# Chapter 11

# Index Development

## 11.1 Overview

Many methods have been developed to assess the condition of water resources from biological data, beginning with the saprobien system in the early 20th century to present-day development of biological markers. This chapter will discuss three methods for analyzing and assessing water body condition from assemblage and community-level biological information:

1. *Multimetric index.*

This is the basis of many indexes used in fresh waters: the Index of Biotic Integrity (IBI; Karr et al. 1986), the Invertebrate Community Index (ICI; Ohio EPA 1987); the Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, Second Edition (RBP; Barbour et al. 1999); and state indexes developed from these (e.g., Southerland and Stribling 1995). More recently, multimetric IBI - type indexes have been developed for estuarine assemblages (e.g. Cape Cod fish, Deegan et al. 1997; Chesapeake Bay macroinvertebrates, Weisberg et al. 1997; Carolinian Province macroinvertebrates, Hyland et al. 1998). The Chesapeake Bay development (Weisberg et al. 1997) will be used to illustrate the method.

2. *Discriminant model index.*

This is the basis of stream bioassessment in Maine (Davies et al. 1993), and of the estuarine invertebrate indexes developed by the EMAP-NC program in the Virginian and Louisianan provinces (Weisberg et al. 1993; Schimmel et al. 1994; Strobel et al. 1994; Summers et al. 1993, 1994; Engle et al. 1994). The EMAP-NC, Louisianian and Virginian Province examples will be used to illustrate the method.

3. *Index derived from multivariate ordination.*

Smith et al. (2000) and Allen and Smith (2000) have developed a pollution tolerance index for near-coastal sites of Southern California, using species composition of benthic macroinvertebrates and demersal fish. Other approaches using ordination have demonstrated differences in composition between reference and stressed sites (e.g., Warwick and Clarke 1991). The approach of Smith et al. uses ordination of species composition to develop a numeric index on a scale of 0-100, that can be used directly for biocriteria. The Smith et al. example will be used to illustrate the method.

Many other methods are possible, as well as permutations of the three methods above, all of which are beyond the scope of this document. These three were selected because:

- They use community and assemblage data;
- The methods are not restricted to any one assemblage. The examples all use benthic macroinvertebrates, but any other assemblage could also be used, such as fish,

phytoplankton, zooplankton or macrophytes;

- The examples used to illustrate the methods have been carried out over wide geographic areas with many sites, demonstrating the generality of the methods;

- The examples used to illustrate the methods are concise, the methods were fully documented, and have been carried to completion, that is, assessment of biological impairment and non-impairment.

All three of the methods use the same general approach: sites are assessed by comparing the assemblage of organisms found at a site to an expectation derived from observations of many relatively undisturbed reference sites. The expectations are modified by classifying the reference sites to account for natural variability, and each assessment site is classified using non-biological (physical, chemical, geographic) information. Finally, metrics (methods 1 and 2) or the species ordination (method 3) are tested for response to stressors by comparison of reference and known impaired sites. An example of the assessment process is summarized in Figure 11-1.

This chapter will first discuss methods of classification, with emphasis on those that have been successful in estuaries and coastal waters. The remainder of the chapter then discusses the three assessment methods. This chapter is not intended to be an instruction manual on using the different statistical methods; it is intended to show, with selected examples, techniques that have been used to develop biological indexes. Details of applications and methodology can be found in the cited documents and articles and in statistical textbooks and manuals (e.g., Ludwig and Reynolds 1988, Reckhow and Warren-Hicks 1996).

## 11.2 Classification and Characterization of Reference Condition

The objective of characterization is to finalize the classification of reference sites and to describe (characterize) each of the reference classes in terms of metrics, assemblage composition, and physical-chemical variables. As outlined in Chapter 4, classification may be a physical rule-based classification, or an analytical data interpretation where rules are derived from the data. The analytical approach requires a relatively large reference data set to derive the classes and rules, with many sites and both biological and physical-chemical data from each site.

The basic assumption of classification is that biogeography, physical habitat, and water quality largely determine attributes such as taxa richness, abundance, and species dominance in estuarine and coastal marine biological communities. In other words, if habitats are classified adequately, reference biological communities should correspond to the habitat classification.

Several statistical tools can assist in site classification, but there is no one set procedure. If the rule-based classification is based on well-developed prior knowledge and professional judgment, graphical analysis of metrics, followed by any necessary modifications and tests of the resultant classification, it is usually sufficient. If necessary, the classification is refined until an optimal classification emerges that satisfactorily accounts for variation in reference site biological data.

If a physical classification is not self-evident, it may be necessary to develop an alternative classification from the data using one or more of several

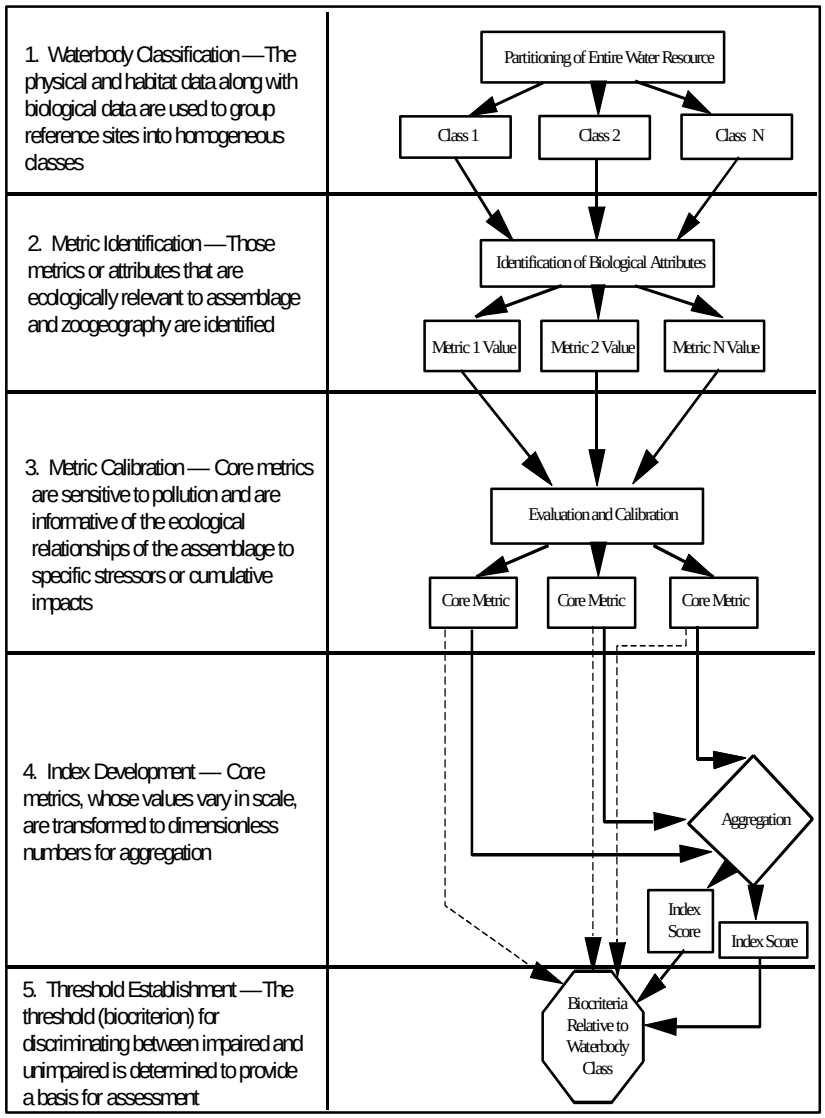



**Figure 11-1**
The process for progressing from the classification of an estuary to assessing the health of the estuary. Adapted from Paulsen et al. 1991.

classification methods. These methods include cluster analysis and several ordination methods such as: principal components analysis, correspondence analysis, and multidimensional scaling.

### 11.2.1 Existing Classifications

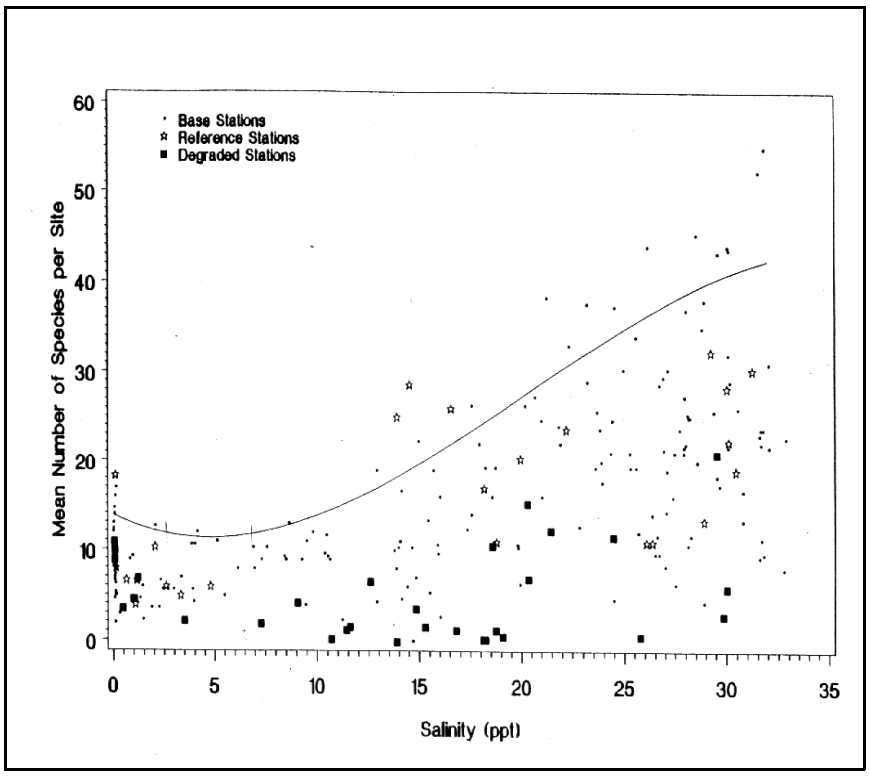
With the growth of efforts to improve environmental monitoring and develop biocriteria, several successful classifications of estuarine and near-coastal biological assemblages have been developed. Here, we summarize several of these and integrate their findings on classification of North American estuarine assemblages.

***EMAP Virginian Province***

Natural environmental factors affecting species composition were examined in the EMAP Virginian Province Project (Paul et al. 1999, Strobel et al. 1995, Weisberg et al. 1993). Salinity has been known to control estuarine organisms since the early days of marine biology. Over 75% of the candidate measures were related significantly to salinity distributions (Figure 11-2). Correlation analysis was used to examine associations of habitat factors with candidate biological metrics. Of the correlations between candidate measures and habitat factors,

**Figure 11-2**

Mean number of species and salinity at EMAP-Estuaries sampling stations in the Virginian Province (from Weisberg et al. 1993). The regression line shown is the expected number of species based on the polynomial regression, and was used to estimate salinity-adjusted species richness measures.




species richness was most strongly correlated with salinity.

In addition to salinity, the physical characteristics of estuarine sediments and depth also influence benthic infaunal distribution and the accumulation of contaminants in sediments (Rhoads 1974, Plumb 1981). EMAP collected sediment grain size, silt-clay content, latitude, and depth data to help interpret benthic response. Although silt-clay content and depth were statistically significant, only salinity was deemed to have a biologically significant influence on benthic macroinvertebrates ($r^2$ >0.025; Weisberg et al. 1993). Estuary type was stratified in the project design, but community differences due to estuary type were not reported.

### *Chesapeake Bay*

Weisberg et al. (1997) developed an estuarine benthic index of biotic integrity for the Chesapeake Bay. Cluster analysis of benthic infauna indicated seven distinct habitats defined by substrate and salinity. Polyhaline sand and mud, (salinities (≥18 ‰) had the highest mean Shannon-Wiener diversities, at 4.0 and 3.55, respectively (Weisberg et al. 1997).

### *EMAP Carolinian Province*

From July - September 1995, a study was conducted to assess the environmental condition of estuaries in the EMAP Carolinian Province (Hyland et al. 1998, see Chapter 13). The program sampled water depth, salinity, and substrate classifications (% silt-clay) as habitat indicators.

Species richness showed highly significant correlations with latitude, bottom salinity, and silt-clay/TOC sediment content. The Shannon-Wiener index, H′(a combination of species richness and evenness), also showed highly significant correlations ($p \leq$ 0.0030) with bottom salinity as well as

with silt-clay fractions. As with diversity, infaunal abundance showed highly significant correlations ($p \leq 0.0016$) with the silt-clay and TOC sediment content (Hyland et al. 1998).

### *North Carolina*

The North Carolina study was designed to compare biological metrics derived from three sampling methods (Ponar, epibenthic trawl, and sweep net). Salinity was the only habitat characteristic that was significantly correlated with biological metrics. Total taxa showed a positive correlation with salinity (Eaton 1994a; see Chapter 13).

### *Puget Sound*

The objective of the Puget Sound study was to characterize benthic macroinvertebrate communities into habitats classified as degraded and habitats that are relatively unimpaired, which can then be classified as reference sites for the Sound (Llansó 1999).

The diverse assemblages sampled were mainly associated with sediment type and water depth, reinforcing results from previous studies (Lie 1974). The classes of sediments defined for the Puget Sound estuaries were: sands, clays, and mixed. These three classes did not have exact boundaries, but instead overlapped at both ends of their spectrums (Llansó 1999). Stations with finer substrates had fewer species than those with coarser substrates. On average, sand substrates supported more species and abundance than did clay, with deep sites having the lowest abundance levels. Overall, clay stations in the southern part of Puget Sound supported fewer species than many other shallow clay locations. The majority of species were not restricted to only one substrate, instead they were widely distributed in different types of sediment showing the most abundance in sand, mixed sediment, or muddy bottoms.

### *EMAP Louisianian Province*

Prior studies in the Gulf of Mexico had shown salinity and sediment type to be among the most important factors that determine benthic infaunal relationships in Gulf of Mexico estuaries (Flint and Kalke 1985, Gaston et al. 1988, Rabalais 1990, Rakocinski et al. 1991). Of the 182 total sites sampled, Pearson correlations were performed between all candidate measures and salinity, longitude of sampling site (as a measure of geographical gradient), percent silt-clay, and total organic carbon content of sediments. Many of the correlations were statistically significant at $p < 0.05$, however, only salinity accounted for 20% or more of the variation (Summers et al. 1993).

### *Southern California Bight*

The Southern California Coastal Water Research Project sampled megabenthic invertebrate assemblages, benthic infaunal assemblages, and demersal fish assemblages to determine their relationship to depth, latitude, and sediment types in the Southern California Bight. There was no salinity gradient because the entire study area was nearshore marine. Overall, depth was found to be the defining factor in the organization of each assemblage (Allen et al. 1999, Bergen et al. 1999).

Sediment type was found to be a secondary factor in the organization of benthic infaunal assemblages. This finding could be attributed to the large study area. In fact, within a constrained depth range, sediment type may be a more important factor (Bergen et al. 1999). These findings are consistent with those of Snelgrove and Butman

(1994) which suggested that the hydrodynamic environment and the amount of organic material in the sediment are more likely to be primary driving forces, with depth and sediment grain size as secondary correlates.

***Conclusions***

Three habitat indicators have been demonstrated repeatedly to influence biological assemblages of estuaries and near coastal environments. In studies where there was a salinity gradient, salinity was found to be the most important habitat indicator. Depth and substrate are also important and usually correlated, especially if there is a large depth gradient at the sample site. The physical type of estuary (e.g., fjord, lagoon, tidal river) has not been demonstrated to be vital in wide geographic studies, such as those conducted by EMAP in the Virginian, Louisianian, and Carolinian provinces, but may not have been adequately tested. Therefore, the importance of measuring estuary type, subregion, or subprovinces is still questionable.

Lessons learned from both EMAP and other independent studies conclude that the basic classification of an index should be by biogeograpical province, salinity, substrate (silt-clay content, sediment grain size), and depth. The effects of salinity, substrate, and depth should be tested within the study area to determine whether all are required as habitat indicators in an individual area. Moreover, decisions need to be made as to the use of discrete classes or continuous covariates in statistical analysis. If other classifications are suspected to be important indicators of the health of a system, they should also be tested (e.g., estuary type).

### 11.2.2 Assessing *a priori* Classifications

Although there is no serious doubt over the influence of salinity, sediment, and depth on estuarine biota, the effects must be characterized or calibrated to establish reference conditions. Several approaches have been used, as outlined in the examples in this chapter. Often, one of the first steps is a cluster analysis of the species composition of the sites to determine if sites can be broken down into groups (e.g., Weisberg et al. 1997, Smith et al. 2000). Sites may be divided into groups defined by the important variables (e.g., salinity and sediment; Weisberg et al. 1997, depth; Smith et al. 1999), or the groups may be separated by discriminant function analysis (DFA) if simple, single relationships are not sufficient (e.g., Engle and Summers 1999).

Another approach is to examine correlations between environmental variables and biological metrics calculated from the species data, so that reference expectations can be calibrated accordingly. For example, species richness in estuaries is strongly affected by salinity (refer to Figure 11-2). Weisberg et al. (1993) used the relationships of Figure 11-2 to develop a nonlinear regression of maximum expected species richness on salinity. Species richness was then adjusted by the salinity-specific maximum in further development of their model of impairment.

## 11.3 Index Development

An index for assessing sites can be developed after classification of sites of the region is completed. Index development using the three approaches followed in this chapter is discussed here.

### 11.3.1 Multimetric Index

*Step 1. Identify Potential Measures For Each Assemblage.*

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) ecological relevance to the biological assemblage or community under study and to the specified program objectives, (2) sensitivity to stressors and provide 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. 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 taken to identify the candidate metrics that are most informative, and therefore, warrant further analysis.

Representative metrics should be selected from each of four primary categories: (1) richness measures for diversity or variety of the assemblage; (2) composition measures for identity and dominance; (3) tolerance measures that represent sensitivity to perturbation; and (4) trophic or habit measures for information on feeding strategies and guilds. Table 11-1 further illustrates metrics for various assemblages that have been useful in estuaries. Components of Step 1 include:

- Review of the value ranges of potential metrics, and elimination of those that have too many zero values in the population of reference sites to calculate the metric at a large enough proportion of sites;
- Descriptive statistics (central tendency, range, distribution, outliers) to characterize metric performance within the population of reference sites of each site class;
- Elimination of metrics that have too high variability in the reference site population such that they cannot discriminate among sites of different condition.

*Step 2. Select Robust Measures.*

Core metrics are those that will discriminate between good and poor quality ecological conditions. Discriminatory ability of biological metrics is 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 11-3).

Criteria are established to identify a population of "known" stressed sites based on physical and chemical measures of degradation. Criteria for

**Table 11-1**. Potential metrics for macrophytes, benthic macroinvertebrates, and fish that could be considered for estuaries. Redundancy can be evaluated during the calibration phase to eliminate overlapping metrics.

| | Richness | Composition | Tolerance | Trophic/Habitat |
|---|---|---|---|---|
| Macrophytes | ▸ Not applicable | ▸ Not applicable | ▸ TSS<br>▸ light attenuation<br>▸ Chlorophyll *a*<br>▸ DIN<br>▸ DIP | ▸ % cover<br>▸ density of new shoots<br>▸ biomass<br>▸ stem counts |
| Benthic Macroinvertebrates | ▸ dominant taxa<br>▸ taxa richness<br>▸ Shannon-Wiener Diversity Index<br>▸ mean # of species<br>▸ Pielou's Evenness Index | ▸ # amphipods per event<br>▸ amphipod biomass<br>▸ mean abundance of bivalves/site<br>▸ # of gastropods per event | ▸ % polychaetes<br>▸ polychaete biomass | ▸ % or biomass epibenthic<br>▸ % or biomass deposit feeders<br>▸ % or biomass suspension feeders |
| Fish | ▸ dominant taxa<br>▸ taxa richness<br>▸ # of estuarine spawners<br>▸ # anadronomous spawners<br>▸ total fish exclusive of Atlantic menhaden | ▸ total # of species<br>▸ # species in bottom trawl<br>▸ # species comprising 90% of individuals | ▸ #, % or biomass of menhaden | ▸ Proportion of planktivores<br>▸ Proportion of benthic feeders<br>▸ Proportion of piscivores |

stressed sites can include (Weisberg et al. 1993):

- Any sediment contaminant exceeding the Long et al. (1995) effects range-median (ER-M) concentration;
- Survival in toxicity tests less than 80% of controls;
- Low dissolved oxygen;
- Total sediment organic carbon > 3%.

Following identification of reference and stressed sites, the biological metrics that best discriminate between them are determined.

Those metrics having the strongest discriminatory power provide the most confidence in assessing biological condition of unknown sites. Metrics can be easily compared by estimating their discrimination efficiency (DE) or the percentage of stressed sites below a threshold representing the reference sites. For example, DE could be measured as the percentage of stressed sites below the 25$^{th}$ percentile of reference sites, for a given metric.

Several studies have used tests of statistical significance between reference and stressed sites to select metrics (e.g., Weisberg et al. 1997, Hyland et al. 1998). Significance tests should only be used if the sample size (number of reference and stressed sites) is large enough that the test has sufficient power to detect a meaningful difference.

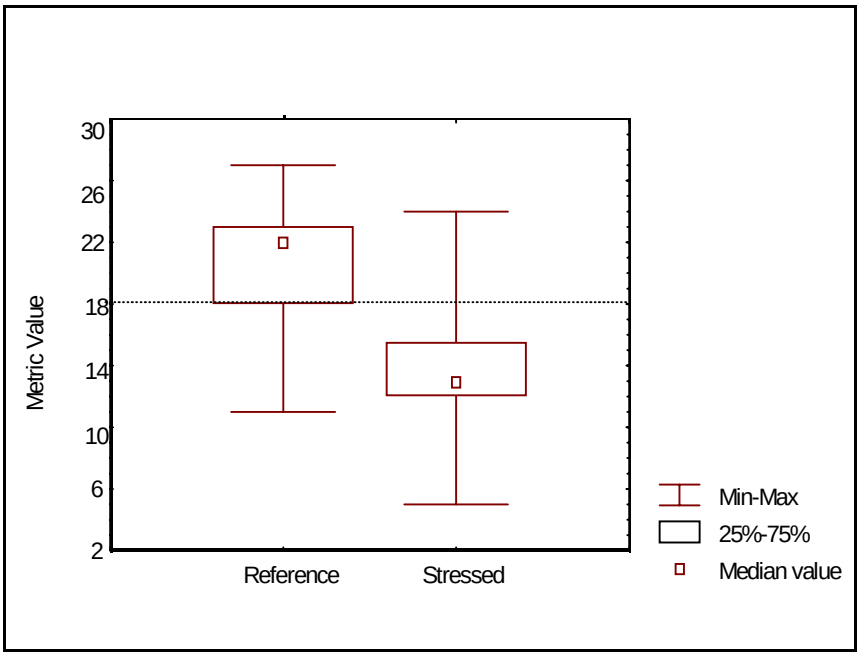



**Figure 11-3**
Hypothetical box plot illustrating how a successful metric discriminates between reference and stressed sites.

*Step 3. 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 sites, which include reference sites; for example, the 95$^{th}$ percentile of the data distribution is commonly used to eliminate extreme outliers. From this upper percentile, the range of the metric values can be standardized as a percentage of the 95$^{th}$ percentile value, or other (e.g., trisected or quadrisected), to provide a range of scores. Those values that are closest to the 95$^{th}$ percentile receive higher scores, and those having a greater deviation from this percentile receive lower scores. For those metrics whose values *increase* in response to perturbation the 5$^{th}$ percentile is used to remove outliers and to form a basis for scoring.

Alternative methods for scoring metrics are currently in use in various parts of the U.S. for multimetric indexes. A "trisection" of the scoring range has been well documented (Karr et al. 1986, Ohio EPA 1987, Weisberg et al. 1997, Hyland et al. 1998). More recent studies are finding that a standardization of all metrics as percentages of the 95$^{th}$ percentile value yields the most sensitive index, because more information of the component metrics is retained (e.g., Hughes et al. 1998).

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. 1996b). Components of Step 3 include:

- Development of scoring criteria for each metric (within each site class) from the appropriate percentile of the data distribution (Figure 11-4). If the metric is associated with a significant covariate such as estuary size, depth, or salinity a scatterplot of the metric and covariate and a moving estimate of the appropriate percentile, are used to determine scoring criteria as a function of the covariate (e.g., Weisberg et al. 1993);

- Testing the ability of the final index to discriminate between populations of reference and anthropogenically affected (stressed) sites.

*Step 4. 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 site class and index period, depending on the maximum possible score of the metrics (Barbour et al. 1996a). 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 classes, any decision on subdivision should reflect the distribution of the scores for the reference sites.

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 10$^{th}$ 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 "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 sites.

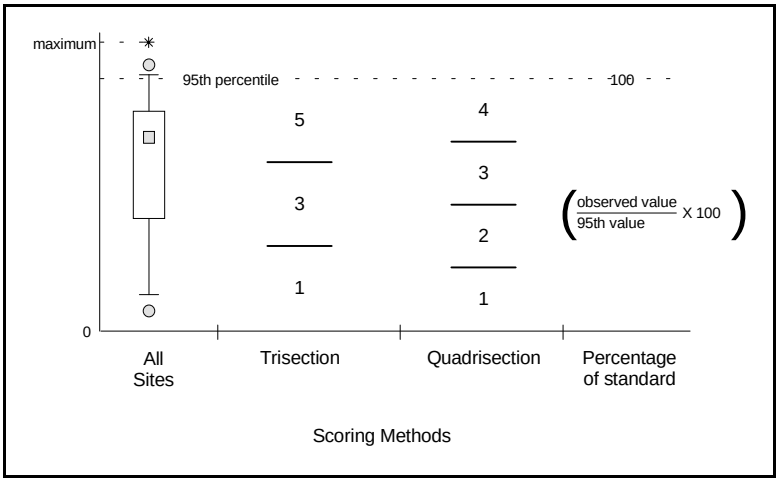



**Figure 11-4**
Basis of metric scores using the 95th percentile as a standard.

Components of Step 4 include:

- Assessment categories are subdivided from the range of possible scores for each site class. Categories should be proportional to the interquartile range (or standard deviation) of total scores in the reference sites. Thus, reference sites with a small interquartile range (small s.d.; small coefficient of variation) would yield more assessment categories than a more variable reference population;

- The validity of biological condition categories is evaluated by comparing the index scores of the reference and known stressed sites to those categories. If reference sites are not rated good or very good, then some adjustment in either the biological condition designations or the listing of reference sites may be necessary;

- Confidence intervals are estimated for the multimetric index to help determine biological condition for sites that fall in close proximity to a threshold. Precision and sensitivity are determined from replicate samples, and are important for estimating the confidence of individual assessments.

Once the framework for bioassessment is in place, conducting bioassessments becomes relatively routine. Either a targeted design that focuses on site-specific problems or a probability-based design, which 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 for cost efficiencies in sampling while monitoring the status of the reference condition. Potential reference sites of each 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.

*Example 1: Chesapeake Bay Index Development*

For the Chesapeake Bay, a separate benthic index was developed for each of seven habitats: tidal fresh, oligohaline, low mesohaline, high mesohaline sand, high mesohaline mud, polyhaline sand, and polyhaline mud (Weisberg et al. 1997). These habitats had been identified as separate assemblages in the classification step.

Reference and stressed sites were identified by the following: from existing Chesapeake Bay data, no reference sites could be in highly developed (urban) watersheds or near known point-source discharges, no reference site could have organic carbon content > 2%, no reference site could have any sediment contaminants exceeding the Long et al. (1995) effects range-median (ER-M) concentration, no reference site could have low dissolved oxygen, and no reference site could exhibit any sediment toxicity. Stressed sites were defined as those with any contaminant exceeding the ER-M concentration and measured sediment toxicity, or total organic carbon exceeded 3%, or dissolved oxygen was low, < 2-mg$L^{-1}$ (Weisberg et al. 1997).

Index development proceeded through the steps:

*Step 1.* 17 candidate metrics were identified based on the paradigms of Pearson and Rosenberg (1978).

*Step 2.* 15 of the 17 metrics could distinguish between reference sites and stressed sites in one or more of the seven habitats.

*Step 3.* Four to seven of the metrics were used for an index specific to each habitat type. Scoring of metrics was on a 5-3-1 scale, with metric values greater than the reference site median scored as 5; between the 5$^{th}$ and the median of the reference sites scored as 3; and below the 5$^{th}$ percentile scored as 1.

*Step 4.* The index was able to correctly classify as reference or stressed 93% of an independent validation data set that had not been used to develop the index.

*Example 2: Louisiana and Maryland Fish Indexes*

Several states are developing fish indexes of biotic integrity (IBI) for estuarine species. The multimetric Index of Biotic Integrity (IBI) concept was originally developed for fresh water streams (Karr 1981), and has been modified and applied to a Louisiana estuary (Thompson and Fitzhugh 1986). The strength of this index is that many factors affecting biological integrity can be measured in fish (e.g., community composition, relative abundance, health, etc.). This proposed estuarine IBI maintains the same three main categories as those of the fresh water IBI: species composition, trophic composition, and fish condition. However, the metrics are modified to reflect estuarine habitats and fish assemblages. In addition, because estuarine systems exhibit a high degree of seasonality in their fish fauna, a measure of seasonal variability was incorporated. The metrics for estuaries are based on life history and habitat requirements similar to those of the fresh water IBI. Proposed metrics from Thompson and Fitzhugh (1986) for estuarine communities are listed in Table 11-2. A similar fish Index of Biotic Integrity is being adapted for application in estuarine and coastal marine habitats on the Gulf Coast of Texas (Guillen 1995).

The state of Maryland has also developed a fish Index of Biotic Integrity that is more rapid and less expensive to apply (Jordan et al. 1992). This fish IBI is comprised of nine metrics (Table 11-3) that can be compared to measurements of the physical environment such as dissolved oxygen and land use.

**Table 11-2.** Estuarine fish IBI metrics proposed by Thompson and Fitzhugh (1986).

| Community Structure/Function | Metric |
|---|---|
| Species composition | Total number of fish species<br>Number and identity of resident estuarine species<br>Number and identity of marine species<br>Number and identity of sciaenids<br>Number and identity of freshwater species<br>Proportion of individuals as bay anchovy<br>Measure of seasonal overlap of fish community<br>Number of species needed to make up 90% of collection |
| Trophic composition (for adults of species) | Proportion of individuals as generalized benthic feeders<br>Proportion of individuals as generalized plankton grazers<br>Proportion of individuals as top carnivores |
| Fish abundance and condition | Proportion of young of year in sample or number of individuals in sample<br>Proportion of individuals with disease, tumors, fin damage, and other anomalies |

The results of some preliminary analyses from areas in the Chesapeake Bay with salinities ranging from 0-to-16 ppt indicate that the Maryland IBI can be used to identify large scale spatial and temporal trends in biological integrity and that the index responds to water quality (DO) and land use impacts.

### 11.3.2 Discriminant Model Index

#### *Discriminant Model Approach*

The discriminant model approach was used by EMAP to develop benthic condition indexes for the Virginian Province (Mid-Atlantic) and for the Louisianian Province (Gulf Coast) (Engle et al. 1994, Summers et al. 1993, Weisberg et al. 1993, Paul et al. 1999) based on defined reference sites. Sets of minimally impaired sites; i.e., "reference" and impaired sites were identified; impaired sites were affected by either hypoxia (DO <2 $mgL^{-1}$); toxic sediments; or sediment contamination above the ER-M threshold. Minimally impaired sites were defined to have DO >5 $mgL^{-1}$ and no detectable toxicity or contamination. The two site types represented the ends of a continuum, with intermediate sites not used for discriminant model building (Engle et al. 1994, Weisberg et al. 1993).

The classification step for the EMAP discriminant models consisted of examining associations between benthic macroinvertebrate metrics and physical habitat measures of salinity, sediment grain size, and depth. Only salinity had a strong relationship with the taxa richness metric; taxa richness was estimated as the percent of taxa expected, adjusted for salinity (refer to Figure 11-2).

#### *Discriminant Model Analysis*

The discriminant model analysis is a multivariate procedure that attempts to build a model that will predict the membership of a site into two or more predetermined classes. In the example used in EMAP, the classes were reference and impaired sites (by low DO, toxicity or metal contamination). The model procedure attempts to find a linear combination of input variables (biological metrics) that best predicts membership in the class. Alternative

**Table 11-3.** Maryland estuarine fish IBI metrics.

| Community Structure/Function | Metric | Response to Impairment |
|---|---|---|
| Species composition | Total number of species<br>Number of species in bottom trawl<br>Number of species comprising 90 percent of individuals | reduced<br>reduced<br>reduced |
| Trophic composition (for adults of species) | Proportion of planktivores<br>Proportion of benthic feeders<br>Proportion of piscivores | increased<br>decreased<br>decreased |
| Fish abundance and condition | Number of estuarine spawners<br>Number of anadromous spawners<br>Total fish exclusive of Atlantic menhaden | decreased<br>decreased<br>decreased |

models are tested by estimating the proportion of sites (from the model-building data set) that are misclassified. The best model usually has the lowest misclassification rate. A test of a model requires an independent test data set that was not used to build the model.

EMAP built discriminant models using benthic metrics in a stepwise model building approach. The models used three to five metrics in the Louisianian and Virginian provinces respectively, and both models used taxa richness (Engle et al. 1994, Weisberg et al. 1993). The benthic indexes were then calculated as the discriminant score of a site and standardized on a scale of 1 to 10.

Performance of the discriminant models was good in distinguishing reference from impaired sites in the calibration data: 100% for the Gulf of Mexico sites (Engle et al. 1994; n = 16 sites) and 86-93% for the Virginian Province sites (Weisberg et al. 1993; n = 33 sites). When tested with validation data collected in subsequent years, however, both sets or models failed to predict adequately and had to be redeveloped (Engle and Summers 1999, Strobel et al. 1994, 1995, Paul et al. 1999). Inclusion of several years of monitoring data in both provinces produced more robust and reliable models. In the Virginian Province, the robust calibration data set consisted of 60 sites (30 each).

An improved index was created to be applicable across a variety of estuarine environments in the Gulf of Mexico (Engle and Summers 1999). The statistical approach described in Engle and Summers (1999) proved to be applicable throughout the estuaries in the northern Gulf of Mexico. This benthic index was also validated independently by Rakoncinski (1997), who compared results of canonical correspondence analysis (CCA) with data from EMAP-E (1991-1992), using the index developed in Engle et al. 1994 (Engle and Summers 1999).

### 11.3.3 Index Derived from Multivariate Ordination

An index for biocriteria was derived by Smith et al. (2000) using multivariate ordination to derive a pollution gradient, which in turn was used to develop an index. The approach was developed with benthic macroinvertebrates from the Southern

California Bight (Smith et al. 2000; see also 11.2.1, p. 11-5), and is currently being applied to demersal fish from the same waters (Allen and Smith 2000). The approach is computationally intensive and rather complex. We will describe the result first (the index and its components), and then briefly describe how the components themselves are derived.

The central assumption of this approach is that each species has a tolerance for pollution, and that if the pollution tolerance is known for sufficiently large set of species, it is possible to infer the degree of degradation from species composition and the tolerances. This is the basis of the familiar Hilsenhoff Biotic Index (HBI; Hilsenhoff 1987) of freshwater bioassessment, as well as of several metrics in the multimetric approach. For example, capitellid polychaetes are known to be tolerant to organic pollution (BOD). The index used by Smith et al. is a weighted average tolerance value of all species found in a sample, weighted by abundance of the species:

**Equation 11-1.**

$$I_s = \frac{\sum_{i=1}^{n} a_{si}^{f} \; p_i}{\sum_{i=1}^{n} a_{si}^{f}}$$

where $I_s$ is the index value for sample $s$, $n$ is the number of species in sample $s$, $a_{si}$ is the abundance of species $i$ in sample $s$, $p_i$ is the tolerance value of species $i$, and the exponent $f$ is used to downweight extreme abundances. If $f$ is zero, then the index is not weighted by abundance (Smith et al. 2000, Allen and Smith 2000).

The index of equation (11-1) is computationally almost identical (except for the introduction of the transformation exponent $f$) to the Hilsenhoff Biotic Index. Biocriteria can be assigned to index values; for example, if the index is defined in the range from 0 (unpolluted) to 100 (severely polluted), then a criterion for Class A estuarine waters might be values $\leq 25$.

The steps below outline the derivation of the tolerance values $p_i$. A data set is required with sites that span a range from unpolluted to severely polluted. In the Southern California Bight, these were defined by sediment contaminant levels above and below the effects range median (ER-M) and effects range low (ER-L) concentrations (Long et al. 1995). Levels of a contaminant below the ER-L, between the ER-L and ER-M, and above the ER-M are rarely, occasionally, and frequently, respectively, associated with adverse effects. Impacted sites had six of eight selected contaminants (Cu, Pb, Ni, Zn, Cd, Cr, PCB, and DDT) above the ER-M. Reference sites consisted of stations lying outside of POTW discharge areas and with no more than one selected contaminant above the ER-L for a contaminant.

The data must be divided into two sets: a calibration subset and a test subset.

*Step 1. Ordination analysis of species abundance (calibration data).*

Ordination analysis produces a plot of sites in ordination space (Figure 11-5). Distances between pairs of points are proportional to the dissimilarity of species composition in the corresponding samples: samples with very similar composition will be close together in the ordination diagram. If the species are associated with the pollution gradient, the sites will define a gradient, with polluted sites at one end

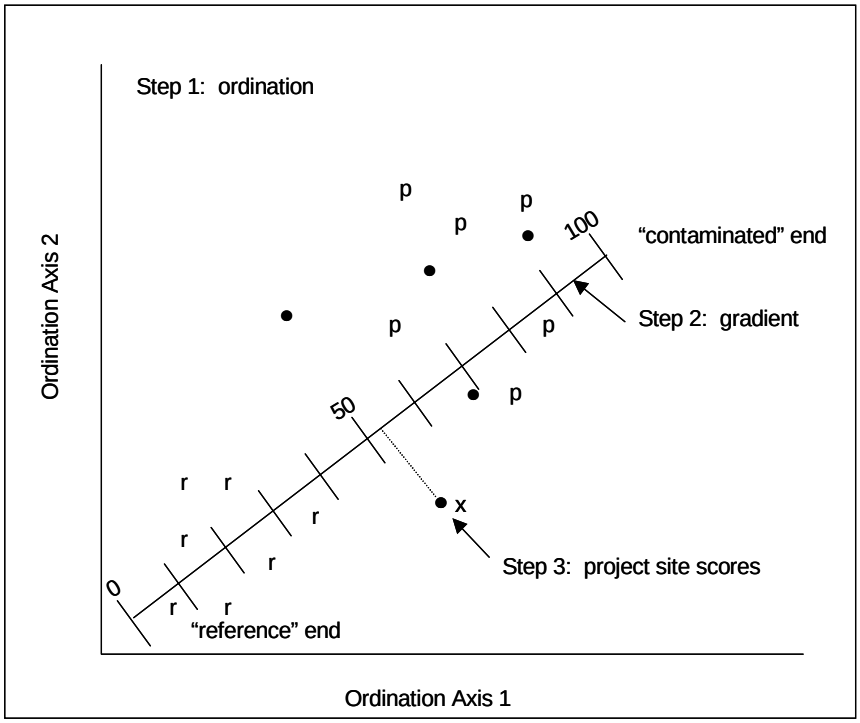



**Figure 11-5**

Steps 1-3. Establishing site scores on a contamination gradient. The gradient is established between reference ("r") and contaminated ("p") sites as plotted in ordination space. Dots are sites not designated as either reference or contaminated. The projection of site "x" on the gradient (dotted line) yields its site score. Adapted from Smith et al. 2000.

and unpolluted sites at the other (Figure 11-5).

*Step 2. Find the pollution gradient.*

The two ends of the pollution gradient are defined as the average positions in ordination space of the unpolluted and polluted sites, respectively. These ends are connected by a line, which represents the pollution gradient as expressed by the observed species compositions.

*Step 3. Project all calibration observations onto the pollution-effects gradient.*

The position of each site in the ordination space is projected onto the gradient. This projection is the site score of the calibration sites (Figure 11-5).

*Step 4. Rescale the projections.*

Site scores are scaled from 0 ("least polluted") to 100 ("most polluted").

*Step 5. Compute tolerance values for each species.*

Each species has an "average" position on the pollution gradient. These species positions are the tolerance values ($p_i$) of Equation 11-1. Site scores calculated in Step 4 give each site a position on the pollution gradient. The abundance of a species at each observation can be plotted against the site scores (Figure 11-6). The species position on the gradient, or the tolerance value $p_i$, is the abundance-weighted average position for the species over all sites.

*Step 6. Compute the f parameter.*

The f-parameter is iterated simultaneously with the $p_i$ in an optimization procedure (Smith et al. 2000).

The species tolerance scores were in turn used to predict the Benthic Response Index (BRI) according to Equation 11-1. The BRI is the position of a site on the contamination gradient, or the predicted value of the site score

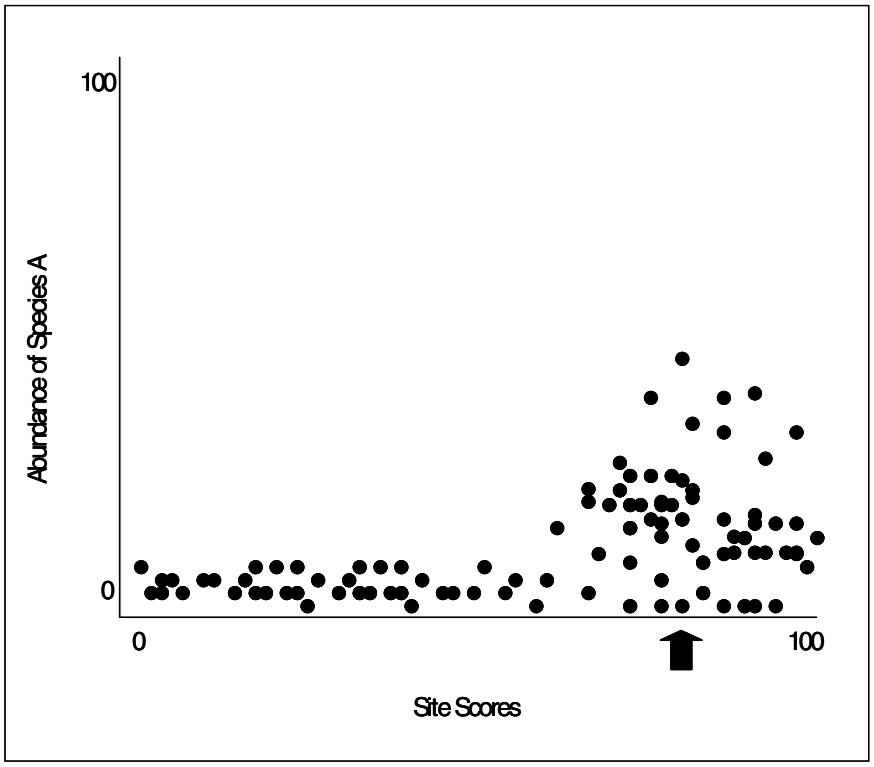



**Figure 11-6**
Step 5. Computing the tolerance values. Abundance of Species A and site scores (from Figure 11-5) of all sites where Species A occurs. The abundance-weighted average score over all sites is Species A's pollution tolerance score (arrow). This example shows a highly tolerant species, which occurs in greatest abundances at the most polluted sites. Adapted from Smith et al. 2000.

calculated in Step 4. Actual site scores (Step 4) are calculated only for the calibration data; site score is predicted as BRI for all assessment sites.

The BRI was developed separately for 3 depth zones: 10-35-m, 25-130-m, and 110-324-m. Earlier work had shown that benthic communities off Southern California could be classified by depth and sediment type (see Section 11.2.1, p. 11-5). Sediment type was secondary, and was not deemed to have a strong enough effect to justify further categorization of the data set.

The tolerance index developed by Smith et al. was then tested with the independent data (not used to develop the index). The independent test showed that the model was largely correct in predicting position along the contamination gradient. For further details of calculations and formulas, see Smith et al. (2000). The approach is currently being extended to demersal fish from the same region (Allen and Smith 2000).

Smith et al. estimated tolerance values for over 450 marine species from southern California. The BRI contamination score can be calculated for any new site from species abundance data at the site. The BRI has a range from 0 (unpolluted) to 100 (severely polluted) and biocriteria can be set at selected values for specific aquatic life uses of coastal waters of Southern California. Reference sites had BRI values < 25, and all severely contaminated sites had BRI > 36 (Smith et al. 2000). The reference values could form the basis of biocriteria for the region.