DOI: 10.1007/s00267-002-2945-7

Development and Evaluation of a Macroinvertebrate Biotic Integrity Index (MBII) for Regionally Assessing Mid-Atlantic Highlands Streams

DONALD J. KLEMM*
KAREN A. BLOCKSOM
FLORENCE A. FULK
U.S. Environmental Protection Agency
National Exposure Research Laboratory
26 W. Martin Luther King Drive
Cincinnati, Ohio 45268, USA

ALAN T. HERLIHY
Oregon State University
c/o U.S. Environmental Protection Agency
200 SW 35th St.
Corvallis, Oregon 97333, USA

ROBERT M. HUGHES
Dynamac Corporation
200 SW 35th St.
Corvallis, Oregon 97333, USA

PHILIP R. KAUFMANN
DAVID V. PECK
JOHN L. STODDARD
U.S. Environmental Protection Agency
National Health and Environmental Effects Research Laboratory
200 SW 35th St., Corvallis, Oregon 97333, USA

WILLIAM T. THOENY
SoBran, Inc.
c/o U.S. Environmental Protection Agency
National Exposure Research Laboratory
26 W. Martin Luther King Drive
Cincinnati, Ohio 45268, USA

MICHAEL B. GRIFFITH
U.S. Environmental Protection Agency
National Center for Environmental Assessment
26 W. Martin Luther King Drive
Cincinnati, Ohio 45268, USA

WAYNE S. DAVIS
U.S. Environmental Protection Agency
Office of Environmental Information, Environmental Analysis Division
1200 Pennsylvania Avenue
Washington, DC 20460, USA

ABSTRACT / The Macroinvertebrate Biotic Integrity Index (MBII) was developed from data collected at 574 wadeable stream reaches in the Mid-Atlantic Highlands region (MAHR) by the U.S. Environmental Protection Agency's (USEPA) Environmental Monitoring and Assessment Program (EMAP). Over 100 candidate metrics were evaluated for range, precision, responsiveness to various disturbances, relationship to catchment area, and redundancy. Seven metrics were selected, representing taxa richness (Ephemeroptera richness, Plecoptera richness, Trichoptera richness), assemblage composition (percent non-insect individuals, percent 5 dominant taxa), pollution tolerance [Macroinvertebrate Tolerance Index (MTI)], and one functional feeding group (collector-filterer richness). We scored metrics and summed them, then ranked the resulting index through use of independently evaluated reference stream reaches. Although sites were classified into lowland and upland ecoregional groups, we did not need to develop separate scoring criteria for each ecoregional group. We were able to use the same metrics for pool and riffle composite samples, but we had to score them differently. Using the EMAP probability design, we inferred the results, with known confidence bounds, to the 167,797 kilometers of wadeable streams in the Mid-Atlantic Highlands. We classified 17% of the target stream length in the MAHR as good, 57% as fair, and 26% as poor. Pool-dominated reaches were relatively rare in the MAHR, and the usefulness of the MBII was more difficult to assess in these reaches. The process used for developing the MBII is widely applicable and resulted in an index effective in evaluating region-wide conditions and distinguishing good and impaired reaches among both upland and lowland streams dominated by riffle habitat.

Benthic macroinvertebrates exhibit highly diverse taxonomic, morphological, trophic level, and physiological responses to changes in stream ecosystems (Hynes 1970, Allan 1995, Merritt and Cummins 1996).

KEY WORDS: Bioassessment; Biological monitoring; Benthic macroinvertebrates; Metrics; Multimetric; EMAP; IBI

*Author to whom correspondence should be addressed; *email:* klemm.donald@epa.gov

However, anthropogenic disturbances and pollution stressors of landscapes and streams often elicit predictable responses by benthic macroinvertebrates (Hynes 1960, 1970, Warren 1971, Allan 1995, Simon 2002). Following Karr's (1981) introduction of a multimetric index of biotic integrity on the basis of fish assemblages, similar indices were developed for benthic macroinvertebrates, fish, and periphyton (Kerans and Karr 1994, DeShon 1995, Barbour and others 1996, USEPA 1996, Fore and others 1996), fish (Miller and others 1988, Simon and Lyons 1995, Hughes and Oberdorff 1999, McCormick and others 2001), periphyton (Hill and others 2000), and riparian birds (Bryce and others 2002). The rationale behind such indices, whether developed for birds, fish, benthic macroinvertebrates, or periphyton, is that a set of variables or metrics representing community structure (taxa richness, relative abundance, dominance), pollution tolerance, functional feeding groups and habitat occurrences, life history strategies, disease, and density offers robust and sensitive insights into how an assemblage responds to natural and anthropogenic stressors. Some might argue that taxonomic composition alone suffices for making assessments, through use of multivariate analyses (Reynoldson and others 1997, Hawkins and others 2000, Norris and Hawkins 2000), but others favor a multimetric approach (Karr and Chu 1999, 2000). We believe both approaches have merit, and both require regional adaptations suited to the landscapes, biota, and stressors common to particular areas.

There are four substantial differences between our approach to index development and those presented by Plafkin and others (1989), Kerans and Karr (1994), DeShon (1995), and Fore and others (1996). First, we used data collected mostly from reaches chosen through use of a probability design from across a large geographic area, as opposed to hand-selected sites from a limited area representing a single stressor gradient. Second, we rigorously evaluated five characteristics (range, precision, relationship to catchment area, responsiveness to disturbance, and redundancy with other metrics) of a large number of candidate metrics. Third, we scored the metrics and resulting index on a continuous scale rather than on an interval scale (e.g., 1–3–5). Fourth, because of the probability design, we extrapolated the survey results to the total length of all target streams in the study region.

The Environmental Monitoring and Assessment Program-Surface Waters (EMAP-SW) was designed and implemented by the U.S. Environmental Protection Agency (USEPA) to assess status and trends in ecological conditions of streams on a regional scale (Messer and others 1991, USEPA 1998). The objective of this paper is to document the development and application of the Macroinvertebrate Biotic Integrity Index (MBII) for assessing the condition of all mapped wadeable streams in the Mid-Atlantic Highlands region (MAHR) of the eastern USA.

Methods

Study Area and Survey Design

The MAHR of the United States includes the states of PA, MD, VA, WV, and DE outside the coastal Plains and the Catskill Mountains of New York (Figure 1). Much of the region consists of low forested mountains or hills and agricultural valleys, with a long history of logging, coal mining, urbanization, and agriculture. Data were collected from 506 reaches that were selected via a randomized systematic design with a spatial component (Herlihy and others 2000). Reaches were chosen randomly, but were restricted to wadeable streams, defined as first, second, and third order blue lines on 1:100,000 scale USGS maps (Strahler 1957). Inclusion probabilities were set so that roughly equal numbers of first, second, and third order streams would be sampled in this study. These probability reaches were supplemented with 68 hand-picked reaches chosen by USEPA Region 3 and state biologists. These supplemental sites were selected to increase representation of the best and worst conditions in the region. Thus, a total of 574 wadeable stream reaches were sampled for this study (Figure 1).

Sample Collection and Processing

Samples were collected (Lazorchak and others 1998) during the spring base-flow period, late April to June, 1993-1995. Sample reaches consisted of a length of stream equal to 40 times the mean wetted width (minimum of 150 m and maximum of 500 m) delineated around the randomly selected reach midpoint. Water chemistry samples were collected from the midpoint of the reach and analyzed using EMAP-SW protocols (USEPA 1987, Klemm and others 1990, Klemm and Lazorchak 1994). Field crews recorded physical habitat data using EPA RBP qualitative methods (Plafkin and others 1989, Barbour and Stribling 1991, Barbour and others 1999) at all reaches, and quantitative methods (Lazorchak and others 1998) at 102 probability reaches and the 68 hand-picked reaches. Landscape condition of catchments was assessed using land cover data (McCormick and others 2001), human population density, road density, and pollution source density data.

Within the sampling reach, benthic macroinvertebrates were collected with a kick net (595 μm mesh

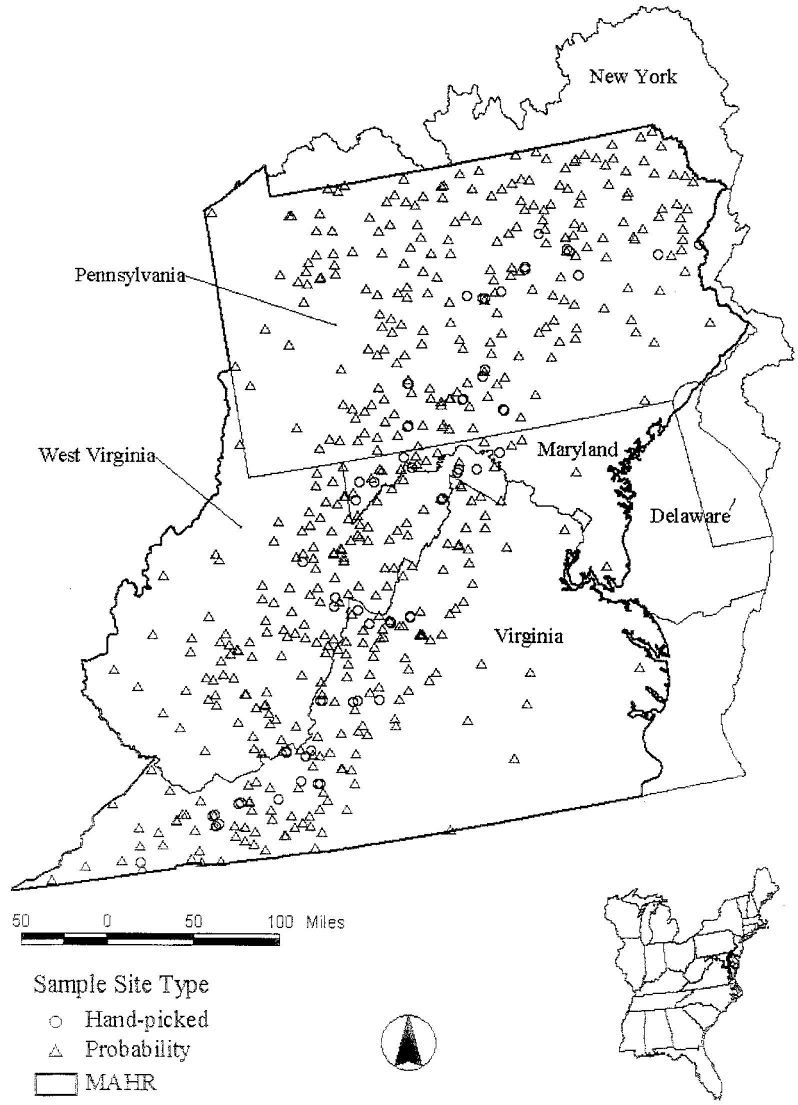

Figure 1. Distribution of surveyed stream reaches in the Mid-Atlantic Highlands Region.

size) from the nine inner transects of eleven evenly spaced transects (Lazorchak and others 1998), with each transect being identified as either riffle or pool habitat. Macroinvertebrate samples for riffles and pools were collected, composited, and processed separately. Thus, from one to nine kick samples might be combined into a single composite sample for each habitat type per stream reach (depending on the number of transects with that habitat type). In the laboratory, a benthic macroinvertebrate sample was placed in a sorting pan with a numbered grid, and squares from which organisms were sorted were chosen at random (Klemm and Lazorchak 1994). Squares were chosen and sorted in this way until at least 270 (300 ± 10%) organisms or the entire sample was sorted. All organisms were identified to the lowest possible taxon.

Data Used for Index Development

Of the 574 reaches sampled, 546 had riffle composite samples and 223 had pool composite samples. Thirty reaches produced only pool composite samples and 351 reaches produced only riffle composite samples.

We divided the data into a calibration data set and a validation data set. The calibration data set consisted of the 68 hand-picked reaches and the 404 probability reaches lacking quantitative physical habitat data. The validation data set consisted of the remaining 102 probability reaches with quantitative physical habitat data. Only the calibration data set was used to assess responsiveness to disturbance, but both data sets were combined for catchment area correction of metrics and several other evaluation steps. Final scoring of individual metrics and the final index were on the basis of the calibration data set. To ensure sufficient sampling effort for each reach, we excluded 53 composite samples (4 riffle, 49 pool) because they were comprised of fewer than 4 transects and fewer than 150 organisms in the entire sample. For all analyses except precision, the first visit to a reach was used. For evaluation of metric precision, we used 34 pairs of within-year revisits.

We used water chemistry, qualitative habitat and minimum organism count criteria (Waite and others 2000) to identify reference and impaired reaches from among all reaches sampled. Reference reaches were defined by meeting all of the following criteria: sulfate $<$400 μeq/L, Acid Neutralizing Capacity (ANC) $>$50 μeq/L, chloride $<$100 μeq/L, total phosphorus $<$20 μg/L, total nitrogen $<$750 μg/L, mean qualitative habitat score $>$15 (of a possible 20), and at least 150 organisms. Impaired reaches, or those reaches with a recognized impairment, were defined as those for which any of the following criteria were met: pH $<$5, chloride $>$1000 μeq/L, sulfate $>$1000 μeq/L, total phosphorus $>$100 μg/L, total nitrogen $>$5000 μg/L, or a mean qualitative habitat score $<$10. Sites not meeting all reference criteria or any impairment criteria were considered as part of the middle range of conditions existing in the region and were included in analyses unless otherwise noted. On the basis of these criteria, we identified 86 (73 calibration and 13 validation) reference and 150 (112 calibration and 38 validation) impaired reaches among riffle samples. There were 20 (17 calibration and 3 validation) reference and 68 (48 calibration and 20 validation) impaired reaches identified from the pool samples.

We calculated and evaluated over 100 candidate metrics representing pollution tolerance values (PTVs) (Klemm and Hiltunen 1992, Klemm and others 2002), taxon groups, and trophic functional groups. Whenever appropriate, we tested three derivations of each metric: taxa richness, percent of total taxa in the composite sample, and percent of total individuals counted in the composite sample. One pollution tolerance metric, the Macroinvertebrate Tolerance Index (MTI), was calculated similarly to the Hilsenhoff Biotic Index (HBI) (Hilsenhoff, 1987), except that PTVs applied to general pollution (acidity, domestic waste, organics, heavy metals, agriculture, and sedimentation) rather than just organic pollution. In addition, the MTI is modified from the HBI in that it includes all macroinvertebrates, not just insects.

Index Development

Five major steps were involved in developing the MBII, following the process reported by Hughes and others (1998). First, we screened metrics in a stepwise manner for range, precision, relationship to catchment area, responsiveness to disturbance, and redundancy with other metrics. On the basis of these results, we selected a set of non-redundant metrics that responded to a variety of disturbance types and included different metric classes. In order to combine the selected metrics into an index, we had to transform each metric into a dimensionless number by scoring or rescaling it. The scored metrics were then summed to obtain the final index score. We used reference stream reach distributions to help establish condition categories (Good, Fair, and Poor) for the final index for the purpose of assessing biological condition in each reach.

Although riffles were much more commonly sampled than pools, many reaches were dominated by pools (N = 48), and 30 reaches lacked riffle composite samples. An index based solely on riffle habitat samples would preclude assessment of pool-only reaches. Thus, we developed an index that could be applied to both pool and riffle composite sample data, with slightly different scoring for the two habitat types.

Many level III and level IV ecoregions (Omernik 1987) exist in the MAHR, but we classified reaches as either uplands (Central Appalachian, North and North Central Appalachians, Ridge, and Blue Ridge ecoregions), or lowlands (Valley, Piedmont, and Western Appalachians), on the basis of these ecoregions. Although there were fewer reference reaches in the lowlands, we compared reference reach results between the uplands and lowlands and concluded that classification of reaches into two ecoregion groups was unnecessary.

Metric Screening

Range test. We applied a range test to all metrics for riffles and pools separately. We began by eliminating richness metrics with a range of 5 or less. We eliminated percentage metrics with a range of less than 10%. If 90% or more of values were 0 for any metric, we eliminated that metric. The following metrics were eliminated for both riffles and pools on the basis of range criteria: *Corbicula* (all forms), Crustacea richness, Hydropsychidae/Trichoptera ratio (richness and individuals), Megaloptera richness, omnivore richness, parasites (all forms), piercer-herbivores (all forms), Physidae (all forms), *Pteronarcys* richness and percent taxa, and Simuliidae richness. The following metrics were eliminated for pools only: *Pteronarcys* percent individuals and taxa, Simuliidae percent individuals, Mollusca richness, and Hydropsychidae richness.

Signal-to-Noise test. Next, we calculated the signal-to-noise ratio (S/N) for each metric as a measure of precision, using the among-reach variance as the signal and within-year revisit variance as the noise (Kaufmann and others 1999). These ratios were calculated for pools and riffles separately. A ratio of at least 1.50 was required for a metric to pass the S/N test because this was judged to represent the minimum acceptable "signal" to discern among stream sites relative to the "noise" due to measurement variation (Kaufmann and others 1999). However, because there were only 13 within-year pool revisit samples, if a metric failed the test for pools but passed for riffles, the metric was retained. If the metric passed for pools but failed for riffles, it was eliminated. Only 15 metrics passed the S/N screen for both riffles and pools, but 28 metrics passed for pools, and 37 metrics passed the S/N screen for riffles (Table 1). The 13 metrics that passed the S/N test for pools, but not riffles, were eliminated.

Catchment area adjustment. Certain characteristics of macroinvertebrate assemblages are expected to change with increasing stream size according to the River Continuum Concept (Vannote and others 1980). Thus, whether streams are disturbed or not, some metrics may vary with catchment area. We calibrated metrics by catchment area when they showed a significant linear relationship with $\log_{10}$ of catchment area (km^2) on the basis of Pearson correlations ($p < 0.01$) and visual inspection of scatter plots. Visual inspection was used to ensure that a particular correlation was not due to just a few points and that upper or lower values (depending on variable type) did indeed vary with catchment area. The relationship was evaluated with pool and riffle data combined, and included only reference reaches from the calibration and validation data sets. A simple linear regression was applied to the reference reach data, and residuals from regression lines were calculated for all sites. To maintain all positive values for simplicity, a constant equal to the predicted value for the largest catchment area (500 km^2) was added to each residual, resulting in all positive values for metrics adjusted for catchment area. Several metrics were adjusted for catchment area before screening for responsiveness (Table 1).

Responsiveness. We evaluated responsiveness of metrics to several disturbance gradients for pools and riffles separately. Gradients included general habitat (mean qualitative habitat metric score, qualitative channel alteration metric score), general disturbance (chloride, percent catchment land use in agriculture, urban, and mining), sedimentation (qualitative embeddedness score, turbidity, qualitative epifaunal substrate metric score, percent fine substrate), riparian habitat (qualitative riparian vegetation metric score, canopy density at bank, riparian disturbance, Riparian Habitat Condition Index (Kaufmann and others 1999), acidity and mine drainage (pH and sulfate), and nutrients (total phosphorus, and total nitrogen). Responsiveness was assessed using visual inspection of scatter plots and Spearman rank correlations. As an approximation of a Bonferroni correction to an overall type I error rate of 0.05 for the very large number of correlations, only those with a p-value ≤ 0.0001 (the smallest p-value available in SAS v. 8.0) were considered significant. All remaining metrics were responsive to at least one type of stressor for riffles, but several pool metrics did not show a significant correlation with any stressor (Table 1).

Redundancy. We evaluated redundancy in the remaining metrics via Pearson correlation coefficients. Metrics with a correlation coefficient (r) ≥ 0.7 were considered redundant. Higher correlations would indicate that over 50% of the variability of one metric could be accounted for by the other metric (i.e., $r^2 \geq 0.50$, Neter and others 1996). Metrics were evaluated for both pools and riffles and included the entire data set. Several sets of metrics were redundant, and only one metric from a group of redundant metrics was included in the final index.

Metric Selection

In selecting metrics from those that passed the screening process, we incorporated metrics of different general types (i.e., richness, pollution tolerance, taxonomic groups, trophic functional groups). From redundant groups we chose the metric with the greater S/N ratio, responsiveness, ecological relevance, and interpretability. We also tried to include metrics that responded to different disturbance types and that responded well in both riffles and pools and in uplands and lowlands.

We chose Ephemeroptera richness, Plecoptera richness, Trichoptera richness, collector-filterer richness, percent non-insect individuals, percent individuals in the 5 most dominant taxa, and the MTI for inclusion in the index. The MTI is an index that incorporates ranks for each taxon with respect to pollution tolerance, weighted by taxon abundance, and results in higher scores as the proportion of taxa tolerant to general pollution increases. Ephemeroptera richness and Plecoptera richness were each redundant with Ephemeroptera + Plecoptera + Trichoptera (EPT) richness, EPT-Hydropsychidae richness, EPT-tolerant Ephemeroptera richness, and intolerant taxa richness, though not with each other. Trichoptera richness was redundant with Trichoptera percent taxa, as well as all the EPT richness and intolerant taxa richness metrics. Collector-filterer richness was redundant with total taxa

Table 1. Metric screening results, beginning with metrics that passed the signal-to-noise test

Metric	Passed riffle S/N (≥ 1.50)	Passed pool S/N	Watershed adj. for riffles	Responsiveness (riffles)	Responsiveness (pools)
Collector-filterer richness	Yes	No	Yes	GH, S, A	None
% *Cricotopus* individuals	No	Yes		‡	‡
% *Cricotopus* taxa	No	Yes		‡	‡
% Crustacea/Mollusca individuals	Yes	Yes		GH, S, RH	None
% Crustacea/Mollusca taxa	Yes	No		GH, S, RH	None
Crustacea/Mollusca richness	No	Yes		‡	‡
% Crustacea individuals	Yes	*		S	*
% Crustacea taxa	Yes	No		S	None
% Dominant 5 taxa	Yes	No		A	None
% Ephemeroptera individuals	No	Yes		‡	‡
Ephemeroptera richness	Yes	Yes	Yes	GH, S, A, N	N
EPT:Chironomidae indiv. Ratio	Yes	No		GH, GD, S, RH, A, N	GH, GD, S, A, N
% EPT individuals	No	Yes		‡	‡
EPT richness	Yes	No		GH, GD, S, RH, A, N	GH, GD, S, N
% EPT-Hydropsychidae individuals	No	Yes		‡	‡
EPT-Hydropsychidae richness	Yes	Yes		GH, GD, S, RH, A, N	H, GD, S, N
EPT-Non-intolerant Ephemeroptera richness	Yes	Yes		GH, GD, S, RH, A, N	GH, GD, S, N
Macroinvertebrate tolerance index (MTI)	Yes	Yes		GH, GD, S, RH, A, N	GH, GD, S, A, N
Shannon diversity	Yes	No		A	None
Hydropsychidae/Trichoptera richness ratio	Yes	No		GD, A, N	A
% Hydropsychidae individuals	No	Yes		‡	‡
% Hydropsychidae taxa	Yes	No		S	None
Hydropsychidae richness	Yes	*		GH, S, A	*
No. individuals per taxon	Yes	No	Yes	GD, RH, A, N	None
Intolerant taxa richness	Yes	No		GH, GD, S, RH, A, N	GH, GD
% Megaloptera individuals	Yes	No		None	None
% Megaloptera taxa	Yes	No		None	None
% Mollusca taxa	No	Yes		‡	‡
% Non-insect individuals	Yes	Yes		GH, GD, S, RH, A, N	GD, S, N
% Oligochaete and leech individuals	Yes	Yes		GH, GD, S, RH, N	GH, GD, S, A, N
% Omnivore individuals	Yes	Yes		S	None
% Omnivore taxa	Yes	Yes		S	None
% Plecoptera individuals	Yes	Yes	Yes	GH, GD, S, RH, A, N	GH, S, N
Plecoptera richness	Yes	No	Yes	GH, GD, S, RH, A, N	GH
Predator richness	No	Yes		‡	‡
% *Pteronarcys* individuals	Yes	*		GH, GD, S, RH	*
% Scavenger individuals	Yes	Yes		GH, GD, S, RH, N	GH, GD, S, A, N
Scraper:Filterer individuals ratio	Yes	Yes		A	A
% Shredder individuals	Yes	No	Yes	GD, A, N	GD, N
Simpson Diversity Index	Yes	No		A	None
% Simuliidae taxa	Yes	*		N	*
% Super-tolerant individuals (PTV > = 8)	Yes	Yes		GH, GD, S, N	GH, GD, S, RH, A, N

Continued

Table 1. (Continued)

Metric	Passed riffle S/N (≥ 1.50)	Passed pool S/N	Watershed adj. for riffles	Responsiveness (riffles)	Responsiveness (pools)
% Super-tolerant taxa	No	Yes		‡	‡
% Tolerant individuals	Yes	Yes		GH, GD, S, RH, A, N	GH, GD, A, N
% Tolerant taxa	No	Yes		‡	‡
% Tanytarsini individuals	No	Yes		‡	‡
Tanytarsini richness	No	Yes		‡	‡
Number of distinct taxa	Yes	No	Yes	A	None
% Trichoptera taxa	Yes	Yes		GH, GD, S, RH, A	GH, GD, A
Trichoptera richness	Yes	No		GH, GD, S, RH, A	GH, GD

*The metric failed the range test for that stream habitat type (riffle or pool).

‡The metric passed the S/N test for pools but not for riffles and was eliminated.

GH = General habitat, GD = General disturbance, S = Sedimentation, RH = Riparian habitat, A = Acidity, N = Nutrients

richness. Percent non-insects was redundant with percent crustacean individuals, percent crustacean and mollusk individuals, and percent omnivore individuals. The MTI was redundant with percent tolerant individuals and percent tolerant taxa. Finally, 5% dominant taxa was redundant with the Simpson diversity index and total taxa richness.

Metric Scoring

Floor and ceiling values were determined separately for pool- and riffle-based metrics (Table 2). Each metric was scored on a continuous scale from 0 (poor) to 10 (good), using lower and upper expectation limits (Hughes and others 1998). For positive metrics (i.e., those that increased with improving conditions), the upper expectation (ceiling) was the 75th percentile of the distribution of reference reaches, while the lower expectation (floor) was the 25th percentile of the distribution of impaired reaches. Metrics with a value above the ceiling received a score of 10, while those below the floor scored 0. All other values were linearly scaled along the range between the high and the low. In other words, a raw metric value that was half way between the floor and ceiling values would be scored as 5.

For negative metrics, those that decreased with improving condition, the ceiling was the 75th percentile of the distribution of impaired reaches, and the floor was the 25th percentile of the distribution of reference reaches. Negative metrics with a value above the ceiling scored a 0, while those below the floor scored 10. All other values were linearly scaled along the range between the low and high as for positive metrics. To calculate the MBII value, we added the metric scores together and scaled the sum by (100/70) for a range of 0 to 100 for the MBII.

Condition Categories

We established three condition categories for the MBII. These were established on the basis of three successively more restrictive reference reach distributions using data from only the dominant habitat at each reference reach. The initial set of reference stream reaches was identified using the criteria provided above (Data Used for Index Development). We included additional criteria to further screen the reference reach data sets from the original set of all reference reaches (n = 86). For the second subset of reference reaches (n = 23), we added a quantitative physical habitat criterion. For the third subset (n = 14), we added both the quantitative physical habitat criterion and a catchment condition screen. Only those reference reaches with quantitative physical habitat data were included in the latter two reference subsets. The average of the 25th percentile values across the three groups of reference reaches (74 for riffles) was used to separate Good from Fair condition and the 1st percentile of the distribution of reference reaches (39 for riffles) was used to distinguish Fair from Poor. A single set of condition scoring criteria was used for both pool- and riffle-based index scores because scoring of individual metrics was done separately for riffle and pool data. However, the MBII score and condition category for each reach was evaluated using the dominant habitat type in the reach.

Results

Index Responsiveness and Precision

We evaluated the final index by examining its responsiveness and precision using the validation data set. Among riffle samples, the S/N ratio was 4.804, but

Table 2. Floor and ceiling values for metric scoring expectations for pools and riffles

Metric	Riffle floor	Riffle ceiling	Pool floor	Pool ceiling
Positive metrics				
Ephemeroptera richness (watershed-adjusted)	4.21	14.11	4.31	9.72
Plecoptera richness (watershed-adjusted)	3.56	9.46	2.95	6.03
Trichoptera richness	1	7	0	3
Collector-filterer richness (watershed-adjusted)	3.88	10.37	3.75	8.98
Negative metrics				
% Non-insect individuals	0	17.89	0.69	27.03
MTI	3.65	5.36	4.32	6.21
% Individuals in 5 dominant taxa	42.39	80.18	55.37	78.85

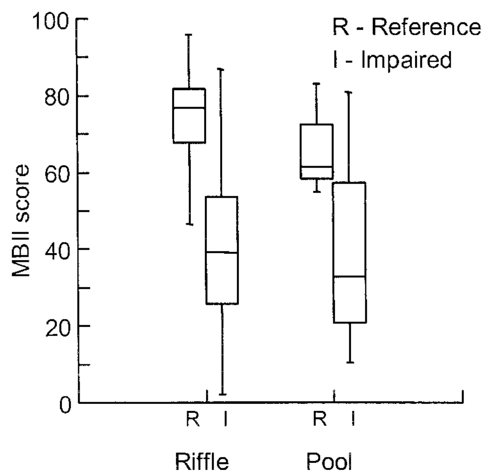

Figure 2. MBII scores in reference and impaired reaches for pool and riffle habitats. Reaches are from a validation data set (n = 102) of probability reaches. The center line represents the median, and the upper and lower edges of the box represent the 75th and 25th percentiles, respectively. The lower and upper vertical lines (whiskers) represent values that fall within the inner fences [25th percentile - 1.5* (median - 25th percentile) and 75th percentile + 1.5* (75th percentile - median)] (SPSS 1998).

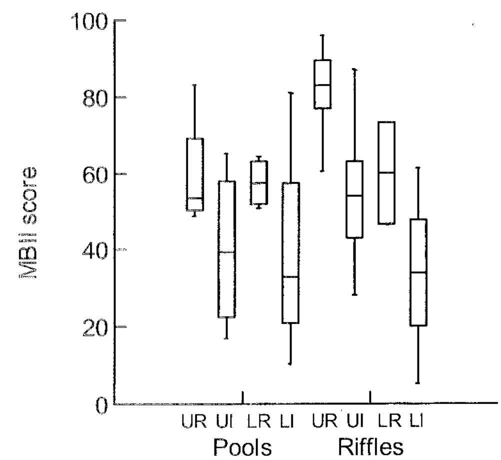

Figure 3. MBII scores by ecoregion group and habitat type for riffle and pool samples using the entire data set.

among pool samples, it was only 0.975. Riffle scores were significantly correlated with general habitat, general disturbance, sedimentation, riparian habitat, and acidity measures ($p = 0.0031$, when a Bonferroni correction for multiple tests is applied). Pool index scores were significantly correlated with measures of general habitat, general disturbance, sedimentation, and nutrients ($p < 0.0031$). Neither riffle- nor pool-based scores were correlated with catchment area ($p > 0.50$).

Based only on the validation data set, riffle scores discriminated well between reference and impaired reaches (Figure 2). However, there was considerable variability among pool samples from impaired reaches, and the index did not distinguish well between reference and impaired reaches for pools. Scores for uplands reference reach riffles were higher than for pools, but lowlands reference reach scores did not differ between riffles and pools (Figure 3, all data included in plots). In addition, scores for lowlands reference and impaired riffles were shifted downward relative to scores from uplands reference and impaired stream reaches, although this was not true for pools. However, when only the sample from the dominant habitat type (*i.e.,* riffles or pools) for each reach (using all reference and impaired reach data) was used to calculate the MBII, there was excellent discrimination of reference from impaired reaches in both the uplands and lowlands (Figure 4). Pools were the dominant habitat in only one lowland reference stream reach, and they dominated in 12 uplands and 10 lowlands impaired reaches.

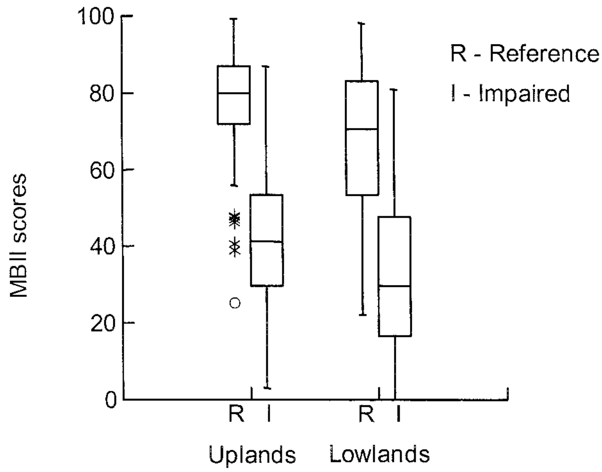

Figure 4. Ability of MBII scores to distinguish reference and impaired sites in the uplands and lowlands, on the basis of only the dominant habitat type (*i.e.,* riffles or pools) within each reach.

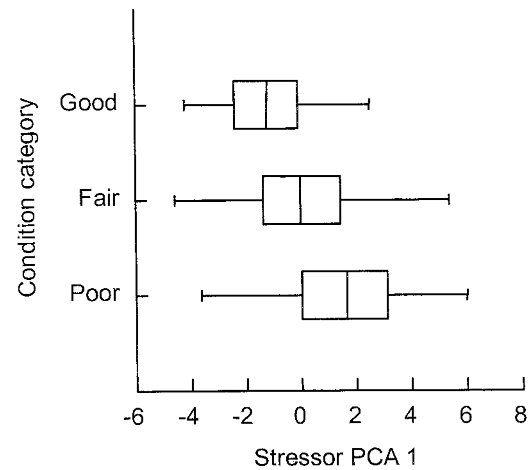

Figure 5. Stream condition category versus a multivariate stressor axis (PCA-1).

Table 3. Weights on the first two stressor principal components for MAHR sampling reaches

Variable	Axis 1	Axis 2
Qualitative mean habitat score	−0.375	0.402
Qualitative channel alteration score	−0.243	0.335
Qualitative embeddedness score	−0.326	0.234
Qualitative epifaunal substrate score	−0.273	0.371
Qualitative riparian vegetation width score	−0.310	0.233
pH	0.206	0.127
Turbidity	0.274	0.114
Sulfate	0.134	0.175
Total nitrogen	0.309	0.351
Total phosphorus	0.299	0.283
Chloride	0.301	0.349
% Catchment as urban, agriculture, and mining land use	0.335	0.307

We ran a Principal Components Analysis (PCA) on the stressor variables used for testing responsiveness. We used only those variables common to most reaches, and thus excluded variables associated with quantitative physical habitat data. We log-transformed turbidity, sulfate, total nitrogen, total phosphorus, chloride ion, and percent catchment land use in agriculture, urban, and mining data (percent disturbed catchment). The first PCA axis accounted for approximately 38% of the variation (eigenvalue = 4.58, Table 3) and the second accounted for another 15% (eigenvalue = 1.81, Table 3). Both of these axes were significant, with eigenvalues much larger than those predicted with the broken stick model (Frontier 1976). On the first axis, the mean qualitative habitat, qualitative embeddedness, and qualitative riparian vegetation width scores had the highest negative loadings, and the percent disturbed (as urban, agriculture, or mining land use) catchment, total nitrogen, and chloride ion had the highest positive loadings. Note that for all qualitative habitat scores, a higher value indicates better condition. On the second axis, the mean qualitative habitat score, qualitative epifaunal substrate score, total nitrogen, and chloride ion all had relatively high positive loadings. The first axis clearly reflects a water quality, land cover, and habitat gradient. However, the interpretation of the second axis was less obvious. The first axis was significantly correlated with the MBII ($r = -0.483$, $p < 0.001$), but the second was uncorrelated with index score ($r = 0.042$, $p = 0.342$). MBII condition class and score, both based on the dominant habitat type at a reach, progressively decreased with increasing anthropogenic stress (Figures 5 and 6). Both uplands and lowlands ecoregional groupings and reach types exhibited substantial variation at any location for both the stressor gradient and the MBII score (Figure 6). This provides evidence that separate metrics and/or scoring of metrics is not needed for these two ecoregional groups.

Regional Assessment of Stream Condition

We used the MAHR sampling design and MBII scores from probability reaches to estimate the target stream length and percent of target stream length in each condition category. We calculated 90% confidence bounds for these measures over the entire MAHR and separately for the uplands and lowlands subsets. The condition of each stream reach was based

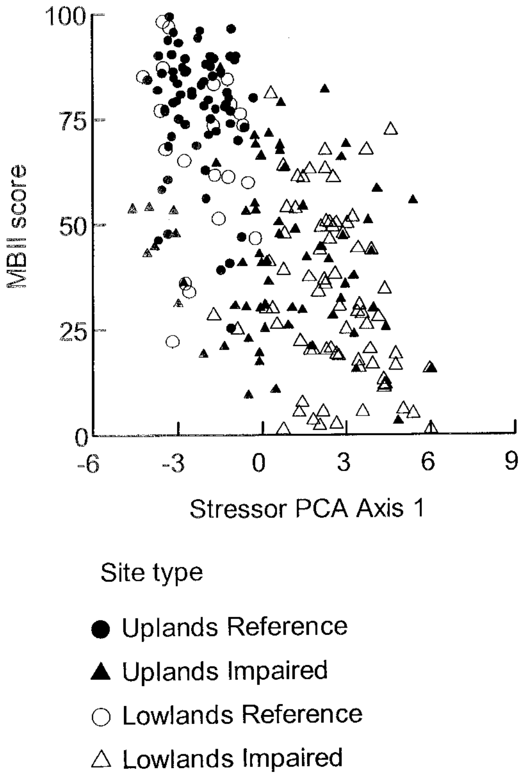

Figure 6. MBII score as a function of a multivariate stressor axis (PCA-1).

on the dominant habitat type (riffle or pool) as determined by the proportion of transects of each habitat type in each stream reach sampled. Approximately 92% of uplands stream length and 85% of lowlands stream length (88% of stream length overall) was dominated by riffle habitat (i.e., represented by riffle composite samples).

The total target sampleable, wadeable stream length mapped at a 1:100,000 scale in the MAHR is estimated to be 169,008 km. Approximately 1211 km (0.72%) of stream length was not assessed because of a lack of sufficient samples, resulting in assessment of approximately 167,797 km of stream length according to the MBII. Over 50% of the total target stream length was estimated to be in Fair condition in the uplands, lowlands, and the entire MAHR (Table 4), with 90% confidence interval lengths ranging between 12% and 30%. In the lowlands, however, the second most common condition was Poor (between 20% and 46%), whereas it was Good in the uplands (between 23% and 34%). Over the entire MAHR, between 17% and 34% of the total target stream length was estimated as being in Poor condition on the basis of our MBII.

Table 4. Estimated percentage and length of stream kilometers in each condition category

Ecoregion grouping	Condition category[a]	% of stream km (90% confidence interval)	Length of stream km
Uplands	Good	28.2 ± 5.4	18237
	Fair	56.5 ± 6.4	36492
	Poor	15.2 ± 4.0	9832
Lowlands	Good	9.6 ± 10.3	9955
	Fair	57.7 ± 14.6	59551
	Poor	32.7 ± 13.1	33730
Entire MAHR	Good	16.8 ± 6.6	28192
	Fair	57.2 ± 9.3	96043
	Poor	26.0 ± 8.1	43562

[a]Condition based on the dominant stream habitat at each sampling site.

Discussion

Biological integrity has been assessed at a regional scale using macroinvertebrates for the Mid-Atlantic Coastal Plains (Maxted and others 2000) and using fish for the MAHR (McCormick and others 2001), but such assessments are uncommon. Our results differed slightly from those of McCormick and others (2001). They estimated that 27% (± 13%) of the stream length in the MAHR was good or excellent, 38% (± 10%) was fair, and 14% (± 5%) was poor on the basis of a fish assemblage IBI. We concluded that approximately 16.8% (± 6.6%) of the MAHR was in good condition, 57.2% (± 9.3%) was in fair condition, and 26% (± 8.1%) was in poor condition. However, McCormick and others (2001) were unable to estimate condition on the basis of fish assemblages in 21% of the stream length because many small catchments were too small to support fish (<10 individual fish were caught). The MBII scores in these stream reaches were very similar to those of the whole population. The MBII was able to assess biological condition for over 99% of the targeted stream length, making this index a particularly useful tool in such situations.

The MBII was developed for the MAHR, and its application in other parts of the country must be with caution. The set of anthropogenic influences may differ in other areas, and this index may not respond to those influences. Alternatively, the metrics selected for this index may be responsive in other areas, but the metric scoring expectations set using the MAHR data may be inappropriate for other areas, either too low or too high to represent reference conditions well. However, the process used to develop the MBII is widely applicable as long as minimally disturbed and highly disturbed sites can be identified from available data.

Our regional-scale assessment was at a considerably larger scale than the stream reach, catchment, or state scales at which biological indices have traditionally been applied (Ohio EPA 1988, Lenat 1993, Kerans and Karr 1994, DeShon 1995, Barbour and others 1996, Lewis and others 2001). This larger spatial scale afforded us a wider range in the intensity and variety of human disturbances against which to measure index responses. However, in such a heterogeneous region, efforts to set scoring expectations also had to incorporate wider natural variation in macroinvertebrate assemblages. In addition, many topographic and edaphic characteristics that distinguish the uplands from the lowlands have also influenced historical and current human land uses in these areas. Lowland areas in the MAHR are more suitable for agricultural or urban land uses. The steeper-sloped, thinner-soiled catchments in the uplands are less suitable for urban and agricultural land use, but support intensive coal mining and silviculture, and are more sensitive to acidic deposition. Ecoregional differences such as these frequently lead scientists to treat ecoregions as separate strata with unique expectations, an option that we explored. The number of sample reaches with evidence of high anthropogenic stress (PCA values >0) (Figure 6) was greater in the lowland ecoregional group than in the uplands. Despite the greater frequency of disturbance and the paucity of minimally disturbed reference reaches in the lowlands, the full range of PCA axis 1 occurs in both ecoregion groupings (Figure 6). The disturbance gradient in Figure 5 incorporates a wide variety of disturbance types and intensities, both singly and in combination. The MBII responded to this generalized measure without any clear indication that the response to a given level of disturbance depended on the ecoregional group, as evidenced by the overlap between upland and lowland sites in Figure 6 and the consistent discrimination of reference from impaired reaches in both ecoregional groups (Figure 4).

Our selection of minimally disturbed reference reaches in the lowlands was limited. However, a large amount of the ecoregional variation in macroinvertebrates can be attributed to ecoregional variation in stream gradient. Our use of differential scoring in riffle-dominated and pool-dominated stream reaches apparently made it unnecessary to score reaches differently according to their ecoregional grouping. Waite and others (2000) also found that macroinvertebrate assemblage composition in the MAHR was more strongly associated with local aquatic habitat characteristics (stream size and gradient) than with ecoregional classification. In addition, McCormick and others (2001) found that fish assemblage composition was not strongly related to ecoregion. Given separate index scoring on the basis of local habitat features related to stream size and gradient, we found that the MBII responded well to a generalized measure of disturbance (PCA axis) representing a wide variety of combined stressors across a heterogeneous upland and lowland landscape.

The separation of samples into pool and riffle components allowed us to base our assessment of a particular reach on only the dominant habitat at a reach. We were also able to determine that pools in the MAHR consistently exhibited smaller ranges between the ceiling and floor values used in scoring metrics, typically setting lower expectations for the maximum scores for each metric. By and large, we were able to correct for this difference in dominant habitat by separately scoring pool and riffle samples. Reach assessment may differ depending on the dominant habitat, particularly when pools and riffles are both very common. However, we did not observe a difference in scores between the two habitat types that would influence the condition rating greatly (Figure 2). The proportion of reaches with pool habitat was larger in the lowlands than the uplands, but the proportion of reaches dominated by pools was very small overall, particularly among reference reaches, so differences among riffles and pools is probably a very minor concern in terms of whole reach assessments.

Several of the metrics selected for the MBII are not used typically in other biological indices, though variations are. Although EPT richness is commonly used in biological assessments (Barbour and others 1995), the use of a separate richness metric for each order is less common (Fore and others 1996, DeShon 1995). Breaking the EPT into separate orders may allow more effective diagnosis of stressors among impaired waters because each order responds differently to different stressors (Clements 1994, Fore and others 1996, Karr and Chu 1999). The use of feeding guilds in assessment has been suggested because these measures can reflect fundamental differences in trophic patterns and nutrient sources among reaches (Kerans and Karr 1994, Hannaford and Resh 1995). However, few indices include feeding guild metrics. The Florida Department of Environmental Protection has successfully included percent filterers in its Stream Condition Index (SCI) for wadeable streams (Barbour and others 1996). The use of collector-filterer richness in the MBII may allow better detection of impairments related to increased sedimentation because this metric was related negatively to turbidity and the qualitative habitat score.

Percent non-insects is not commonly used (Barbour and others 1995), although variations, such as percent-non-Tanytarsini dipterans plus non-insects (DeShon 1995) and percent oligochaetes (Kerans and Karr 1994), have been used because these metrics seem to respond more strongly to organic pollution (DeShon 1995). Percent 5 dominant taxa is also not very common (Barbour and others 1995), although some indices do currently include variations of this metric, particularly percent dominant taxon (Barbour and others 1995). Klemm and others (2002) evaluated several characteristics of a subset of the metrics included in this study but used only riffle data from the MAHR. Their results indicated that all of the metrics selected for the MBII except collector-filterer richness, which was not evaluated, were suitable for inclusion in a biotic index.

Overall, the MBII was an effective tool for assessing stream condition on a regional scale in the MAHR. The index responded to a variety of stressors affecting streams in the region, although it did rely on separate sampling of riffle and pool habitats in those streams. In addition, although the index performed slightly better in upland streams, it did not appear to require separate determinations in various ecoregions. Because data on pool habitat among reference streams were more limited, this index is probably most useful in streams dominated by riffle habitat. For those streams, the ability of the index to distinguish impaired conditions is excellent in both uplands and lowlands, making it very useful for region-wide and basin-wide assessments.

Acknowledgments

We thank Phil Lewis, Jim Green, James Lazorchak, Phil Larsen, and Ian Waite for suggestions on field, laboratory, and analytical methods. In addition, we thank Erica Grimmett, Brad Autrey, Shawn Mills, EMAP field crews, and Mary Schubauer-Berigan for assistance in various phases of sample collection, sample processing, computer data input and analyses, and data interpretation. Finally, we thank Don Schloesser, Mark Munn, John Hutchens, Gary Collins, and Bob Angelo for suggestions provided in critical reviews of the manuscript. This work was supported by the U.S. Environmental Protection Agency through its Office of Research and Development, in part through cooperative agreement 824682 with Oregon State University, contract 68D0I005 with Dynamac, and contract 68C60019 with SoBran, Inc. (c/o USEPA). It has been subjected to Agency review and approved for publication.

Literature Cited

Allan, J. D. 1995. Stream ecology: structure and function of running waters. Kluwer Academic, Boston, MA, 388 pp.

Barbour, M. T., and J. B. Stribling. 1991. Use of habitat assessment in evaluating the biological integrity of stream communities. Pages 25–38 *in* Biological criteria: Research and regulation. EPA-440/5-91/00 5. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Barbour, M. T., J. B. Stribling, and J. R. Karr. 1995. The multimetric approach for establishing biocriteria and measuring biological conditions. Pages 63–77 *in* W.S. Davis and T. Simon (eds.) Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

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* 15:185–211.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA 841-B-99-002.

Bryce, S. A., R. M. Hughes, and P. R. Kaufmann. 2002. Development of a bird integrity index: using bird assemblages as indicators of riparian condition. *Environmental Management* 30(2):294–310.

Clements, W. H. 1994. Benthic invertebrate community response to heavy metals in the Upper Arkansas River Basin, Colorado. *Journal of the North American Benthological Society* 13:30–44.

DeShon, J. E. 1995. Development and application of the invertebrate community index (ICI). Pages 217–243 *in* W.S. Davis and T.P. Simon (eds.) Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Fore, L. S., J. R. Karr, and R. W. Wisseman. 1996. Assessing invertebrate responses to human activities: Evaluating alternative approaches. *Journal of the North American Benthological Society* 15:212–231.

Frontier, S. 1976. Étude de la décroissance des valeurs propres dans une analyse en composantes principales: comparaison avec le modèle du bâton brisé. *Journal of Experimental Marine Biology Ecology* 25:67–75.

Hannaford, M. J., and V. H. Resh. 1995. Variability in macroinvertebrate rapid-bioassessment survey and habitat assessment in a northern California stream. *Journal of the North American Benthological Society* 14:430–439.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.

Herlihy, A. T., D. P. Larsen, S. G. Paulsen, N. S. Urquhart, and B. J. Rosenbaum. 2000. Designing a spatially balanced, randomized reach selection process for regional stream surveys: The EMAP Mid-Atlantic pilot study. *Environmental Monitoring and Assessment* 63:95–113.

Hill, B. H., A. T. Herlihy, P. R. Kaufmann, R. J. Stevenson, F. H. McCormick, and C. B. Johnson. 2000. Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19:50–67.

Hilsenhoff, W. L. 1987. An improved biotic index of organic stream pollution. *The Great Lakes Entomologist* 20:31–39.

Hughes, R. M,, and T. Oberdorff. 1999. Application of IBI concepts and metrics to waters outside the United States and Canada. Pages 79–93 *in* T.P. Simon (ed.), Assessing the sustainability and biological integrity of water resources using fish communities. Lewis Publishers, Boca Raton, Florida.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, T. M. Kincaid, L. Reynolds, and D. P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1618–1631.

Hynes, H. B. N. 1960. The biology of polluted waters. Liverpool University Press, Liverpool, 202 pp.

Hynes, H. B. N. 1970. The ecology of running waters. University of Toronto Press, Toronto, Ontario, Canada, 555 pp.

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

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters. Island Press, Washington, D.C., 206 pp.

Karr, J. R., and E. W. Chu. 2000. Sustaining living rivers. *Hydrobiologia* 422– 423: 1–14.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Surface Waters: Quantifying physical habitat in wadeable streams. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Environmental Monitoring and Assessment Program, Research Triangle Park, North Carolina. EPA/620/R-99/003.

Kerans, B. L., and J. R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* 4:768–785.

Klemm, D. J., P. A. Lewis, F. A. Fulk, and J. M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. U.S. Environmental Protection Agency, Cincinnati, OH. EPA.600/4-90/030.

Klemm, D. J., and J. K. Hiltunen. 1992. The freshwater Annelida (Polychaeta, naidid and tubificid Oligochaeta, and Hirudinea) of the Great Lakes region—an overview. Pages 35– 48 *in* Proceedings of the 1991 Midwest Pollution Control Biologists Meeting: Environmental Indicators: Measurements and Assessment Endpoints, 19–22 March 1991. Lincolnwood, Illinois, U.S. Environmental Protection Agency, Region V, Chicago, Illinois. EPA 90-5/R-92/003.

Klemm, D. J., and J. M. Lazorchak. 1994. Environmental Monitoring and Assessment Program, Surface Waters and Region 3 Regional environmental Monitoring and Assessment Program, 1994 Pilot Laboratory Methods Manual for Streams. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio. EPA/620/R-94/003.

Klemm, D. J., K. A. Blocksom, W. T. Thoeny, F. A. Fulk, A. T. Herlihy, P. R. Kaufmann, and S. M. Cormier. 2002. Methods development and use of macroinvertebrates as indicators of ecological condition for streams in the Mid-Atlantic Highlands Region. *Environmental Monitoring and Assessment* 78(2): 169–212.

Lazorchak, J. M., D. J. Klemm, and D. V. Peck. 1998. Environmental Monitoring and Assessment Program-Surface Waters: Field operations and methods for measuring the ecological condition of wadeable streams. U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory and National Health and Environmental Effects Research Laboratory, Research Triangle Park, North Carolina. EPA/620/R-94/004F.

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. *Journal of the North American Benthological Society* 12:279–290.

Lewis, P. A., D. J. Klemm, and W. T. Thoeny. 2001. Perspectives on use of a multimetric lake bioassessment integrity index using benthic macroinvertebrates. *Northeastern Naturalist* 1:233–246.

Maxted, J. R., M. T. Barbour, J. Gerritsen, V. Poretti, N. Primrose, A. Silvia, D. Penrose, and R. Renfrow. 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 19:128–144.

McCormick, F. H., R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands Region. *Transactions of the American Fisheries Society* 130:857– 877.

Merritt, R. W., and K. W. Cummins. 1996. An introduction to the aquatic insects of North America. Kendall/Hunt, Dubuque, Iowa, 722 pp.

Messer, J. J., R. A. Linthurst, and W. S. Overton. 1991. An EPA program for monitoring ecological status and trends. *Environmental Monitoring and Assessment* 17:67–78.

Miller, D. L., P. M. Leonard, R. M. Hughes, J. R. Karr, P. R. Moyle, L. H. Schrader, B. A. Thompson, R. A. Daniels, K. D. Fausch, G. A. Fitzhugh, J. R. Gammon, D. B., Halliwell, P. L. Angermeier, and D. J. Orth. 1988. Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13(5):12–20.

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. Applied linear statistical models, fourth edition. McGraw-Hill, Boston, Massachusetts, 1408 pp.

Norris, R. H., and C. P. Hawkins. 2000. Monitoring river health. *Hydrobiologia* 435: 5–17.

Ohio EPA. 1988. Biological criteria for the protection of aquatic life. Volume I: The role of biological data in water quality assessment. Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment, Columbus, Ohio.

Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77: 118–125.

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. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, D.C. EPA 841-D-97-002.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16:833– 852.

Simon, T. P. 2002. Biological response signatures: Indicator patterns using aquatic communities. CRC Press, Boca Raton, Florida, 584 pp.

Simon, T. P., and J. Lyons. 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. Pages 245–262 *in* W.S. Davis and T.P. Simon (eds.) Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

SPSS. 1998. Systat 8.0 Graphics. SPSS, Inc., Chicago, Illinois, 432 pp.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union* 38:913–920.

USEPA. 1987. Handbook of methods for acid deposition studies: Laboratory analysis for surface water chemistry. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. EPA/600/4-87/026.

USEPA. 1996. Summary of state biological assessment programs for streams and rivers. U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Washington, D.C. EPA230-R-96-007.

USEPA. 1998. Environmental Monitoring and Assessment Program (EMAP): research plan 1997. U.S. Environmental Protection Agency, Washington, D.C. EPA/620/R-98/002.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37:130–137.

Waite, I. R., A. T. Herlihy, D. P. Larsen, and D. J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. *Journal of the North American Benthological Society* 19:429–441.

Warren, C. E. 1971. Biology and water pollution control. W.B. Saunders, Philadelphia, Pennsylvania, 434 pp.