Integration and Assessment of Non-Tidal Potomac River Data: a Review of Analytical Methods Used to Assess Benthic Macroinvertebrate Data

ICPRB Report Number 97-5

October 1997

Prepared by

Michelle Fox and Claire Buchanan
Interstate Commission on the Potomac River Basin
Suite 300, 6110 Executive Blvd.
Rockville, MD 20850

and

R. Christian Jones
Biology Department, George Mason University
4400 University Drive
Fairfax, VA 22030

for

Virginia Department of Environmental Quality
629 East Main Street
P.O.Box 10009
Richmond, VA 23240

Acknowledgments

This publication was prepared by the Interstate Commission on the Potomac River Basin, an interstate compact river basin commission. Funds for this project were provided by a 604(b) grant from the Virginia Department of Environmental Quality and by the Interstate Commission on the Potomac River Basin. The authors gratefully acknowledge the invaluable literature and data searches performed by Mary Pierce and Jan Ducnuigeen of the Commission, and their assistance in preparing this report.

Disclaimer

The opinions expressed are those of the authors and should not be construed as representing the opinions or policies of the Virginia Department of Environmental Quality, the several jurisdictions, or the signatories or Commissioners to the Interstate Commission on the Potomac River Basin: Maryland, Pennsylvania, Virginia, West Virginia and the District of Columbia.

Introduction

Between 1986 and 1988 a five-fold increase in stream and river miles not attaining designated uses as dictated by the U.S. Clean Water Act was noted. This increase was in large part due to a greater reliance on biological data rather than chemical data alone (Resh et al., 1995). This increase brought to the forefront the need for selecting reliable measures of biological integrity in streams and rivers.

In this report we focus on the accuracy of these measures as well as the diversity of bioassessment methods that now exist and their potential incompatibilities. The EPA's Rapid Bioassessment Protocols and the USGS's National Water Quality Assessment Program (NAQWA) were two attempts to standardize bioassessment methods in the United States (Diamond et al., 1996). Nevertheless, many bioassessment techniques have been developed in addition to these to accommodate specific research objectives and ecoregional variability among other things. Furthermore, environmental agencies have taken the steps towards customizing common methods to more accurately discriminate between impacted and nonimpacted sites in their specific region. Some of these "customized" methods have been based on scientists' long-term experience with benthic populations in their region, while other methods have used various statistical tools to test the discriminating power of different combinations of metrics. Over the years, a number of monitoring programs have developed their own sampling methods, measures of biological integrity, and techniques for data interpretation. As a result, these groups find it very difficult to share data. Researchers are alarmed by sampling method discrepancies that potentially yield different representations of biological populations and by metrics that may yield conflicting interpretations at the same sites (Diamond et al., 1996). Many researchers are also reluctant to combine data from different monitoring programs because of the "quality" (i.e. accuracy and precision of the data) of data produced is often documented.

Attempts to integrate data sets into a uniform database has raised a number of the data compatibility issues mentioned above. In addition to differences in the raw data resulting from using different sampling gear, sampling from different habitats, etc, differences in the way various environmental agencies analyze these data make grasping the entire basin's health an elusive goal. Metrics and multimetrics chosen for analysis often are selected for their sensitivity to specific types of impairment. Thus, one bioassessment method could conclude that a steam is impaired while another may not detect impairment at all. Additionally, multimetric scores are often incomparable because the scores are evaluated on different scales. Finally, some agencies do not evaluate their data using a multimetric, but instead, evaluate their streams on metric by metric basis. These streams are not readily comparable to streams in other watersheds evaluated by a multimetric that produces a single score.

Diamond et al., and the Interagency Task Force on Water Quality Monitoring (ITFM) have recommended an approach termed the "Performance Based Method System"(PBMS) to resolve data compatibility issues. This system encourages the development of "data quality characteristics" for every available bioassessment method. "Data quality characteristics" include such information as data precision, method biases, method sensitivity, and the range of conditions over which a method yields satisfactory data. Knowing this type of information allows researchers to combine data collected with different methods if they produce data of similar quality. Additional sampling to understand variabilities within methods is sometimes required to determine the "data quality characteristics", but once they are established for individual monitoring protocols monitoring results can be confidently combined with other comparable monitoring results.

Report Objectives

For purposes of conducting a basin-wide assessment, the Interstate Commission on the Potomac River Basin has been integrating biological datasets from various researchers and monitoring groups from around the basin. One objective of this report is to take the first steps towards resolving some of the data compatibility issues among monitoring programs in the Basin. The report compiles a list of benthic assessment methods currently in use in the Potomac River Basin and their "data quality characteristics as encouraged by Diamond et al. (1996) and the ITFM. In other words, for each bioassessment method used to collect and/or analyze data in the basin, we have reviewed those components of the bioassessment method important in determining the comparability of different data sets as well as the comparability of metrics/multimetrics produced from these data sets. Accompanying each bioassessment method description is a table containing information on the specific data collection methods used (season sampled, habitat sampled, gear method, sampling frequency, subsampling, and taxonomic level to which data identified).

A second objective of the report is to compile a list of recently sampled sites of ecoregions represented in the Potomac River Basin and include status (i.e. impaired/unimpaired) as determined by the agencies monitoring the region. At the start of the project, we hoped to attain enough raw data and metric data to first rigorously and independently select reference sites for the three main regions in the Potomac (Coastal, Piedmont, and Mid-Atlantic Highlands) and second, apply specific scoring criteria (e.g. DEQ's) to the metric data. However, key information for some monitoring programs was not forthcoming in time (e.g. latitude and longitude which is needed to determine ecoregion, metric documentation to explain metric calculations, and scoring criteria). The report provides a list of potential reference sites and impaired sites recently identified in ecoregions represented in the Potomac River Basin. These will eventually be pooled and used for basin-wide assessments. A larger number of reference sites will increase the statistical power to differentiate between impaired and unimpaired sites regardless of the assessment method used.

Stream Classification and Selection of Reference Sites

Classification of sites according to ecoregion is fundamental to the EPA's Rapid Bioassessment Protocols (Plafkin et al. 1989). Ecoregions are large geographic areas that are intended to group contiguous, naturally similar ecosystems. Biological assemblages are presumed to be a function of their ecoregions characteristics, and to reflect physiographic, ecosystem and other differences. Biological variability between sites is expected to be significantly lower within ecoregions, facilitating development of "regionally attainable, quantitative, chemical, and biological goals" (Hughes and Larsen 1988). Olmernik (1987) developed a map of ecoregions in the conterminous United States which was recommended by the U.S. EPA (Plafkin et al. 1989) and which has been since used extensively to stratify data.

The Olmernik ecoregions and subecoregions overlapping the Potomac River Basin are shown in Figure 1 and can be obtained on the world wide web at http://www.epa.gov/nsdi/projects/ecoreg.html.

The premise underlying the ecoregion concept - that biological communities at reference sites are similar within ecoregion and dissimilar to those outside the ecoregion - has been tested in ecoregions in the Potomac River Basin, sometimes with surprising results. For the Mid-Atlantic Highlands area, Smith and Voshell (1997) used reference sites to determine that data from subecoregions were significantly ($F=3.77$, $p\text{-value} < 0.0001$) different but not enormously so. Separation of reference site data grouped by ecoregions was also significant but, interestingly, not as significant as between subecoregions ($F=2.47$, $p\text{-value} < 0.0001$). Closer examination of the reference site data reveal that the North Central Appalachian ecoregion (62), the three Blue Ridge Mountain subecoregions (66A-C) and three of the four Central Appalachian Ridge and Valley subecoregions (67B-D) cluster. Similarly, the two Central Appalachian subecoregions (69A-B) and the Northern Appalachian Plateau and Uplands ecoregion (60) cluster. The two clusters separate weakly. However, the Limestone and Dolomite Valley subecoregion of the Central Appalachian Ridge and Valley ecoregion (67A) and the Western Alleghany Plateau stand apart. NAWQA is also aware that the Limestone and Dolomite Valley subecoregion stands apart. However, NAWQA refers to this subecoregion as the Calcareous Valley Region or the Great Valley subprovince. Smith and Voshell (1996) could have followed this cluster scheme, but they concluded that for their purpose (i.e. developing a Highlands multimetric), classifying streams into their original ecoregions without further classification into subregions was best, with the possible exception of subregion 67A. Reference sites were not empirically determined with habitat and land use criteria, but rather with best professional judgement of monitoring program staff.

Using ecoregions as a basis for analysis assumes that biological communities are similar within these ecoregions and that their assemblages are in some way a function of ecoregion characteristics. However, Resh et al. list a number of researchers (Cummins et al., 1989; Corkum, 1990; Richards et al., 1993; Sweeney, 1993) who have provided evidence that benthic assemblages are dominated by local conditions rather than regional ones (Resh et al., 1995). The original ecoregions that are defined by Omernik (1987) may overlap a number of distinct physiographic features that individually support very different benthic assemblages. The spatial extent of each ecoregion as specified by Omernik (1987) may not minimize expected variation resulting from natural environmental differences, thus necessitating a finer delineation to yield biologically homogeneous regions. It may be necessary to include instream and riparian characteristics to adequately divide stream systems into groupings with hhomogeneous biological populations.

A step closely related to selecting biologically homogeneous streams is the selection of reference sites (sites that most closely represent what the biotic population of a stream would look like without impairment). Reference sites have been identified by a number of researchers and agencies and have been compiled in Appendix A. Many of these sites are selected by professionals who have been working in the region and thus have prior knowledge of the stream reaches where impairment is most likely to have occurred. Preliminary metrics and multimetrics are often calculated for these sites in order to eliminate outliers (sites that may not be visibly impaired but whose biological indicators suggest less than ecologically healthy conditions). We have described how particular groups around the Basin have approached reference site selection in the sections below.

An alternative to grouping reference sites by ecoregion is being used in the United Kingdom. Reference sites are instead grouped by ordination of their faunal communities. These ordinations reveal the relationship between environmental characteristics and invertebrate assemblages. Then, by observing the environmental characteristics at impacted sites, one can predict the resulting invertebrate assemblage. This approach has received little attention in the U.S. (Resh et al., 1995).

Metrics As Analytical Tools

Metrics are responsive at different ranges of impairment. Weaknesses that an individual metric may have for indicating impairment, can be minimized when used in conjunction with other metrics (Ohio EPA 1987a,b). Metrics that have a monotonic response to a biotic integrity gradient are considered to be the best metrics for assessing impairment. Metrics that show a nonmonotonic response can still be useful for assessing impairment but must evaluated in light of other metrics. A metric that is only responsive at the highest levels of impairment can be useful for discriminating *between* impaired sites so that restoration priorities can be better set. Metrics that are able to discriminate between sites at the extremes of impairment are important for a thorough bioassessment (Plafkin et al, 1989). Additionally, by using metrics that overlap in their ranges of sensitivity, the bioassessment conclusion is reinforced (Karr, 1991) (i.e. if four out of ten metrics used had overlapping sensitivity ranges then an impaired station would have the chance to receive four consistently low scores, thus bringing down the overall metric score). By integrating multiple metrics of biological degradation, a broader range of impairment can be detected (Barbour et al., 1995).

Table 1 lists the metrics currently being used in bioassessment methods in the Potomac River Basin including comments on their specific sensitivities to impairment and usefulness in analyses. This table is still in development but will ultimately be very useful for understanding the relative sensitivity of metrics to impairment.

Environmental Protection Agency’s Rapid Bioassessment Protocols

EA Engineering, Science and Technology and North Carolina Division of Environmental Management (DEM) conducted a pilot study of the Ararat and Mitchell Rivers. Using data attained from a 100 organism subsample of two kick net samples in riffle areas, thirteen “candidate” metrics were applied to the data. Cluster analysis was used to determine the “unique information contributed by each metric to an integrated bioassessment”. Seven of these metrics (in bold) appeared to add information to the biological assessment whereas the other six metrics were found to be somewhat redundant.

Table 1 Metrics currently used in bioassessments in the Potomac River Basin

Metric	**Calculation of metric/What Metric Indicates**	H	P	C	**Comments**
EPT Index	Absence of the pollution-sensitive taxa Ephemeroptera, Plecoptera, and Tricoptera may indicate biological impairment (Plafkin et al., 1989).		X	X	--LOW variability (Resh, 1988) --significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)
Taxa Richness	Low taxa richness may indicate biological impairment (Plafkin et al., 1989).		X	X	--LOW variability (Resh, 1988) --this metric should only be compared with taxa richness metrics from subsamples taken from similar stream orders or drainage area (Montgomery County, 1996) --significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)
Percent Contribution of the Dominant Taxon	Communities dominated by few taxa indicate impairment (Montgomery County, 1996). Measures redundancy and evenness. Major abundance by a single taxon can indicate impairment (Barbour, 1992).		X	X	--significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)
Abundance of scrapers/(scrapers +filtering collectors)	A predominance of a particular feeding type may indicate an over abundance of a particular food source. For example, organic enrichment would increase the filtering collectors. Scrapers however, increase with an increase in diatoms and decrease with an increase in filamentous algae and aquatic mosses (Montgomery County, 1996).		X		
Coefficient of Community Loss	Measures the loss of benthic taxa between reference station and station of interest (Barbour, 1992).			X	
Modified Hilsenhoff Family Biotic Index (FBI)			X		--significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)

Ratio of EPT and Chironimidae Abundances			X		--significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)
Shredders/Total # Organisms	Shredders are sensitive to riparian zone impacts		X		
Sorenson’s Index			X		--significantly correlated with percentage of low intensity land use (Jones and Kelso, 1997)
# Mayfly families		X			
% EPT	Percent abundance of mayfly nymphs, stonefly nymphs, and caddisfly larvae and pupae (Smith and Voshell, 1997)	X	X	X	
% Mayflies	Percent abundance of mayfly nymphs (Smith and Voshell, 1997)	X			
% 5 Dominant taxa	% Abundance of the 5 most abundant taxa combined (Smith and Voshell, 1997)	X			
Simpson Diversity Index	Integrates richness and evenness into a measure of general diversity (Smith and Voshell, 1997)	X			
# Intolerant Taxa	Number of macroinvertebrate families with tolerance values of 5 or less (Smith and Voshell, 1997)	X			
HBI (Hilsenhoff Biotic Index)	Weighted sum of total taxa by pollution tolerance (Smith and Voshell, 1997)	X		X	
% Scrapers	Percent abundance of macroinvertebrates scraping and feeding upon periphyton (Smith and Voshell, 1997)	X			
% Haptobenthos	Percent abundance of macroinvertebrates requiring clean, coarse, firm substrates (Smith and Voshell, 1997)	X			

(Hydropsyche+Cheumatopsyche)/Total EPT individuals			X		
% Chironomidae				X	
% Diptera				X	
% Non-Insect				X	
% Oligochaetes				X	
North Carolina Biotic Index				X	
Shannon-Wiener Index				X	

“Candidate Metrics” (“Best” metrics in **BOLD)**

- **Taxa Richness**
- **%Contribution of Dominant Taxon**
- Shredders/Collectors
- Pinkham and Pearson Community Similarity Index
- **Scrapers/Filterers**
- Jaccard Coefficient of Community Similarity
- Weighted Pinkham and Pearson Community Similarity Index
- **Community Loss Index**
- Filterers/Gatherers
- **Abundance EPT/Abundance Chironomid**
- **EPT Index**
- **Modified Hilsenhoff Biotic Index (HBI)**

These seven metrics are the foundation of the RBP methods. An eighth metric was added (ratio of shredder functional feeding group to total number of organisms) with the addition of a CPOM sample. The Environmental Protection Agency’s Rapid Bioassessment Protocols are often the basis for multimetrics developed by state and local agencies. Table 2 is a summary of the EPA RBP sampling methods. We will provide a brief description of sampling methods for each bioassessment method since sampling methods are largely chosen according to the data analysis objectives and in the case of analyses based on historical data sets, the data analysis objectives are largely constrained by the chosen sampling method.

Variability in metrics (Table 2) was accounted for in RBP II and RBP III by the Biological Condition Scoring Criteria (Table 3). Ten to twenty percent differences between the station of interest and the reference station for the metrics, taxa richness, FBI, and the EPT index were considered nominal. Metrics that are ratios are inherently more variable and thus require broader scoring increments to more accurately characterize the site (Plafkin et al., 1989).

Once scoring criteria have been applied and the metric scores totaled for both the reference site and the site of interest, “biological condition categories” are assigned based on the criteria in Table 4.

EPA’s RBPS emphasize community metrics reflecting habitat and water quality impairment. Barbour et al. (1992) tested seventeen metrics for their ability to discriminate between macroinvertebrate communities in two ecoregion groupings, montane and valley/plains. The EPA’s RBP metrics are eight of the seventeen. Metrics were tested using reference station data from 10 ecoregions in Oregon, Colorado, and Kentucky. Statistical analyses included coefficient of variance, analysis of variance, correlation, principal components analysis, and step-wise discriminant analysis. Taxa richness and the EPT index were found to be highly correlated with each other. The authors however note that this redundancy may have occurred because in this particular database mayflies, stoneflies, and caddisflies constituted a major proportion of the

Table 2 Summary of EPA's Rapid Bioassessment Protocols II & III

HABITAT SAMPLED	riffle/run sample; course particulate organic material (CPOM) sample recommended (used only for calculation of ratio of shredders to total number of individuals collected)
GEAR METHOD	**Riffle/run sample:** 2 - 1m^2 samples using a kicknet. One sample should be from an area of fast current velocity and the other from an area of slow current velocity. The two samples are then composited. **CPOM sample:** Upper surface of litter accumulation in depositional areas (i.e. leaf packs, shore zones).
SUBSAMPLING	**Riffle/run sample:** 100 count subsampling procedure (Hilsenhoff,1987b) **CPOM sample:** A representative sampling of 20-60 organisms
TAXONOMIC LEVEL	RBP II: Family level RBP III: Genus or species level
BASIS FOR METRIC SCORING	evaluations of pilot study results; compliance monitoring requirements; discussions with aquatic biologists

Table 3 Scoring Criteria For Characterization of Biological Condition (Taken from Plafkin et al.,1987).

Metric	**Score '6'**	**Score '3'**	**Score '0'**
Taxa Richness(a)	>80%	40%-80%	<40%
Family Biotic Index(b)	>85%	50%-85%	<50%
Ratio of scrapers to filtering collectors(a,c)	>50%	25%-50%	<25%
Ratio of EPT abundance to Chironomid abundance(a)	>75%	25%-75%	<25%
% Contribution of Dominant Family(d)	<30%	30%-50%	>50%
EPT Index(a)	>90%	70%-90%	<70%
Community Loss Index(e)	<.5	.5-4.0	>4.0
Ratio of Shredders to Total(a,c)	>50%	25%-50%	<25%

(a) Score is a ratio of study site to reference site x 100
(b) Score is a ratio of reference site to study site x 100
(c) Determination of Functional Feeding Group is independent of taxonomic grouping
(d) Scoring criteria evaluate actual percent contribution, not percent comparability to the reference station.
(e) Range of values obtained. A comparison to the reference station is incorporated in these indices.

Table 4 Bioassessment (Taken from Plafkin et al., 1989)

% Comp. to Reference Score[(a)]	**Biological Condition Category**
>79%	Non-impaired
29%-72%	Moderately impaired
<21%	Severely impaired

(a) Percentage values obtained that are intermediate to the above ranges will require subjective judgement as to the correct placement. Use of the habitat assessment and the physicochemical data may be necessary to aid in the decision process.

total population. Other databases or databases from other ecoregions would most likely not show this redundancy. Redundancy was additionally found between the Pinkham and Pearson Index and the Jaccard Coefficient. Boyle showed the Pinkham and Pearson Index to be more consistent in detecting impairment (Boyle, 1988). Metrics that were ratios showed high variation about the mean (i.e. shredders/total, scrapers/filterers, EPT/Chironomidae abundances). The use or disuse of each of these metrics should be done on an ecoregion by ecoregion basis since variation around the mean was quite different for each of the ten ecoregions (Barbour et al., 1992).

Progress to Date in Developing Metrics

Mid-Atlantic Highlands Region

Recently, Eric Smith and J. Reese Voshell at Virginia Polytechnic Institute made new inroads into the development of biological indicators of ecological condition for the Mid-Appalachian Highlands. The multimetric they developed is called the Macroinvertebrate Aggregated Index for Streams (MAIS) (Smith and Voshell, 1997). The following summarizes the steps Smith and Voshell took to develop a bioassessment method in the Mid-Atlantic Highlands region.

Stream Classification:

Ecoregions and subecoregions as defined by Omernik(1987) were analyzed for their ability to yield comparable biological assemblages.

Reference sites:

Data for this initiative came from water quality agencies in Maryland, Pennsylvania, Virginia, and West Virginia. Biological condition (reference, impacted, or unknown) was assigned to each station in the biological database by the state regional biologists who supplied the data. The designations were based on knowledge of chemical and physical attributes of the site rather than knowledge of the macroinvertebrate populations.

Source of Functional Feeding Groups and Habit:

Functional Feeding Groups and Habit designations were primarily taken from Merritt and Cummins (1996). Smith and Voshell did, however, create one new for habit which they called “crawlers”. They defined crawlers as those organisms that “move about regularly, but slowly, on or within spaces in solid substrata that is relatively clean.”(Smith and Voshell, 1997). Examples of crawlers include most stoneflies and some mayflies. While Merritt and Cummins (1996) categorizes these organisms as clingers or sprawlers, Smith and Voshell find the "crawler" habit to more accurately describe their movements and habitat preferences.

Tolerance Values:

For metric calculations, Hilsenhoff's family level tolerance values (Hilsenhoff, 1988) were used with some modifications based on Smith and Voshell's research and the research of state regional biologists in Virginia.

The following table (Table 5) summarizes the data that Smith and Voshell used to develop their bioassessment method. They resolved apparent data incompatibilities by simply limiting the data they used for their analyses to very specific kinds of data. For instance, because population samples can be biased towards particular species or sizes of organisms depending on the type of gear used for sampling, Smith and Voshell simply limited their database to samples collected using open net devices.

Table 5 Summary of data used by Smith and Voshell for analyses (Smith and Voshell, 1997)

SEASON SAMPLED	Only used data collected between June and early October in order to reduce seasonal variation in macroinvertebrate samples.
HABITAT SAMPLED	Only used data collected from riffle areas of wadeable streams.
GEAR METHOD	Only used data collected using open net devices (i.e. D-frame dip net or kick screen) used on natural substratum.
SAMPLING FREQUENCY	Not applicable
SUBSAMPLING	Data were derived from 200 organism or 100 organism subsamples. In West Virginia, data represented sampling from 1 square meter for approximately 60 seconds.
TAXONOMIC LEVEL	Family level or truncated to the family level when data was collected with more specificity.

Data Evaluation:

Smith and Voshell began with 69 "candidate" metrics and through a rigorous selection process narrowed the list of metrics down to 10 "candidate" metrics. This "narrowing" was done through a variety of methods. Metric data were plotted using empirical probability plots that allowed comparisons of the medians of impact versus reference plots as well as showed the separation between impacted and reference plots. Metrics that had a robust coefficient of variation among reference sites of less than 50 percent and the separation statistic between reference and impacted sites received values greater than +1 or less than -1 were seen as potential good candidate metrics. Step-wise discriminant analysis was used to eliminate redundancy among metrics per ecoregion. Additionally, in selecting metrics, an attempt was made to maintain a balance among six categories that the metrics can be divided into: richness, composition, balance, tolerance, trophic relations, and habits. The sixth category, habits, was developed by Smith and Voshell. By Merritt and Cummins' definition "sprawlers" inhabit the surface of fine sediments. However, they give some mayflies and stoneflies this designation whose habits do not truly fit this definition. Thus, a new habit was established, "crawlers", defined as "moving around slowly in clean, firm substratum such as loose rocks, leaves, and branches" (Smith and Voshell, 1996). The two categories of habits are haptobenthos and herpobenthos. Haptobenthos is now calculated as the sum of clingers and crawlers and herpobenthos as the sum of sprawlers and burrowers.

Using the multiple methods described above, the ten "best" metrics were selected for further consideration (Table 6). The authors noted that since these metrics are highly correlated with other metrics, substitution is possible.

Table 6 "Best" Metrics (Taken from Smith and Voshell, 1997)

Category	Metric	Explanation	Expected response to perturbation
Richness	EPT Index	Number of mayfly, stonefly, and caddisfly families	neg
Richness	# Mayfly Families	Number of mayfly families	neg
Composition	% EPT	Percent abundance of mayfly nymphs, stonefly nymphs, and caddisfly larvae and pupae.	neg
Composition	% Mayflies	Percent abundance of mayfly nymphs	neg
Balance	% 5 Dominant Taxa	Percent abundance of the 5 most abundant taxa combined	pos
Balance	Simpson Diversity Index	Integrates richness and evenness into a measure of general diversity	neg
Tolerance	# Intolerant Taxa	Number of macroinvertebrate families with tolerance values of 5 or less	neg
Tolerance	HBI (Hilsenhoff Biotic Index)	Weighted sum of total taxa by pollution tolerance	pos
Trophic	% Scrapers	Percent abundance of macroinvertebrates scraping and feeding upon periphyton	neg
Habit	% Haptobenthos	Percent abundance of macroinvertebrates requiring clean, coarse, firm substrates	neg

The above metrics were chosen based upon the assumption that stations within ecoregions are comparable. Smith and Voshell conducted additional analyses to see if there was a more suitable way to divide stations than by ecoregion. Ecoregions and subregions are listed in Table 7. Data were analyzed by the original classification, ecoregion, and by rearranging subregions into "bioregions". They concluded that the reclassification of streams by "bioregion" did not improve results. Also, removing classification schemes altogether did not improve results. Their results did indicate that separating the Limestone and Dolamite Valley subregion from the Central Appalachian Ridges and Valley ecoregion may result in a better classification.

Table 7 Ecoregion and Subregion in the Mid-Atlantic Highlands (Taken from Smith and Voshell, 1997)

Ecoregion	**Subregion**
Northern Appalachian Plateau and Uplands	None
North Central Appalachians	None
Blue Ridge Mountains	Non-Calcareous
	Shale-Dominated Ridges
	Interior Plateau
Central Appalachian Ridges and Valleys	Limestone and Dolomite Valleys
	Shale and Slate Valleys
	Sandstone Ridges
	Shale Ridges
Central Appalachians	Forested Hills and Mountains
	Uplands and Valleys of Mixed Land Use
Western Alleghany Plateau	None

Metric Transformation:

Smith and Voshell investigated several different ways to transform metrics into unitless scores in order to combine scores into a multimetric. They found that assigning metric values to one of four unitless scores provided too many categories for data with so much variation and overlap. They also chose to use sequential metric scoring (i.e. 0,1,2) to eliminate score inflation that occurs when metrics are scored with only even or odd numbers as is often the case. This score inflation falsely implies the multimetric has a high sensitivity to impairment.

The methods they focused on are summarized in Table 8 and Figure 2. Method 1 stands apart from the other three methods because scoring criteria are based on reference site scores alone. Smith and Voshell feel that scoring criteria can be most accurately assigned when based on data from both the reference sites and impacted sites rather than just reference sites. The second method uniquely assigns the lowest score to metric values less than or equal to the fence (RF1) for the reference sites (the fence is the metric value among reference sites that is observed the most if the data is normal). The third method was developed to broaden the range of metric values receiving intermediate scores by assigning the highest score to metric values falling either above the 75th percentile of impacted sites or the 25th percentile of the reference sites, whichever is highest. This is to accommodate for those instances when there is considerable overlap between reference and impacted sites. The intermediate score is then assigned to metric values falling above the 25th percentile up to the criterion for the highest score. Unlike the first three methods, method 4 does not assign criteria to metric values based on quartiles but instead scales the metric based on the range of scores (reference maximum - impact minimum).

Table 8 Methods for assigning scores to metrics. Taken from Smith and Voshell (1997) RQ1, RF1, IQ1, IQ3 are defined in Figure 2

Method	Score ‘0'	Score ‘1'	Score ‘2'
1	<=.5 *RQ1	between .5*RQ1 and RQ1	>RQ1
2	<=RF1	between RF1 and IQ3	>IQ3
3	<=IQ1	>IQ1, <= max(RQ1, IQ3)	>IQ3
4	(metric-impact min)/(reference max-impact min)		

Tests of these four transformation/standardization methods showed that the methods performed comparably for accurately classifying reference and impact sites.

Multimetric development:

Contrary to what they suspected, Smith and Voshell found that the number of individual metrics, whether 5, 6, 7, 8, 9 , or 10 and regardless of which transformation method used, resulted in little difference in the misclassification rates (error rate for sites being designated in the appropriate reference or impacted condition) for either impacted or reference sites. However, they do feel that it is best to use a high number of metrics as long as the misclassification rate does not increase since then the multimetric may be capable of responding to a greater diversity of impacts.

Biocriteria development:

The threshold to discriminate impaired versus unimpaired biological conditions was determined by taking the average of the multimetric mean for impacted sites and the multimetric mean for reference sites. Beyond this, they recommend that two categories be established on either side of this threshold. Thus, multimetric values greater than of equal to the third quartile or reference sites would then be considered “very good” and multimetric values less than or equal to the first quartile of impacted sites are considered “poor”. However, they feel these cutoffs need to be further refined.

Based on the classification rate and biological considerations, Smith and Voshell recommend of the six multimetrics tested that a multimetric containing the metrics listed in Table 6 with the exclusion of “% EPT” be used. They also recommend that individual metric scores be transformed according to Method 2 above. Table 9 shows the metric scoring criteria using Method 2.

Table 9 Established cutoff values using method 2. Taken from Smith and Voshell (1997)

Metric	Score ‘0'	Score ‘1'	Score ‘2'
EPT	<=2	>2	>7
% EPT	<=20.25	>20.25	>64.65
% 5 most dominant	>=100	<100	<79.13
HBI	>=5.56	<5.56	<4.22
# mayfly	<=0	>0	>3

% haptobenthos	<=51.98	>51.98	>83.26
% mayfly	<=.1	>.1	>17.52
# intolerant	<=1	>1	>9
% scraper	<=.1	>.1	>10.7
Simpson	<=.66	>.656	>.823

Piedmont Region

Table 10 summarizes the sampling methods used by the Maryland Department of Environmental Protection in Montgomery County, MD.

Table 10 Summary of data collected by Maryland Department of Environmental Protection (Montgomery County, MD)

HABITAT SAMPLED	One riffle-pool-riffle reach at a minimum.
GEAR METHOD	2 - 1m^2 samples using a 530 micron kicknet from a 1 m^2 collection area. One sample should be from an area of fast current velocity and the other from an area of slow current velocity. The two samples are then composited.
SAMPLING FREQUENCY	Once during the spring (March 15 to April 15) and once during the fall (October 15 and November 15).
SUBSAMPLING	Sample is evenly distributed over 20 square grid. Grids are randomly selected for subsampling. The grid containing the 200th organism is completely picked.
TAXONOMIC LEVEL	Genus level or to the lowest positively identified taxonomic level.

MD DEP assessed the following “candidate” metrics for incorporation into a multimetric.

<u>“Candidate” Benthic Metrics</u>

1. Taxa Richness
2. Density
3. Percent Contribution of the Dominant Taxon
4. EPT Index (total number of distinct taxa and total number of individuals from the EPT orders)
5. Biotic Index (DEP uses the New York Biotic Index values (Bode, 1991) for genus tolerance values and modifies the Maryland Family Level Biotic Index values (Primrose, 1993) for use as family level values (The New York Biotic (0-10) was on a different scale than the Maryland Family Level Biotic Index (0-5). To make the scales comparable, the Maryland Family Level Biotic Index was multiplied by 2).
6. Total Hydropsychidae/EPT
7. Abundance of scrapers/(scrapers+filtering collectors)

MD DEP consider the total number of individuals from the orders Ephemeroptera, Plecoptera, and Trichoptera to be important as well as the traditional EPT index (number of distinct taxa from the EPT orders). By calculating both, sites can be distinguished that have the same number of EPT taxa yet dramatically different abundances within these orders (Van Ness, 1996). The ratio of total Hydropsychidae to the total number of individuals within the EPT orders further helps to distinguish between impaired and nonimpaired streams; while impaired and unimpaired sites may have the same number of EPT individuals, the impaired stream will tend to have a higher ratio of total Hydropsychidae to total EPT individuals. Van Ness notes that since filtering collectors can be sensitive to toxicants bound in the FPOM, this metric may not be suitable for point-source monitoring (Van Ness et al, 1996).

Metric Evaluation:

Metrics were evaluated within two subecoregions, channery silt loam and silt loam, and by stream order groupings (1st and 2nd order streams; 3rd and 4th order streams). Metrics were evaluated individually using box and whisker plots. Metrics were selected that had low variability among designated reference sites but showed a wide range of values either above or below reference site values when populations from impaired sites were sampled.

Reference stream selection:

Reference stream station IBI scores, categorized by soil type and stream order, were analyzed for outliers (outliers indicated possible impairment). Once reference station were removed from the database, individual metric were reevaluated and given new point scores.

Chosen Metrics

1. Taxa Richness
2. EPT Index
3. Proportion of the dominant taxon
4. Biotic Index
5. Abundance of scrapers/(scrapers+filtering collectors)
6. Abundance of shredders/total
7. (Hydropsyche + Cheumatopsyche)/total EPT individuals
8. Total EPT individuals/ total individuals in the subsample

Comments on refinements:

Metric 2 from the list of "candidate" metrics, density, was dropped because counting all individuals in a sample was not practical. Metric 3, Total Hydropsychidae/EPT individuals, was modified to produce metric 7 from the list of "chosen metrics", (Hydropsyche + Cheumatopsyche)/total EPT individuals. This metric was modified because the Hydropsychidae genus Diplectrona, a relatively sensitive genera, was found in many reference streams. Since Hydropsychidae, containing mostly relatively pollution tolerant taxa was being used as an indicator of pollution, the presence of Diplectrona at unimpaired sites was reducing the discrepancy between metric scores of impaired and unimpaired sites. The eighth metric, Total EPT individuals/ total individuals in the subsample, was added since less impaired sites will not only have an increase in the number of EPT taxa but also an increase in the number of individuals within these taxa (Van Ness, 1996).

Scoring Criteria:

Metric values decreasing in the presence of stressors:

Metric values that fall above the 25th percentile of reference site metric scores receive a point score of 5. Metric values falling below the 25th percentile were equally divided between point scores 3 and 1.

Metric values increasing in the presence of stressors:

Metric values that fall below the 75th percentile of reference site metric scores receive a point score of 5. Metric values that fall above the 75th percentile are equally divided between point scores 3 and 1.

Tables 11-14 present scoring criteria per soil type and stream order.

Biological Integrity Classes:

Reference stream station IBI scores were graphed according to soil type and stream size with box and whisker plots. The lowest median score among reference stations when soil types and stream orders were grouped together was the cutoff for "excellent" stations. IBI scores for all stations (reference and impaired) below the "excellent" cutoff were equally divided among "good", "fair", and "poor" designations (Table 15).

Table 11 Scoring Criteria for Channery Silt Loam Region, 1st and 2nd Order Streams (Taken from Van Ness, 1996). Scoring criteria are based on 1995 and 1996 reference streams.

Metric	**Score '5'**	**Score '3'**	**Score '1'**
Taxa Richness	>19	10-19	<10
Biotic Index	<3.7	3.7-6.8	>6.8
Ratio of scrapers to scrapers+filtering collectors	>24%	13%-24%	<13%
Proportion of hydropsyche and cheumatopsyche/total EPT individuals	<10%	10%-54%	>54%
Proportion of dominant taxa	<47%	47%-73%	>73%
Total EPT taxa	>10	5-10	<5
Proportion of total EPT individuals	>57%	29%-57%	<29%
Proportion of shredders	>5%	3%-5%	<3%

Table 12 Scoring Criteria for Channery Silt Loam Region, 3rd and 4th Order Streams (Taken from Van Ness, 1996). Scoring criteria are based on 1995 and 1996 reference streams.

Metric	**Score '5'**	**Score '3'**	**Score '1'**
Taxa Richness	>18	10-18	<10

Biotic Index	<4.2	4.2-7.0	>7.0
Ratio of scrapers to scrapers+filtering collectors	>6%	4%-6%	<4%
Proportion of hydropsyche and cheumatopsyche/total EPT individuals	<27%	27%-63%	>63%
Proportion of dominant taxa	<50%	50%-75%	>75%
Total EPT taxa	>11	6-11	<6
Proportion of total EPT individuals	>33%	17%-33%	<17%
Proportion of shredders	>9%	5%-9%	<5%

Table 13 Scoring Criteria for Silt Loam Regions, 1st and 2nd Order Streams (Taken from Van Ness, 1996). Scoring criteria are based on 1995 and 1996 reference streams.

Metric	**Score '5'**	**Score '3'**	**Score '1'**
Taxa Richness	>23	12-23	<12
Biotic Index	<3.86	3.86-6.93	>6.93
Ratio of scrapers to scrapers+filtering collectors	>20%	10%-20%	<10%
Proportion of hydropsyche and cheumatopsyche/total EPT individuals	<15%	15%-57%	>57%
Proportion of dominant taxa	<33%	33%-67%	>67%
Total EPT taxa	>11	6-11	<6
Proportion of total EPT individuals	>55%	28%-55%	<28%
Proportion of shredders	>5%	3%-5%	<3%

Table 14 Scoring Criteria for Silt Loam Region, 3rd and 4th Order Streams (Taken from Van Ness, 1996). Scoring Criteria are based on 1995 and 1996 reference streams.

Metric	**Score '5'**	**Score '3'**	**Score '1'**
Taxa Richness	>22	11-22	<11
Biotic Index	<3.78	3.78-6.89	>6.89

Ratio of scrapers to scrapers+filtering collectors	>18%	9%-18%	<9%
Proportion of hydropsyche and cheumatopsyche/total EPT individuals	<17%	17%-59%	>59%
Proportion of dominant taxa	<47%	47%-74%	>74%
Total EPT taxa	>12	7-12	<7
Proportion of total EPT individuals	>55%	28%-55%	<28%
Proportion of shredders	>5%	3%-5%	<3%

Table 15 Biological Integrity Classes

Score	**Rating**
46-50	Excellent
34-45	Good
22-33	Fair
10-21	Poor

Other researchers have contributed to the development of bioassessments in the Piedmont Region in addition to the bioassessment research conducted by MD DEP. A benthic macroinvertebrate survey is one part of The Rapid Stream Assessment Technique (RSAT), developed by the Department of Environmental Programs, Metropolitan Washington Council of Governments(Galli, 1997). The RSAT survey is a much more rapid analytical method than most of the other methods being used in the Basin. It most closely resembles the Environmental Protection Agency's Rapid Bioassessment Protocol I. Analysis is for the most part, restricted to two metrics, taxa richness and the abundance of caddisflies, mayflies, and stoneflies. The distribution of tolerant and intolerant taxa is also factored into the assessment. Tolerance values are derived from Bode et al.(1991) and Lenat (1993).

In a bioassessment conducted in Prince William County Watershed, Jones and Kelso used a modification of the EPA Rapid Bioassessment Protocol II(Jones and Kelso, 1997). Previous work they had done in this area led to the following modifications:

- Deleted the scrapers/ filter collectors metric
- Used Sorenson's index (SI) for community similarity with the following criteria
    - 6: $SI < .5$
    - 3: $.5 < SI < .3$
    - 0: $SI < .3$
- Deleted shredders/total since CPOM was not available at many sites

Metrics Selected by Jones and Kelso(1997)

1. Taxa Richness
2. Sorenson's Index
3. Family Biotic Index
4. Ratio of EPT to Chironomidae abundance
5. Percent Contribution of Dominant Family
6. EPT Index

Jones and Kelso found that each of the above metrics was significantly correlated with the percentage of low intensity land use although the highest correlation was with the EPT index and taxa richness and the lowest correlation with EPT/chironomid abundance. The multimetric score correlated with percentage of low intensity land use better than any of the individual metrics. Also, EPT/chironomid abundance was the only metric to be significantly correlated with watershed area (Jones and Kelso, 1997).

In order to do the RBP II calculations, Jones and Kelso chose the reference site that appeared best in terms of taxa richness, the EPT index and the FBI from a larger group of reference sites they were sampling in that region.

Coastal Region

The Potomac River estuary travels for a short distance across the Mid-Atlantic Coastal Plain ecoregion before emptying into the Chesapeake Bay. A few small, nontidal Potomac tributaries flow through Virginia and Maryland's Coastal Plain (Eastern Shore and Norfolk area watersheds are also in Virginia's Coastal Plains). The Mid-Atlantic Coastal Streams (MACS) workgroup, consisting of representatives from all the coastal states from Delaware to South Carolina and from the U.S. EPA, have developed excellent guidelines for measuring biological and habitat conditions in the Mid-Atlantic coastal streams (U.S. EPA, 1997). The MACS workgroup incorporated RBP concepts and general approaches into protocols specifically designed for the Mid-Atlantic Coastal Plain since RBP protocols are designed for streams with shallow riffles and cannot be applied to slow moving, low-gradient streams almost devoid of riffles, with sandy and muddy substrates. Table 16 summarizes the sampling protocol recommended by MACS.

Table 16 Summary of MACS Recommended Sampling Protocols

HABITAT SAMPLED	Perennial streams; woody snags, banks with roots and snag material, submerged macrophytes; NOT channel bottom (sand, mud, detritus) which is relatively unproductive
GEAR METHOD	Gear: 1-foot wide D-frame dip net, 0.3 m width, 650 µm mesh; 600 µm mesh sieve bucket; 70% ethanol (final strength) preservative. Method: Habitats are sampled in proportion to their occurrence at the site. Snags: scrape the net along large woody surfaces or jab in smaller sized snag material. Banks: scrape or jab. Submerged macrophytes: draw net through bed.
SAMPLING FREQUENCY	Fall, Winter and Spring are preferred; Summer should be avoided
SUBSAMPLING	20 scrapes or jabs composited for a total of approximately 6.2 m^2 sampled.

TAXONOMIC LEVEL	Sorting done in laboratory following U.S. EPA RBP guidelines. Organisms identified to lowest practicable taxonomic level, generally species for most groups, and counted (100-120 organism target count). Metric calculations made at genus level.

Metrics:

Most benthic macroinvertebrate in the Mid-Atlantic Coastal Plain streams are collectors, so metrics reflecting other functional groups are not useful (e.g. %Shredders, %Filter-Feeders). Coleoptera, Oligochaete and Odonata metrics are also not useful because taxa richness in these groups is low. The MACS Workgroup identified 9 benthic macroinvertebrate metrics presently used in the Coastal Plain by Workgroup States, and Gerritsen et al. (1996) identified 11 that could possible contribute to an invertebrate index for Coastal Plain streams (Table 17). The metrics are similar to ones being tested in Delaware and Florida Coastal Plain ecoregions (Gerritsen 1996). In the Mid-Atlantic region, the metrics appear most responsive to habitat quality, and in particular to a group of habitat metrics reflecting human disturbances. Non-point source impacts from agriculture were minimized by a good riparian zone of 18 m or more.

Multimetric:

Efforts are underway to develop a macroinvertebrate index (multimetric) by MACS and others. As with the RBP protocols, reference sites need to be identified before the selected metrics can be scored and combined in an Index. Gerritsen et al (1996) have made a preliminary selection of metrics for an Index using a few candidate reference sites selected on the basis of land use and habitat quality. A final selection cannot be made until reference and impaired sites are better characterized with respect to natural variation and response to anthropogenic impacts (enrichment, urban effects, severe habitat disruption). The candidate metrics for the Index can then be further tested for discriminatory ability.

Table 17 Metrics in use and being considered for inclusion in macroinvertebrate multimetrics or indexes for Coastal Plain ecoregions

Metric	MACS Workgroup	Gerritsen et al. 1996	Delaware	Florida	Exp. Resp. to Perturb.
Taxonomic Richness	X	X	X	X	decrease
EPT Taxa	X	X	X	X	decrease
Crustacean plus Mollusc Taxa		X		X	decrease
Chironomidae Taxa					decrease
%EPT Abundance	X	X	X		decrease
%Chironomidae	X		X		increase
%Diptera		X		X	increase
%Dominant Taxon	X	X	X	X	increase
%non-insect	X				
%Crustacean plus Mollusc		X		X	decrease

%Oligochaetes		X			increase
North Carolina Biotic Index	X				increase
Community Loss Index	X				increase
Shannon-Wiener Index		X			decrease
Hilsenhoff Biotic Index	X	X	FBI[1]		increase
Florida Index				X	
%Filterers				X	
%Shredders				X	

[1] Family Biotic Index

Gerritsen et al (1996) have made a preliminary attempt at selecting reference sites in Maryland's Coastal Plain. They used stringent criteria to evaluate land use and habitat disturbance and to select the least impaired sites: less than 10% urban or residential; less than 50% agricultural; and a disturbance index score of ≥80. The disturbance index is the sum of five habitat metrics (channel alteration, bank vegetation, bank stability, riparian vegetation zone width, sediment deposition) each scored from 1(poor) - 20 (excellent). There were no striking differences between either the eastern and western shore, or between northern and southern sites. Division into Coastal Plain subecoregions does not appear to be warranted at this time.

Data Integration Challenges

Gear Method Comparability

When applying a multi-metric to potentially impacted stations and reference stations, we assume that the gear method used yields an adequate representation of the biota at these stations. We are concerned that different multimetrics when applied to station samples of equivalent biological integrity will falsely indicate that the stations have different levels of biological integrity. There is the additional concern, however, that the samples themselves are supplying different representations of the biota depending on the gear method used to collect them. A multimetric that has been adopted by a particular group is tweaked and modified to best analyze the biotic representation that their gear method yields. Sampling with a slightly different gear method might have changed the distribution of organisms enough that individual metrics would have to weighted in the multimetric in a different way to discriminate between reference and impacted stations as effectively. This same multimetric may not assess stream impairment as well when used on samples collected by other gear methods. In order then to develop or adopt a multimetric for an entire ecoregion, we must choose a multimetric per ecoregion that discriminates between reference and impaired sites but also is comprised of individual metrics that are not subject to known gear method biases.

Integration of these various data is one of the greatest challenges in avoiding duplication of efforts and fully utilizing all the monitoring data collected in an area. However, an organization can leverage and enhance the value of its own results if it does find ways to successfully integrate other data sets with its own.

A variety of sampling gears and protocols have been used to perform biological surveys of benthic macroinvertebrate. Variations of the simple kick method are most common because of their cost-effectiveness. Lenz and Miller (1996) compared samples collected with different, standardized kick methods used by organizations monitoring 2nd - 5th order streams in Wisconsin, namely the Wisconsin Department of Natural Resources (WDNR), the U.S. Geological Survey National Water Quality Assessment Program (USGS-NAWQA), the U.S. Department of Agriculture Forest Service (USDA-FS) and the Water Action Volunteer Water Quality Monitoring Program (WAV). Samples were preserved and counted by the same laboratory to reduce variability due to counting protocol.

The authors concluded:

- Sampling methods used by each organization tended to assess the macroinvertebrate community structure differently and sharing of macroinvertebrate data may not be feasible when specific species assemblages is required. Differences could be attributed to dissimilar capture efficiencies due to gear (e.g. 1400µ versus 425µ and 589µ mesh nets), different microhabitats sampled (e.g. riffle versus snag), and differences in kicking and scrubbing technique.
- Replicates from several riffle areas resulted in a higher total number of taxa collected and subsequently a greater reported taxa richness. Replicates collected with the same method from the same riffle contained very similar proportions of the macroinvertebrate taxa, however replicates collected by different methods showed biases.
- Applying Hilsenhoff's Biotic Index (HBI) (Hilsenhoff, 1987) to the dissimilar data produce very similar interpretations or ratings of biotic condition. The HBI appears to be a robust measure of biotic condition that is not differentially influenced by the collection methods tested in the study. The Family Level Biotic Index (FBI) (Hilsenhoff 1988) and the mean tolerance value measure (Lillie and Schlesser 1994) were also robust.
- Other measures such as %EPT, Margalef's diversity index and certain measures of trophic function were not as consistent, but the variability did not appear to be due to sampling method.

In conclusion, field collection method differences can preclude the direct comparison of data, however the HBI, FBI, Mean Tolerance Value Measure, and possibly other indices seem to overcome the effects of the method differences and produce similar interpretations and ratings of biotic condition. Further testing these metrics on diverse data sets could identify a suite of indices that can be successfully applied to data collected with specific criteria (e.g. riffle habitat, fine mesh net) to produce comparable results.

Old Versus New Taxonomy

Taxonomic anomalies are found when integrating historical and recently collected data, and can demand a considerable amount of time to resolve. Superceded or incorrectly spelled taxon names are frequently found in older data (Fox et. al. 1996). NOAA's National Oceanographic Data Center (NODC) taxon list (http://www.nodc.gov/NODC-products.html) or the Interagency Taxon System taxon list (http://www.itis.usda.gov/itis/access.html) are often helpful in resolving and updating names. Replacing superceded names with their modern names and associating taxa with their NODC codes, a taxonomically based hierarchical coding system, will facilitate data integration and later data interpretation.

Taxonomic Level

The macroinvertebrate metrics Taxa Richness, FBI, Scraper-Filter Ratio, % Dominant Family, and % Shredder all require taxonomic identification to at least family level. Similarly, many fish metrics require family level taxonomic identification before they can be calculated. However, because monitoring programs have different sampling requirement, not all organisms are identified to the same taxonomic level. Thus, some metrics can not be used for certain datasets. When doing a basin-wide analysis, metrics and multimetrics should be chosen with the data requirements of these metrics kept in mind. If an analysis uses metrics dependent on species level data, the analysis may be based on very few datasets since a large majority of macroinvertebrate data, for instance, is only collected to family level. However, only selecting metrics that require order level identifications may compromise the sensitivity of the analysis.

Tolerance Values

A number of different research groups have developed tolerance values for use in Biotic Indices. The tolerance value for a particular family or species tends to change with geographic area. For the Biotic Index to be an accurate measure of impairment, tolerance values that truly reflect the relative tolerance of organisms to stream impact must be selected. Currently, several different tolerance value lists are being used for stream analyses in the Basin.

Status of Monitoring Stations in Basin

Appendix A is a listing of current monitoring stations throughout the ecoregions represented in the Potomac River Basin. This table contains the following information: 1. agency name, 2. station name, 3. the ecoregion designation if known, 4. the stream on which the station is located, 5. an indication as to whether the station is used as a reference station, 6. station latitude and longitude, and 7. the status of the monitoring station using the rating system of that station's monitoring agency. We have included Maryland Biological Stream Survey stations in this listing, however, a station "rating" for MBSS stations is unavailable at this time.

Discussion and Conclusions

A number of standard metrics have been developed to assess benthic macroinvertebrates in streams and rivers. The Environmental Protection Agency's document *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish* (Plafkin et al. 1989) provided States with one standard method for performing biological assessments. Several State water quality agencies were initially involved in developing the benthic protocols. These protocols have since been incorporated into many State biomonitoring programs. The benthic protocols, named Rapid Bioassessment I, II and III, recommend standard methods for collecting and counting samples as well as for calculating and scoring metrics from the data. They are intended to be interpreted in conjunction with habitat assessments and measured against conditions at selected reference sites. Reference sites represent the natural range of variation in "least disturbed" water chemistry, habitat and biological conditions (Plafkin et al 1989).

Using the protocols as a starting point, many jurisdictions have modified key components to better suit their region (e.g. changing pollution tolerance and functional feeding group classifications required to calculate key metrics, developing different scoring approaches and criteria, and adding and deleting metrics). Presently, various monitoring programs assess biotic condition in a generally consistent fashion in the Potomac River Basin, but because of modifications made, results are not directly comparable. They cannot be simply merged to obtain, for example, an interstate assessment. Federal, state, and local jurisdictions in the Potomac Basin have each identified reference sites to use in scoring their data. However, protocol differences often prevent using other jurisdictions reference sites for the same ecoregion. The added statistical power for distinguishing impaired and unimpaired sites that comes from a larger subset of reference sites is therefore lost.

Diamond et al. (1996) as well as the Interagency Task Force on Water Quality Monitoring (ITFM) encourage the development of "data quality characteristics" for every available bioassessment method. "Data quality characteristics" include such information as data precision, method biases, method sensitivity, and range of conditions over which a method yields satisfactory data. Knowing this type of information allows researchers to combine data collected with different methods if they produce data of similar quality. This approach is referred to as the "Performance Based Method System"(PBMS). While initiating a performance based approach involves a considerable amount of effort, the benefits of such an approach far outweigh the costs. Some of these benefits include 1. Inciting monitoring agencies to identify what their specific data quality objectives are in order to then identify the appropriate methods, 2. Encouraging agencies to compare data collected by their various monitoring programs to avoid duplicated efforts, and 3. Encouraging flexibility in choice of monitoring methods as long as the chosen method meets the data quality objectives of the study at hand. Diamond et al (1996) state that although we may never truly know whether a particular assessment is accurately, characterizing a site, bioassessment methods using a performance-based approach will help us discern objectively the level of impairment we can reliably detect using different methods.

As recommended in the "performance based method system" (Diamond et al., 1996), the Comission has begun to document in this report the "data quality characteristics" of regional monitoring data. The report summarize the various assessments methods, or metrics, currently applied to benthic data collected in the Potomac River Basin by monitoring programs. It also compiles a list of recently sampled locations in Potomac ecoregions and includes status (e.g. impaired, unimpaired) as calculated by methods selected for basin-wide assessments. A thorough search of "gray" and periodical literature focussed on Virginia identified historical and ongoing monitoring programs in the region and various efforts to develop metrics and multimetrics. Raw data and metric/multimetric values were requested from the region's ongoing benthic monitoring programs. The Commission was only partially successful in obtaining the data and metric information. A number of monitoring programs could not make their latest data available, or if they did the data were incomplete (e.g. latitude/longitude information missing). We needed the metric data, or the raw data with which to calculate the metric data, and a list of accepted reference sites before we could apply a standard scoring approach. We have listed stations that are accepted by the monitoring programs as reference sites when possible. The list is still not complete.

The following obstacles must be overcome before a basin-wide assessment is possible.

- Data assembly needs to be completed.
  Although not available in time for this project, we expect raw data and metric data will become available within a year.

- Critical station information such as latitude and longitude must be attained or determined.
  Without locational information, we are unable to determine ecoregion, stream order, or surrounding land use. All of these are necessary to minimize environmental variability in analyses and to score sites against the appropriate subset of reference sites.

- A concensus is necessary on tolerance values and functional feeding goup designations.
  These are needed to calculate key metrics use to evaluate site impairement.

- Reference sites need to be clearly identified on an ecoregion basis.
  Reference sites from the individual monitoring programs need to be evaluated in the manner described in Gerritsen et al. (1996). A comparable, basin-wide subset of reference sites will be obtained using predetermined habitat and land use criteria and biological metric scores.

References

Angermeier, P.L. and J. R. Karr. 1986. Applying an index of biotic integrity based on stream-fish communities : considerations in sampling and interpretation. North American Journal of Fisheries Management 6: 418-429.

Angermeier, P.L., and I.J. Schlosser. 1987. Assessing biotic integrity of the fish community in a small Illinois stream. North American Journal of Fisheries Management 7: 331-338.

Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M. L. Bastian. 1996. A Framework for Biological Criteria for Florida Streams using Benthic Macroinvertebrates. Journal of the North American Benthological Society 15(2): 185-211.

Barbour, M.T., J. Gerritsen. 1996. Subsampling of benthic samples: A defense of the fixed-count method. Journal of the North American Benthological Society 15(3): 386-391.

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: metric redundancy and variability among reference stream sites. Environmental Toxicology and Chemistry 11(4): 437-449.

Barbour, M.T., J. Gerritsen, J.S. White.1996. Development of the Stream Condition Index (SCI) for Florida. Florida Department of Environmental Protection, Tallahassee, FA

Biological Criteria: Research and Regulation. Proceedings of a Symposium. Office of Water, U.S. EPA. December 12-13, 1990.

Bode, R.W., M.A. Novak, and L.E. Abele. 1991. Methods for rapid biological assessment of streams. Stream Biomonitoring Unit. New York State Department of Environmental Conservation. Albany, NY.

Boyle, T.P., G.M. Smillie, J.C. Anderson, and D.P. Beeson. 1990. A sensitivity analysis of nine diversity and seven similarity indices. Journal of the Water Pollution Control Federation 62: 749-762.

Brooks, R.P., E.D. Bellis, C.S. Keener, M.J. Croonquist, and D.E. Arnold. Pages 387-398 in Wetlands and River Corridor Management. J. A. Kusler and S. Daly (eds.) Association of Wetland Managers.

Cairns, J. 1977. Quantification of biological integrity. Pages 171-187 in The Integrity of Water, R.K. Ballentine and L.J. Guarraia. U.S. EPA 055-001-01068-1.

Cairns, J., D.W. Albaugh, F. Busey, and M.D. Chanay. 1968. The sequential comparison index-a simplified method for non-biologists to estimate relative differences inbiological diversity in stream pollution studies. J. Water Poll. Contr. Fed. 40: 1607-1613.

Cairns, J. and E.P. Smith. 1989. Randomizing data and the sequential comparison index. Research Journal WPCF 61:1733-1738.

Courtemanch, D. L. 1996. Commenting on the subsampling procedures used for rapid bioassessments. Journal of the North American Benthological Society 15 (3) 381-385.

Cummins, Kenneth W.1993 Bioassessment and Analysis of functional Organization of Running Water Ecosystems. *In: Biological Monitoring of Aquatic Systems. Stanford L. Loeb and Anne Spacie, eds.,155-169.*

Davis, W.S. and A. Lubin. 1989. Statistical validation of Ohio"s EPA's Invertebrate Community Index. Pages 23-32 in W.S. Davis and T.P. Simon (editors), Proceedings of the 1989 Midwest Pollution Control Biologists Meeting, 14-17 February 1989. U.S. Environmental Protection Agency, Region 5, Environmental Sciences Division, Chicago, Illinois.

Diamond, J. M.,M.T. Barbour, Stribling, J.B. 1996. Characterizing and comparing bioassessment methods and their results: A perspective. Journal of the North American Benthological Society 15(4): 713-727.

Fausch, K.D., J.R. Karr, and P.R. Yant. 1984. Regional application of an index of biotic integrity based on stream communties. Trans. Amer. Fish. Soc. 113: 39-55.

Fausch, K.D., J. Lyons, J.R. Karr, and P.L. Angermeier. 1990. Fish communities as indicators of environmental degradation. Pages 1-68 in Biological indicators of stress in fish. Amer. Fish. Soc. Symp. No. 8. Bethesda, MD.

Final Report of the Intergovernmental Task Force on Monitoring Water Quality. "The Strategy for Improving Water-Quality Monitoring in the United States." IGTFMWQ, Feb. 1995. USGS Office of Water Data Coordination.

Fore, L. S., and J. R. Karr. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15(2): 212-231.

Fore, L. S., J. R. Karr and L. L. Conquest. 1994. Statistical properties of an index of biotic integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences 51: 1077-1087.

Galli, J. 1988. A limnological study of an urban stormwater management pond and stream ecosystem. M.S. Thesis. George Mason University.

Galli, J. and K. Corish. 1997. Rapid Stream Assessment Technique (RSAT) Survey of the Sugarland Run Watershed. PhaseI: Sugarland Run Mainstem. Fairfax and Loudoun Counties, Virginia. Department of Environmental Programs. Metropolitan Washington Council of Governments.

Gurtz, M.E. 1994. Design of biological components of the National Water-quality Assessment (NAWQA) Program. Pages 323-354 in Loeb, S.L. and A. Spacie. Biological Monitoring of Aquatic Systems. Lewis Publishers.

Gurtz, M.E. and T.A. Muir. 1994. Report of the Interagency Biological Methods Workshop. U.S. Geological Survey Open-File Report 94-490. Raleigh, NC.

Heckman, C.W., H. Kamieth, and M. Stohr. 1990. The usefulness of various numerical methods for assessing the specific effects of pollution on aquatic biota. Int. Revue ges. Hydrobiol. 75: 353-377.

Hellawell, J.M. 1986. Biological Indicators of Freshwater Pollution and Environmental Management. *In: Field Assessments of Environmental Quality. 398-434.*

Herricks, Edwin E. and David J. Schaeffer. Compliance Biomonitoring -- Standard Development and Regulation Enforcement Using Biomonitoring Data. *In: Freshwater Biological Monitoring. Conference Proceedings. D. Pascoe and R.W. Edwards, Eds. Cardiff, U.K. September 1984.*

Hilsenhoff, W.L. 1982. Using a biotic index to evaluate water quality in streams. Technical Bulletin 132. Department of Natural Resources, Box 7921, Madison, WI 53707.

Hilsenhoff, W.L. 1987. An Improved Biotic Index of Organic Stream Pollution. The Great Lakes Entomologist.20(1): 31-39

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. J. N. Am. Benthol. Soc. 7: 65-68.

Hughes, R.M. and D.P. Larsen 1988. Ecoregions: An approach to surface water protection. Journal of the Water Pollution Control Federation 60:486-93.

Hughes, R.M., Heiskary, S.A., W. J. Matthews, and C.O. Yoder. 1994. Use of Ecoregions in Biological Monitoring. *In: Biological Monitoring of Aquatic Systems. S.L. Loeb and anne Spacie, eds. 125-151.*

Jones, R.C. and C.C. Clark. 1987. Impact of watershed urbanization on stream insect communities. Water Resources Bulletin 23: 1047-1055.

Kaesler, R.L., J. Cairns, and J.S. Crossman. 1974. Redundancy in data from stream surveys. Water Research 8: 637-642.

Karr, J.R. 1981. Monitoring of biological integrity: an evolving approach to the assessment and classification of water resources. Manuscript.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6: 21-27.

Karr, J.R. 1987. Biological monitoring and environmental assessment: a conceptual framework. Environmental Management 11: 249-256.

Karr, J. R. and B.L. Kerans. 1991. Components of biological integrity: their definition and use in development of an invertebrate IBI. Pages 1-16 in W.S. Davis and T.P. Simon (editors). Proceedings of the 1991 Midwest Pollution Control Biologists Meeting. EPA 905-R-92-003. U.S. Environmental Protection Agency, Environmental Sciences Division, Chicago, Illinois.

Karr, J.R., L.A. Toth, and D.R. Dudley. 1985. Fish communities of midwestern rivers: a history of degradation. Bioscience 35: 90-95.

Karr, J.R., P.R. Yant, D.D. Fausch, and I.J. Schlosser. 1987. Spatial and temporal variability of the index of biotic integrity in three Midwestern streams. Trans. Amer. Fish. Soc. 116: 1-11.

Kerans, B.L. J.R. Karr, and S.A. Ahlstedt. 1992. Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. Journal of the North American Benthological Society 11(4): 377-390.

Kreuger, H.O., J.P. Ward, and S. H. Anderson. 1988. A resource manager's guide for using aquatic organisms to assess water quality for evaluation of contaminants. Biological Report 88 (20). U.S. Fish and Wildlife Service.

Lang, C., G.L'Eplattenier, and O. Reymond. 1989. Water Quality in rivers of western Switzerland: application of an adaptable index based on benthic invertebrates. Aquatic Sciences 51: 224-234.

Lenat, D.R. 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. J. N. Am. Benthol. Soc. 7: 222-233.

Lenat, D.R. 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water-quality ratings. J. N. Am. Benthol. Soc. 12: 279-290.

Lenat, David R. and Michael T. Barbour, 1994. Using Benthic Macroinvertebrate Community Structure for Rapid, Cost-Effective, Water Quality Monitoring: Rapid Bioassessment. Pp. 187-216 *In: Loeb, S.L. and A. Spacie, eds. Biological Monitoring of Aquatic Systems.* Lewis Publishers, Boca Raton, Florida.

Lenz, Bernard N. and Michael A. Miller. Dept. of Interior, USGS Fact Sheet FS-216-96. Mid-Atlantic Coastal Streams (MACS) Workgroup. 1997. Field and Laboratory Methods for Macroinvertebrate and Habitat Assessment of Low Gradient, Nontidal Streams.

Miller, D.L. et al. 1988. Regional applications of an index of biotic integrity for use in water resource management. Fisheries 13: 12-20.

Norris, R. H. and A. Georges. 1992. Analysis and Interpretation of Benthic Macroinvertebrate Surveys. In: *In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D. M. Rosenberg and V. H. Resh, Ed. 234-285.*

Osborne, L.L., R.W. Davies, and K.J. Linton. 1980. Use of hierarchical diversity indices in lotic community analysis. Journal of Applied Ecology 17: 567-580.

Osborne, L.L. et al. 1991. Stream habitat assessment programs in states of the AFS North Central Division. Fisheries 16: 28-35.

Petersen, R.C. 1992. The RCE: a riparian, channel, and environmental inventory for small streams in the agricultural landscape. Freshwater Biology 27:295-306.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. EPA/444/4-89-001. U. S. Environmental Protection Agency, Office of Water, Washington, D.C.

Resh, V. H. 1988. Variability, accuracy, and taxonomic costs of rapid assessment approaches in benthic biomonitoring. Presented at the 36th annual North American Benthological Society meeting at Tuscaloosa, Alabama, 17-20 May 1988.

Resh, V.H. and J. K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. Pages 195-233 in D.M. Rosenberg and V.H. Resh (editors).

Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York.

Resh, V. H. and E. P. McElravy. 1992. Contemporary Quantitative Approaches to Biomonitoring Using Benthic Macroinvertebrates. *In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D. M. Rosenberg and V. H. Resh, Eds. 159-194.*

Resh, V.H., R. H. Norris, and M.T. Barbour. 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology*. 20: 108-121.

Richards, C., G.E. Host and J.W. Arthur. 1993. Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. *Freshwater Biology* 29: 285-94.

Shackleford, B. 1988. Rapid bioassessments of lotic macroinvertebrate communities:Biocriteria development. Arkansas Department of Pollution Control and Ecology, Little Rock, AR.

Shelor, M.H. and R.W. Ayers. 1984. Virginia State Water Control Board. Procedure for Conducting Qualitative Biological Surveys. Bureau of Surveillance and Field Studies. Division of Ecological Studies.

Smith, E.P. and J.R. Voshell. 1997. Studies of Benthic Macroinvertebrates and Fish in Streams within EPA Region 3 for Development of Biological Indicators of Ecological Condition. Part I. Benthic Macroinvertebrates. Final Report.

Vinson, Mark R., and C.P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15(3): 392-399.

Wallace, J. B., J.W. Grubaugh, and M.R. Whiles. 1996. Biotic indices and stream ecosystem processes: results from an experimental study. Ecological Applications (6): 140-151.

Walski, T.M., and F.L. Parker. 1974. Consumers water quality index. Journal of the Environmental Engineering Division of ASCE EE3: 593-611.

Whittier, R.R., R.M. Hughes and D.P. Larsen. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Can. J. fish. Aquat. Sci. 45: 1264-1278.

Wright, J.F., D. Moss, P.D. Armitage, and M.T. Furse, 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshwater Geology 14:221-256.

Wright, J.F. , P.D. Armitage, M. T. Furse, and D. Moss. 1989. Prediction of invertebrate communities using stream measurements. Regulated Rivers: Research & Management 4: 147-155.