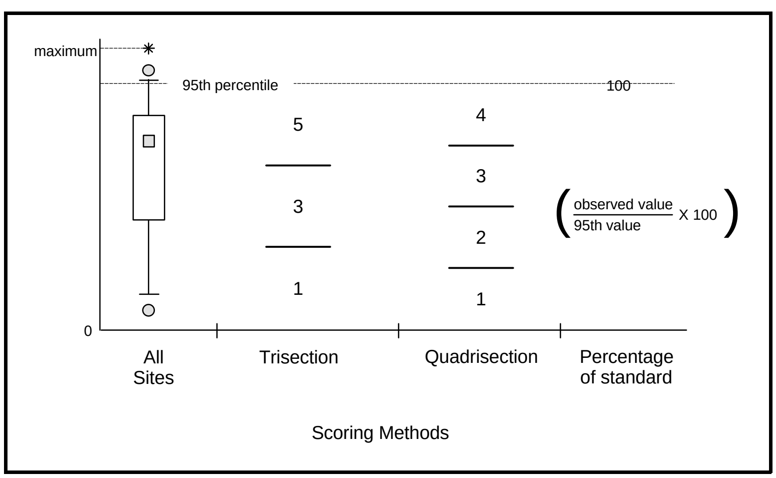

Figure 9-7. Basis of metric scores using the 95th percentile as a standard.

scatterplot of the metric and covariate (Figure 9-3) and a moving estimate of the appropriate percentile, are used to determine scoring criteria as a function of the covariate (e.g., Fausch et al. 1984, Plafkin et al. 1989).

- ! Test the ability of the final index to discriminate between populations of reference and anthropogenically affected (stressed) sites (Figure 9-8). Generally, indices (aggregate of metrics) discriminate better than individual metrics (e.g., total taxa is generally a weak metric because of inconsistency in taxonomic resolution). Those sites that are misclassified with regard to "reference" and "stressed" can be identified and evaluated for reassignment.

Step 5. Index thresholds for assessment and biocriteria

The multimetric index value for a site is a summation of the scores of the metrics and has a finite range within each stream class and index period depending on the maximum possible scores of the metrics (Barbour et al. 1996c). This range can be subdivided into any number of categories corresponding to various levels of impairment. Because the metrics are normalized to reference conditions and expectations for the stream classes, any decision on subdivision should reflect the distribution of the scores for the reference sites. For example, division of the Wyoming benthic IBI range (aggregation of metric scores) within each stream class provides 5 ordinal rating categories for assessment of impairment (Stribling et al. 1999, Figure 9-8).

Biocriteria are based on *thresholds* determined to differentiate impaired from non-impaired conditions. While these thresholds may be subjective, the performance of the *a priori* selected reference sites will ultimately verify the appropriateness of the threshold.

The 5 rating categories are used to assess the condition of both reference and non-reference sites. Most of the reference sites should be rated as *good* or *very good* in biological condition, which would be as expected. However, a few reference sites may be given the rating as *poor* sporadically among the collection dates. If a "reference" site consistently receives a fair or poor rating, then the site should be re-evaluated as to its proper assignment. Putative reference sites may be rated "poor" for several reasons:

- **Natural variability** — owing to seasonal, spatial, and random biological events, any reference site may score below the reference population 10th percentile. If due to natural variability, a low score should occur 10% of the time or less.

- **Impairment** — stressors that were not detected in previous sampling or surveys may occur at a

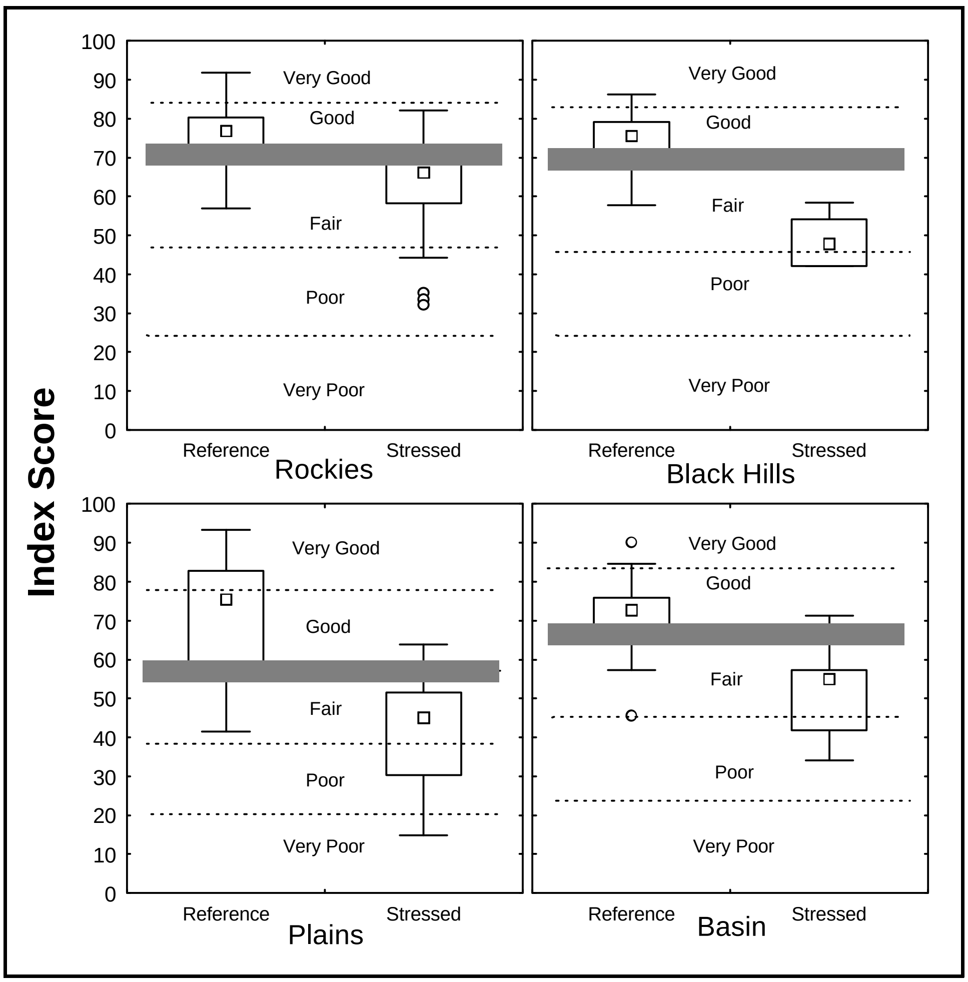

Figure 9-8. Discriminatory power analysis of the Wyoming Benthic Index of Biotic Integrity. The population of stressed sites was determined *a priori*. The 25th percentile of the reference distribution determined the threshold, or separation between "good" and "fair" condition ratings. All other condition ratings resulted from equidistant sectioning of the remaining index range. The shaded region represents the 90% confidence limits around a single observation (no replication) falling near the critical threshold.

"reference" site; for example, episodic non-point-source pollution or historical contamination may be present at a site.

- **Non-representative site** — reference sites are intended to be representative of their class. If there are no anthropogenic stressors, yet a "reference" site consistently scores outside the range of the rest of the reference population the site may be a special or unique case, or it may have been misclassified and actually belong to another class of streams.

An understanding of variability is necessary to ensure that sites that are near the threshold are rated with known precision (discussed in more detail in Chapter 4). To account for variance associated with measurement error in an assessment, replication is required. The first step is to estimate the standard deviation of repeated measures of streams. The standard deviation is calculated as the root mean square error (RMSE) of an analysis of variance (ANOVA), where the sites are treatments in the ANOVA.

As an example, the question of precision was tested for the Wyoming Benthic IBI scores in the stream classes. This study showed that the 95% confidence interval (CI) around a single sample is ±8 points, on a scale of 100 (Table 9-2). What if a single site was sampled with no replication and found to be points below the biocriterion? The rightmost column (Table 9-2) shows that a triplicate sample is required for a 95% CI less than 5 points. These conclusions make 3 assumptions:

- measurement error is normally distributed,
- measurement error is not affected by subecoregion or impairment, and
- the sample standard deviation of repeated measures is an unbiased and precise estimate of population measurement error.

Components of Step 5 include:

- The range in possible scores for each stream class is the minimum number of metrics (if a score of 1 is assigned to greatest level of degradation) to the maximum aggregate of scores. Pentasect, quadrisect, or trisect this range, depending on how many biological condition categories are desired.
- Evaluate the validity of these biological condition categories by comparing the index scores of the reference and known stressed sites to those categories. If reference sites are not rated as good or very good, then some adjustment in either the biological condition designations or the listing of reference sites may be necessary.
- Test for confidence in multimetric analysis to determine biological condition for sites that fall within close proximity to threshold. Calculate precision and sensitivity values to determine repeatability and detectable differences that will be important in the confidence level of the assessment.

Table 9-2. Statistics of repeated samples in Wyoming and the detectable difference (effect size) at 0.10 significance level. The index is on a 100 point scale (taken from Stribling et al. 1999).

Metric	Standard Deviation for Repeated Measures	Approx. Mean[a]	Approx. Coefficient of Variation (%)	Detectable Differences (p = 0.10)		
				Single Sample	Duplicate Samples	Triplicate Samples
Total Taxa	4.1	35.9	11.5	7 taxa	5 taxa	5 taxa
Ephemeroptera taxa	0.9	6.8	13.3	2 taxa	1 taxa	1 taxa
Plecoptera taxa	1.0	4.8	21.2	2 taxa	1 taxa	1 taxa
Trichoptera taxa	1.1	6.9	15.3	2 taxa	1 taxa	1 taxa
% non-insects	3.8	8.9	42.9	6.3 %	4.4 %	4.3 %
% diptera (non-chironomid)	1.3	5.1	25.0	2.1 %	1.5 %	1.4 %
HBI	0.27	3.43	7.85	0.44 units	0.31 units	0.26 units
% 5 dominant taxa	4.3	64.2	6.7	7.1 %	5.0 %	4.1 %
% scrapers	4.8	25.5	18.9	7.9 %	5.6 %	4.6 %
Index	2.0	70.0	2.9	3.3 units	2.3 units	1.9 units

a: Mean of 25 replicated sites; population means may differ.

9.1.2 Assessment of Biological Condition

Once the framework for bioassessment is in place, conducting bioassessments becomes relatively straightforward. Either a targeted design that focuses on site-specific problems or a probability-based design, which has a component of randomness and is appropriate for 305(b), area-wide, and watershed monitoring, can be done efficiently. Routine monitoring of reference sites should be based on a random selection procedure, which will allow cost efficiencies in sampling while monitoring the status of the reference condition of a state’s streams. Potential reference sites of each stream class would be randomly selected for sampling, so that an unbiased estimate of reference condition can be developed. A randomized subset of reference sites can be resampled at some regular interval (e.g., a 4 year cycle) to provide information on trends in reference sites.

A reduced effort in monitoring reference sites allows more investment of time into assessing other stream reaches and problem sites. Through use of Geographical Information System (GIS) and station location codes, assessment sites throughout the state can be randomly selected for sampling as is being done for the reference sites. This procedure will provide a statistically valid means of estimating attainment of aquatic life use for the state’s 305(b) reporting. In addition, the multimetric index will be helpful for targeted sampling at specific problem areas and judging biological condition with a procedure that has been calibrated regionally (Barbour et al. 1996c). To evaluate possible influences on the biological condition of sites, relationships among total bioassessment scores and physicochemical variables can be investigated. These relationships may indicate the influence of particular categories of stressors on the biological condition of individual sites. For example, a strong negative correlation between total bioassessment score and embeddedness would suggest that siltation from nonpoint sources could be affecting the biological condition at a site. Considerations relevant to assessment and diagnostics of biological condition are as follows:

! Evaluate the relationship of biological response signatures such as functional attributes (reproduction, feeding group responses, etc.) to specific stressors.

- Hold physical habitat relationships constant and look for associations with other physical stressors (e.g., hydrologic modification, streambed stability), chemical stressors (e.g., point-source discharges or pesticide application to cropland), biological stressors (i.e., exotics), and landscape measures (e.g., impervious surface, Thematic mapper land use classes, human population census information, landscape ecology parameter of dominance, contagion, fractal dimension).

- Explore the relationship between historical change in biota and change in landscape (e.g., use available historical data from the state or region).

9.2 DISCRIMINANT MODEL INDEX

Discriminant analysis may be used to develop a model that will divide, or discriminate, observations among two or more predetermined classes. Output of discriminant analysis is a function that is a linear combination of the input variables, and that obtains the maximum separation (discrimination) among the defined classes. The model may then be used to determine class membership of new observations. Thus, given a set of unaffected reference sites, and a set of degraded sites (due to toxicity, low DO, or habitat degradation), a discriminant function model can identify variables that will discriminate reference from degraded sites.

Developing biocriteria with a discriminant model requires a training data set to develop the discriminant model, and a confirmation data set to test the model. The training and confirmation data may be from the same biosurvey, randomly divided into two, or they may be two consecutive years of survey data, etc. All sites in each data set are identified by degradation class (e.g., reference vs stressed) or by designated aquatic life use class. To avoid circularity, identification of reference and stressed, or of designated use classes, should be made from non-biological information such as quality of the riparian zone and other habitat features; presence of known discharges and nonpoint sources, extent of impervious surface in the watershed, extent of land use practices, etc.

One or more discriminant function models are developed from the training set, to predict class membership from biological data. After development, the model is applied to the confirmation data set to determine its performance: The test determines how well the model can assign sites to classes, using independent data that were not used to develop the model. More information on discriminant analysis is in any textbook on multivariate statistics (e.g., Ludwig and Reynolds 1988, Jongman et al. 1987, Johnson and Wichern 1992).

An example of this approach is the hierarchical decision-making technique used by Maine DEP. It begins with statistical models (linear discriminant analysis) to make an initial prediction of the classification of an unknown sample by comparing it to characteristics of each class identified in the baseline database (Davies et al. 1993). The output from analysis by the primary statistical model is a list of probabilities of membership for each of four groups designated as classes A, B, C, and nonattainment (NA) of Class C (Table 9-3). Subsequent models are designed to distinguish between a given class and any higher classes as one group, and any lower classes as a second group.

One or more discriminant models to predict class membership are developed from the training set. The purpose of the discriminant analysis here is not to test the classification (the classification is administrative rather than scientific), but to assign test sites to one of the classes.

Stream biologists from Maine DEP assigned a training set of streams to four life use classes. In operational assessment, sites are evaluated with the two-step hierarchical models. The first stage linear discriminant model is applied to estimate the probability of membership of sites into one of the four classes (A, B, C, or NA). Second, the series of two-way models are applied to distinguish the membership between a given class and any higher classes, as one 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 has gone down in class, or for reclassification to a higher class if the site has improved.

Table 9-3. Maine's water quality classification system for rivers and streams, with associated biological standards (taken from Davies et al. 1993).

Aquatic Life Use Class	**Management**	**Biological Standard**	**Discriminant Class**
AA	High quality water for recreation and ecological interests. No discharges or impoundments permitted.	Habitat natural and free flowing. Aquatic life as naturally occurs.	A
A	High quality water with limited human interference. Discharges restricted to noncontact process water or highly treated wastewater equal to or better than the receiving water. Impoundments allowed.	Habitat natural. Aquatic life as naturally occurs.	A and AA are indistinguish-able because biota are "as naturally occurs."
B	Good quality water. Discharge of well treated effluent with ample dilution permitted.	Habitat minimally impaired. Ambient water quality sufficient to support life stages of all indigenous aquatic species. Only nondetrimental changes in community composition allowed.	B
C	Lowest water quality. Maintains the interim goals of the Federal Water Quality Act (fishable/swimmable). Discharge of well-treated effluent permitted.	Ambient water quality sufficient to support life stages of all indigenous fish species. Change in community composition may occur but structure and function of the community must be maintained.	C
NA			Not attaining Class C

Maine biocriteria thus establish a direct relationship between management objectives (the three aquatic life use classes and nonattainment) and biological measurements. The relationship is immediately viable for management and enforcement as long as the aquatic life use classes remain the same. If the classes are redefined, a complete reassignment of streams and a review of the calibration procedure would be necessary. This approach is detailed by Davies et al. (1993).

See Maine DEP's website for more information
http://www.state.me.us/dep/blwq/biohompg.htm

9.3 RIVER INVERTEBRATE PREDICTION AND CLASSIFICATION SCHEME (RIVPACS)

RIVPACS and its derivative, AusRivAS (Australian Rivers Assessment System) are empirical (statistical) models that predict the aquatic macroinvertebrate fauna that would be expected to occur at a site in the absence of environmental stress (Simpson et al. 1996). The AusRivAS models predict the invertebrate communities that would be expected to occur at test sites in the absence of impact. 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” invertebrate community to measure the success of any remediation measures taken to rectify identified impacts. The type of taxa predicted by the AusRivAS models may also provide clues as to the type of impact a test site is experiencing. This information can be used to facilitate further investigations e.g., the absence of predicted Leptophlebiidae may indicate an impact on a stream from trace metal input.

These models are the primary ecological assessment analysis techniques for Great Britain (Wright et al. 1993) and Australia (Norris 1995). 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. Regional applications of the AusRivAS model, in particular, have been developed for the Australian states and territories (Simpson et al. 1996), and for streams in the Sierra and Cascade mountain ranges in California (Hawkins and Norris 1997). Users of these models claim rapid turn around of results is possible and output can be tailored for a range of users including community groups, managers, and ecologists. These attributes make RIVPACS and AusRivAS likely candidate analysis techniques for rapid bioassessment programs.

Although the same procedures are used to build all AusRivAS models, each model is tailored to specific regions (or states) to provide the most accurate predictions for the season and habitat sampled. The stream habitats for which these models have been applied include the edge/backwater, main channel, riffle, pool, and macrophyte stands. The multihabitat sampling techniques used in many RBP programs have not yet been tested with a RIVPACS model. The models can be constructed for a single season, or data from several seasons may be combined to provide more robust predictions. To date the RIVPACS/AusRivAs models have only been developed for the benthic assemblage. Discussion of RIVPACS and AusRivAS is taken from the *Australian River Assessment System National River Health Program Predictive Model Manual* by Simpson et al. (1996). As is the case with the multimetric approach, a more thorough treatment of the RIVPACS/AusRivAS models can be obtained by referring to the citations of the supporting documentation provided in this discussion.

The reader is directed to the AusRivAS website for more specific information and guidance regarding these multivariate techniques.
http://ausrivas.canberra.edu.au/ausrivas