

October 2015

Technical Memorandum #4

Applying Benthic Macroinvertebrate Multimetric Indexes to Stream Condition Assessments

James B. Stribling and Steven A. Dressing. 2015. Technical Memorandum #4: Applying Benthic Macroinvertebrate Multimetric Indexes to Stream Condition Assessments, October 2015. Developed for U.S. Environmental Protection Agency by Tetra Tech, Inc., Fairfax, VA, 14 p. Available online at <https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/watershed-approach-technical-resources>.

Introduction

The primary objective of the Federal Water Pollution Control Amendments of 1972—commonly known as the Clean Water Act (CWA)—“is to restore and maintain the chemical, physical, and biological integrity of the Nation’s waters” (Title 33 of the *United States Code* section 1251). To strengthen the scientific foundation of the CWA, “biological integrity” has been defined as “the ability (of a water body) to support and maintain a balanced, integrated, biological community having a species composition, diversity, and functional organization comparable to that of natural habitats in the region” (Karr and Dudley 1981; Schneider 1992). The capacity for aquatic organisms to survive and reproduce in nature is controlled by both basic biological and physiological processes of the organisms and characteristics of their immediate environment. A water body with chemical and physical characteristics that are close to those found in a naturally occurring habitat can be considered to represent chemical and physical integrity, and therefore potentially supportive of a healthy biological condition.

Waterbodies may be physically or chemically degraded by a variety of forces. Environmental disturbances are caused either by natural extremes (e.g., hurricanes, volcanic eruptions, earthquakes) or by human activities (e.g., waste discharge, cropland erosion, stormwater runoff), and both types can lead to changes in the natural environment. Many human activities lead to direct introduction of chemical pollution, changes in physical habitat, or alteration of hydrologic processes. These changes may diminish the availability of food and habitat area—including refuge from natural extremes such as stream flow or temperature—and can alter reproductive behavior. Such factors are termed “stressors.” For purposes of this technical memorandum, a stressor is defined as any agent that limits the biological capacity for survival and reproduction. Biological responses to stressor exposure are varied, but can include direct mortality (e.g., from acute toxicity), longer term chronic effects, reduced reproductive success, increased incidence of disease and/or predation, and elimination of a local population from loss of habitat.

Degraded biological conditions result from the buildup of stressors, which results from environmental degradation. As stewards, we must recognize and effectively document environmental degradation and accept the responsibility to take appropriate actions to address it and its causes. Using aquatic organisms as indicators of changes to chemical and physical components of the environment can enhance our ability to identify the level of degradation and the actions to take (Figure 1). Rigorous and defensible measurement techniques do not exist for all pollutants, combinations of pollutants, or other potential detrimental factors. But, because aquatic organisms are directly exposed to their immediate surroundings, biological characteristics will reflect

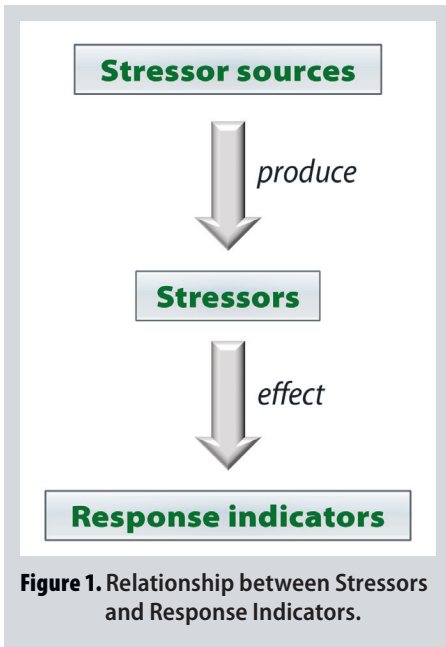


Figure 1. Relationship between Stressors and Response Indicators.

environmental suitability, including factors that are unknown, unmeasured, and often, unmeasurable. In addition, the community of aquatic biota naturally and cumulatively integrates the effects of multiple and complex stressors over time, whether exposure is continuous, intermittent, or episodic. For those pollutants, or stressors, that are less than continuous, use of biological indicators sidesteps the need to sample specifically when the pollutants are present.

Over the past 20–30 years, methods and protocols have been developed to support characterization of biological condition through direct field sampling, laboratory and data analysis, and stream and watershed assessment. Protocols are typically customized to meet technical needs, with individual monitoring programs synchronizing capabilities and staff resources within budgetary constraints. Simultaneously, new quality control (QC) procedures have been developed to ensure that data and assessments are of known and acceptable quality (Flotemersch et al. 2006; Stribling 2011).

The organism groups most commonly used in routine biological monitoring and assessment programs in the United States are benthic macroinvertebrates (BMIs)—aquatic insects, snails, mollusks, crustaceans, worms, and mites; fish; and/or algae—with indicators most often taking the form of a multimetric Index of Biological Integrity, or IBI (Karr et al. 1986; Hughes et al. 1998; Barbour et al. 1999; Hill et al. 2000, 2003). An IBI combines several individual metrics, each of which describes a different aspect of assemblage structure and function (Barbour et al. 1995). Note that IBIs are occasionally called multimetric indexes (MMIs). The acronym “IBI” is a direct connection of the indicator to biological integrity of the CWA, whereas “MMI” is a generic term simply meant to communicate, in part, the numeric structure of the index.

Although fish and algae also are often used for evaluating the biological condition of water bodies, this technical memorandum focuses solely on BMIs. This group of organisms has continually been proven for use in routine biological monitoring because field sampling protocols are well-established, the level of effort required for field sampling is reasonable (Barbour et al. 1999), and taxonomic expertise is relatively easily accessible. An IBI has a straightforward numeric structure that serves to summarize and scale complex biological data. Within boundaries defined by the sampling and analysis protocol, the BMIs in a stream sample represent the full taxonomic diversity of the assemblage present at a site. Taxonomic diversity, compositional abundance, and autecological characteristics (e.g., feeding types, habits, and stressor tolerance values) each conveys information potentially useful in detecting and understanding degradation and biological responses to it. An effort is typically made during IBI development to have one or more individual metrics represent as many different information types, or metric categories, as possible associated with the biological assemblage, including taxonomic richness, community composition, stressor or pollution tolerance, functional feeding types, and locomotory habit (Barbour et al. 1999). This diversity of categories increases the probability that an index can detect the effects of multiple and complex stressors. The multimetric IBI approach allows numeric consideration of the composite sample content (via the

biological index score) and places the sample result on the same scale with other samples from the same or different sample locations and/or from different water bodies.

The structure of IBIs allows them to be easily disaggregated into individual metrics as well as individual taxa. Properly calibrated indexes will be made up of metrics that have been tested for direction of change in the presence of stressors and are based on taxa with which autecological attributes are associated. It also is possible to evaluate the specific component metrics and taxa that have the most influence on the overall index score and individual metric values and, thus, the site assessment. Also, since typically, multiple and complex stressors are present in streams to which the assemblage is exposed, examining overall index scores and assessments, metric values, and taxa can often provide evidence pointing to stressor types and intensities.

In a very broad sense, two components make up the “life cycle” of an IBI: 1) calibration or development, and 2) application. Driving the need for calibration is the recognition that one size does not fit all and that customizing the index makes each indicator more effective in detecting degradation. Natural variation is observed in benthic macroinvertebrate assemblages across all spatial (e.g., regions, landscapes, water body types) and temporal (e.g., daily, monthly, seasonally, and within and among years) scales (Figure 2). Calibrating a biological index defines protocols and procedures that help control variability in sample data so that differences in index scores and assessments can be interpreted as indicative of real environmental change, including detection of degradation due to stressors from both nonpoint and point sources. Variability in data resulting from their collection, analysis, and documentation—known as “systematic error”—is minimized by using standard operating procedures and ensuring acceptability for application through use of routine and consistent QC (Stribling 2011).

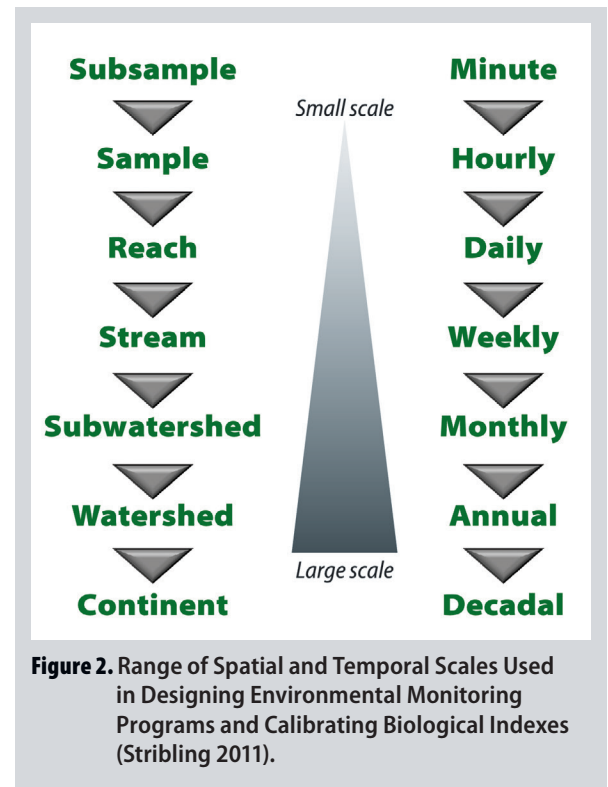


Figure 2. Range of Spatial and Temporal Scales Used in Designing Environmental Monitoring Programs and Calibrating Biological Indexes (Stribling 2011).

Purpose and Audience

The purpose of this technical memorandum is to describe the second component of the IBI life cycle: its application to assessment of and reporting on aquatic ecological condition of a water body. The content of the memorandum is presented with the assumption that an index has been appropriately calibrated for the region and water body type of interest. It describes field sampling for benthic macroinvertebrates, laboratory processing (i.e., sorting/subsampling and taxonomic identification), IBI calculation, and site assessment. This technical memorandum should be most helpful to staff responsible for implementing biological monitoring and assessment programs for nontidal streams and rivers.

The biological assessment results presented in this technical memorandum do not tell environmental managers what actions to take; rather, they stop with the documentation of the presence/absence and severity of a problem (in this case, biology was not degraded, and thus, no problem exists). Additional analysis and interpretation is often necessary to increase confidence when determining necessary and appropriate actions. Stressor identification and causal analysis (USEPA 2010) is a systematic approach for determining the stressors and stressor sources most likely to be causing the problem, and improve support for decisions on the types and locations of stressor control activities (e.g., best management practices, stream restoration, stormwater management).

Quality Control and the Use of Measurement Quality Objectives

Reliability of biological assessment results can be impaired by errors introduced throughout the process, beginning with field sampling and sample sorting and subsampling and ending with metric and index calculation, and site condition assessment. Use of a series of QC checks—for example, measurement quality objectives (MQOs) (Stribling et al. 2011)—integrated into the process helps the assessor to recognize, control, and minimize errors (Table 1). If observed values exceed the MQOs, it does not automatically indicate the existence of invalid or unacceptable data points. Rather, the values are targeted for closer scrutiny to determine possible reasons for the exceedance and might indicate a need for corrective actions.

Table 1. Key Measurement Quality Objectives (MQO) Used for QC Evaluation and Tracking Data Quality

Performance Characteristic	MQO
Field sampling precision (multimetric index)	CV < 10%, for a sampling event (field season, watershed, or other strata)
	CI90 ≤ 15 index points (on a 100-point scale)
	RPD < 15
Field sampling completeness	Completeness > 98%
Sorting/subsampling bias	PSE ≥ 90 (for ≥ 90% of externally QC'd sort residues)
Taxonomic precision	Median PTD ≤ 15% for overall sample lot (samples with PTD ≥ 15% examined for patterns of error)
	Median PDE ≤ 5% (samples with PDE ≥ 5% should be further examined for patterns of error)
Taxonomic completeness	Median PTC ≥ 90% (samples with PTC ≤ 90% should be examined and taxa not meeting targets should be isolated) mAbs diff ≤ 5%

Notes:

CI90 = 90% confidence interval; CV = coefficient of variability; mAbs diff = median absolute difference;
 PDE = percent difference in enumeration; PSE = percent sorting efficiency; PTC = percent taxonomic completeness;
 PTD = percent taxonomic disagreement; QC = quality control; RPD = relative percent difference.

IBI Application

Field Sampling

Sites to be sampled are typically preselected and defined in the office using maps and other available data, regardless of whether they are targeted or probability-based. Assessments at targeted sites are for the specific locations from which the samples are taken, whereas those from probability sites can also be aggregated to broader-scale assessments, such as watershed. Reach length sampled varies among programs, but 100 meters is widely used for wadeable streams (generally, 1st–4th order Strahler). Some programs sample reach lengths that are a multiple of the median wetted channel width (MWCW), such as 20x or 40x (Flotemersch et al. 2011).

For its national surveys of streams and rivers, the U.S. Environmental Protection Agency defined a sample reach as 40x the mean wetted width (USEPA 2004) and collected BMI samples along 11 transects evenly distributed throughout the reach. Using the transect approach helps minimize and control bias by the field crews in selecting the specific habitats sampled in the reach. An alternative approach—proportional distribution—has the field crews estimate the proportion of different habitat types in a defined reach (e. g., 100m, and distribute a fixed level of sampling effort in proportion to their frequency of occurrence throughout the reach) (Barbour et al. 1999; Stribling 2011). Samples from both approaches result in a biological descriptor (i.e., using BMI) of the full sampling reach.

Carter and Resh (2013) reported that D-frame nets are commonly used across the United States by biological monitoring programs for composite sampling of BMI. Organic and inorganic sample material (e.g., leaf litter, small woody twigs, silt, and sand) are composited in one or more containers (Figure 3), preserved in 95 percent denatured ethanol, and delivered to laboratories for processing.

Repeat sampling for a monitoring and assessment program provides data for both field QC evaluation and for calculating different measures of field sampling precision (Stribling 2011; Flotemersch et al. 2006). A rate of 10 percent duplication is relatively common among programs (Carter and Resh 2013), and we recommend collecting duplicate samples at a *minimum* of three reaches. A duplicate sample in streams and other flowing waters would be one taken from an adjacent reach (i.e., one that begins at either the upstream or downstream extent of the primary reach). Reaches for which duplicate samples are taken are randomly selected prior to initiation of the field effort. Sample labeling should be identical between primary and duplicate samples, with the exception

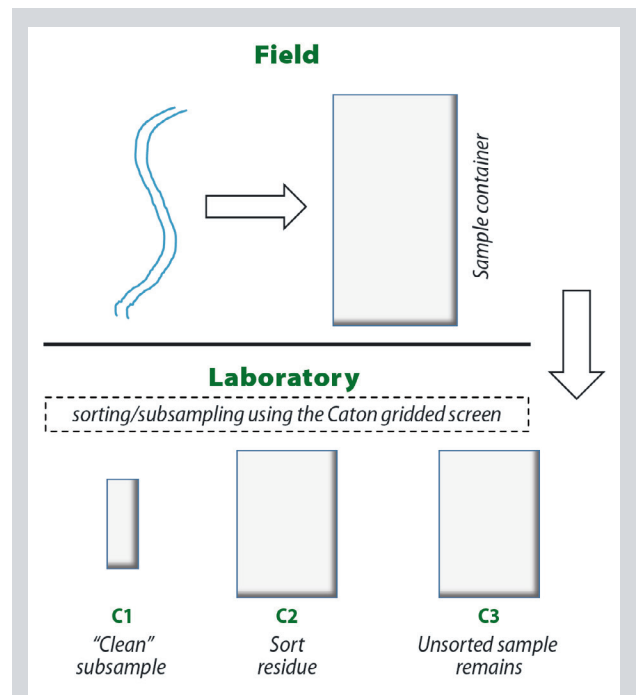


Figure 3. Benthic macroinvertebrate samples are field-collected, composited, and preserved in one or more containers with 95% ethanol. They are transferred to the laboratory, and the sorting/subsampling process results in three sample units: containers C1, C2, and C3.

that the latter should be clearly labelled as such (e. g., with “DUP” or “QC”). Data quality indicators for field sampling cannot be calculated until after laboratory processing is completed (see the Sample Processing and Data Preparation section). Performance characteristics recommended for quantifying field sampling precision include relative percent difference (RPD), coefficient of variation (CV), and 90 percent confidence intervals (CI90). For relevant formulas, see Stribling (2011) and/or Flotemersch et al. (2006).

Sample Processing and Data Preparation

Several aspects of BMI sample processing must be completed prior to calculating performance measures or indicator values (i.e., metrics and index). First, the field samples are sorted and subsampled, which serves to separate individual organisms from nontarget material such as leaf litter and other detritus, bits of woody material, silt, and sand. Simultaneously, a randomly selected subset of organisms (the subsample) is isolated for taxonomic identification. Names (accepted nomenclature) are matched to individual specimens in the subsample, resulting in a list of organisms by name and the number of individuals of each taxon.

Sorting/subsampling. Recommended equipment and supplies for sorting and subsampling in the laboratory include the Caton gridded screen and accessories (Figure 4), forceps, sample jars or vials, and 70–80-percent ethanol. Typically, sorting is done solely with the naked eye or using some combination of an illuminated ring lamp with 3–5x magnification and a binocular dissecting microscope (10–40x magnification). The process involves the following steps:

- Step 1. Empty material from sample jar onto gridded subsampling tray; fully and evenly spread material on tray, giving special attention to the corner squares.
- Step 2. Randomly select four grid squares.
- Step 3. Lift screen from outer box, offset (as shown in Figure 4).
- Step 4. Remove material from the four selected grids and place in separate picking tray (use cookie cutter and scoop; if necessary, use scissors to cut leaf litter, algae, small twigs)
- Step 5. Pick all specimens from the four grids' worth of sample material (use forceps and/or eye dropper and keep track of rough count); place specimens in one or more vials (Figure 3, C1) with approximately 70–80 percent ethanol.
- Step 6. After all specimens are removed from picking tray, pour remaining material from picking tray into a separate container (Figure 3, C2), clearly labelled as “Sort residue.”

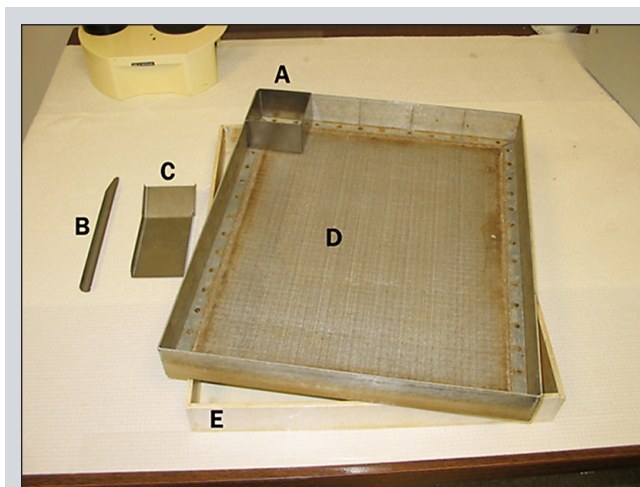


Figure 4. Caton Gridded Screen Photo with All Accessories, including “Cookie Cutter” Frame (A), Spatula (B), Scoop (C), 480-Micron Screen (D), and Outer Tray (E).

- Step 7. If rough count is less than target count, randomly select another grid (1), repeat from step 4; repeat with single grids until target count is exceeded; make sure to pick final grid to completion.
- Step 8. Transfer sample material remaining on the gridded screen back into original or other sample container (Figure 3, C3), labelling it as “Unsorted sample remains.”

Various target counts are used for monitoring, consistent within programs, with the majority being 200–300 (Carter and Resh 2013). This range must be consistent within programs, but can range from 100–600 among programs. The subsampling process results in at least three sample containers for each sample (Figure 3)¹. The first container (C1) has the clean subsample of specimens that will be given to the taxonomist for identification and counting; this is generally one or more glass vials, typically 7–9 drams (25–33 ml). Note that some labs prefer primary sorters to segregate certain organism groups into isolated vials—for example, midges (Chironomidae), snails, clams and mussels (Mollusca), or scud, sowbugs, and crayfish (Crustacea). The second container (C2) holds the sort residue—that is, the material remaining in the separate picking tray after all organisms have been removed. This container will most likely be a 1L Nalgene jar with a plastic screw cap or similar vessel. The sort residue is retained for use in a QC check for missed specimens. The third container (C3) holds the unsorted sample remains. If there are problems with the specimens removed by the sorter or some other QC issue, additional sorting from C3 might be necessary. Stribling (2011) provides more detail on the sorting and subsampling procedure using the Caton screen. QC is performed by having an independent sort-checker go through the sort residue (C2) to check for missed specimens. This should be done for a randomly selected set of 10 percent or a minimum of three of the samples. *Percent sorting efficiency* (PSE) is the performance characteristic used to quantify sorting bias (Stribling 2011).

Taxonomic identification. Identification of BMI is the process of associating a single name with each specimen in the subsample using a technical taxonomic key or comparing it to a reference collection of preserved organisms. Personnel extensively experienced in taxonomic identification often use *sight identification*, or *sight ID*, to recognize the taxon or morphological and anatomical features of a specimen without consulting the literature. Whether using morphology-based dichotomous keys, technical diagnostic literature, or sight ID, the most important factor is that the taxonomist record the result using the nomenclatural standard specified for the project. The end result of the overall identification process for a sample is a list of taxa and the number of individuals of each (Table 2). Quality control for the taxonomic identification step of the assessment process is achieved through confirmation by the QC taxonomist of a randomly selected subset of the samples that were already identified by the primary taxonomist (Stribling 2011). Performance characteristics for quantifying taxonomic precision and consistency are percent taxonomic disagreement (PTD), percent difference in enumeration (PDE), and percent taxonomic completeness (PTC) (Stribling 2011) (Table 1).

¹ Each of C1-3 can actually be multiple containers. Following identifications by the primary taxonomist, C1 will potentially include slide-mounted specimens (midges and worms). In any case, the different sets of sample containers should be considered sample subunits.

Table 2. Raw Data from One Benthic Macroinvertebrate Sample (Example), including Autecological Designations

Family	Taxon	Count ^a	STV	FFG	Habit
Enchytraeidae	Enchytraeidae	3	4.9	CG	BU
Lumbricidae	Lumbricidae	2	8.3	CG	BU
Cambaridae	Cambaridae	2	6.3	SV	SP
Asellidae	<i>Lirceus</i>	2	7.3	SV	SP
Elmidae	<i>Ancyronyx</i>	3	2.0	OM	CN
Elmidae	<i>Macronychus</i>	2	2.4	OM	CN
Elmidae	<i>Microcylloepus</i>	1	1.9	CG	BU
Elmidae	<i>Stenelmis</i>	7	4.8	SC	CN
Ptilodactylidae	<i>Anchytarsus</i>	1	4.0	SH	CN
Athericidae	<i>Atherix</i>	1	2.0	PR	SP
Ceratopogonidae	Ceratopogonidae	8	4.7	PR	SP
Chironomidae	<i>Kloosia</i>	1			
Chironomidae	<i>Pagastiella</i>	1	0.0	CG	SP
Chironomidae	<i>Polypedilum</i>	13	4.1	SH	Cb
Chironomidae	<i>Tribelos</i>	1	2.9	CG	BU
Chironomidae	<i>Brillia</i>	1	2.9	SH	BU
Chironomidae	<i>Corynoneura</i>	3	3.2	CG	SP
Chironomidae	<i>Cricotopus/Orthocladius</i>	1	5.8	SH	SP
Chironomidae	<i>Parametriocnemus</i>	11	3.1	CG	SP
Chironomidae	<i>Pseudorthocladius</i>	3	1.1	CG	SP
Chironomidae	<i>Rheosmittia</i>	20	7.0	CG	BU
Chironomidae	<i>Ablabesmyia</i>	1	5.0	PR	SP
Chironomidae	<i>Nilotanytus</i>	1	3.0	PR	SP
Chironomidae	<i>Thienemannimyia</i> genus gr.	10	6.0	PR	SP
Chironomidae	<i>Micropsectra</i>	1	1.5	CG	CN
Chironomidae	<i>Rheotanytarsus</i>	1	3.3	CF	CN
Chironomidae	<i>Stempellinella</i>	1	1.6	CG	CN
Chironomidae	<i>Tanytarsus</i>	19	3.5	CF	CN
Empididae	<i>Hemerodromia</i>	3	4.2	PR	SP
Tabanidae	<i>Tabanus</i>	1	7.4	PR	SP
Tipulidae	<i>Tipula</i>	2	5.0	SH	BU
Tipulidae	<i>Hexatoma</i>	1	0.0	PR	BU
Heptageniidae	<i>Maccaffertium</i>	4		SC	CN
Neophemeridae	<i>Neophemera</i>	1	2.1	CG	SP
Coenagrionidae	<i>Argia</i>	1	6.5	PR	CN

Table 2. Raw Data from One Benthic Macroinvertebrate Sample (Example), including Autecological Designations

Family	Taxon	Count ^a	STV	FFG	Habit
Corduliidae	<i>Neurocordulia</i>	1	5.0	PR	CB
Corduliidae/Libellulidae	Corduliidae/Libellulidae	1	5.0	PR	CB
Chloroperlidae	<i>Perlinella</i>	1	2.0	PR	CN
Leuctridae	<i>Leuctra</i>	7	0.0	SH	SP
Perlidae	<i>Perlesta</i>	25	1.8	PR	CN
Calamoceratidae	<i>Anisocentropus</i>	3	2.0	SH	SP
Hydropsychidae	<i>Cheumatopsyche</i>	6	5.8	CF	CN
Hydropsychidae	<i>Diplectrona</i>	1	2.0	CF	CN
Hydroptilidae	<i>Hydroptila</i>	1	3.8	PI	CN
Leptoceridae	<i>Oecetis</i>	11	2.4	PR	CN
Leptoceridae	<i>Triaenodes</i>	9	0.7	SH	SW
Limnephilidae	<i>Pycnopsyche</i>	1	1.4	SH	SP
Philopotamidae	<i>Chimarra</i>	3	1.2	CF	CN
Polycentropodidae	<i>Cernotina</i>	1	1.2	PR	CN
Polycentropodidae	<i>Neureclipsis</i>	2	2.7	CF	CN
Pisidiidae	Pisidiidae	2	5.4	CF	

Notes:

a The number of individuals in the sample.

BU = burrower; CB = climber; CF = collector-filterer; CG = collector-gatherer; CN = clinger; FFG = functional feeding group; OM = omnivore; PI = piercer; PR = predator; SC = scraper; SH = shredder; SP = sprawler; STV = stressor tolerance value; SV = scavenger; SW = swimmer.

Autecological attributes. “Autecology” is the study of the relationship between an individual taxon (usually a species) and its immediate environmental factors. Attributes related to autecology are used for calculating some individual metrics and also can contribute to additional interpretation of disaggregated indexes and metrics. Attributes used for BMI are functional feeding groups (FFG), habit, and stressor tolerance values (TV). Attributes should be assigned to the taxa list after taxonomic identification and any necessary QC corrective actions. Merritt et al. (2008) is a good source for FFG and habit attributes, as is Barbour et al. (1999) for TV. Many states, however, have done additional TV analyses to calibrate more specifically to their region, variable water body types and conditions, and dominant stressor characteristics. Most current monitoring programs use relational databases (e.g., Microsoft Office Access® customized to the program structure) that include data tables with attributes already assigned to individual taxa.

At this point in the process, the data are ready for metric and index calculations and site assessments.

Metric Calculation and IBI Aggregation

Two of the main guidelines used for metric selection during index calibration are to: (1) have at least one metric in each of the metric categories (Barbour et al. 1995) and (2) test the direction of response in the presence of stressors. Evaluation and selection of metrics typically involves testing of many more metrics than end up in the final index. They also are scaled to 100 points, which enables them to be averaged together as the full, multimetric biological index on each sample. As an example, calibration performed on data from Mississippi (Bioregion East) resulted in a benthic index made up of six metrics (Table 3). Only the Hilsenhoff Biotic Index (HBI) requires a formula that is not evident from the description (Hilsenhoff 1982). It is calculated using the following formula:

$$HBI = \sum \frac{n_i a_i}{N}$$

where:

n_i and a_i = the number of individuals and the stressor tolerance value of taxon i , respectively.

N = the total number of individuals in the sample.

Table 3. Benthic Macroinvertebrate Metrics, Text Description, Category, and Trend with Increasing Stressor Load

Metric Name (abbreviation)	Description	Category	Direction of Change with Increasing Stressors
Total number of taxa (TotalTax)	Number of distinct taxa identified in the subsample	Richness	Decrease
Total number of EPT taxa (EPTTax)	Number of distinct taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)	Richness	Decrease
Percent individuals as <i>Cricotopus</i> / <i>Orthocladius</i> / <i>Chironomus</i> of total Chironomidae (COC2ChiPct)	Count of individuals in these relatively tolerant genera as percent total Chironomidae individuals in the sample	Composition	Increase
Percent individuals as sensitive EPT (PSensEPT)	Of all individuals in the sample, the percentage of individuals in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies), except individuals in the families Caenidae, Baetidae, Hydropsychidae, and Hydroptilidae	Composition	Decrease
Number of taxa, as shredders (ShredTax)	Number of distinct taxa in the sample that are considered shredders (i.e., they use coarse organic material—primarily leaf litter—for food)	Functional feeding group	Decrease
Hilsenhoff Biotic Index (HBI)	Composite of total relative sensitivity of all organisms in the sample, calculated as the average tolerance value of all individuals in the sample	Tolerance	Increase

Following is an example metric calculation for Bioregion East using raw sample data that consists of a list of taxa, number of individuals of each, and each taxon’s associated autecological attributes (i.e., FFG, habit, and TV) (Table 2). This sample result has 51 taxa represented by 209 specimens. The sample is dominated by midges (Chironomidae; 89 specimens, ~43 percent of the whole sample), but also has a substantial number of caddisflies (Trichoptera, 10 genera in seven families; 38 specimens, ~19 percent) and stoneflies (Plecoptera; 33 specimens in three genera, ~16 percent).

Individual metric values are calculated using the sample data, but because the metrics are mostly on different scales, they cannot be directly aggregated. Formulas developed during the index calibration process (not shown in this technical memorandum) allow the individual metric values to be converted to a 100-point scale (Table 4). The overall benthic IBI score for a single sample at this site is the mean value of the six metrics, which in this example is 96.3.

Table 4. Metric Calculations from Example Sample

	Metric ^a	Value ^b	Formula	Score ^c
1	TotalTax	51	$100*(metric\ value)/51.5$	99
2	EPTTax	15	$100*(metric\ value)/14$	100
3	COC2ChiPct	1.1	$100*(45-(metric\ value))/45$	97.6
4	PSensEPT	32.4	$100*(metric\ value)/39$	83.1
5	ShredTax	9	$100*(metric\ value)/7$	100
6	HBI	3.6	$100*(8.5-(metric\ value))/5$	98
IBI			<i>Sum of metric scores/6</i>	96.3
Narrative Rating			Non-degraded	

Notes:

a Metric abbreviations defined in Table 3.

b Numbers calculated or compiled directly from sample data, list of taxa, and counts of individuals of each.

c Normalized to a 100-point scale using formulas developed during the index calibration process.

Site Assessment

Continuing with our example site for Bioregion East, the IBI impairment threshold is 65.7 on a 100-point scale, meaning that sites with aggregated scores (mean values) falling above it are considered “non-degraded,” or similar to reference conditions. This threshold was defined during calibration based on the 75th percentile of the reference site distribution. It has a discrimination efficiency² of 87.3 percent and a 90-percent confidence interval of ±12.5 points. The biological index score of 96.3 for this site sample falls well above the threshold, providing good confidence in the finding that the site is non-degraded. By disaggregating the index, the investigator can see that individual metrics all have high scores, the lowest being 83 (percent sensitive EPT). In this instance,

² Sometimes abbreviated as DE, this is an estimate of the accuracy of indexes and metrics characterized during calibration. It is a statement of their capacity to correctly identify stressor conditions, and is quantified using the formula: $DE = (a/b) \times 100$, where *a* is the number of *a priori* stressor sites identified as being below the quantified biological impairment threshold of the reference distribution (25th percentile, 10th, or other), and *b* is the total number of stressor sites.

none of the metrics are interpreted as more heavily influencing the narrative of being non-degraded than any other. A sample with 51 (primarily) genus-level taxa is a solid taxonomic diversity for macroinvertebrates. More detailed evaluation of the raw sample data, however, including autecological characteristics of individual taxa and their numerical dominance in the sample, can provide useful information. For example, the vast majority of individual organisms (75 percent) and taxa (67 percent) have $TV < 5$, indicating a dominance of stressor-sensitive biota. Caddisflies (Trichoptera) are generally considered to be stressor-sensitive (i.e., intolerant). The sample has 38 individuals representing 10 genera in seven caddisfly families. The full range of TV is from 0 to 10 with higher numbers indicating greater stressor tolerance. All but two of the 10 genera (*Cheumatopsyche* and *Hydroptila*) have $TV < 3$. This level of stressor tolerance (sensitivity) would also be reflected partly by the metric 'percent individuals as sensitive EPT (PSensEPT)' (Table 2). Although not as diverse in the sample, stoneflies (Plecoptera) are similarly considered to be sensitive to stressors; the three genera have TVs of 0.0, 1.8, and 2.0. Thus, the assessment summary is that the site is not degraded. It has an IBI score of 96.3 and the sample is dominated by stressor-sensitive organisms that would most likely not be present if the location was substantially polluted.

Summary/Recommendations

The information presented here covers the background and procedures for field sampling and laboratory processing of BMI samples from freshwater streams as well as how to calculate and use a multimetric index as an indicator of water resource quality. The example biological assessment presented in this technical memorandum shows the result from a stream site that is non-degraded. The stressors that may be present have not been substantially detrimental to the survival or reproduction of the benthic community. Following these guidelines will provide rigorous and defensible assessments of ecological condition. The IBI provides a technique for summarizing complex biological data into a format that can be scaled and ranked relative to data from other sites and samples and then translated into an assessment narrative. It is important, however, to use metrics, metric aggregation techniques, and scoring criteria—especially the degradation decision threshold—that have been appropriately calibrated for the water body type, site class, and region of the project. Use of an IBI that has not been so calibrated increases the potential for spurious or misleading results. Whereas this document takes the process to the point of demonstrating the presence or absence and severity of a problem, further more detailed analysis of stream biological assessment results may be needed to identify the stressors and sources causing the degradation. Application of stressor identification and causal analysis would inform decisions on management actions to be taken to address identified water quality problems.

References/Additional Resources

- Barbour, M.T., J.B. Stribling, and J.R. Karr. 1995. The Multimetric Approach for Establishing Biocriteria and Measuring Biological Condition. Chapter 6 in *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*, ed. W.S. Davis and T.P. Simon, pp. 63–77. Lewis Publishers, Boca Raton, FL.
- Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*. Second Edition. EPA/841-D-97-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed September 30, 2015. <http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm>.
- Carter, J.L., and V.H. Resh. 2013. *Analytical Approaches Used in Stream Benthic Macroinvertebrate Biomonitoring Programs of State Agencies in the United States*. Open-File Report 2013-1129. U.S. Geological Survey, Reston, VA.
- Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006. *Concepts and Approaches for the Bioassessment of Non-Wadeable Streams and Rivers*. EPA/600/R-06/127. U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Cincinnati, OH. Accessed September 30, 2015. http://www3.epa.gov/region1/npdes/merrimackstation/pdfs/ar/AR-1157.pdf#_ga=1.80358807.1889529594.1444762539.
- Flotemersch, J.E., J.B. Stribling, R.M. Hughes, L. Reynolds, M.J. Paul, and C. Wolter. 2011. Reach length for biological assessment of boatable rivers. *River Research and Applications* 27(4):520–535. doi: 10.1002/rra.1367.
- 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.
- Hill, B.H., A.T. Herlihy, P.R. Kaufmann, S.J. Decelles, and M.A. Vander Borgh. 2003. Assessment of streams of the eastern United States using a periphyton index of biotic integrity. *Ecological Indicators* 2:325–338.
- Hilsenhoff, W.L. 1982. *Using a Biotic Index to Evaluate Water Quality in Streams*. Technical Bulletin 132. Wisconsin Department of Natural Resources, Madison, WI.
- 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.
- Karr, J.R., and D. Dudley. 1981. Ecological perspective on water quality goals. *Environmental Management* 5:55–68.

- Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. *Assessing Biological Integrity in Running Waters: A Method and its Rationale*. Special Publication 5. Illinois Natural History Survey, Champaign, IL.
- Merritt, R.W., K.W. Cummins, and M.B. Berg, ed. 2008. *An Introduction to the Aquatic Insects of North America*. Fourth Edition. Kendall/Hunt Publishing Company, Dubuque, IA.
- Schneider, E.D. 1992. Monitoring for Ecological Integrity. Chapter 79 in *Ecological Indicators*, Vol. 2, ed. D.H. McKenzie, D.E. Hyatt, and V.J. McDonald, pp. 1403-1419. Elsevier Applied Science, New York.
- Stribling, J.B. 2011. Partitioning Error Sources for Quality Control and Comparability Analysis in Biological Monitoring and Assessment. Chapter 4 in *Modern Approaches to Quality Control*, ed. A.B. Eldin, pp. 59–84. INTECH Open Access Publisher. doi: 10.5772/22388.
- USEPA (U.S. Environmental Protection Agency). 2004. *Wadeable Stream Assessment: Field Operations Manual*. EPA 841-B-04-004. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.
- USEPA (U.S. Environmental Protection Agency). 2010. Causal Analysis/Diagnosis Decision Information System (CADDIS). U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC. Last updated September 23, 2010. Accessed September 30, 2015. <http://www.epa.gov/caddis>.