4 Biological Monitoring of Aquatic Communities

By J. B. Stribling, C.J. Millard, J.B. Harcum, and D.W. Meals

4.1 Overview

Biological monitoring uses surveys of resident biota (e.g., fish, benthic macroinvertebrates, periphyton, amphibians) to characterize the structure and function of the assemblage and assess the condition of a waterbody (USEPA 2013). The central purpose of assessing biological condition is to determine the degradation level of an ecosystem (or water body) and the cumulative effects of physical, chemical, hydrologic or biological stressors on aquatic biota. Resident biota reflect the integrated effects of variable magnitudes of these different stressors and stressor types, and thus provide an overall measure of environmental quality. As such, biological assessment is a crucial monitoring tool used by all 50 states and increasingly by tribes (USEPA 2002, 2011).

Biological assessments conducted by state water resource quality programs are often designed to assess regional or state-wide condition as well as conditions in smaller targeted watersheds or projects. During the past 20 years, these water quality agencies have invested resources to develop biological assessment capabilities within their state or tribe. These capabilities include using scientifically defensible and documented field and laboratory methods/protocols, establishing reference sites, and evaluating/developing metrics and indexes most suitable for assessing biological condition, and implementing them in routine monitoring programs or projects. "Index calibration" is a term encompassing data analyses leading to establishing scoring criteria and testing/selection of the suite of metrics making up multimetric indexes (MMI; further discussed below). To the extent practical, watershed projects or NGOs should use MMI that have been regionally calibrated based on broader datasets of known quality. Use of MMI previously established by local, state, or regional agencies requires that the same or similar methods be used for field sampling and laboratory processing for other streams and sites being evaluated relative to BMP or other issues. This approach of using accepted protocols and calibrated metrics and indexes, coupled with sufficient and appropriate quality control (QC) checks, improves defensibility of assessment results and increases confidence in natural resource management decision-making.

We recommend that biological monitoring be coupled with physical/chemical monitoring, which includes the stressor(s) of concern and focus on smaller watersheds (i.e., sub- hydrologic unit code [HUC] 12 watershed level) to document effectiveness of an individual BMP. Biological assessment is a useful tool for evaluating overall ecological condition because it integrates multiple stressors over time; however, it does not directly measure changes in a specific stressor (e.g., decreased sediment loading resulting from riparian buffers or point source discharge). The BMP could still be evaluated as effective in reducing stressor load, even though a positive biological response might not be detected. And, linking detectable changes in broad scale biological condition to a particular small scale BMP is more difficult as the potential increases for unknown and multiple stressors. Case Study 1 from Wisconsin illustrates the use of several types of indicators for monitoring BMP effectiveness (*Evaluating Effectiveness of Best Management Practices for Dairy Operations in the Otter Creek Watershed, Wisconsin*).

Case Study 1: Evaluating Effectiveness of Best Management Practices for Dairy Operations in the Otter Creek Watershed, Wisconsin

- ✓ Dairy farms, barnyard runoff
- ✓ Manure storage, barnyard runoff control, stream bank protection, stream fencing and crossings, stabilization, buffer strips
- ✓ Fish assemblage monitoring
- ✓ Effectiveness evaluation

Located in eastern Wisconsin, the Otter Creek watershed is a 24.6 km^2 sub-basin tributary to the Sheboygan Creek watershed (Figure CS1-1), the latter ultimately feeding into Lake Michigan. Otter Creek has a total stream length of approximately 21 km and is a third-order watershed with a low to medium slope (2.5 - 5.4 m/km) throughout the area with a median wetted width of 4.2 m (Wang et al. 2006). Stream bottom is composed mostly of sand, silt, and clay, with riffle areas of medium gravel. Land use and land cover during the study period was dominated by row crop agriculture (62 percent), forests (14 percent), and to lesser degrees by grasslands and wetlands (10 percent and 6 percent, respectively). Corsi et al. (2005) noted that the basin was home to 64 farms averaging approximately 0.5 km^2 in production. In the basin, there were eight barnyards associated with dairy operations, with an average herd size of 45 animals. The Wisconsin Department of Natural Resources reported in 1993 that the primary problems in the basin were direct livestock access to streams, resulting in elimination of bank vegetation, fish habitat degradation, accelerated and extensive bank erosion, and water temperature modification. Other acute problems associated with the livestock were barnyard manure runoff, upland delivery of sediment, and runoff from areas of winter manure-spreading. In addition to degradation of physical habitat, these

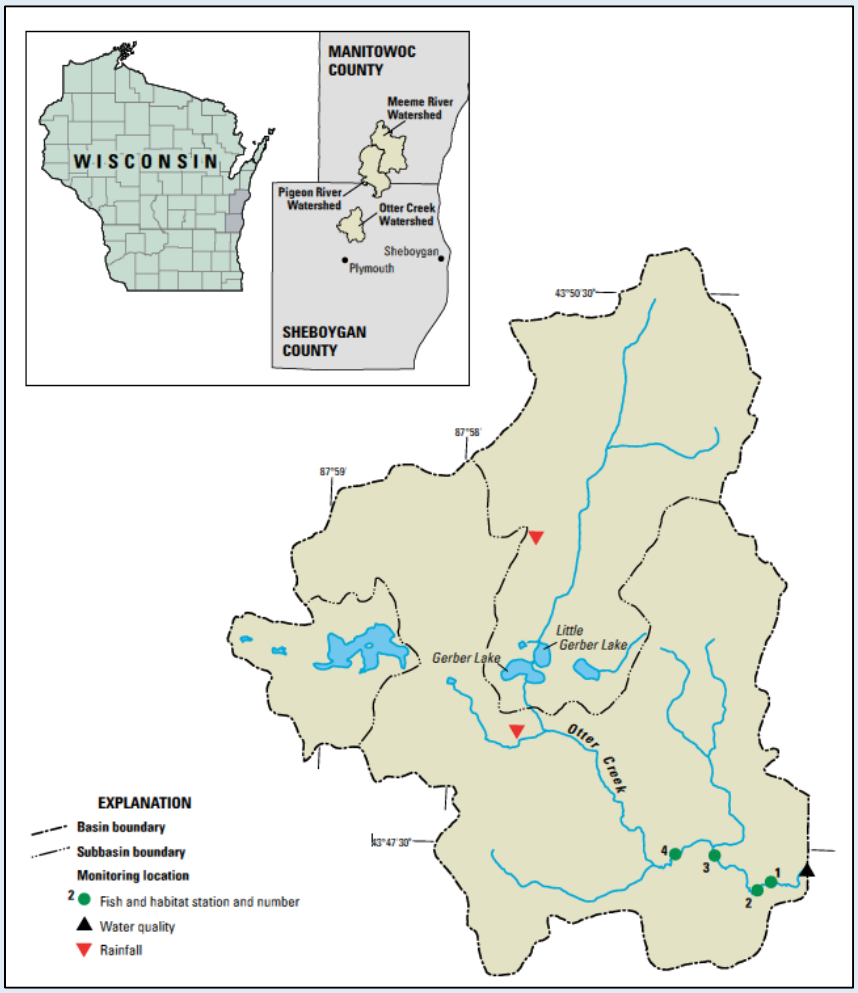

Figure CS1-1. Otter Creek watershed (Corsi et al. 2005) sources and stressors led to organic and inorganic nutrient over-enrichment, and occasional, if not often, severe depletion of dissolved oxygen (Wang et al. 2006).

The BMPs designed and installed were focused on buffering, reducing, or otherwise eliminating the stressors and included a combination of animal waste management, stream bank protection, and upland management. Waste management practices were developed as facilities that provided improved manure storage, better control of barnyard runoff, and treatment of milkhouse wastewater. Four different types of BMPs were implemented for protection of >1,900 m of stream bank: fencing (2,800 m), stream crossings, grade stabilization, and buffer strips. Upland management included 635 ha (6.4 km^2) of nutrient management and sediment reduction of 250 metric tons/year achieved with changes in crop rotation, reduced tillage, critical area stabilization, grass waterways, and pasture management. Riparian and upland BMPs were implemented during 1993 and 1999.

Monitoring and Sampling Design

The goal of this monitoring project was to evaluate the effectiveness of multiple BMPs on the biological, habitat, and water chemistry characteristics in Otter Creek. Changes, if any, in instream habitat and biological characteristics relative to timing of BMP installation would be interpreted as BMP effectiveness at the watershed scale. Water chemistry evaluated using data from both base- and stormflow sampling would be used to establish changes in stressors impacting the fish assemblage.

Annual fish and habitat evaluations occurred relatively continuously from 1990 to 2002, providing for pre- (1990 to 1992), during (1993 to 1999), and post-installation (2000 to 2002) monitoring at four stations. Stations 1, 3, and 4 were located in areas where streambanks were trampled from cattle with free access while station 2 was located in a wooded riparian area (Figure CS1-1). Streambanks at station 1 and 3 were fenced in 1993 and 1996, respectively, while station 4 was not fenced (Corsi et al. 2005). There were four streams in similar watersheds that also flow to Lake Michigan that were also monitored. Two watersheds, the Meeme and Pigeon Rivers, were monitored as two control watersheds in a paired-watershed monitoring design with a single sampling station each (see section 2.4.2.3). Neshota and Trout Creeks (tributaries to the West River and Duck Creek, respectively) were also monitored with single stations.

Fish sampling and physical habitat evaluations were generally conducted each year at four locations in the lower part of the watershed, during roughly a six-week period spanning August and September. Gradually increasing funding resulted in more complete sampling over time, and thus, the record for some of the sites is better than others. In Otter Creek, 1 station was sampled in 1990, 2 in 1991, and all 4 from 1992-2002. Fish were sampled from a reach length of 35 times the median wetted channel width, which was an actual range from 105 to 234 m. Fish sampling used a single-tow barge electrofisher to cover the sampling reach in a single-pass/no block net approach. All fish captured were identified to species, counted, weighed (total, by species), and released unharmed back to the stream. Data were summarized by reach and sampling event as species lists, as proportions of individuals in the samples omnivores, insectivores, carnivores, simple lithophils, relative stressor tolerance, and as an IBI. The IBI had been previously calibrated for Wisconsin warm-water streams. Habitat features recorded for each sampling event at each location included turbidity, dissolved oxygen, specific conductance, and flow, along with 30 habitat variables, encompassing channel morphology, bottom substrates, cover for fish, bank conditions, riparian vegetation, and land use. These data were summarize by calculating mean and variance for each reach and sampling event, and used to calculate width/depth ratio and total habitat quality index.

Monitoring also included a single-watershed, before/after study of water chemistry. A stream gage installed in 1990 at the base of Otter Creek just upstream of its confluence with the Sheboygan River was equipped to continuously record data on stage and for automatic activation to collect and refrigerate samples with every 0.2-ft. increase and 0.3-ft decrease in stage. Water chemistry constituents measured from these grab samples included TSS, TP, dissolved ammonia nitrogen (NH_3-N), 5-day biochemical oxygen demand (BOD_5), and fecal coliforms. Fixed interval grab samples, analyzed for the same constituents, were also taken throughout the pre- and post-BMP study period. Precipitation was measured at three locations in the watershed, one near the gaging station, and two others further up in the headwaters.

Results

For both fish and habitat data, analysis of covariance (ANCOVA) (see section 7.8) was used to relate the fish assessments and habitat variables and conditions of the Otter Creek sites to the other watersheds over time. The statistical significance of changes was evaluated using the Wilcoxon rank-sum nonparametric test at 95 percent confidence. Changes were measured from different subsets of base- and stormflows, including some investigation of seasonal effects (vegetative vs. nonvegetative).

Post-BMP implementation baseflow samples showed statistically significant lower concentrations of TSS and BOD_5, higher fecal coliform, and no differences in dissolved NH_3-N and TP for the combined seasons (vegetative and nonvegetative), whereas samples from nonvegetative periods exhibited lower BOD_5 and no changes in the other four analytes. TSS concentration was lower during the vegetative season. From several different analyses, BOD_5 was demonstrated to have decreased substantially (by 45 percent median concentrations) in baseflow from the pre- to post-BMP periods. TSS was also found to be lower in post-BMP samples relative to those of pre-BMP, however, only actually evident in the full dataset, and not in the smaller data subsets (i.e., vegetative vs. nonvegetative). Dissolved NH_3-N concentrations decreased from pre- to post-BMP baseflow for the full dataset and nonvegetative season, but there were no differences for the vegetative season. Analyses showed none of these difference to be statistically significant ($p<0.05$), even with a measured 32 percent decrease in dissolved NH_3-N. There was a significant ($p<0.05$) increase in median fecal coliform concentrations of 260 percent for pre- to post-samples, demonstrated in both the full and stratified (vegetative and nonvegetative) datasets.

Analyses using pre- and post-BMP regressions across all monitored storms showed reductions in stormflow constituents, whether combined or stratified by vegetative vs. nonvegetative. For TSS, the pre/post reduction was 58 percent for the combined nonvegetative and vegetative dataset, 41 percent for the vegetative season, and 73% for the nonvegetative season. TP predictions for reductions were 48 percent, 34 percent , and 61 percent , while for dissolved NH_3-N they were 41 percent , 40 percent , and 42 percent , respectively.

Although considerable variation in stream physical habitat was observed in the reference streams, there was clear post-BMP improvement in Otter Creek (Corsi et al. 2005). In particular, there were significant increases ($p<0.05$) in percent cobble and percent gravel, and significant decreases in percent embeddedness, sand, and silt for stations 1, 2, and 3 (Wang et al. 2006). These changes reflected the natural woody buffer (station 2) or exclusion fences (stations 1 and 3). Stream width-to-depth (W/D) ratio and percent bank erosion decreased significantly ($p<0.05$) for stations

1 and 3 where exclusion fences were installed. Other habitat variables that showed significant improvement for one or more of the sample reaches were sediment depth and percent riffles. ANCOVA showed that overall habitat quality improved in Otter Creek only for those reaches that had a natural riparian buffer or had exclusion fences installed. For stations 1, 2, and 3, most of the habitat variables (8 out of 10) improved significantly ($p<0.05$) with BMP implementation, with substrate embeddedness, bank erosion, sediment depth, silt and sand substrates, and W/D ratio decreasing and conductivity, gravel and rubble substrates, and overall habitat scores increasing.

Cumulative fish species were similar for each of the Otter Creek locations, the two control watersheds (Meeme and Pigeon Rivers), and the two additional watersheds (Neshota and Trout Creeks) (33, 29, and 31 species, respectively), and six species dominated in all streams. From pre- to post-BMP installation, fish abundance decreased by 79 percent in Otter Creek and by 65 percent in the control watersheds. When sampling years were considered in combination, the percentages of stressor-tolerant fishes and omnivores *increased* in Otter Creek, while stressor tolerant fishes *decreased* and insectivores increased in control watersheds. In neither Otter Creek nor the control watersheds was there an obvious directional change in percentage of stressor intolerant fishes or IBI scores. ANCOVA showed significance ($p<0.05$) in a decrease of abundance and an increase in percent omnivores; and, for one of the Otter Creek sample locations where riparian pasture was dominant, there were significant decreases in number of fish species and percentage of darter individuals.

Characteristics of physical habitat and water chemistry improved, apparently substantially so, as a result of the several BMPs that were implemented in Otter Creek. However, there was no significant improvement in biological condition as measured by the fish community. This status of the fish assemblage even with habitat and water chemistry suitable for supporting much higher quality led the authors to conclude that there are some broader scale watershed or regional factors preventing more evident, positive changes in the fish community. Wang et al. (2006) speculate that pollution intolerant species might have been largely eliminated in the larger watershed and thus not able to colonize Otter Creek.

Literature

Corsi, S.R., J.F. Walker, L. Wang, J.A. Horwatich, and R.T. Bannerman. 2005. *Effects of Best-Management Practices in Otter Creek in the Sheboygan River Priority Watershed, Wisconsin, 1990-2002*. Scientific Investigations Report 2005–5009. U.S. Geological Survey (in cooperation with the Wisconsin Department of Natural Resources). Accessed February 9, 2016. http://pubs.usgs.gov/sir/2005/5009/.

Wang, L., J. Lyons, and P. Kanehl. 2006. Habitat and fish responses to multiple agricultural best management practices in a warm water stream. *Journal of the American Water Resources Association* 42(4):1047-1062.

Whether indicators being monitored are based on biological, physical, chemical, or hydrologic data, or targeted vs. probability-based site selection, questions required for monitoring designs are similar. Targeted sampling designs are needed to answer questions on BMP effectiveness at a particular location typically using a paired (treatment/control) watershed or upstream/downstream approach. Alternatively, probability-based designs could also be used to evaluate the difference in biological condition from randomly selected sample locations where a particular BMP (treatment) was implemented versus a sample of non-treated (i.e., control) areas. Unlike traditional physical/chemical monitoring programs that take measurements and collect samples frequently throughout a year, most biological monitoring programs take samples once (or, rarely, twice) during specific periods annually because biota do not usually vary dramatically in response to individual transient events. Biological data are converted into indicator values (e.g., such as an MMI, also known as an Index of Biological Integrity [IBI]). As a result of many of these protocol characteristics, variability that could otherwise cause data analysis issues related to seasonality, autocorrelation, and non-normality is controlled (Fore and Yoder 2003). Disaggregation of indexes to individual metrics or even taxa can allow for more detailed interpretation of subtle biological changes relative to BMP effectiveness. However, that level of interpretation requires access to a biologist with appropriate training and experience.

The concept of reference conditions has begun to emerge in the analysis of traditional physical/chemical monitoring data. Reference conditions are those observed in unimpaired or minimally impaired waterbodies in the region of interest and are used as a benchmark against which to measure changes. As previously stated, many states and tribes have established regional reference sites in support of their ongoing assessment programs, and it may be useful and efficient to use them in a watershed monitoring program.

This chapter presents basic information about biological monitoring and its applicability to NPS and watershed projects. Section 4.2 introduces the different types of biological monitoring and common terms used in this chapter with an emphasis on benthic macroinvertebrates and periphyton. An overview of monitoring design and assessment protocols is provided in sections 4.3 and 4.4, respectively.

4.2 Background

Natural biological communities are often diverse, comprising multiple species at various trophic levels (e.g., primary producers, secondary producers, carnivores) and varying degrees of sensitivity to environmental changes. Adverse impacts from NPS pollution or other stressors, such as habitat alteration, can reduce the diversity of the biota, change the relative abundances of different taxa, or alter the trophic structure. Biological surveys of resident biota particularly sensitive to stressors, such as fish, benthic macroinvertebrates, or periphyton, take advantage of this sensitivity as a means to evaluate the collective influence of the stressors on the biota (Cummins 1994).

The central purpose of biological assessment is to characterize the condition of resident biota relative to cumulative effects of stressors as the principal indicator of stream (or water body) condition. Monitoring changes in biological condition can be particularly useful for determining the impacts, depending on the frequency and duration of exposure, of episodic stresses (e.g., spills, dumping, treatment plant malfunctions), toxic nonpoint source pollution (e.g., agricultural pesticides), cumulative pollution (i.e., multiple impacts over time or continuous low level stress), non-toxic mechanisms of impact (e.g., trophic structure changes due to nutrient enrichment), or other impacts that periodic chemical sampling might not detect (USEPA 2011).

4.2.1 Types of Biological Monitoring

Different kinds of biological monitoring are defined by the particular indicators being used and the spatial and temporal scales of questions being addressed. The most common biological indicator groups (or assemblages) used for routine biological monitoring and assessment of freshwater ecosystems in North America are benthic macroinvertebrates, fish, and periphyton (algae). Additionally, programs will often collect data on physical, chemical, and hydrologic features of the systems being evaluated to provide information on environmental factors potentially affecting the biota. The scales of the questions being addressed drive (or should drive) the number and distribution of locations, and the frequency and duration of sampling, that is, the monitoring and sampling design. Further, efforts to control or better understand the natural variability of the biota have led to different kinds of specific field and laboratory methods for taking the samples or performing measurements. Controlling variability of indicator values means that assessment data are of known quality, and leads to improved confidence in and defensibility of management decisions.

The purpose of this document is to provide users the information necessary for applying existing biological indicators (metrics and indexes) to their ecosystems or water bodies of concern. It is not intended to give a comprehensive review of all methods and procedures used for biological assessments. By existing indicators, we mean those MMI or observed/expected (O/E) ratios that have been appropriately calibrated for the region and water body type. The remainder of this section, and those that follow, provide an overview of selected methods.

4.2.1.1 Benthic Macroinvertebrates

Stream environments contain a variety of macro- and microhabitat types including pools, riffles, and runs of various substrate types; snags; and macrophyte beds (Hawkins et al. 1993). Relatively distinct assemblages of benthic macroinvertebrates inhabit various habitats, and it is unlikely that most sampling programs would have the time and resources to sample all habitat types. Decisions on the habitats selected for sampling should be made with consideration of the regional characteristics of the streams. For instance, higher-gradient streams (slope roughly >1:1) often have riffle habitat with hard bottom substrate (cobble or gravel riffles) that serve as excellent habitat for a diversity macroinvertebrates, whereas low-gradient coastal streams lack riffles but have a very productive habitat including woody debris snags (Figure 4-1), leaf packs, undercut banks, and shorezone vegetation. These two different stream types could be sampled with different methods, during different times of the year, or with different biological index periods, but consideration should be given to protocols already being applied in the water body or region. Many routine biological monitoring programs use multi-habitat composite sampling techniques, e. g., the 20-jab method, because the intent is to characterize the biota of the stream reach.

4.2.1.2 Fish

Fish surveys yield a representative sample of the species present at all habitats within a sampling reach that is representative of the stream. If comparable physical habitat is not sampled at all stations, it will be difficult to separate degraded habitat from degraded water quality as the factor limiting the fish community (Klemm et al. 1992).

At least two of each of the major habitat types (i.e., riffles, runs, and pools) should be incorporated into the sampling as long as they are typical of the stream being sampled. Most species will be successfully sampled in areas where there is adequate cover, such as macrophytes, boulders, snags, or brush.

Figure 4-1. Using a D-frame net to sample woody snag habitat for stream benthic macroinvertebrates

Sampling near modified sites, such as channelized stretches or impoundments, should be avoided unless it is conducted to assess the impact of those habitat alterations on the fish community. Sampling at mouths of tributaries entering larger waterbodies should be avoided because these areas have habitat characteristics more typical of the larger waterbody (Karr et al. 1986) and non-resident visitors from the larger waterbody may be captured. Sampling reach lengths range from 100 to 200 m for small streams and 500 to 1000 m for rivers. Some agencies identify their sampling reach by measuring a length of stream that is 20 to 40 times the stream width.

For biological assessments of the entire assemblage, the gear and methods used should ensure that a representative sample is collected.

Fish can be collected actively or passively. Active collection methods involve the use of seines, trawls, electrofishing equipment, or hook and line. Passive collection can be conducted either by entanglement using gill nets, trammel nets, or tow nets, or by entrapment with hoop nets or traps. For a discussion on the advantages and limitations of the different gear types, see Klemm et al. (1992). IBI emphasizes active gear, and electrofishing is the most widely used active collection method. Ohio EPA (1987) discusses appropriate electrofishing techniques for bioassessment. Other sources for sampling method discussions are Allen et al. (1992), Dauble and Gray (1980), Dewey et al. (1989), Hayes (1983), Hubert (1983), Meador et al. (1993), and USFWS (1991). Fish are generally identified to the species or subspecies level.

4.2.1.2.1 *Length, Weight, and Age Measurements*

Length and weight measurements can provide estimations of growth, standing crop, and production of fish. The three most commonly used length measurements are standard length, fork length, and total length. Total length is the measurement most often used.

Age may be determined using the length-frequency method, which assumes that fish increase in size with age. However, this method is not considered reliable for aging fish beyond their second or third growing season. Length can also be converted to age by using a growth equation (Gulland 1983).

Annulus formation is a commonly used method for aging fish. Annuli (bands formed on hard bony structures) form when fish go through differential growth patterns due to the seasonal temperature changes of the water. Scales are generally used for age determination, and each species of fish has a specific location on the body for scale removal that yields the clearest view for identifying the annuli. More information on the annulus formation method and most appropriate scale locations by species can be found in Jerald (1983) and Weatherley (1972).

4.2.1.2.2 *Fish External Anomalies*

The physical appearance of fish usually indicates their general state of well-being and therefore gives a broad indication of the quality of their environment. Fish captured in a biological assessment should be examined to determine overall condition such as health (whether they appear emaciated or plump), occurrence of external anomalies, disease, parasites, fungus, reddening, lesions, eroded fins, tumors, and gill condition. Specimens may be retained for further laboratory analysis of internal organs and stomach contents, if desired.

4.2.1.3 Periphyton

Periphyton is an assemblage of organisms that adhere to and form a surface coating on stones, plants, and other submerged objects in aquatic habitats. These can take the form of soft algae, algal or filamentous mats, or diatoms. The advantages of using the periphyton assemblage as an indicator include:

- Rapid reproduction rates and short life cycles and thus quick response to perturbation, which makes them valuable indicators of short-term impacts.
- Primary producers are ubiquitous in all waters, and they are directly affected by water quality.
- Rapid periphyton sampling requiring few personnel, with easily quantifiable results.
- A list of the taxa present and their proportionate abundance can be analyzed using several metrics or indices to determine biotic condition and diagnose specific stressors.
- The periphyton community contains a naturally high number of taxa that can usually be identified to species.
- Tolerance of or sensitivity to changes in environmental conditions that are known for many species or assemblages of diatoms.
- Periphyton is sensitive to many abiotic factors that might not be detectable in the insect and fish assemblages.

The state of Kentucky, for example, has developed a Diatom Bioassessment Index (DBI), currently used in water quality assessments. Metrics used to construct the DBI include diatom species richness, species diversity, percent community similarity to reference sites, a pollution tolerance index, and percent sensitive species. Scores for each metric range from 1 to 5. The scores are then translated into descriptive site bioassessments that are used to determine aquatic life use.

In Montana, Bahls (1993) developed metrics for diatoms that included a diversity index, a pollution index, a similarity index, and a siltation index. Three other metrics—dominant phylum, indicator taxa, and number of genera—were used for soft-bodied algae to support the diatom assessment. Further study and refinement of these metrics has led to the development of diatom biocriteria (Teply and Bahls 2005) and an evaluation of their discrimination efficiency (Teply and Bahls 2007). Periphyton data collections currently comprise routine monitoring of the Statewide Monitoring Network (SWM) and support the determination of designated aquatic life uses.

4.2.2 Linkages to Habitat

The quality of the physical habitat is an important factor in determining the structure of benthic macroinvertebrate, fish, and periphyton assemblages (Southwood 1977). The physical features of a habitat include substrate type, quantity and quality of organic debris (leaf litter, woody materials) in the waterbody, exposure to sunlight, flow regime, and type and extent of aquatic and riparian vegetation. Even though there might not be sharp boundaries between habitat features in a stream, such as riffles and pools, the biota inhabiting each feature are often taxonomically and biologically distinct (Hawkins et al. 1993). Habitat quality is assessed during biological assessment.

Habitat features are generally associated with biological diversity, (Southwood 1977, Raven et al. 1998) and their quality largely determines both the structure and function of benthic macroinvertebrate and fish assemblages. Habitat quality refers to the extent to which a suitable environment for a healthy biota exists. It encompasses five factors: habitat structure, flow regime, energy source, biotic interactions (such as invasive species and disease), and chemical water quality. *Habitat structure* refers to the physical characteristics of stream environments. It comprises channel morphology (width, depth, sinuosity), floodplain shape and size, channel gradient, instream cover (boulders, woody debris), substrate types and diversity, riparian vegetation and canopy cover, and bank stability. *Flow regime* is defined by the patterns of velocity and volume of water moving through a stream over time. *Energy* enters streams as the input of nutrients in runoff or ground water, as organic debris (e.g., leaves) falling into streams, or from photosynthesis by aquatic plants and algae. *Biotic interactions* include issues such as invasive species and disease while *water quality* includes a range of issues from nutrient sources to toxicants.

These factors are interrelated and make stream environments naturally heterogeneous. Habitat structural features that determine the assemblages of macroinvertebrates can differ greatly within small areas—or microhabitats—or in short stretches of a stream. For instance, woody debris in a stream affects the flow in the immediate area, provides a source of energy, and offers cover to aquatic organisms. Curvature (sinuosity) in a stream affects currents and thereby deposition of sediment on the inner and outer banks, in turn influencing the character of the streambed. Rocks and boulders create turbulence, which affects dissolved oxygen levels. Shallow stream reaches where water velocity is relatively high (riffles) provide areas of high dissolved oxygen and a gravel or cobble bottom ; deep, wide portions (pools) are areas of lowered velocity where material can settle out of the water, streambeds are composed of soft sediments, and increased decomposition occurs.

Aspects of habitat structure are separated into primary, secondary, and tertiary groupings corresponding to their influence on small-, medium-, and large-scale aquatic habitat features (Barbour et al. 1999). The status or condition of each aspect of habitat is characterized along a continuum from optimal to poor. An optimal condition would be one that is in a natural state. A less than optimal condition, but one that satisfies most expectations, is suboptimal. Slightly worse is a marginal condition, where degradation is for the most part moderate, but is severe in some instances. Severe degradation is characterized as a poor condition. Habitat assessment field data sheets (see Plafkin et al. 1989) provide narrative descriptions of the condition categories for each parameter. Habitat can be assessed visually, and a number of biological assessment methods incorporate assessments of the surrounding habitat (Ball 1982, Barbour et al. 1999, OEPA 1987, Platts et al. 1983).

A clear distinction between impacts due to watershed (i.e., large-scale habitat), stream habitat, and water quality degradation is often not possible, so it is difficult to determine with certainty the extent to which biological condition will improve with specific improvements in either habitat or water quality. Generally, if the biological community condition varies directly with the habitat quality, then water quality is not the principal factor affecting the biota. The opposite is considered true if the biological condition is degraded relative to the potential of its habitat. Some measures of biological condition in comparison to habitat condition can be used as indicators of organic enrichment or energy source alteration (Barbour and Stribling 1991).

Biological and habitat data collected from numerous sites in good or near-natural condition can be used to determine the type of biological community that should be found in a particular aquatic habitat. This natural condition has been referred to as reflecting biological integrity, defined by Karr and Dudley (1981) as “the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region.” Highly detailed biological assessments are comparisons of biological conditions at a test site to the expected natural community and are thus a measure of the degree to which a site supports (or does not support) its “ideal” or potential biological community (Gibson et al. 1996). Other types of biological assessment involve comparisons of impacted sites to control sites, the latter being sites that are similar to monitored sites but are not affected by the stresses that affect the monitored sites (Skalski and McKenzie 1982). Knowledge of the natural condition is still valuable for accurate data interpretation when control sites are used (Cowie et al. 1991).

4.2.3 Limitations of Biological Assessments

Although biological assessment is useful for detecting and prioritizing severity of aquatic ecosystem degradation, it does not necessarily provide a direct linkage to or measurement of specific stressors. Thus, it usually does not provide definitive information about the cause of observed water body degradation, i.e., the specific pollutants or their sources. Monitoring of chemical water quality and toxicity may also be necessary to design appropriate pollution control programs.

Prior to routine application in monitoring and assessment, it is necessary to calibrate biological indicators for the water body types, geographic and ecological regions, and to structure sampling and analysis programs to appropriate address management objectives. Thus, establishing a biological assessment and monitoring program can require a significant investment of time, staff, and money. However, the majority of these are one-time, up-front investments dedicated to the establishment of reference conditions, standard operating procedures, and a programmatic quality assurance and quality control plan.

Finally, there can often be a lag between the time at which a toxic contaminant or some other stressor is introduced into a water body and a detectable biological response. Consequently, biological monitoring may not always be appropriate for determining system response due to short-term stresses, such as storms. Similarly, there is often a lag time in the improvement of biological condition following habitat restoration or pollution abatement. The extent of this lag time is difficult to predict; but, it should be recognized and anticipated. Other factors also determine the rate at which a biological community recovers, e.g., the availability of nearby populations of species for recolonization following pollution mitigation and the extent or magnitude of ecological damage done during the period of perturbation (Richards and Minshall 1992). Both the possibility of the lack of detectable recovery from perturbation and unpredictable lag times before improvement is noticeable have obvious implications for the applicability of biological monitoring to some NPS pollution monitoring objectives. Table 4-1 summarizes the strengths and limitations of the biomonitoring approach.

Table 4-1. General strengths and limitations of biological monitoring and assessment approaches

Strengths	Limitations
Properly developed methods, metrics, and reference conditions (i. e., calibration) provide a means to assess the ecological condition of a waterbody	Development of regional methods, metrics, and reference conditions takes considerable effort and an organized and well-thought-out design
Biological assessment data can be interpreted based on regional reference conditions where reference sites for the immediate area being monitored are not available	Rigorous bioassessment can be expensive and requires a high level of training and expertise to implement
Bioassessments using two or more organism groups at different trophic levels provide improved confidence in interpretation of assessment results	Basic biological assessment information does not provide information on specific cause-effect relationships
Biological condition is an indicator of cumulative effects from both short- and long-term stressors	There may be a lag time between pollution abatement or BMP installation and community recovery, so monitoring over time is required for trend detection
	Biological assessment does not always distinguish between the effects of different stressors in a system impacted by more than one stressor

4.2.4 Reference Sites and Conditions

Biological condition assessments often compare metrics of observed assemblages to the expectations for those assemblages in the absence of environmental disturbance. Those expectations are based on samples taken from water bodies that are minimally degraded, have low stressor loads, or are considered "best available". They constitute the reference condition, which is often derived from observations collected from reference sites with minimal levels of disturbance (Hughes 1995, Stoddard et al. 2006, Gibson et al. 1996 [see Figure 4-2]).

A reference condition is a composite characterization of the natural biological condition in multiple ecologically homogeneous reference sites. The overall goal of establishing a reference condition is to describe the natural potential of the biota in the context of natural variation. Minimally disturbed reference sites are those with habitats assumed to fully support a natural biota. The greater the difference is between indicator characteristics of reference and monitored samples, the more disturbed the monitored samples are considered. The disturbance responsible for the difference might be a habitat change, pollution, or some other stress. Another approach is describing expectations relative to a complete gradient of disturbance conditions, and thus, interpreting biological conditions relative to that gradient, the biological condition gradient (BCG) (Davies and Jackson 2006 [see section 4.4.3]). Site classification is integral to the reference condition concept (Gerritsen et al. 2000, Hawkins et al. 2000a). Site classification accounts for natural biological variability prior to evaluating potential effects of human disturbance. The objective of classification is to group water bodies with similar reference biological

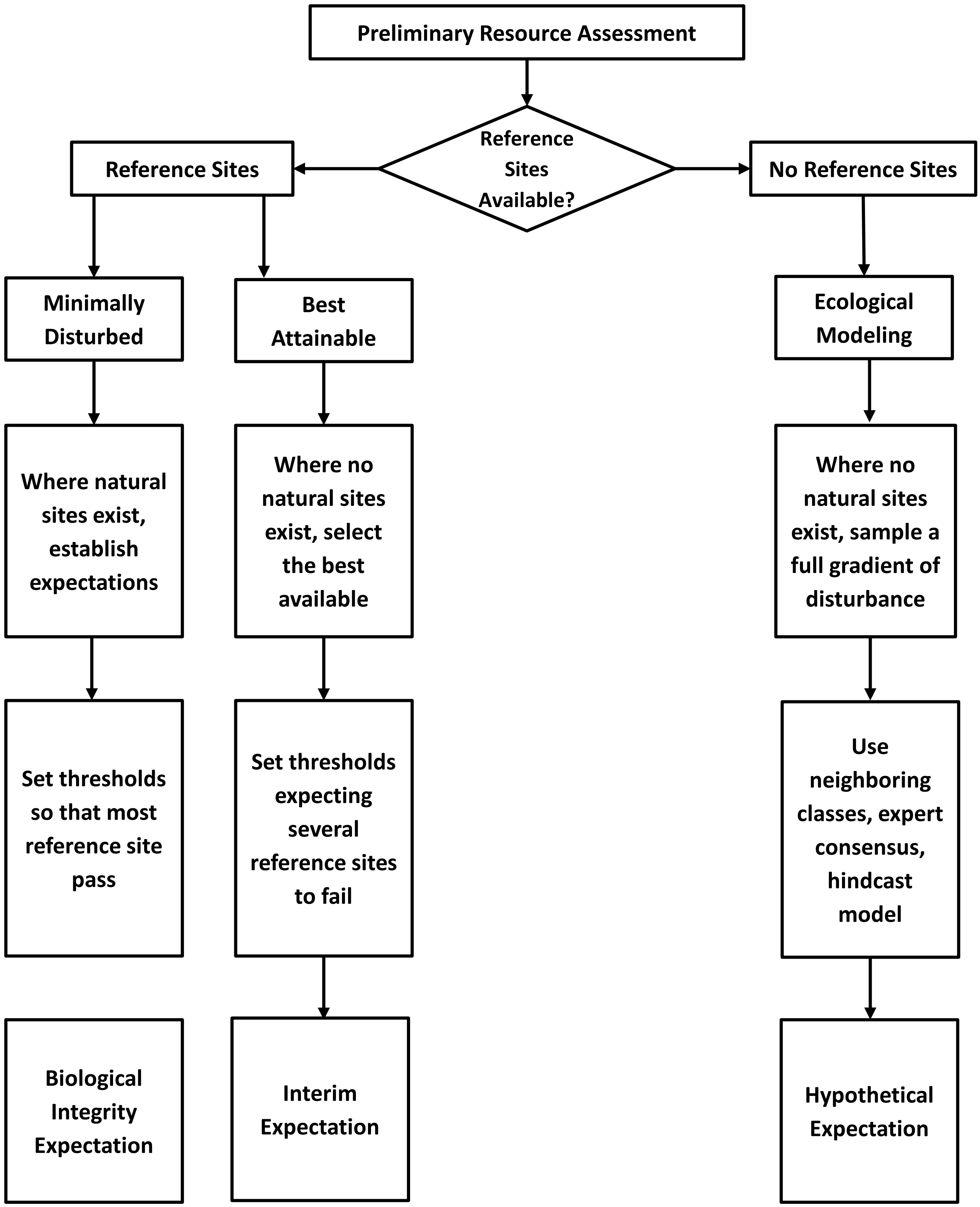

Figure 4-2. Approach to establishing reference conditions (after Gibson et al. 1996)

characteristics together, allowing formulation of precise indicator expectations within each group (Hughes et al. 1986). Site classification can be categorical or continuous. Categorical classification groups distinct site types with similar biological and environmental characteristics. Continuous classification recognizes the gradation of site types. Categorical classification is generally used for MMIs, and continuous classification is often associated with predictive models of observed and expected taxa. Reference sites can be defined identically for either indicator type. Developing reference condition/expectations and site classification are both critical components of index calibration and will have already been accomplished by regional environmental resource management agencies, universities, or other monitoring entities.

In the sense of biological monitoring and assessment, reference sites are traditionally thought of as having been used in the index calibration process, and thus, sampling of them for a particular site specific assessment is not required. In addition to the calibrated metric and index scoring framework used in assessments, individual control sites could add potentially useful information for specific stressor inputs to streams or other water bodies. Control sites could be positioned to be upstream/downstream, or pre- and post-implementation, of input points or zones from the land use/land cover of concern, including BMP or other stressor source control activities. In this situation, it would be necessary to have multiple reaches sampled to represent different assessment areas.

4.3 Biomonitoring Program Design

As described in chapter 2, different monitoring designs might be applicable depending on the objectives of individual monitoring projects with consequent implications for site selection process, number of sites sampled, what features are monitored, and time and frequency of sampling. The sampling design used in NPS biological monitoring might consist of either targeted or probabilistic designs. Targeted monitoring designs are normally chosen for site-specific objectives such as whether biological impairment exists at a given site or whether impairment has been reduced by a watershed project. See section 2.4 for a more detailed discussion of targeted site selection design; however in brief, site locations are selected based on the purpose of the project. Targets might include vulnerable areas with known or suspected perturbations (stressors), planned point source controls, or waterbodies in areas treated with BMPs. NPDES permits, urban stormwater sites, timber harvest areas, rangeland, row crop farming, and construction sites are examples of known stressor sites. Upstream/downstream sampling stations, before-and-after site alterations, or recovery zones (sampling at established distances from sources) are types of sampling locations for known stressor sites. Ecologically sensitive sites that may or may not be affected by stressors and reference sites with minimal disturbance might also be chosen as with targeted monitoring.

While targeted monitoring designs are often used for site-specific purposes, the results from those studies cannot be extended to other sites in the region (Fore and Yoder 2003). Alternatively, probabilistic designs are useful for providing unbiased assessments of conditions across a water body or large geographic area. In a probabilistic sampling program, the entity about which inferences are made is the population or target population and consists of population units. The sample population is the set of population units that are measured. As an example, in a watershed impacted by nonpoint sources, the target population could be the biological condition of all 1st-, 2nd-, and 3rd- order streams. Benthic macroinvertebrates, selected water chemistry, and physical habitat quality are then collected at randomly selected sites drawn from the population of 1st, 2nd, and 3rd order streams. By sampling and statistically evaluating randomly selected population units, inferences can be made about the entire waterbody. The advantages and disadvantages of targeted and probabilistic sampling are summarized in Table 4-2. In some cases, a monitoring program may have a combination of targeted and probabilistic sites.

Table 4-2. Comparison of probability-based and targeted monitoring designs

	Advantages	Disadvantages
Probabilistic Design	Provides unbiased estimates of status for a valid assessment on a scale larger than that of the individual sample location. Can provide large-scale assessment of status and trends of resource or geographic area that can be used to evaluate effectiveness of environmental management decisions for watersheds, counties, or states over time. Stratified random sampling can improve sampling efficiency, provide separate data on each stratum, and enhance statistical test sensitivity by separating variance among strata from variance within strata.	Small-scale problems will not necessarily be identified unless the waterbody or site happens to be chosen in the random selection process. Cannot track restoration progress at an individual site or site-specific management goals.
Targeted Design	Targeted sampling along a stream or river provides an efficient means of detecting pollution sources (Gilbert 1987). Identifies small-scale status and trends of individual sites, which can be used to assess potential improvements due to stressor controls and other management activities. Contributes to understanding of responses of biological resources to environmental impact.	A targeted design will not yield information on the condition of a large-scale area such as the watershed, county, state, or region. It cannot specifically monitor changes from management activities on a scale larger than site-specific. Resource limitations usually make it impossible to monitor effects of all pollutant sources using a targeted design. Targeted sampling can result in biased results if there is a systematic variation in the sampled population.

For probability-based designs, simple random sampling is not optimal. It can produce clusters of sampling sites that might not be representative of the larger scale area of interest (e.g., Hurlbert 1984). Therefore, some sort of stratification is preferred for ensuring a dispersed distribution of site locations (Stevens and Olsen 2004). The approach of stratifying the target population and then randomly sampling, referred to as *stratified-random sampling*, is often more efficient than simple random sampling. This is because a target population is recognized to consist of groups that each have internal homogeneity (relative to other groups), and stratifying the target population will tend to minimize within-group variance and maximize among-group variance (Gilbert 1987, Fore and Yoder 2003). Case Study 2 from Maryland provides an example of secondary uses of data and assessment results from stratified-random sampling. Another approach for randomly selecting sites, used by EPA's national aquatic resource surveys, that allows for random sampling while ensuring representation from all relevant site types and locations is the unequal probability Generalized Random Tessellation Stratified (GRTS) spatially-balanced survey design (Stevens and Olsen 2004).

Case Study 2: Effectiveness of Stormwater Ponds in Enhancing Instream Biological Conditions in Maryland Urban Watersheds

- ✓ Urbanized watersheds, elevated flashiness, channel instability, habitat degradation
- ✓ Stormwater retention ponds
- ✓ Multiple watershed monitoring
- ✓ Benthic macroinvertebrates, physical habitat
- ✓ Random site selection, post-stratification
- ✓ Assessing effectiveness

Prince George's County, Maryland covers 1,290 km^2 of the mid-Atlantic Coastal Plain, and is the suburban jurisdiction immediately east of the District of Columbia (Figure CS2-1). It has more than 965 kilometers of stream channels draining to the Patuxent River on the east, Anacostia River in the west and northwest, Potomac River on the southwest, and Mattawoman Creek in the south. The northeast corridor of the U. S. has undergone heavy and nearly continuous urbanization, resulting in an increasing percentage of landscape covered with impervious surfaces. Increased imperviousness reduces the capacity for land to absorb rainfall and causes storm water to be instantly converted to runoff. This can result in accelerated erosion and severe channel instability in watersheds that contain large proportions of impervious cover. Surface runoff can also transport solid trash and contaminants from parking lot asphalt, sidewalks, and rooftops. These stressors can combine to cause substantial harm to aquatic communities. The primary objective for this project was to assess the effectiveness of stormwater detention/retention ponds in protecting and enhancing instream biological condition.

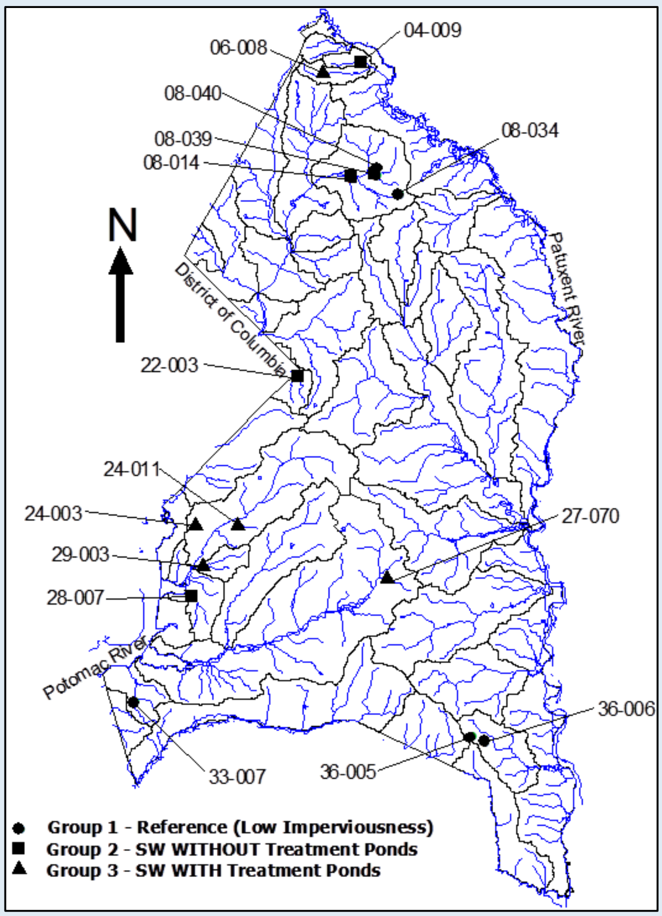

Figure CS2-1. Prince George's County, Maryland. Distribution of sample locations used as part of this analysis.

Stream and watershed assessments have been conducted in Prince George's County for nearly two decades and have fully documented spatial patterns of aquatic biological conditions. The initial round of county-wide assessments (1999 to 2003) showed that more than half of the stream length (52.5 percent) was biologically degraded, with the majority of impaired stream reaches located in the western part of the county near major regional transportation thoroughfares. Higher quality streams predominantly rated as "fair" and "good" were found to be in the eastern part of the county along the drainage to the Patuxent River, and in the south, near the border with Charles County and Mattawoman Creek. This pattern was expected because greater development intensity typically follows transportation corridors with the attendant increases in impervious surfaces.

One of the principal BMPs used for stormwater management in the 1990s was stormwater detention/retention ponds (SWPs). Many managers designing and installing SWPs had multiple objectives including, but not limited to, collecting and slowing runoff from impervious surfaces, allowing suspended particulates to settle out, collecting trash solids, and providing features that helped foster lower water temperatures and elevated levels of dissolved oxygen. BMP siting decisions were often based on stream problems brought to the attention of county managers by the public, citizens' environmental groups, or stream monitoring data. Thus, it was not always straightforward to identify specific management objectives or goals associated with an individual SWP. Nonetheless, the authors of this study (Stribling et al. 2001) assumed that improving instream biological condition was among the objectives for SWP implementation.

Monitoring and Sampling Design

Routine county-wide monitoring performed by the Prince George's County Department of Environmental Resources (PGDER) to assess stream and watershed biological conditions is based on a long-term, probability-based, rotating basin plan. Stratified (by wadeable stream order) random sampling was used to select sampling sites for this study, with the number of potential sites for each stream order within each of the 41 subwatersheds set proportional to the number of stream km in each order (map scale = 1:100,000). There are 50 to 60 sites sampled per year for biological (benthic macroinvertebrates), selected water chemistry, and physical habitat quality variables. Benthic macroinvertebrates are collected over 100-m channel reaches by making 20 1-m linear sweeps (jabs) with a 500 μ mesh D-frame net distributed among different habitat types (such as snags, leaf packs, vegetated/undercut banks, bottom, riffle/cobble) in proportion to their frequency of occurrence at each site. To minimize the effects of seasonal variability, all sampling occurs during the Maryland Department of Natural Resources' Biological Stream Survey (MBSS)-specified index period, which is March 01 to April 30. Ten percent of the sampling segments are randomly selected for replicate reaches, which provides information necessary for quality control and for calculating field sampling precision. Data on physical habitat quality and water chemistry (pH, conductivity, water temperature, and dissolved oxygen) are collected at each site for their potential in explaining biological condition. Benthic data (number of taxa, number of individuals of each, per sample) are used to calculate the benthic-index of biological integrity (B-IBI) developed by the MBSS. In general, the B-IBI and the physical habitat quality index, as applied by the PGDER program, have 90 percent confidence intervals of ± 0.67 points on a 5-point scale, and ±6.7 points on a 200-point scale, respectively.

The purpose of this study (Stribling et al. 2001) was to determine effectiveness of stormwater detention ponds in protecting and/or enhancing in-stream biological condition. The physical habitat and biological data were segregated into the following three treatment groups and directly compared using percentile distributions of measurement values:

Group 1 Streams with minimal stormwater stressors (*0-5 percent impervious surface*),

Group 2 Streams with substantial stormwater stressors (>12 percent impervious surface) and *without* SWP, and

Group 3 Streams with stormwater stressors (>12 percent impervious surface) and *with* SWP.

Using GIS analysis, upstream drainage areas were delineated for all sites sampled in 2000, and land use/land cover (LULC) determined for each, including calculation of impervious surface. From these data and the existing dataset, all sites were screened to define a set of sites that would reasonably represent each of the groups (Table CS2-1), resulting in five sites per treatment group. Group 1 was represented by sites with drainage areas (DAs) ranging from 31 to 585 ha and imperviousness of 1.0 to 4.7 percent. Group 2 had DAs ranging from 248 – 634 ha and imperviousness of 17 to 34 percent, and Group 3 had DAs from 76 to 568 ha and 11.2 to 34.9 percent imperviousness. The ages (time elapsed since installation) of the SWPs of Group 3 ranged from 7 to 13 years, so all SWPs should have had ample opportunity to "mature" through multiple growing seasons. SWP functionality was not assessed as part of this study.

Table CS2-1. Sampling sites in Prince George's County including drainage area, percent imperviousness in the drainage, and proportions of different land use/land cover types in the drainage area

						Percent Land Use/Land Cover								
Site ID	**Grp**	**Site Name**	**DA (ha)**	**% Imper**	**Date of Pond**	AGR	BAR	COM	FOR	HDR	IND	LDR	MDR	OS
08-034	1	Beck Branch	60	1.0	NA	19.5	0	0	80.5	0	0	0	0	0
08-040	1	UT to Upper Beaverdam Creek	148	2.7	NA	27.8	0	1.7	70.5	0	0	0	0	0
33-007	1	UT to Lower Potomac River	284	2.1	NA	12.7	0	0	80.3	0	0	5.2	1.7	0
36-005	1	Black Swamp Creek	585	4.7	NA	18.2	5.9	1.3	64.3	0	0	7.8	2.4	0
36-006	1	UT to Black Swamp Creek	31	2.4	NA	28.3	10.0	0	61.6	0	0	0	0	0
04-009	2	Crows Branch	297	34.1	NA	0	0	7.3	19.9	20.9	0	17.9	29.5	4.4
08-014	2	UT to Upper Beaverdam Creek	476	27.6	NA	2.7	0.7	21.2	57.7	16.9	0	0	0	0.8
08-039	2	UT to Upper Beaverdam Creek	599	17.2	NA	13.1	0.7	21.3	64.7	0	0	0.3	0	0
22-003	2	Watts Branch	248	26.2	NA	0	0	7.5	37.5	2.9	0	9	43.1	0
28-007	2	UT to Broad Creek	634	20.5	NA	6.4	0.1	10.6	47.9	0	0	7.9	26.9	0
06-008	3	Bear Branch	167	20.5	1987	3.9	20.2	3.8	51.3	5.6	15.2	0	0	0
24-003	3	UT to Carey Branch	76	34.9	1987	0	0	0.1	4.6	0	0	13.3	81.9	0
24-011	3	UT to Henson Creek (Broad Creek)	120	31.3	1993	0	0	0	13.6	0	0	13.1	73.3	0
27-070	3	UT to Piscataway Creek	515	11.2	1989	15.7	6.2	0.7	52.2	1.9	0	5.2	18.1	0
29-003	3	Hunters Mill Branch	568	11.7	1987	12.1	0.1	0.5	59.7	0	0.9	2.7	24.1	0

Abbreviations: Grp-treatment group; DA-drainage area; % Imper-percent imperviousness; AGR-agriculture; BAR-bare ground; COM-commercial; FOR-forest; HDR-high density residential; IND-industrial; LDR-low density residential; MDR-medium density residential; OS-open space

Results

As is common in many stream assessments, biological conditions (B-IBI scores and assessments) of the stream groups were plotted against the overall scores for physical habitat quality. Expectations are that biological conditions would be elevated in the presence of good habitat quality, which generally held true with this dataset (Figure CS2-2). Those locations with higher scores and ratings for biological condition tended to be those with better habitat quality and falling in the upper right hand quadrant of the chart; those sites also tended to be in the Group 1 set of sites, that is, with lower stormwater stressors. Those sites with lower biological condition scores and ratings were generally in the central and lower left portions of the chart, and tended to be members of site groups 2 and 3 which were in drainages with substantial impervious surface areas (>11 percent) and the attendant storm water stressors.

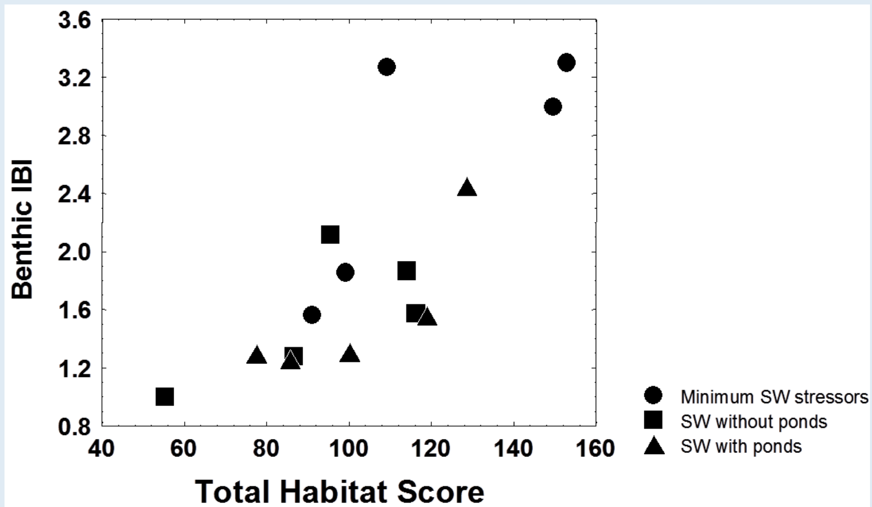

Figure CS2-2. Relationship of biological condition (benthic index of biological integrity [B-IBI]) to physical habitat quality (total habitat score). The 90 percent confidence intervals (CI) are 0.67 points on a 5-point scale for the B-IBI and 6.7 points on a 200-point scale for the habitat quality index.

Both habitat quality and biological condition indicated higher quality for streams with minimal impervious (<5 percent) (Figures CS2-3 and CS2-4). Habitat quality was slightly better in Group 1 streams (with scores ranging from 91 to 150) than in Groups 2 and 3 where scores ranged from 55 to 116 and 78 to 129, respectively (Figure CS2-3). The biological condition of Group 1, however, was more strongly separated from the other two groups (Figure CS2-4). The ranges of B-IBI scores were 1.57 to 3.29 for Group 1, 1.0 to 2.14 for Group 2, and 1.29 to 2.43 for Group 3. There was very little difference between the two groups of streams exposed to stormwater stressors, whether with SWPs (Group 3) or without (Group 2), suggesting that the ponds, in themselves, do not have a strong effect on improving the quality of instream biological conditions. While the SWPs buffered some stressors arising from impervious surface runoff, it is likely that there were other stressors that were unrecognized and not addressed by the SWPs. These other stressors could include upstream or atmospheric chemical contamination, excessive suspended particulates, altered energy input (i.e., leaf litter or woody materials), or habitat alteration. Potential shortcomings of this study include lack of focus on evaluating changes of stressor loads from individual SWPs through, for example, pre- and post-implementation data on stressors for streams where SWPs were implemented; and lack of information regarding the design specifications of the SWPs. The investigators recognized that evaluating the effectiveness of individual BMPs required more intensive load monitoring around the BMP(s) of concern, but, nonetheless, concluded that isolated BMPs were not likely to enhance instream biological condition on their own and that restoration and protection of natural resources requires management on the watershed scale and dealing with multiple and complex stressors and stressor sources.

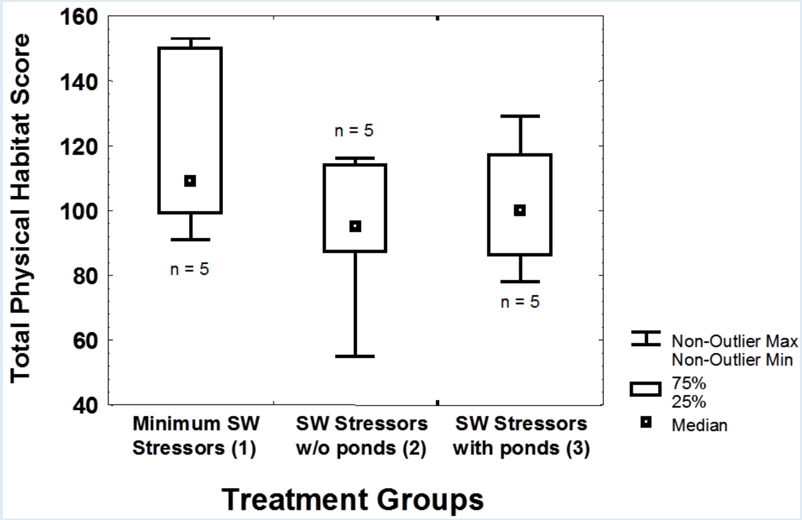

Figure CS2-3. Boxplots of physical habitat quality among each of three stormwater (SW) treatment groups

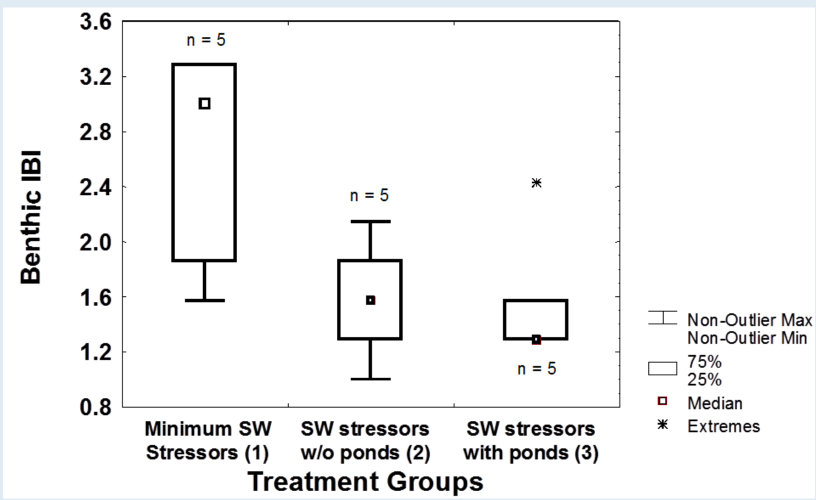

Figure CS2-4. Boxplots of biological condition (benthic index of biological integrity [B-IBI]) among each of three stormwater (SW) treatment group

Literature

Stribling, J. B., E. W. Leppo, J. D. Cummins, J. Galli, S. Meigs, L. Coffman, and M.-S. Cheng. 2001. Relating instream biological condition to BMP activities in streams and watersheds. In *Linking Stormwater BMP Designs and Performance to Receiving Water Impact Mitigation Proceedings of the United Engineering Foundation Conference, August 19-24, 2001, Snowmass Village, Colorado* ed. B. R. Urbonas, pp. 287-304. ISBN 0-7844-0602-2.

Monitoring performed at different spatial scales can provide different types of information on the quality and status of water resources. Conquest et al. (1994) discuss a hierarchical landscape classification system, originally developed by Cupp (1989) for drainage basins in Washington State that provides an organizing framework for integrating data from diverse sources and at different resolution levels. Assessments of waterbodies on a large scale such as an ecoregion, subregion, state, or county provide information on the overall condition of waterbodies in the respective unit. Assessment on a small geographic scale may involve a whole stream, river, or bay or a segment (reach) of the waterbody. A targeted sampling design applies to monitoring waterbodies within a watershed that are exposed to known stressors. Known disturbances, such as point sources, specific NPS inputs, or urban stormwater runoff, can all be targeted for small-scale assessments. At this scale the effectiveness of specific pollution controls, BMP installation/implementation, natural resource management activities, or physical habitat restoration can be monitored. This scale is also where a paired-watershed design can be useful. See Case Study 3 from Pennsylvania for an example of interpreting results from this type of assessment design. Table 4-3 summarizes a waterbody stratification hierarchy for streams and rivers, lakes, reservoirs, estuaries, and wetlands. Further, the sampling site, or the portion of the water body to be sampled, is defined based on technical objectives and programmatic goals of the assessment and/or monitoring activity (Flotemersch et al. 2011). For example, EPA defined a sample reach as 20 times the mean wetted width for its national surveys of lotic waters (streams and rivers) (USEPA 2009); however, many individual states use a fixed 100 m as the sampling reach (Barbour et al. 1999, Carter and Resh 2013).

Depending on the waterbody, subsequent stratification levels may vary in number and may be quite different across waterbodies at a given level in the hierarchy. For example, a state or regional monitoring program designed to assess the status of biological communities in streams might need to be stratified to the level of segments, whereas monitoring to assess the efficacy of specific stream restoration measures might need to be stratified to the macro- or microhabitat level. If data collected by a particular design are so variable that meaningful conclusions cannot be drawn, post-stratification of the data set might be required. If stratification to the level of microhabitat is needed, the sampling and analysis methods used at higher levels might be inappropriate or inadequate.

Case Study 3: Effects of Streambank Fencing on Benthic Macroinvertebrates

- ✓ SE Pennsylvania
- ✓ Pasture animal exclusion from stream access
- ✓ Paired-watershed monitoring
- ✓ Nested experimental design

Big Spring Run Basin, a subbasin of Mill Creek Watershed located in Lancaster County, Pennsylvania (Figure CS3-1) is dominated by agricultural land use, much of which is adjacent to aquatic systems. The Mill Creek Basin falls in the Susquehanna River Basin, which ultimately feeds into the Chesapeake Bay. The most common agricultural NPS control measures implemented in the watershed were barnyard runoff control and streambank fencing. This study was designed to provide land managers information on the effectiveness of streambank fencing in controlling NPS pollution. While the project addressed both water quality and biological condition, the emphasis here is on results associated with macroinvertebrate monitoring.

Monitoring and Sampling Design

The objective of this monitoring program was to document the effectiveness of streambank fencing of pasture land on the quality of surface water and near stream ground water. The primary monitoring design was a paired-watershed design, but above/below monitoring was also included to provide multiple opportunities for comparisons to ensure that the effects of fencing could be documented.

The Big Spring Run Basin consists of two similar subbasins ideal for paired-watershed analysis; one was chosen as the treatment basin and the other as the control. The treatment basin was 3.6 km^2 with 4.5 km of stream, of which approximately 70 percent of the streams run through open pasture. The control basin was 4.7 km^2 with 4.3 km of stream and consisting of approximately 70 percent of streams running through open pasture. Elevation and geologic makeup were also nearly identical for both basins, with stream gradients ranging from 0.3 to 0.6 m elevation change for every 30 m of channel. Temperate zone climate was typical for the study basins, with an average precipitation of 104 cm and an average temperature of 11°C. Agriculture accounted for over 80 percent of the land use in each subbasin.

Surface water monitoring stations for the paired-watershed analysis were located at the outlets of the control (C-1) and treatment (T-1) subbasins (Figure CS3-1). Site T-1 was to also be used with T-3 in the treatment subbasin for an above/below study; streambank fencing was to be installed between T-1 and T-3. A site (T-2) was also added at a visually degraded upstream tributary for comparison with C-1 in another paired-watershed analysis. Site T-4 was added to determine the effects of new construction that began two years into the study. Surface water samples were collected every 10 days from April to November (about 25 to 30 samples per site per year) because this was when dairy cows and heifers were pastured. Monthly base-flow samples were collected during the remaining part of the year. Storm event samples were collected at all sites except T-3, with from 35 to 60 percent of the storm events sampled over the entire study period. Monitoring variables included total and dissolved nutrients, suspended sediment concentration, field parameters (low flow only), fecal streptococcus (low flow only), and discharge.

A nested well approach was used to monitor near-stream groundwater parameters, including water quality and level, flow directions, age dating, and chemical quality. Two nested wells were placed in the treatment basin at sites T-1 and T-2. At each location, three shallow wells and one deep well were clustered together. One of the shallow wells at each ground water monitoring location was placed outside of the fenced area as a control. Water level was measured continuously and wells were sampled monthly during periods of little to no recharge. The resulting samples were analyzed for nutrients and fecal streptococcus.

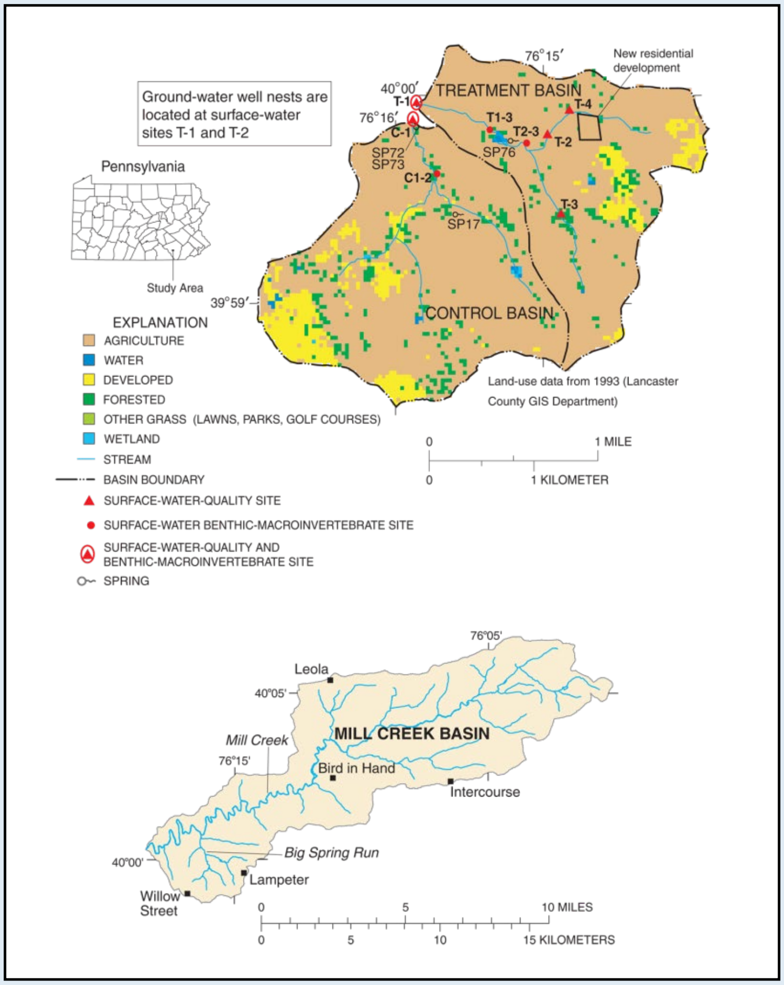

Figure CS3-1. Land-use map of study area and location of surface-water sites, ground-water well nests, and selected springs in the Big Spring Run Basin, Lancaster County, PA

Benthic macroinvertebrate sampling was conducted in May and September of each year at five different locations (Figure CS3-1): one at the outlets of the control and treatment basins (sites T-1 and C-1), two upstream in the treatment basin (sites T1-3 and T2-3), and one upstream in the control basin (site C1-2). Stream pool and riffle habits were sampled using the kick-net method; the USEPA Rapid Bioassessment Protocols (RBP) was used to characterize habitat; and water quality samples and stream measurements were also taken. Most metrics were applied to taxonomic identifications to the family level. The list of metrics includes percent dominant taxon (genus and family), EPT (Ephemeroptera, Plecoptera, and Trichoptera) index, generic EPT/Chironomidae ratio, EPT/total number, percent Chironomidae, shredders/ total taxa ratio, scrapers/filterers ratio, Hilsenhoff Biotic Index (HBI) (genus and family), taxa richness (genus and family), and percent Oligochaeta.

All monitoring was conducted both before and after installation of streambank fencing which occurred in the treatment subbasin from May 1997 through July 1997. About 3.2 kilometers of fencing was installed along riparian zones in pastured areas to create a 1.5- to 3.6-m-wide stream buffer strip. Each pastured fenced had an average of two cattle crossings through the stream to allow the animals to migrate between pasture locations and access a water supply. Monitoring was carried out before and after fencing was installed. The pre-treatment period was from 1993 to 1997 and post-treatment monitoring was carried out from mid-July 1997 through June 2001.

Due to their potential effect on the quality and quantity of the water and habitat, basin-wide covariate data were also collected during the study period, such as precipitation, inorganic and organic nutrient applications, and the number of cows present. Precipitation and agricultural data were obtained, respectively, using precipitation gauges (logging at 15-minute intervals), and from farm operators (monthly records).

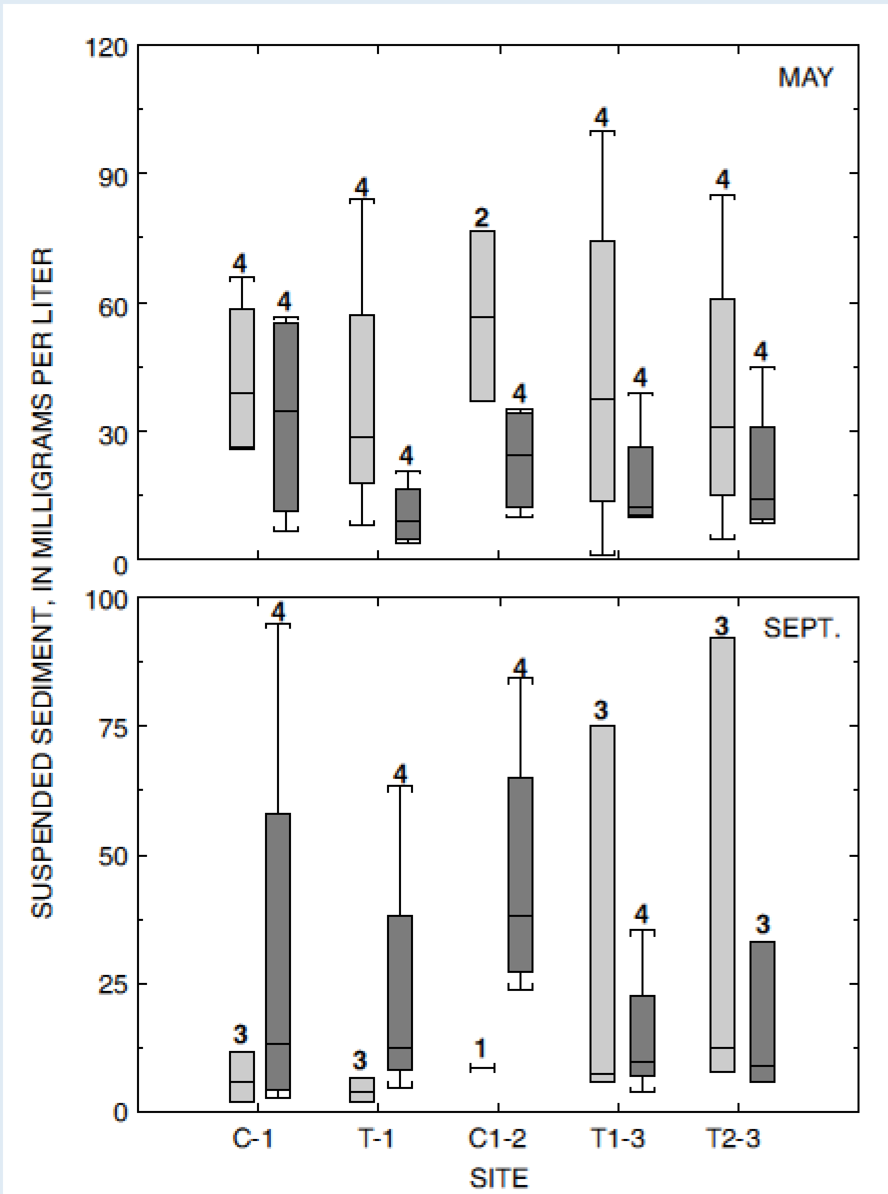

Figure CS3-2. Distribution of concentrations of suspended sediment for May and September sampling events at benthic macroinvertebrate sites for the pre- and post-treatment periods in the Big Spring Run Basin, Lancaster County, PA. Light shaded boxes are pretreatment, dark shaded are post-treatment, C represents Control, and T represents Treatment.

Results

Data collected from the pre-treatment period (1993 to 1997) were compared to those from the treatment and control basins to determine the effectiveness of streambank fencing. Changes in the yields of nutrients and suspended sediment during low flow and storm flow events were quantified using ANCOVA. ANCOVA was also used to quantify changes between pre- and post-treatment concentrations of nutrients, water quality, and fecal streptococcus in collected water samples, as well as nested wells inside and outside of the treatment area. Canonical correspondence analysis (CCA) was used to determine the effects of streambank fencing on instream biological conditions as characterized by benthic macroinvertebrates. A brief overview of major findings from analysis of water chemistry data is presented here, followed by a more detailed summary of results from macroinvertebrate monitoring.

It was concluded that water chemistry results indicated that riparian fencing had fairly consistent effects on suspended sediment but less clear effects on nutrients. Post-treatment period improvements were evident at site T-1 for both nutrients and sediments; however, site T-2 showed reductions only in suspended sediment. The average reduction in suspended sediment yield for the treated sites was about 40 percent. N species at T-1 showed reductions of 18 percent (dissolved NO_3) to 36 percent (dissolved ammonia); yields of TP were reduced by 14 percent. Conversely, site T-2 showed increases in N species of 10 percent (dissolved ammonia) to 43 percent (total ammonia plus organic N), and a 51-percent increase in yield of TP. The different results for nutrients at T-2 and T-1 were attributed to ground water contributions and the failure to implement nutrient management along with the fencing. Shallow ground water flow contributed to stream flow at T-2, but the stream was losing water to the shallow ground water system at T-1. It is believed that an upland agricultural field caused increased dissolved P levels in shallow ground water at T-2, resulting in a transport of P from ground water to the stream that increased stream P levels. In addition, cattle contributed nutrients directly to the stream via excretion at the embedded stream crossing at T-2.

Analysis of the benthic macroinvertebrate samples showed some apparent improvements relative to the control sites in riparian and instream habitat (sites T2-3 and T-1 versus C-1). Some differences in bottom substrate, bank stability, available cover, and scouring and deposition were observed in the downstream and upstream locations within the treatment basin that could potentially be considered slight improvements. Water quality data collected during the benthic macroinvertebrate sampling suggested the overall improvement to instream habitat was due to the decreased load of suspended sediment (Figure CS3-2). The fenced riparian buffer, despite being narrower than what was considered optimal, allowed vegetation to become fully established and bank stability to improve. It was particularly evident at site T2-3 where it became overgrown with vegetation and blocked the stream from view.

The composite benthic index, which combines all metrics, is called the "Macroinvertebrate Aggregated Index" (MAI). For both spring and fall samples, the index showed some improvement for the treatment sites relative to control sites, though, trends were mixed overall. For the treatment basin, sites T1-3 and T2-3 showed no change with spring samples, while the outlet site, T-1, showed a slight 1 unit increase. From the pre-treatment to the post-treatment period, fall index scores changed in the control basin by 1 unit, increasing at C-1 and decreasing at C1-2. Fall scores for the treatment basin also changed over this timeframe, increasing by 2 units for T-1 and T1-3, but decreasing by 1 unit at T2-3.

Disaggregating the index into individual metrics allows evaluation of different components of the benthic macroinvertebrate assemblage. In this dataset, there are different responses by different metrics. No difference was seen for five of the 10 genus-level metrics, which included percent dominant taxa (generic level) (PDTG), EPT taxa, percent EPT taxa, percent shredders, and ratio of scrapers to filterers. Thus, treatment elicited no effect for 50 percent of the metrics. Two of the metrics (EPT/Chironomidae ratio, Hilsenhoff Biotic Index [genus level]; 20 percent) suggested some, or slight, effect of the streambank fencing on treatment sites relative to control; and, distinct effects were seen for the remaining three metrics: percent Chironomidae (Figure CS3-3), taxa richness, and percent Oligochaeta.

Further evaluation of taxa lists for dominance, occurrence, and uniqueness of and by individual taxa can help illuminate differences, in particular, for those taxa that are known to be more pollution-tolerant or sensitive. Spring samples were numerically dominated by worms (Naididae, Tubificidae), scud (Amphipoda: Gammaridae), several different midges (Diptera: Chironomidae: i.e., *Cricotopus, Orthocladius, Dicrotendipes, Micropsectra*), and blackflies (Diptera: Simuliidae: *Simulium*). The fall samples illuminated a shift in the actual taxonomic composition of the site to a greater diversity (i.e., a larger number of taxa) and dominance by largely different taxa, including riffle beetles (Coleoptera: Elmidae: *Dubiraphia, Stenelmis*), net-spinning caddisflies (Trichoptera: Hydropsychidae: *Hydropsyche, Cheumatopsyche*), midges (*Chironomus, Dicrotendipes, Polypedilum, Rheotanytarsus*), and blackflies. The only taxa that retained any kind of dominance for this site across seasons were *Dicrotendipes* and blackflies. The overall differences are driven by seasonality of the system, simply showing a greater diversity during the fall season and not by any changes in stressor load. Although the two outlet sites (C-1, T-1) basically had the same taxa dominating sample data, each had a greater diversity than the upstream sites. Elevated taxonomic/biological diversity can be an indicator of greater diversity and complexity of habitat characteristics; lacking other types of stressor loads, this diversity is likely what is reflected at the outlet sites. The upstream sites in the treatment basin (T2-3, T1-3) consistently showed more diversity than C1-2, but the overall PDTG means for T1-3 and T2-3 increased slightly (<1 percent) and by 13 percent, respectively, from the pre- to post-treatment period. Because differences in assemblage makeup are largely explainable by factors other than what might be introduced by streambank fencing, such as seasonality, and expected physical habitat characteristics, the authors concluded that the treatment did not seem to improve benthic-macroinvertebrate community structure based on PDTG.

Across the full dataset, the dominant family-level taxa in spring samples were Chironomidae, Gammaridae, Naididae, and Tubificidae, all recognized as being semi-tolerant to organic enrichment. The dominant families in fall samples were Gammaridae, Tubificidae, Elmidae, Physidae, Baetidae, Chironomidae, and Simuliidae, all considered as being moderately to very tolerant of organic enrichment. This indicates that the more sensitive taxa were not able to become dominant members of the benthic-macroinvertebrate assemblage after the fences were installed in the treatment basin. Sensitive taxa may not have been present or only a few individuals were present during the post-treatment period because of 1) not enough time for the system to equilibrate to the new conditions, or 2) because these are spring-fed, first- to second-order limestone streams. Limestone streams typically support assemblages including mayflies (Ephemeroptera), midges (Chironomidae), scud (Amphipoda: Gammaridae), and pillbugs/sowbugs (Isopoda: Asellidae), all of which were present in treatment and control basins. Some of the more stressor-sensitive taxa that were present in small numbers included *Promoresia* (Elmidae), *Oxyethira* (Hydroptilidae), *Antocha* (Tipulidae), and two species of Chironomidae (*Pagastia* and *Prodiamesa*).

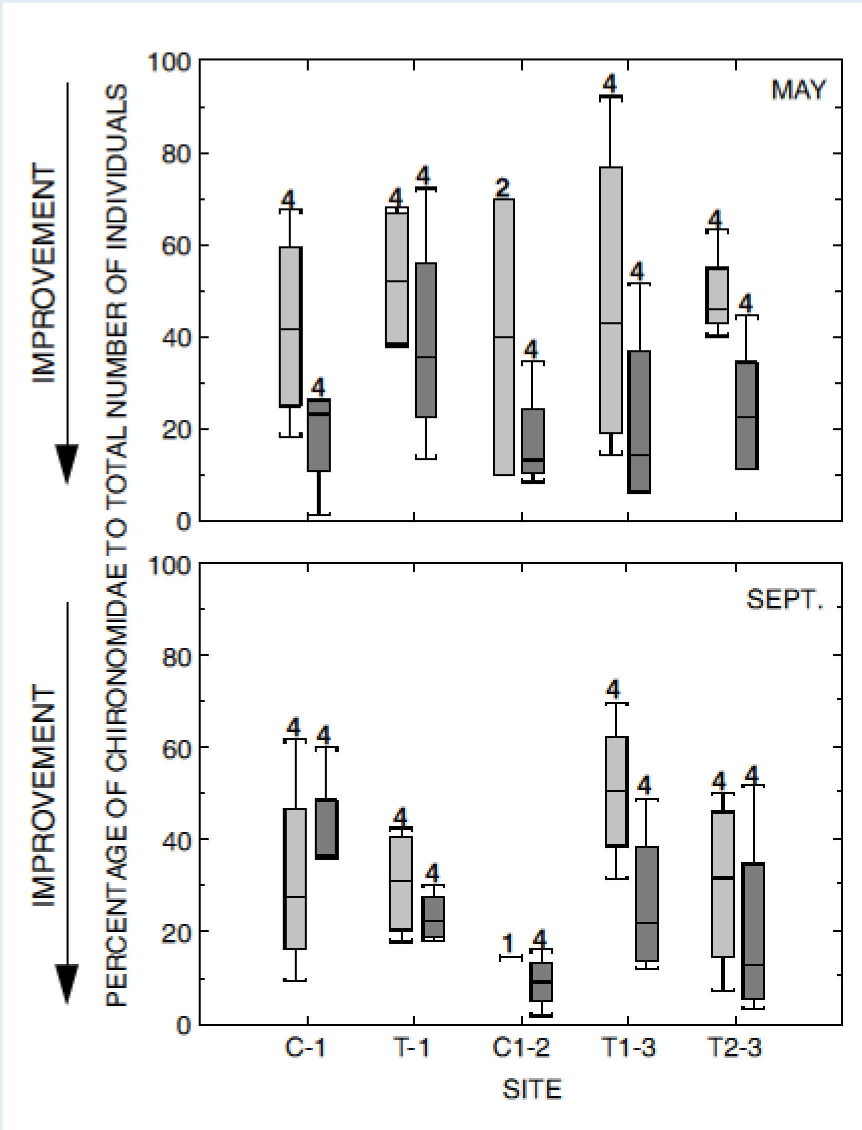

Figure CS3-3. Distribution of the percentage of Chironomidae to total number of individuals for May and September sampling events at benthic-macroinvertebrate sites for the pre- and post-treatment periods in the Big Spring Run Basin, Lancaster County, PA. Light shaded boxes are pretreatment, dark shaded are post-treatment, C represents Control, and T represents Treatment.

Overall, the authors conclude that streambank fencing had a positive influence on the taxonomic diversity of benthic-macroinvertebrates, both at genus and family levels. This positive influence is interpreted as primarily resulting from stabilization of the riparian zone, allowing growth of streamside vegetation to progress, and ultimately allowing better habitat to develop and support more taxa.

Previous studies suggest that for optimal reduction of nutrient loads into nearby aquatic systems, the buffer size should be greater than the 1.5- to 3.6-m buffer used in this study. Because there was such a wide range of what would be considered an adequate buffer size, it was uncertain which nutrient types, if any, could be controlled or reduced with this approach. Study results show that while the fenced streambank buffer was relatively small, it still was substantially effective in improving surface and near-stream shallow ground water quality, and led to some improvement in instream biology. Small-scale stream buffers and exclosure fencing both have limited effectiveness in controlling high nutrient input ultimately transported through subsurface flows; the most pronounced effects of exclosures are in reducing suspended sediment inputs and consequently leading to improved habitat. There may be some effectiveness in controlling excessive nutrient flows, but the benefits are likely minor in comparison to the habitat effects.

Literature

Galeone, D.G., R.A. Brightbill, D.J. Low, and D.L. O'Brien. 2006. *Effects of Streambank Fencing of Pastureland on Benthic Macroinvertebrates and the Quality of Surface Water and Shallow Ground Water in the Big Spring Run Basin of Mill Creek Watershed, Lancaster County, Pennsylvania, 1993-2001*. Scientific Investigations Report 2006-5141. U. S. Geological Survey, Reston, VA.

Galeone, D.G. and E.H. Koerkle. 1996. *Study design and Preliminary Data Analysis for a Streambank Fencing Project in the Mill Creek Basin, Pennsylvania.* Fact Sheet FS-193-96. U. S. Geological Survey, Lemoyne, PA.

Table 4-3. Waterbody stratification hierarchy

Population	Streams	Lakes[b,c]	Reservoirs[b,d]	Estuaries[e]	Wetlands[f]
Operational Sampling Unit (SU)	Channel segment (i.e., a length of river channel into which no tributaries flow)	Self-contained basin	Self-contained basin (hydrologically isolated from other basins)	Self-contained basin	Transect upland or deep water boundaries
Strata (or higher stages) comprising SUs	Ecoregion	Ecoregion	Ecoregion	Biogeographic province	Wetland system type (marine, estuarine, riverine, lacustrine, palustrine)
	Watershed	Size/surface area of lake (km^2)	Size/surface area of reservoir (km^2) (watershed area/basin surface area)	Watershed	Watershed recharge, discharge or both
	Stream/river channel (ordinal/areal)	Lake hydrology (retention time, thermal stratification)	Hydrology (water level fluctuation/ drawdown; retention time stratification)	Watershed area (km^2)	Class (based on vegetative type; substrate and flooding regime; hydroperiod)
	Segment	Characteristic water quality (natural conditions)	Characteristic water quality (natural conditions)	Zones (tidal basin, depth, salinity)	Flooding regime water chemistry soil type
Habitat within SUs	Characteristic water quality (natural conditions) Hydraulic conductivity (homogeneous, heterogeneous, isotropic, anisotropic)				
	Macrohabitat (pool/riffle; Shorezone vegetation; submerged aquatic macrophytes)		Longitudinal zone (riverine; transitional; lacustrine; tail waters (can be more riverine but always associated with dams))		
		Depth zone (eulittoral 6 profundal)	Depth	Substrate/habitat	Subsystem (subtidal, intertidal, tidal, lower perennial, upper perennial, intermittent, littoral, limnetic)
	Microhabitat	Substrate/ microhabitat	Substrate/ microhabitat		Substrate/ microhabitat

[a] Frissell et al. 1986; [b] Gerritsen et al. 1996; [c] Wetzel 1983; [d] Thornton et al. 1990; [e] Day et al. 1989; [f] Cowardin et al. 1979

4.4 Biological Assessment Protocols

Biological indicators are widely recognized as being critical for evaluating ecological conditions, helping identify and prioritize problems, designing controls or other solutions, and in evaluating effectiveness of management efforts. However, prior to being able to make management decisions using these indicators, two things need to occur. First, the indicator must be calibrated; and, second, sampling and analysis must be instituted in a routine and consistent manner to directly address management objectives. To obtain valid assessment results, and to optimize defensibility of decisions based on them, it is imperative that monitoring be founded on data of known quality (Flotemersch et al. 2006b, Stribling 2011). In this section, our discussion focuses on wadeable streams, and we assume that the user has access to indicators that have been calibrated and are applicable to their region and water body (-ies) of concern. Many states have developed MMIs using one or more biological groups, but most typically benthic macroinvertebrates, and sequentially less so, fish and periphyton (algae and diatoms). Carter and Resh (2013) surveyed state agencies about different characteristics of their biological monitoring programs, and, although there are differences in some of the specific techniques, there has also been considerable convergence among methods during the past 10 to 15 years. Monitoring programs that have gone through the index calibration process have worked or are working through technical issues associated with customizing sampling techniques to water body type, prevailing climatic conditions, and programmatic capacity; using field data to characterize environmental/ecological variability; defining mathematical terms of the indicator (metrics and index make-up); and defining thresholds for judging degradation.

Part of customizing field sampling and laboratory analysis methods for a program involves understanding the range of variability of field conditions and the data/assessments that arise from them. One approach for developing such an understanding is to recognize that biological monitoring and assessment protocols are made up of a series of methods (Flotemersch et al. 2006b, Stribling 2011) generally corresponding to different steps of the overall process: field sampling, sample preparation, taxonomic identification, enumeration, data entry, data reduction, and site assessment/interpretation. In this section, we present descriptions of the background, purpose, application, and output of the methods along with relevant procedures for documenting data quality associated with each.

4.4.1 Field Sampling

In the context of the field sampling approach being used for biological assessment, taking or observing organisms from a defined sample location is intended to provide a representation of the biological assemblage supported by that water body, whether benthic macroinvertebrates, fish, or periphyton. These three assemblages are emphasized because they are most commonly used in routine monitoring and assessment programs in the US, and methods for them are relatively well-documented (Barbour et al. 1999, Moulton et al. 2002, Stribling 2011, Carter and Resh 2013).

4.4.1.1 Benthic macroinvertebrates

Benthic macroinvertebrate samples are taken from multiple habitat types, and composited in a single sample container (Figure 4-3). If sampled from transects as per the USEPA national surveys (USEPA 2009), they are collected along 11 transects evenly distributed throughout the reach length, using a D-frame net with 500-μm mesh openings (Klemm et al. 1998, Flotemersch et al. 2006b). An alternative to transects is to estimate the proportion of different habitat types in a defined reach (e.g., 100m), and distribute a fixed level of sampling effort proportional to their frequency of occurrence throughout the reach (Barbour et al. 1999, 2006). Whether using transects or proportional distribution, organic and inorganic sample material (leaf litter, small woody twigs, silt, and sand; also includes all invertebrate specimens) are composited in one or more containers, preserved with 95% denatured ethanol, and delivered to laboratories for processing (Figure 4-4). A composite sample over multiple habitats in a reach is a common protocol feature of many monitoring program throughout the US (Carter and Resh 2013), although some programs choose to keep samples and data from different habitat types segregated.

Figure 4-3. Removing a benthic macroinvertebrate sample from a sieve bucket and placing the sample material in a 1-liter container with approximately 95% ethanol preservative

Figure 4-4. Labelling benthic macroinvertebrate sample containers and recording field data

4.4.1.2 Fish

Fish sampling is designed to provide a sample that is representative of the fish community inhabiting the reach, and which assumed to reasonably represent species richness, guilds, relative abundance, size, and anomalies. The goal is to collect fish community data that will allow the calculation of an IBI and observed/expected (O/E) models. Electrofishing is the preferred method of sampling, involving the operator and (ideally) two netters, and occurs in a downstream direction at all habitats along alternating banks, over a length of 20 times the mean channel width at designated transects (USEPA 2009). Collection of a minimum of 500 fish is the target number of specimens (USEPA 2009), and in the event this is not attained, sampling will continue until 500 individuals are captured or the downstream extent of the site is reached.

4.4.1.3 Periphyton

Periphyton collections are made from shallow areas near each of the sampling locations on the 11 cross-section transects established within the sampling reach and are collected at the same time as the benthic macroinvertebrate samples (USEPA 2009). There is one composite sample of periphyton for each site, from which separate types of laboratory samples can be prepared, if necessary. The different sample type could include a) an ID/enumeration sample to determine taxonomic composition and relative abundances, b) a chlorophyll sample, c) a biomass sample (for ash-free dry mass [AFDM]), or d) an acid/alkaline phosphatase activity [APA] sample). There are potentially other analysis types that could be performed, thus requiring additional sample segregates.

4.4.1.4 Quality control measures

Other than a qualitative judgment that field personnel are adequately trained, have sufficient experience, and have been successfully audited as having completely and accurately applied the correct SOP, the quality of field sampling cannot be determined without sample processing. The consistency of field sampling is a measure of data quality quantified by precision calculations using indicator values – individual metrics, IBI scores, or predictive models - collected from adjacent stream reaches (i. e., two stream channel lengths where the second [B] begins at the endpoint of the first [A]). As a rule-of-thumb, we recommend a site duplication rate of 10 percent, where duplicate locations are randomly selected from the full sample lot, and fieldwork occurs as routine. Terms calculated from the duplicate sample results include median relative percent difference (mRPD), 90 percent confidence intervals/minimum detectable difference (CI90/DD90), and coefficient of variation (CV) (Flotemersch et al. 2006b, Stribling et al. 2008a, Stribling 2011). Depending on programmatic application, natural variability of the landscape the watershed is draining, density and distribution of potential stressor sources, number of field crews, and, of course, budgetary resources, it can be useful to stratify distribution of duplicate reaches. This will allow programmatic measurement quality objectives (MQO) to be established for objective benchmarks for acceptable quality of data. Typical MQO for field sampling precision (Stribling et al. 2008a, Stribling 2011) might be:

- mRPD<15,
- CI90<15 index points on a 100-point scale, and
- CV<10% for a sampling event

Depending on programmatic needs, values exceeding these MQO could highlight samples for more detailed scrutiny to determine causes for the exceedances, and the need for corrective actions.

4.4.2 Sample processing/laboratory analysis

For biological monitoring and assessment programs, sample processing employs procedures for organizing sample contents so that analysis is possible. For benthic macroinvertebrate and periphyton samples, those procedures are laboratory-based; however, for fish, they are performed primarily in the field (USEPA 2004, 2009).

4.4.2.1 Benthic macroinvertebrates

The three aspects of sample processing for benthic macroinvertebrate samples are a) sorting, which serves to separate the organisms from other sample material, specifically, organic detritus inorganic silt, and other materials (Figure 4-5), b) subsampling, which isolates a representative sample fraction from the whole, and c) taxonomic identification, which characterizes the (sub)sample by naming and counting individuals in it.

Figure 4-5. Examining, washing, and removing large components of sample material prior to putting in sample container

4.4.2.1.1 *Sorting and subsampling*

The sorting/subsampling procedure is based on randomly selecting portions of the sample material spread over a gridded Caton screen (Caton 1991, Barbour et al. 1999, Flotemersch et al. 2006b, Stribling 2011), and fully removing (picking) all organisms from the selected fractions. The screen is divided into 30 grid squares, each individual grid square measuring 6 cm x 6 cm, or 36 cm^2 (note that it is *not* 6 cm^2 as indicated in Figure 6-4b of Flotemersch et al. [2006b]). Prior to beginning the sorting/subsampling process, it is important that the sample is mixed thoroughly and distributed evenly across the screen to reduce the effect of organism clumping that may have occurred in the sample container. Depending on the density of organisms in the sample, multiple levels of sorting may be necessary, the purpose of which is to minimize the likelihood that the entire sample to be identified comes from a very small number of grids. Initially, four grids are randomly selected from the 6 x 5 array, removed from the screen, placed in a sorting tray, and coarsely examined. If the density of organisms is high enough that there are many more than the target number in the four selected grids (i. e., greatly exceeding by twofold or more the 100-, 200-, 300-, 500-organisms, or more, depending on the project), that material is re-spread on a second gridded screen and the process repeated (second level sort). This is repeated until it is apparent that the density of specimens will require <u>at least</u> four grids to be sorted to attain the target number (±20%). Once re-spreading is no longer needed, all organisms are removed from the four grids using forceps. If the final rough count is ±20 of the target subsample size, then subsampling is complete; if >20% less than the target subsample size, then additional, single grids of material are moved from the tray, and picked in entirety. This is repeated, one grid at a time, until within 20 percent of the target number. Following picking, the sort residue should be transferred to a separate container labeled with complete sample information, and the words “SORT RESIDUE” clearly visible. Completely record the number of sort levels and grids processed. The sorting and subsampling process should result in at least three containers: a) clean (sub)sample, b) sort residue, and c) unsorted sample remains. Container ‘a’ is provided to the taxonomist for identification and counting, ‘b’ is available for QC sort re-check, and ‘c’ is archived until all QC checks are complete. In the event of certain QC failures, it may be necessary to process portions of the unsorted remains.

Fixed count subsamples - Fixed organism counts vary among monitoring programs (Carter and Resh 2013), with 100, 200, 300 and 500 counts being most often used (Barbour et al. 1999, Cao and Hawkins 2005, Flotemersch et al. 2006a). Flotemersch et al. (2006a) concluded that a 500-organism count was most appropriate for large/nonwadeable river systems, based on examination of the relative increase in richness metric values (< 2%) between sequential 100-organism counts. However, they also suggested that 300-organism count is sufficient for most study needs. Others have recommended higher fixed counts, including a minimum of 600 for wadeable streams (Cao and Hawkins 2005). The subsample count used for the USEPA national surveys is 500 organisms (USEPA 2004); many states use 200 or 300 counts.

4.4.2.1.2 Taxonomic identification

Genus level taxonomy is the principal hierarchical level used by most routine biological monitoring programs for benthic macroinvertebrates (Carter and Resh 2013), although occasionally family level taxonomy is used. For genus level to be attained, most direct observations can be accomplished with dissecting stereomicroscopes with magnification ranges of 7-112x; however, midges (Chironomidae) and worms (Oligochaeta) need to be slide-mounted and viewed through compound microscopes that have magnification ranging 40-1500x, under oil. Slide-mounting specimens in these two groups is usually (though, not always) necessary to attain genus level nomenclature, and sometimes even more coarse level for midges (i.e., less specific). Taxonomic classification is a major potential source of error in any kind of biological monitoring data sets (Stribling et al. 2008b, Bortolus 2008) and the rates of error can be managed by specifying both hierarchical targets and counting rules. Hierarchical targets define the level of effort that should be applied to each specimen but may often not be possible for some specimens due to poor slide mounts, damaged, or their being juvenile (early instars). Further, the requirement for some taxa may be more coarse, such as genus-group, tribe, subfamily, or even family. In any case, the principal responsibility of the taxonomist is to record and report the taxa in the sample and the number of individuals of each taxon. Consistency in the nomenclature used is more important than the actual keys that are used, although, some programmatic SOPs may specify the technical literature. For example, the identification manual “*An Introduction to the Aquatic Insects of North America*” (Merritt et al. 2008) is useful for identifying the majority of aquatic insects in North America to genus level. However, because many taxonomic groups are often (correctly) under perpetual revision and updates, the nomenclatural foundation of many may have changed, thus requiring familiarity of the taxonomist with more current primary taxonomic literature. Merritt et al. (2008) is not applicable to non-insect macroinvertebrate taxa that are often captured in routine sampling, including Oligochaeta, Mollusca, Acari, Crustacea, Platyhelminthes, and others; exhaustive lists of literature for all invertebrate groups are provided by Klemm et al. (1990) and Thorp and Covich (2010). Identification staff may also need information on accepted nomenclature, including validity, authorship, and spelling, all of which could be found in the integrated taxonomic information system (ITIS; http://www.itis.gov/). Although it is a nomenclatural clearinghouse, it should be recognized that it is not completely current for all taxa potentially requiring independent confirmation.

It should be noted that some volunteer monitoring programs, such as the Izaak Walton League (IWL) and the Maryland DNR Stream Waders Program (MDNR), use simpler taxonomic procedures. The IWL uses field identification of a small number of organisms that are of a limited number of kinds, like mayflies, stoneflies, caddisflies, beetles, and mollusks, and just note their presence or absence. The MDNR Stream Waders program, including metrics and index, are based on family level data.

4.4.2.2 Fish (field taxonomic identification)

Identification and processing of fish occur at the completion of each transect (USEPA 2009), where the data recorded include species names, number of individuals of each, length, and DELT anomalies (**D**eformities, **E**roded fins, **L**esions and **T**umors). Taxonomic identification and processing should only be completed on specimens >25 mm total length and by qualified staff. Common names of species should follow those established under the American Fisheries Society's publication, "Common and Scientific Names of Fishes from the United States, Canada and Mexico" (Nelson et al. 2004). Species not positively identified in the field should be separately retained for laboratory identification (up to 20 individuals per species). For programs not using the transect method of sample reach layout, electrofishing will cover all habitat throughout the reach. Further, fish sample vouchers are developed for a minimum of 10% of the sites sampled (USEPA 2012).

4.4.2.3 Periphyton

Two activities making up sample processing for periphyton are further segregated into those for a) soft-bodied algal forms and b) diatoms. Although methods for both are presented by USEPA (2012), diatom procedures are based principally on those of the US Geological Survey National Water Quality Assessment Program (USGS/NAWQA) (Charles et al. 2002). Microscopic diatoms encountered are identified (to lowest possible taxon level), enumerated and recorded. Estimates of the biovolume of dominant species are made using existing parameters, or those found in the literature, and used to determine the biovolume of the sample. Detailed information on the different procedures, especially on the analytical approaches for soft algae using the Sedgewick-Rafter and extended Palmer-Maloney count techniques, can be found in USEPA (2012) and Charles et al. (2002).

4.4.2.4 Quality control measures/data quality documentation

Quality control (QC) for sample processing for these three taxonomic groups is, in some respects, similar, but in others, different. Some of the similarities are that several aspects quality evaluations are based on repeating processes; specifically, duplicating field samples, or repeating of sample processing activities (sorting, identification, and counting). Differences arise out of the fact that there are not always analogous methods for dealing with the different organism types. Specifically, fish are identified in the field, whereas, benthic invertebrates and algae/diatoms are laboratory-identified. Logistical constraints prevent whole-sample re-identification of fish, whereas it is easily done for the other groups. And, subsampling is not done with fish samples, where it is explicitly done for benthic invertebrates, and functionally done for algal/diatom samples.

Sorting QC (benthic macroinvertebrates [only]). – Sorting QC is accomplished through rechecking the sample sort residue from 10% of the samples, randomly selected, and calculating the term 'percent sorting efficiency' (PSE) (Stribling 2011, Flotemersch et al. 2006b). This value reports the number of specimens missed during primary sorting as a proportion of the original number of specimens found. A typical MQO for this is PSE>90%, with the goal of minimizing the number of samples that fail. Individual programs must specify what is acceptable, but generally, the goal should be to have <10% of the samples fail. It is a measure of bias associated with sample sorting.

Taxonomic QC. – As a measure of precision in taxonomic identifications, consistency is quantified by independent re-identification of whole (sub)samples, where those samples are randomly selected (10%, as a rule-of-thumb) from the full sample lot. Sample results from the QC taxonomist are directly compared to those of the primary taxonomist, and differences quantified as 'percent taxonomic disagreement' (PTD) for identifications, as 'percent difference in enumeration' (PDE) for counts, and as 'absolute difference in percent taxonomic completeness' (*abs*[PTC]). Typical MQO for these are PTD<15%, PDE<5%, and *abs*PTC<5% (Stribling 2011).

If MQO thresholds are routinely or broadly exceeded, samples failing should be examined in more detail to determine causes of the problem, and what corrective actions may be necessary.

4.4.3 Data reduction/indicator calculation

Once necessary corrective actions for sample processing and taxonomic identifications have been implemented and effectiveness confirmed, data quality is known and acceptable, sample data are converted into the primary terms to be used for analysis. As stated above, monitoring practitioners usually have access to published MMI for application to sample data, as well as sometimes predictive models and established decision analysis systems. Indicators most often take the form of a multimetric Index of Biological Integrity (IBI; Karr et al. 1986, Hughes et al. 1998, Barbour et al. 1999, Hill et al. 2000, 2003) or a predictive observed/expected (O/E) model based on the River Invertebrate Prediction and Classification System (RIVPACS; Clarke et al. 1996, 2003, Hawkins et al. 2000b, Hawkins 2006). The Illinois Department of Natural Resources used the Macroinvertebrate Biotic Index (MBI) in their analysis of restoration effectiveness in the Waukegan River (see Case Study 4).

4.4.3.1 Multimetric indexes

The purpose of any MMI is to summarize complex biological and environmental information into a form and format that can be used for management decision-making (Karr and Dudley 1981, Karr 1991, Angermeier and Karr 1994), and doing so in a manner that allows uncertainty associated with those decisions to be known and communicated. Index calibration is the empirical process of determining which measures are best suited for that purpose, specifically in terms of their capacity for detecting biological changes in response to environmental variables of concern (pollutants, or stressors). For the purpose of this guidance it is assumed that the calibration procedure (Hughes et al. 1998, Barbour et al. 1999, McCormick et al. 2001) has been completed, and that an MMI is available to monitoring practitioners for application. The reader should be aware that no attempt is made to be comprehensive in discussing either metric diversity or index formulation, or of reviewing supporting technical literature. As such, only selected examples are used below to illustrate different aspects of MMI application.

Case Study 4: Biological and Physical Monitoring of Waukegan River Restoration Efforts, Lake County, Illinois

- ✓ Urbanized watershed
- ✓ Severely degraded stream habitat; channel instability/bank erosion; high velocity runoff
- ✓ Bank stabilization using LUNKERS and riparian re-vegetation
- ✓ Grade control using rock weirs and artificial riffles
- ✓ Benthic macroinvertebrate assemblage monitoring
- ✓ Effectiveness evaluation

The Waukegan River watershed is located on the western shore of Lake Michigan, about 56 km (35 mi) north of Chicago in Lake County (Figure CS4-1). It is approximately 20 km (12.4 mi) long and has a drainage area of 2,994 ha (7,397 ac). The river channel drops from an approximate 222 m (730 ft) headwaters elevation to around 177 m (580 ft) above sea level, before discharging directly into Lake Michigan through Waukegan Harbor on its western shore. The Waukegan River watershed receives a mean annual precipitation of 834 millimeters (mm) (32.8 in) and has a mean annual temperature of 8.8°C (47.8°F). Historical records (circa 1840) indicate substantial marshes in the area, and recent soils studies indicate that wetlands covered approximately 15 percent of the watershed.

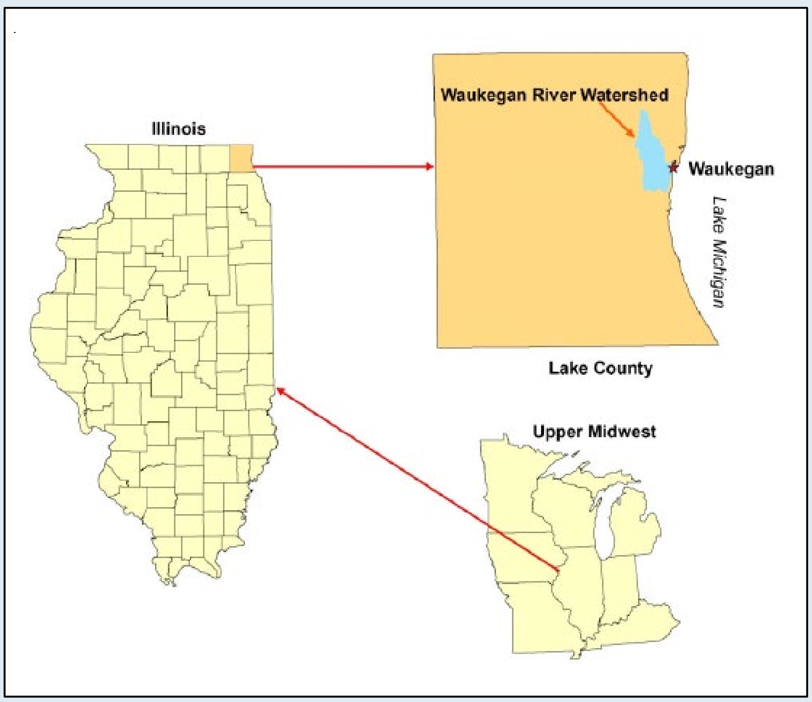

Figure CS4-1. The Waukegan River watershed in northeastern Illinois (White et al. 2010)

The north and south branches of the basin, including the mainstem to Waukegan Harbor, comprise approximately 20 channel km (12.5 mi), excluding Yeoman Creek from the north. The mean channel width ranges from 4.4 to 6.7 m (14.6 – 21.9 ft) and has a mean depth from 0.07 – 0.28 m (0.23 – 0.92 ft) (White et al. 2003). There is more substantial shading from riparian vegetation in the North Branch subwatershed than in the south. The South Branch has a greater discharge than the North Branch, approximately 0.1 cubic meters per second (cms) (3.4 cfs) versus 0.01 cms (0.4 cfs). Dominant substrate types range from sand to large cobble and boulder, with some bedrock. Within the project area, there was one control monitoring site (S2) and three sites where stream restoration was carried out (S1, N1, N2) (Figure CS4-2).

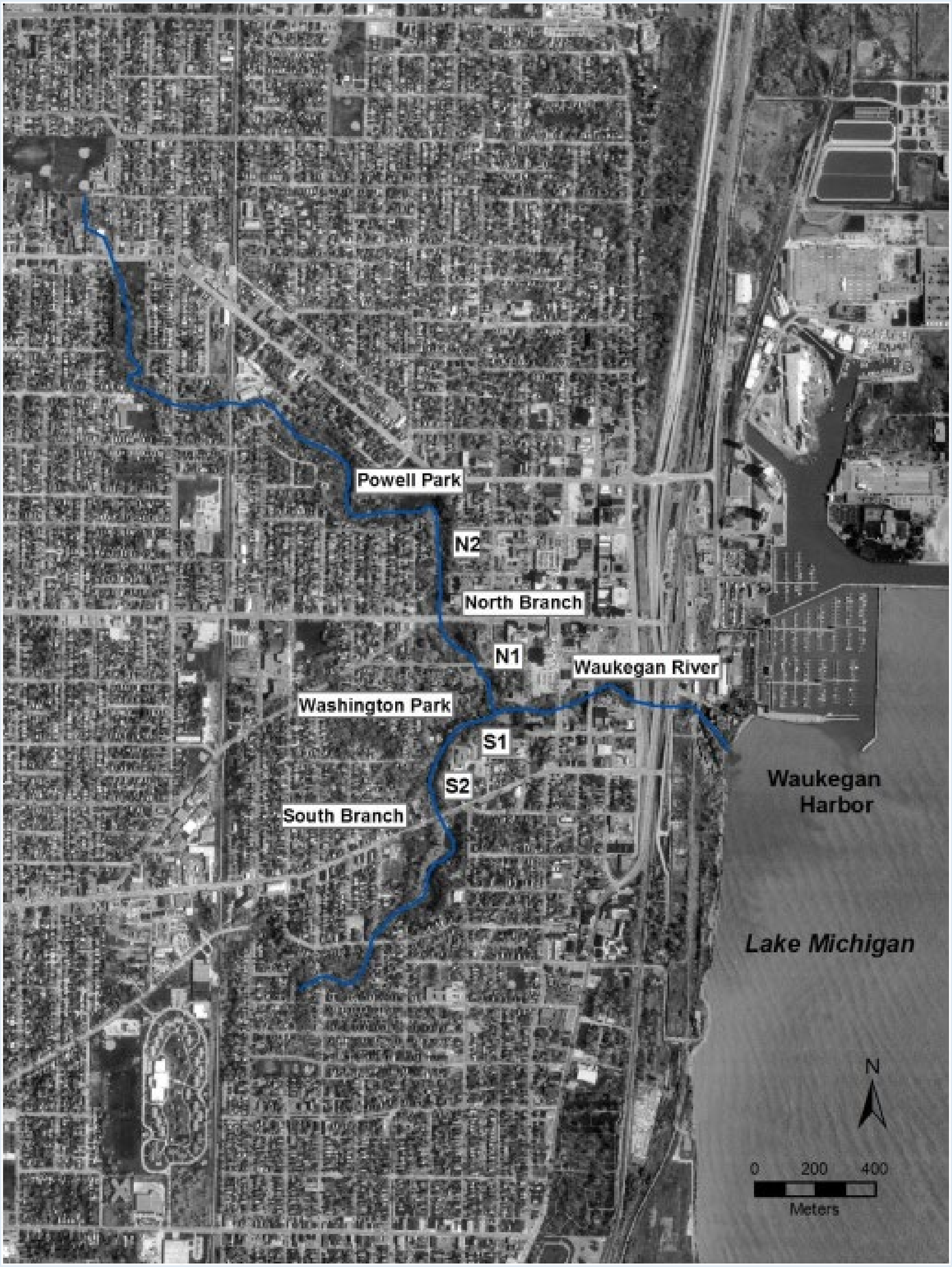

Figure CS4-2. Aerial map taken in 1998 of a portion of the Waukegan River watershed, showing sampling locations and restoration project areas (White et al. 2010)

As of 2005, major land uses in the Waukegan River watershed included approximately 36.7 percent residential; 21.5 percent transportation; 12 percent commercial, retail, government and institutional; and about 20 percent open space, forest, grasslands, and beaches. The remaining land uses are associated with small amounts of disturbed lands (3.6 percent), industrial (2.8 percent), wetlands/water (2.0 percent), and communication/ utilities (1.7 percent) (Lake County SWMC 2008). Approximately 80 percent of the densely urbanized City of Waukegan (2014 population ~89,000) lies within the watershed.

The pace of development and sprawl of Waukegan was substantial throughout the latter part of the 19th century, and until the 1970s and 1980s, when current stormwater regulations began to take effect. Not surprisingly, streams were heavily impacted by urbanization. Minimal management of the stormwater quantity and quality led to flashy stormwater runoff conditions and elevated pollutant loads. Additional water resource issues associated with urban development included combined sewer overflows (CSO), stream channel instability (accelerated vertical and lateral erosion processes), nutrient enrichment, and contamination by metals, pesticides/herbicides, pharmaceuticals and personal care products (PPCPs), and endocrine disruptors (White et al. 2003, 2010). Channel erosion processes were accelerated by flashy stormflows, contributing to degraded physical habitat and decreased capacity of the stream to support the survival and reproduction of stream biota.

Monitoring and Sampling Design

The overall restoration goal in the North Branch and South Branch was to rehabilitate physical habitat and hydrologic conditions to support recovery of benthic macroinvertebrate and fish assemblages (White et al. 2010). Project leaders chose to design and install biotechnical stabilization, a combination of stream bank physical stabilization and riparian re-vegetation, to address the severe channel instability and erosion problems in Washington Park and Powell Park (Figure CS4-2). The objectives of these techniques are complementary. A stream channel experiencing severe lateral and vertical erosion (mass-wasting and down-cutting, respectively), by definition, is losing habitat suitable for stream biota. Damaged or missing riparian vegetation results in diminished root mass to hold soil together, lowered inputs of leaf litter and woody materials (food source and habitat structure), and less shading, which can lead to warmer water and increased photosynthetic activity and algal growth. Combinations of LUNKERS[1], a-jacks, stone, coconut fiber rolls, dogwoods, willows, and grasses were installed at selected locations on the North Branch/Powell Park (1992–93) and on the South Branch/Washington Park (1995).

There were four sample locations, two each on the South Branch and the North Branch (Figure CS4-2). For the South Branch, station S2 was the upstream control reach, and S1 was the downstream treatment reach. On the North Branch, the two sample locations (N1 and N2) were located to coincide with treatment reaches. Wooden LUNKERS were used as the principal rehabilitation feature at N1, while recycled plastic lumber and concrete a-jacks were used for LUNKERS construction at N2. All four locations were sampled annually during spring, summer, and fall over a 13-year period (1994–2006).

This case study focuses solely on responses of benthic macroinvertebrates, although it should be recognized that fish were also evaluated for both branches. Benthic macroinvertebrates, physical habitat, and chemical water quality were sampled, measured, and characterized at each stream location. Three macroinvertebrate samples were taken at each location using a Hess sampler with

[1] 'Little Underwater Neighborhood Keepers Encompassing Rheotaxic Salmonids' (Vetrano 1988)

a 500 micron mesh net. Sample material was preserved in 95 percent ethanol; organisms were sorted to segregate individuals from non-target material, and subsequently identified to genus level.

Sampling data were used to calculate the Macroinvertebrate Biotic Index (MBI). The MBI is based on the Hilsenhoff Biotic Index (HBI; Hilsenhoff 1977, 1982) but uses an 11-point scale, rather than the HBI 5-point scale. Lower MBI scores indicate better or less-degraded water quality. Physical habitat was characterized using the Potential Index of Biological Integrity (PIBI) which incorporates percent substrate particle sizes, magnitude of sediment deposition, pool substrate quality, and substrate stability, a series of hydraulic and morphometric measures, riparian features, and various aspects of instream cover (White et al. 2003). The PIBI was calculated from measurements and field observations made on each of 10 equal-length segments established by the 11-transect method. The purpose of the PIBI is to help illuminate which habitat features, if any, might be limiting survival, growth, and reproduction of stream biota.

Results

Overall, among all four locations, MBI scores ranged from around 5 to just below 10 (good to very poor), with an average around 7.2, or “fair” (see Figure CS4-3). On the South Branch, station S2 exhibited the highest mean score for a single year, 7.5, indicating “fair” stream condition, slightly better than “poor.” MBI scores indicate worsening conditions over time at stations S1 and N1. There was virtually no change over the 13 years for stations S2 and N2, with average MBI scores in the “fair” and “poor” ranges.

All sites were dominated by stressor tolerant taxa, with sample data comprised of 82-89 percent non-biting midges (Insecta: Diptera: Chironomidae), segmented worms (Annelida: Oligochaeta), and aquatic sowbugs (Crustacea: Isopoda: Asellidae). The dominance of these animals in the North and South branches clearly shows stressed or degraded conditions before, during, and after any kind of habitat restoration or other remedial activities. Mean taxa richness (number of distinct taxa) over the sampling period was 10 for site S1, and 8 for the other three locations. Ninety-two percent of the samples fell in the “poor” or “very poor” categories. There were very low numbers of stressor-sensitive EPT taxa (mayflies [Ephemeroptera], stoneflies [Plecoptera], and caddisflies [Trichoptera]) throughout the monitoring period. This is generally indicative of elevated pollutant levels and greater degradation. Some improvement in physical habitat quality was observed at treatment stations S1 and N1, likely due to improvements in bank stability and decreases in overall proportions of percent fines, silt, and mud. The other treatment station, N2, which was bank-armored, remained relatively consistent over the full period of record, as did the non-treatment control, S2.

Improvement in physical habitat quality and overall biological diversity was achieved as a result of these restoration activities, but improvement in biodiversity, primarily relative to the fish assemblage (not discussed in this case study) was only temporary (White et al. 2010). The authors acknowledged that sustainable biological diversity in a damaged watershed will require more complete understanding of landscape and watershed processes, their degree of degradation, and a comprehensive approach to conservation that addresses the system in its entirety. In the case of the Waukegan River watershed, this calls for a systematic approach to correcting other sources of hydrologic and chemical water quality stressors associated with water and sewer management operations, channel and flow alterations, and extensive aquifer drawdown. One result of this project was initiation of a comprehensive watershed plan, selection of a coordinator, development of stakeholder and technical planning committees, and creation of a long-term action plan.

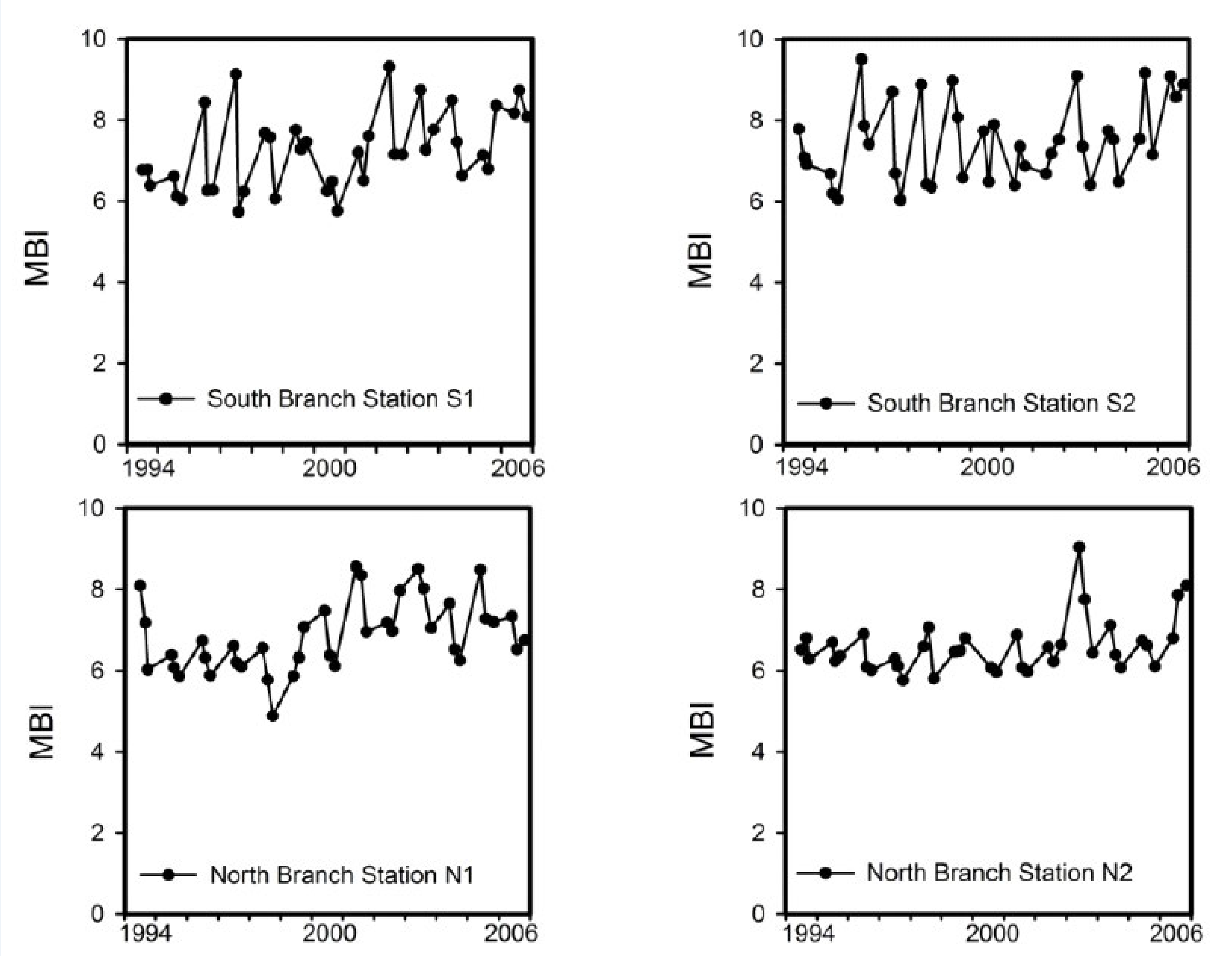

Figure CS4-3. MBI scores from monitoring stations in Waukegan River (White et al. 2010). Assessment classes (narrative ratings) for stream condition based on MBI scores are: very poor, 9.0-11.0; poor, 7.6-8.9; fair, 6.0-7.5; good, 5.0-5.9; very good, <5.0.

Literature

Hilsenhoff, W.L. 1977. *Use of Arthropods to Evaluate Water Quality of Streams.* Technical Bulletin 100. Wisconsin Department of Natural Resources. Madison, WI.

Hilsenhoff, W.L. 1982. *Using a Biotic Index to Evaluate Water Quality in Streams.* Technical Bulletin 132. Wisconsin Department of Natural Resources. Madison, WI.

Lake County SWMC. 2008. *Waukegan River Watershed, Lake County, Illinois. Lake County Stormwater Management Commission, Lake County Department of Information and Technology, GIS & Mapping Division. Map*. Accessed February 9, 2016.

Vetrano, D.H. 1988. *Unit Construction for Trout Habitat Improvement Structures for Wisconsin Coulee Streams.* Administrative Report No. 27. Wisconsin Department of Natural Resources, Madison, WI.

White, W.P., J.D. Beardsley, J.A. Rodsater, and L.T. Duong. 2003. *Biological and Physical Monitoring of Waukegan River Restoration Efforts in Biotechnical Bank Protection and Pool/Riffle Creation.* National Watershed Monitoring Project. Annual Report. *Prepared by:* Watershed Science Section, Illinois State Water Survey, Illinois Department of Natural Resources. *Prepared for:* Illinois Environmental Protection Agency and U.S. Environmental Protection Agency (Region 5).

White, W.P., J. Beardsley, and S. Tomkins. 2010. *Waukegan River, Illinois national nonpoint source monitoring program project. NWQEP NOTES. The NCSU Water Quality Group Newsletter*. No. 133. April 2010. ISSN 1062-9149. North Carolina State University Water Quality Group, Campus Box 7637, North Carolina State University, Raleigh, NC 27695-7637. Accessed February 9, 2016. http://www.ncsu.edu/waterquality/.

4.4.3.1.1 *Metric and index calculations*

Metrics are mathematical terms calculated directly from sample data, with resulting values scored relative to quantitative criteria. The Table 4-5 below presents an example of four different metric sets representing site classes (or bioregions) within a particular US State. Each set of either 5 or 6 metrics forms the basis of an MMI previously calibrated to wadeable streams of the class. Metric values result from direct calculations on raw sample data, taxonomic identifications and counts (list of taxa and number of individuals of each, by sample).

Table 4-4. Metrics and associated scoring formulas for four site classes from an example monitoring and assessment program

Metrics	Scoring formulas
Site class A	
1. Total taxa	100*(*metric value*)/51.5
2. Percent EPT individuals, as sensitive	100*(*metric value*)/39
3. Percent Coleoptera individuals, as sensitive	100*(*metric value*)/10.5
4. Beck's biotic index	100*(*metric value*)/31
5. Percent of taxa, as tolerant	100*(43-[*metric value*])/40
Site class B	
1. Total number of taxa	100*(*metric value*)/51.5
2. Number of EPT taxa	100*(*metric value*)/14
3. Percent individuals *Cricotopus/Orthocladius* + *Chironomus*, of total Chironomidae	100*(45-[*metric value*])/45
4. Percent EPT individuals, as sensitive	100*(*metric value*)/39
5. Number of taxa, as shredders	100*(*metric value*)/7
6. Hilsenhoff Biotic Index	100*(8.5-[*metric value*])/5
Site class C	
1. Total taxa	100*(*metric value*)/51.5)
2. Percent of taxa, as non-insects	100*(46-[*metric value*])/40
3. Percent individuals *Cricotopus/Orthocladius* + *Chironomus*, of total Chironomidae	100*(24-[*metric value*])/24
4. Percent of individuals, as filterers	100*(*metric value*)/70
5. Number of taxa, as sprawlers	100*(*metric value*)/14
6. Hilsenhoff Biotic Index	100*(8.5-[*metric value*])/5
Site class D	
1. Number of Oligochaeta taxa	100*(6-(*metric value*])/6
2. Percent EPT individuals, as sensitive	100*(*metric value*)/15
3. Percent individuals, as Crustacea and Mollusca	100*(*metric value*)/30
4. Percent individuals, as Odonata	100*(16.5-[*metric value*])/16.5
5. Number of taxa, as collectors	100*(20-[*metric value*])/19.5
6 Percent individuals, as swimmers	100*(12-[*metric value*])/12

Many metrics require assigning different characteristics or traits to taxa in the dataset prior to calculation. These characteristics include functional feeding groups (FFGs), habit, and stressor tolerance values. Many states use various literature resources to develop traits databases (e.g., Merritt et al. 2008, Barbour et al. 1999, Vieira et al. 2006, Carter and Resh 2013).

The formulas in the table resulted from the calibration process, and serve to convert the calculated metric value to a normalized, unitless score on a 100-point scale. The multiple metric values are then combined for each sample by simple averaging. Additionally, formulas are developed, in part, so that individual metrics are scored depending on their direction of change in the presence of stressors.

4.4.3.1.2 *Quality control measure*

Metric calculations are typically performed in spreadsheets or relational databases with embedded queries. To ensure that resulting calculations are correct and provide the intended metric values, a subset of them should be recalculated by hand. A reliable approach is to calculate a) one metric across all samples, followed by b) all metrics for one sample. When recalculated values differ from those values in the output matrix, reasons for the disagreement are determined and corrections are made. Reports on performance include the total number of reduced values as a percentage of the total, how many errors were found in the queries, and the corrective actions specifically documented.

4.4.3.2 Predictive models (observed/expected [O/E])

Predictive models are based on the premise that the taxa occurring in a minimally disturbed system can be predicted based on multiple measures of the environmental setting and that if the predicted taxa are not observed in an evaluation site, then disturbance can be suspected. The ratio of the number of observed taxa to that expected to occur in the absence of human-caused stress is an intuitive and ecologically meaningful measure of biological integrity. Low observed-to-expected ratios (O/E << 1.0) imply that test sites are adversely affected by some environmental stressor. The models are commonly called RIVPACS models (River Invertebrate Prediction And Classification System [Wright 1995]) based on observed:expected taxa (Clarke et al. 1996, 2003, Hawkins et al. 2000b). Because they are based on taxa in reference sites, the predictive models are not well suited to assemblages with naturally low diversity (as in oligotrophic fish communities). The loss of reference taxa is difficult to detect when only few taxa are expected.

The number of taxa expected at a site is calculated as the sum of individual probabilities of capture for all taxa found in reference sites in the region of interest. All probabilities greater than a designated threshold are summed to calculate the expected number of taxa (E), and this number is compared to the reference taxa observed (O) at a site. Because these models predict the actual taxonomic composition of a site, they also provide information about the presence or absence of specific taxa. If the sensitivities of taxa to different stressors are known, this information can lead to derived indices and diagnoses of the stressors most likely affecting a site. In addition, taxa can be identified as increasers or decreasers with respect to general environmental stress encountered in the model development data set. A variation of the O/E models measures the Bray-Curtis compositional dissimilarity between an observed and expected assemblage directly, which detects stress-induced shifts in taxonomic composition that leave assemblage richness unchanged (Van Sickle 2008).

The steps that go into building a predictive model include 1) classifying reference sites into biologically similar groups, 2) creating discriminant functions models to estimate group membership of sites from environmental data, 3) establishing taxon-specific probabilities of capture for individual sites,

4) identifying taxa expected and comparing to those observed, 5) estimating model error, and 6) applying the model in test sites. These steps can be automated to allow exploration of model performance for multiple subsets of environmental predictor variables.

4.4.3.3 Quantitative decision analysis systems (biological condition gradient [BCG])

The Biological Condition Gradient (BCG) was described by USEPA (2011) as "a conceptual model that describes how biological attributes of aquatic ecosystems might change along a gradient of increasing anthropogenic stress." The model can serve as a template for organizing field data (biological, chemical, physical, landscape) at an ecoregional, basin, watershed, or stream segment level. The BCG was developed by EPA and other agencies to support tiered aquatic life uses in state water quality standards and criteria. It was developed through a series of workshops, and described fully by Davies and Jackson (2006). The BCG includes a narrative description of ecological condition that can be translated across regions, assemblages, and assessment programs. The descriptions recognize six levels of quality in ecological condition ranging from "1" (most desirable) where natural structural, functional, and taxonomic integrity is preserved to "6" (least desirable) in which there are extreme changes in structure and ecosystem function and wholesale changes in taxonomic composition.

The quantitative decision analysis systems approach explicitly uses the BCG as a scale for biological assessment. It differs from the multimetric and predictive model approaches in that it is not dependent on definition of reference sites (although that can be useful), and development relies on consensus of experts instead of an individual or a few analysts. It is similar to the multimetric approach in its reliance on distinct site classes. It is similar to the predictive modeling approach in its examination of individual taxa (though metrics are also incorporated in the models).

Calibrating a BCG to local conditions begins with the assembly and analysis of biological monitoring data. Following data assembly, a calibration workshop is held in which experts familiar with local biotic assemblages of the region review the data and the general descriptions of each of the BCG levels. The expert panel then uses the data to define the ecological attributes of taxa, and to develop narrative statements of BCG levels based on sample taxa lists. The expert panel is usually convened multiple times to refine decisions, to react to interim results, and to assign BCG levels to new sites. The steps typically taken during a calibration workshop include the following:

1) An overview presentation of the BCG and the process for calibration;
2) A "warm-up" data exercise to further familiarize participants with the process;
3) Assignment of taxa to BCG taxonomic attributes (based on known tolerance and rarity);
4) Description of biota in undisturbed conditions (best professional judgment [BPJ]; regardless of whether such conditions still exist in observed reference sites);
5) Assignment of sites in the data set to BCG levels; and
6) Elicitation of rules used by participants in assigning sites to levels.

Documentation of expert opinion in assigning attributes to taxa and BCG levels to sites is a critical part of the process. Facilitators elicit from participants sets of operational rules for assigning levels to sites. As the panel assigns example sites to BCG levels, the members are polled on the critical information and criteria they used to make their decisions. These form preliminary, narrative rules that explained how panel members make decisions. Rule development requires discussion and documentation of BCG level assignment decisions and the reasoning behind the decisions. During these discussions, records are kept on each participant's decision ("vote") for the site; the critical or most important information for the decision; and any confounding or conflicting information and how this is resolved.

A decision model is then developed that encompasses the taxa attributes and quantitatively replicates the rules used by the expert panel in assigning BCG levels to sites. The decision model is tested with independent data sets as a validation step. A quantitative biological assessment program can then be developed using the rule-based model for consistent decision-making in water quality management.

The decision analysis models can be based on mathematical fuzzy-set theory (citation) to replicate the expert panel decisions. Such models explicitly use linguistic rules or logic statements, e.g., "If taxa richness is high, then condition is good" for quantitative, computerized decisions. The models can usually be calibrated to closely match panel decisions in most cases, where "closely matched" means the model either exactly matched the panel, or selected the panel's minority decision as its level of greatest membership. The decision analysis models can also be cross calibrated to other assessment tools, such as the MMI. Models can be developed as spreadsheet tools to facilitate programmatic application.

4.4.4 Index scoring and site assessment

The site-specific MMI score, as calculated above in section 4.4.3, is compared to degradation thresholds (Table 4-6) to determine whether biological degradation exists relative to minimally degraded reference conditions (Barbour et al. 1999, Stoddard et al. 2006). The range of potential scores in Table 4-6 is 0 (most degraded) to 100 (least degraded). The 90 percent confidence intervals (CI90) are calculated using sample repeats (see section 4.4.2.4). Defining the numeric values of degradation thresholds is an integral phase of index calibration, and is affected by regional and climatic conditions, along with the overall level and consistency of landscape alteration and available data to characterize the broad range of degradation.

Table 4-5. Degradation thresholds to which MMI score are compared for determination of status

Site class	Degradation threshold
A	52.3
B	65.7
C	66.0
D	55.9

The confidence interval (CI), also known as detectable difference (DD) (Stark 1993), is associated with individual MMI scores and represents the magnitude of separation between two values before they can be considered truly different (Stribling et al. 2008a). Reported values falling below the threshold are considered degraded, those above are non-degraded, while site index values falling near a threshold may require additional samples to determine final rating category (Stribling et al. 2008a, Zuellig et al. 2012). Some programs, if not most, also subdivide value ranges above and below the degradation threshold to allow communication of multiple levels of non-degradation and degradation, e.g., very good, good, fair, poor, or very poor.

4.4.5 Reporting assessment results at multiple spatial scales

Depending on the monitoring design, it is possible to use assessment results for several purposes, some of which may have been previously unanticipated. For example, a probability-based design provides assessments that can be aggregated for assessments broader than the individual location from which a sample was taken (Larsen 1997, Urquhart et al. 1998). Simultaneously, each sample from that kind of design provides information useful for interpreting conditions at individual sample locations. Assessments from a targeted design provide information about the sites sampled, and although they cannot be used for broader scale assessments, can assist with confirming the effects of known stressors and stressor sources.

4.4.5.1 Watershed or area-wide

For programs using a stratified random monitoring design, a simple inference model similar to that described by Olsen and Peck (2008) and Van Sickle and Paulsen (2008), can be used to estimate the number of degraded stream miles (D) for a watershed or area-wide region with the formula:

$$D = (N/T)x\ L$$

where:

> N is the number of sites rated by the MMI as degraded,
>
> T is the total number of sites assessed for the sampling unit (subwatershed or watershed group), and
>
> L is the total number of stream miles in the sampling unit.

Total stream channel miles (L) should be estimated with GIS using the National Hydrography Dataset (NHD), or other stream data layer appropriate to the watershed or region of interest. Note that replicate samples taken for QC purposes are not included in these calculations. Results can also be presented as percent degradation (%D) by using the calculation:

$$\%D = (N/T)x100$$

For the Lake Allatoona/Upper Etowah River watershed, site selection and monitoring was stratified by the 53 HUC subwatersheds, and cumulative assessments showed distinctive patterns of degradation (Figure 4-6). More intensive development and imperviousness are closer to transportation corridors.

Trends in %D over time can be evaluated using test such as the Kendall tau test (Helsel and Hirsch 2002). It should be noted that for very small sample sizes (i.e., 3 or 4), all values would need to be consecutively decreasing to reject a one-sided null hypothesis with a p equal to 0.167 and 0.042, respectively.

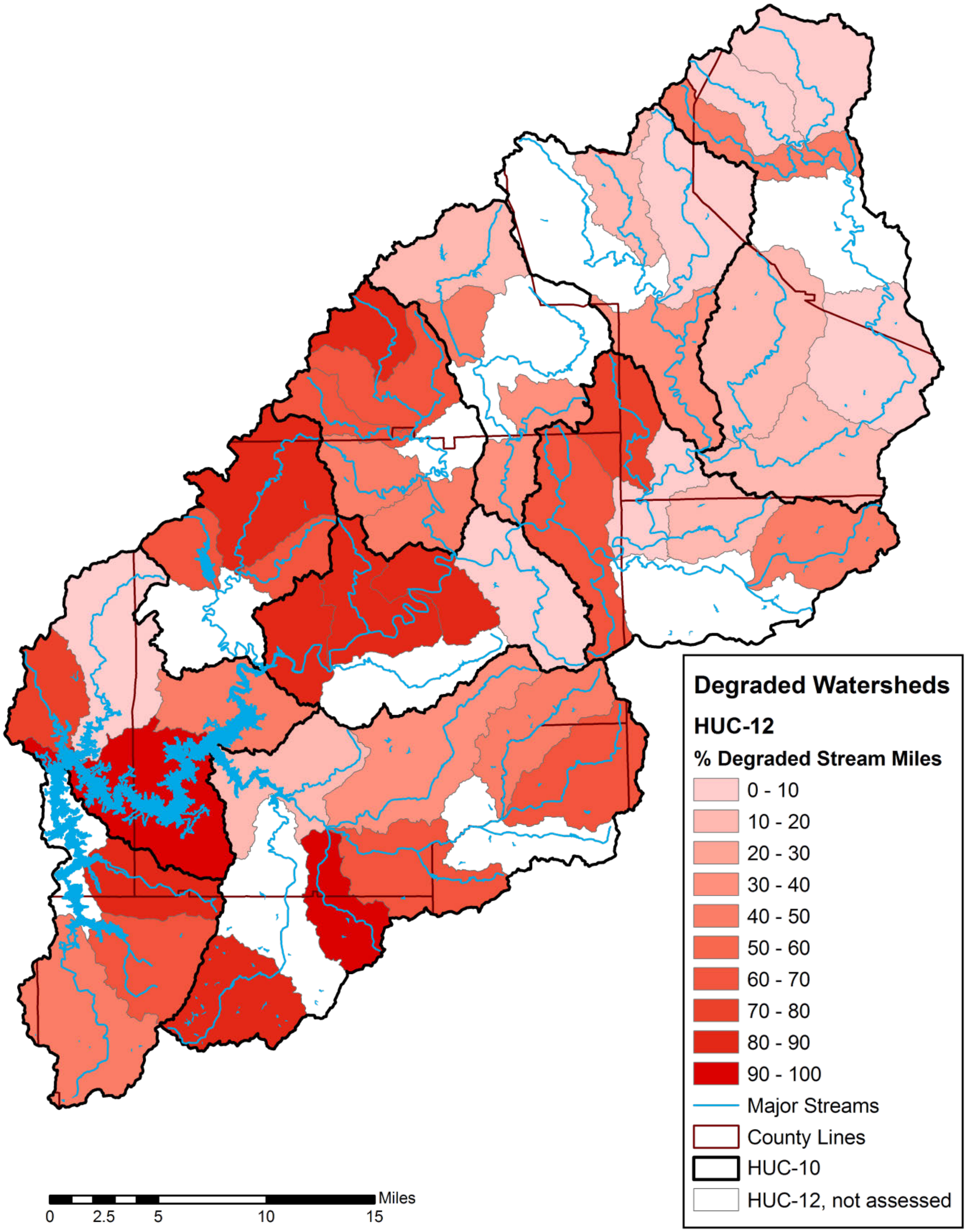

Figure 4-6. Percent degradation of subwatersheds as measured by biological monitoring and assessment, Lake Allatoona/Upper Etowah River watershed (Millard et al. 2011)

4.4.5.2 Stream- or site-specific

The MMI scores and status ratings are approaches useful for summarizing and communicating site specific conditions; this is the application for which they are designed, and for which they are ideally suited. However, a well-organized and functional database allows the index to be disaggregated where individual metrics and even taxa can be evaluated by biologists to help determine those which are most influencing an assessment. Presenting site-specific point assessments from the previous example (Figure 4-7) shows the specific distribution of the most- and least-degraded streams, and more detailed examination can begin to reveal proximity of potential stressor sources (Figure 4-8). At this stage of evaluating watershed-based stream assessments, if necessary, the assessor can turn to the USEPA stressor identification process, also known as "The Causal Analysis/Diagnosis Decision Information System", or CADDIS (http://www.epa.gov/caddis/), to assist in determining the most probable causes of biological degradation. It is using this process, including evaluating the relative dominance of the various taxa that taxon-specific environmental requirements, stressor tolerances, feeding types, and habits, which can lead to more defensible decisions on stressor control actions, such as BMPs or stream/watershed restoration activities. MMI confidence intervals can be computed and used for point comparisons in the same manner as other water quality variables (see section 7.3).

4.4.5.3 Relative to specific sources

Monitoring objectives requiring documentation of instream biological condition relative to a specific and known source of stressors require that sample data be drawn from one or more locations exposed to those stressors. In particular, if the source is an area of specific land use, a type of BMP, or a point source, confidence in the result will likely be enhanced by a thorough and quantitative description of the source. However, it should be recognized that lack of a clear site-specific response to either measured or assumed exposure to stressors does not mean that the biota are non-responsive. Neither does it mean that the BMP is ineffective. The BMP can likely be proven effective at reducing the single or multiple stressors for which it is intended and designed.

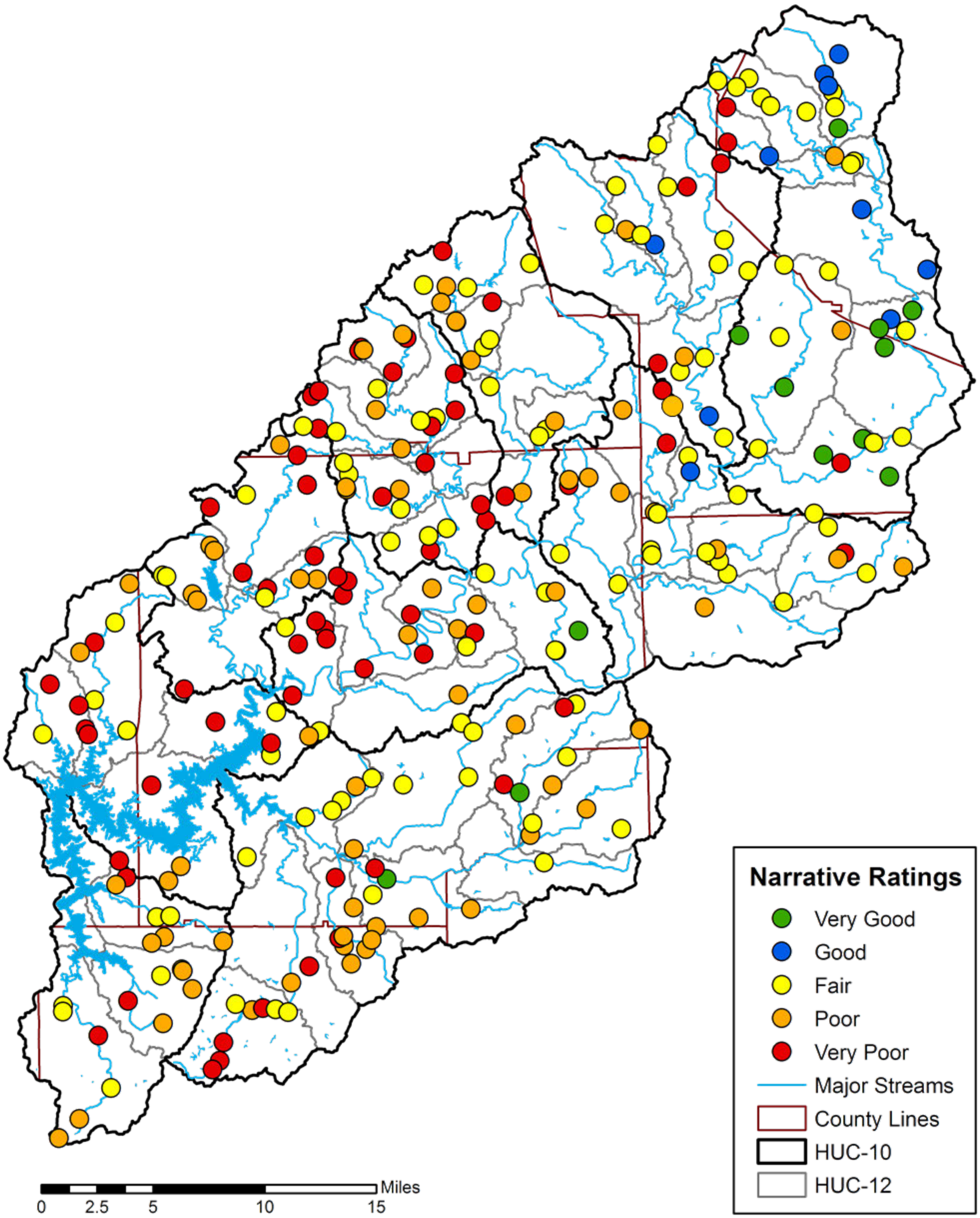

Figure 4-7. Distribution of stream biological assessments in the Lake Allatoona/Upper Etowah River watershed, using a benthic MMI developed by the Georgia Environmental Protection Division (Millard et al. 2011)

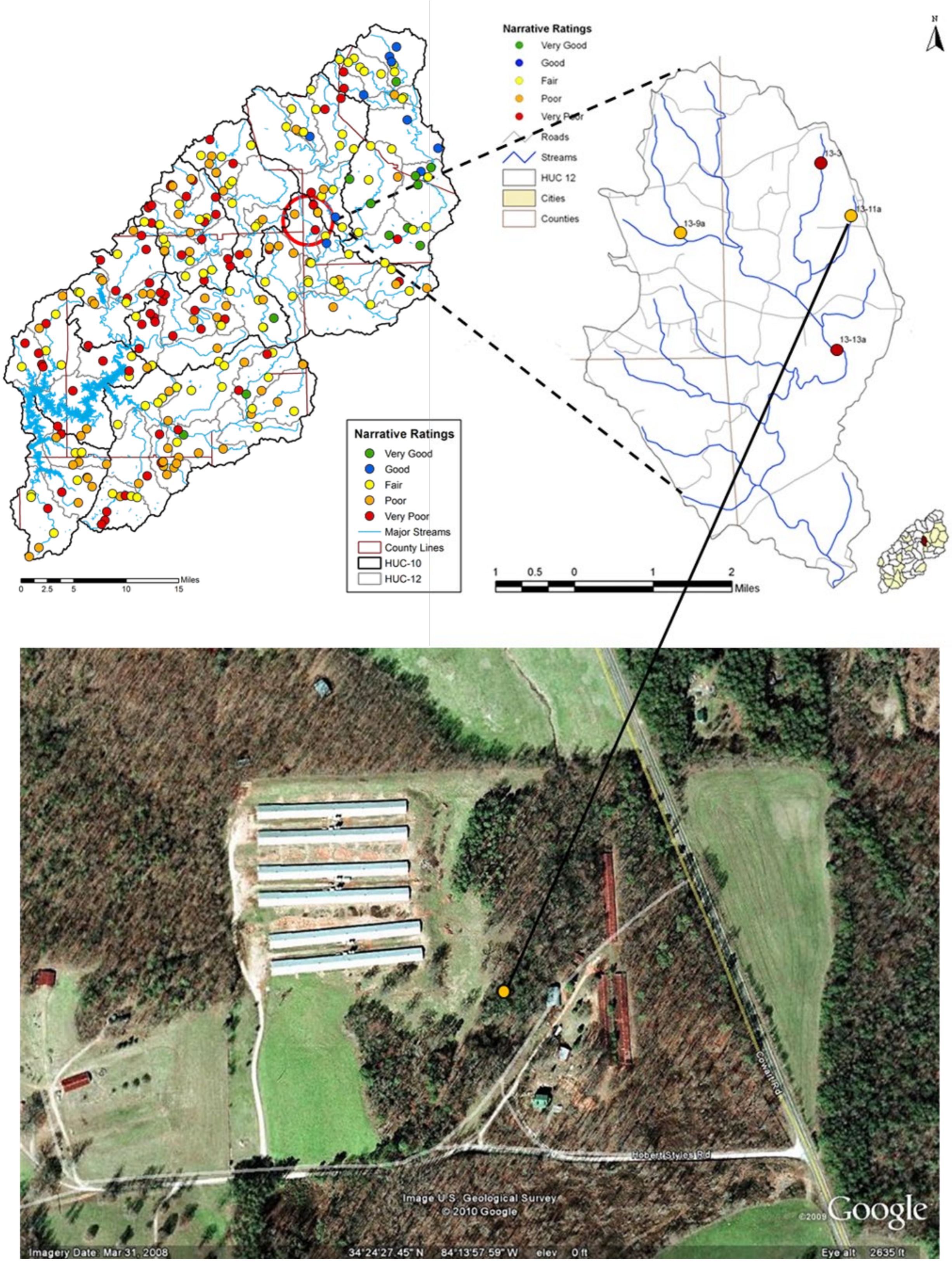

Figure 4-8. More detailed examination of the Yellow Creek subwatershed, Lake Allatoona/Upper Etowah River watershed, Georgia, reveals a sample location, rated biologically as "poor," is on a stream flowing through a poultry production operation (Millard et al. 2011)

4.5 References

Allen, D.M., S.K. Service, and M.V. Ogburn-Matthews. 1992. Factors influencing the collection efficiency of estuarine fishes. *Transactions of the American Fisheries Society* 121(2): 234–244.

Angermeier, P., and J. Karr. 1994. Biological integrity versus biological diversity as policy directives: protecting biotic resources. *BioScience* 44(10): 690–697.

Bahls, L. 1993. *Periphyton Bioassessment Methods for Montana Streams*. Department of Health and Environmental Science, Water Quality Bureau, Helena, MT.

Ball, J. 1982. *Stream Classification Guidelines for Wisconsin*. Technical Bulletin. Wisconsin Department of Natural Resources, Madison, WI.

Barbour, M.T., and J.B. Stribling. 1991. Use of Habitat Assessment in Evaluating the Biological Integrity of Stream Communities. In *Biological Criteria: Research and Regulation Proceedings of a Symposium*. EPA-440/5-91-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

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*. 2nd ed. EPA/841-D-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC. Accessed February 9, 2016. http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm.

Barbour, M.T., J.B. Stribling, and P.F.M. Verdonschot. 2006. The multihabitat approach of USEPA's rapid bioassessment protocols: benthic macroinvertebrates. *Limnetica* 25(3–4): 229–240.

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* 37(2): 114–118.

Cao, Y., and C.P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. *Journal of Applied Ecology* 42: 954–965.

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. Accessed April 22, 2016. http://pubs.usgs.gov/of/2013/1129/pdf/ofr20131129.pdf.

Caton, L. R. 1991. Improved subsampling methods for the EPA rapid bioassessment protocols. *Bulletin of the North American Benthological Society* 8: 317–319.

Charles, D.F., C. Knowles, and R.S. Davis, ed. 2002. *Protocols for the Analysis of Algal Samples Collected as Part of the U.S. Geological Survey National Water-Quality Assessment Program*. Report No. 02-06. Prepared for U.S. Geological Survey, by The Academy of Natural Sciences, Patrick Center for Environmental Research–Phycology Section, Philadelphia, PA.

Clarke, R.T., M.T. Furse, J.F. Wright, and D. Moss. 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics* 23: 311–332.

Clarke, R.T., J.F. Wright, and M.T. Furse. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modeling* 160: 219–233.

Conquest, L.L., S.C. Ralph, and R.J. Naiman. 1994. Implementation of Large-Scale Stream Monitoring Efforts: Sampling Design and Data Analysis Issues. In *Biological Monitoring of Aquatic Systems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Boca Raton, FL.

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe. 1979. *Classification of Wetlands and Deepwater Habitats of the United States*. FWS/OBS-79/31. U.S. Department of the Interior, U.S. Fish and Wildlife Service, Biological Services Program, Washington, DC.

Cowie, G.M., J.L. Cooley, and A. Dutt. 1991. *Use of Modified Benthic Bioassessment Protocols for Evaluation of Water Quality Trends in Georgia*. Technical Completion Report USDI/GS Project 14-08-0001-G1556 (03); ERC-06-91; USGS/G-1556-03. University of Georgia, Institute of Community and Area Development, Athens, GA, and Georgia Institute of Technology, Environmental Resources Center, Atlanta, GA.

Cummins, K.W. 1994. Bioassessment and Analysis of Functional Organization of Running Water Ecosystems. In *Biological Monitoring of Aquatic Ecosystems*, ed. S.L. Loeb and A. Spacie. Lewis Publishers, Ann Arbor, MI.

Cupp, C.E. 1989. *Stream Corridor Classification for Forested Lands of Washington*. Washington Forest Protection Association, Olympia, WA.

Dauble, D.D., and R.H. Gray. 1980. Comparison of a small seine and a backpack electroshocker to evaluate near shore fish populations in rivers. *Progressive Fish-Culturist* 42: 93–95.

Davies, S.P., and S.K. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16(4): 1251–1266.

Day, J.W., Jr., C.A.S. Hall, W.M. Kemp, and A. Yáñnez-Arancibia. 1989. *Estuarine Ecology*. John Wiley and Sons, Inc., New York, NY.

Dewey, M.R., L.E. Holland-Bartel, and S.T. Zigler. 1989. Comparison of fish catches with buoyant pop nets and seines in vegetated and nonvegetated habitats. *North American Journal of Fisheries Management* 9: 249–253.

Flotemersch, J.E., K.A. Blocksom, J.J. Hutchens, and B.C. Autrey. 2006a. Development of a standardized large river bioassessment protocol (LR-BP) for macroinvertebrate assemblages. *River Research and Applications* 22: 775–790.

Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006b. 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 February 10, 2016.

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.

Fore, L., and C.O. Yoder. 2003. *The Design of a Biological Community Trend Monitoring Program for Michigan Wadeable Streams* (DRAFT). MI/DEQ/WD-03/086. Prepared for Michigan Department of Environmental Quality, by Statistical Design and Midwest Biodiversity Institute.

Frissell, C.A., W.J. Liss, C.E. Warren, and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* 10(2): 199–214.

Gerritsen, J., J.S. White, E.W. Leppo, J.B. Stribling, and M.T. Barbour. 1996. *Examination of the Effects of Habitat Quality, Land Use, and Acidification on the Macroinvertebrate Communities of Coastal Plain Streams*. CBWP-MANTA-TR-95-2. Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD.

Gerritsen, J., M.T. Barbour, and K. King. 2000. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. *Journal of the North American Benthological Society* 19(3): 487–496.

Gibson, G.R., M.T. Barbour, J.B. Stribling, J. Gerritsen, and J.R. Karr. 1996. *Biological Criteria: Technical Guidance for Streams and Small Rivers*. EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Company, New York, NY.

Gulland, J.A. 1983. *Fish Stock Assessment: A Manual of Basic Methods*. FAO/Wiley Series, Vol. 1. Wiley & Sons, New York, NY.

Hawkins, C.P. 2006. Quantifying biological integrity by taxonomic completeness: evaluation of a potential indicator for use in regional- and global-scale assessments. *Ecological Applications* 16:1277–1294.

Hawkins, C.P., J.L. Kershner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman, and M.K. Young. 1993. A hierarchical approach to classifying stream habitat features. *Fisheries* 18(6): 3–12.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, and R.J. Stevenson. 2000a. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society* 19(3): 541–556.

Hawkins, C.P., R.H. Norris, J.N. Hogue, and J.W. Feminella. 2000b. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10(5): 1456–1477.

Hayes, M.L. 1983. Active Fish Capture Methods. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 123–145. American Fisheries Society, Bethesda, MD.

Helsel, D.R., and R.M. Hirsch. 2002. Statistical Methods in Water Resources. Book 4, Chapter A3 in *Techniques of Water-Resources Investigations*. U.S. Geological Survey, Reston, VA. Accessed February 10, 2016. http://pubs.usgs.gov/twri/twri4a3/.

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.

Hubert, W.A. 1983. Passive Capture Techniques. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 95–122. American Fisheries Society, Bethesda, MD.

Hughes, R.M. 1995. Defining Acceptable Biological Status by Comparing with Reference Conditions. In *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*, ed. W.S. Davis and T.P. Simons, pp. 31–47. Lewis Publishers. Boca Raton, FL.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.

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.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2):187-211.

Jerald, A., Jr. 1983. Age Determination. In *Fisheries Techniques*, ed. L.A. Nielsen and D.L. Johnson, pp. 301–324. American Fisheries Society, Bethesda, MD.

Karr, J.R. 1991. Biological integrity: a long neglected aspect of water resource management. *Ecological Applications* 1: 66–84.

Karr, J.R., and D.R. 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.

Klemm, D.J., P.A. Lewis, F. Fulk, and J.M. Lazorchak. 1990. *Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. EPA-600-4-90-030. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Klemm, D.J., Q.J. Stober, and J.M. Lazorschak. 1992. *Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*. EPA/600/R-92/111. U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Klemm, D.J., J.M. Lazorchak, and P.A. Lewis. 1998. Benthic Macroinvertebrates. In *Environmental Monitoring and Assessment Program—Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams*, ed. J.M. Lazorchak, D.J. Klemm, and

D.V. Peck, pp. 147–182. EPA/620/R-94/ 004F. U.S. Environmental Protection Agency, Washington, DC.

Larsen, D.P. 1997. Sample survey design issues for bioassessment of inland aquatic ecosystems. *Human and Ecological Risk Assessment* 3(6): 979–991.

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.

Meador, M.R., T.F. Cuffney, and M.E. Gurtz. 1993. *Methods for Sampling Fish Communities as Part of the National Water Assessment Program*. Open File Report 93-104. U.S. Geological Survey, Raleigh, NC.

Merritt, R.W., K.W. Cummins, and M.B. Berg, ed. 2008. *An Introduction to the Aquatic Insects of North America*. 4th ed. Kendall/Hunt Publishing Company, Dubuque, IA. ISBN 978-0-7575-5049-2.

Millard, C.J., N. Jokay, C. Wharton, and J.B. Stribling. 2011. *Ecological Assessment of the Lake Allatoona/Upper Etowah River Watershed - Year 5 of the Long-Term Monitoring Program, 2009–2010*. Prepared for Lake Allatoona/Upper Etowah River Watershed Partnership, Canton, GA, by Tetra Tech, Inc., Owings Mills, MD, and Atlanta, GA.

Moulton, S.R. II, J.G. Kennen, R.M. Goldstein, and J.A. Hambrook. 2002. *Revised Protocols for Sampling Algal, Invertebrate, and Fish Communities as Part of the National Water-Quality Assessment Program*. Open File Report 02-150. U.S. Geological Survey, Reston, VA.

Nelson, J.S., E.J. Crossman, H. Espinosa-Perez, L.T. Findley, C.R. Gilbert, R.N. Lea, and J.D. Williams. 2004. *Common and Scientific Names of Fishes from the United States, Canada, and Mexico*. Special Publication 29. American Fisheries Society, Bethesda, MD.

OEPA (Ohio Environmental Protection Agency). 1987. *Biological Criteria for the Protection of Aquatic Life: Volumes I-III*. Updated 1988, 1989, and 2015. Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, OH. Accessed February 10, 2016. http://www.epa.state.oh.us/dsw/bioassess/BioCriteriaProtAqLife.aspx.

Olsen, A.R., and D.V. Peck. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. *Journal of the North American Benthological Society* 27(4): 822–836.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish*. EPA/440/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Platts, W.S., W.F. Megahan, and G.W. Minshall. 1983. *Methods for Evaluating Stream, Riparian, and Biotic Conditions*. General Technical Report INT-138. U.S. Department of Agriculture, Forest Service, Ogden, UT.

Raven, P.J., N.T.H. Holmes, F.H. Dawson, P.J.A. Fox, M. Everard, I.R. Fozzard, and K.J. Rowen. 1998. *River Habitat Quality: the physical character of rivers and streams in the UK and Isle of Man*. Environment Agency, Bristol, England. ISBN1 873760 42 9.

Richards, C., and G.W. Minshall. 1992. Spatial and temporal trends in stream macroinvertebrate communities: the influence of catchment disturbance. *Hydrobiologia* 241(3): 173–184.

Skalski, J.R., and D.H. McKenzie. 1982. A design for aquatic monitoring programs. *Journal of Environmental Management* 14: 237–251.

Southwood, T.R.E. 1977. Habitat, the templet for ecological strategies? *Journal Animal Ecology* 46: 337–365.

Stark, J.D. 1993. Performance of the macroinvertebrate community index: effects of sampling method, sample replication, water depth, current velocity, and substratum on index values. *New Zealand Journal of Marine and Freshwater Research* 27: 463–478.

Stevens, D.L., and A.R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99(465): 262–278.

Stoddard, J. L., D.P. Larsen, C.P. Hawkins, R.K. Johnson, and R.H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16(4): 1267–1276.

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. ISBN 978-953-307-971-4. doi: 10.5772/22388. Accessed February 10, 2016. http://www.intechopen.com/books/modern-approaches-to-quality-control/partitioning-error-sources-for-quality-control-and-comparability-analysis-in-biological-monitoring-a.

Stribling, J.B., B.K. Jessup, and D.L. Feldman. 2008a. Precision of benthic macroinvertebrate indicators of stream condition in Montana. *Journal of the North American Benthological Society* 27(1): 58–67.

Stribling, J.B., K.L. Pavlik, S.M. Holdsworth, and E.W. Leppo. 2008b. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society* 27(4): 906–919.

Teply, M., and L. Bahls. 2005. *Diatom Biocriteria for Montana Streams*. Prepared for Montana Department of Environmental Quality, by Larix Systems, Inc., Helena, MT. 2005.

Teply, M., and L. Bahls. 2007. *Statistical Evaluation of Periphyton Samples from Montana Reference Streams*. Prepared for Montana Department of Environmental Quality, by Larix Systemes, Inc., Helena, MT. 2007.

Thornton, K.W., B.L. Kimmer, and F.E. Payne, ed. 1990. *Reservoir Limnology: Ecological Perspectives*. John Wiley and Sons, Inc., New York, NY.

Thorp, J.H., and A.P. Covich, ed. 2010. *Ecology and Classification of North American Freshwater Invertebrates*. Academic Press (Elsevier).

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8(2): 246–257.

USEPA (U.S. Environmental Protection Agency). 2002. *Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers*. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2004. *Wadeable Stream Assessment: Benthic Laboratory Methods*. EPA 841-B-04007. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2009. *National Rivers and Streams Assessment: Field Operations Manual*. EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2011. *A Primer on Using Biological Assessments to Support Water Quality Management*. EPA-810-R-11-01. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2012. *National Rivers and Streams Assessment 2013-2014: Laboratory Operations Manual*. EPA-841-B-12-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2013. *Biological Assessment Program Review: Assessing Level of Technical Rigor to Support Water Quality Management*. EPA-820-R-13-001. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USFWS (U.S. Fish and Wildlife Service). 1991. *Principles and Techniques of Electrofishing*. U.S. Fish and Wildlife Service, Fisheries Academy, Office of Technical Fisheries Training, Kearneysville, WV.

Van Sickle, J. 2008. An index of compositional dissimilarity between observed and expected assemblages. *Journal of the North American Benthological Society* 27(2): 227–235.

Van Sickle, J., and S.G. Paulsen. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. *Journal of the North American Benthological Society* 27(4): 920–930.

Vieira, N.K.M., N.L. Poff, D.M. Carlisle, S.R. Moulton II, M.K. Koski, and B.C. Kondratieff. 2006. A Database of Lotic Invertebrate Traits for North America. Data Series 187. U.S. Geological Survey, Reston, VA. Accessed February 10, 2016. http://pubs.water.usgs.gov/ds187.

Weatherley, A.H. 1972. *Growth and Ecology of Fish Populations.* Academic Press, New York, NY.

Wetzel, R.G. 1983. *Limnology*. Saunders College Publishing, New York, NY.

Wright, J.F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology* 20:181–197.

Zuellig, R.E., D.M. Carlisle, M.R. Meador, and M. Potapova. 2012. Variance partitioning of stream diatom, fish, and invertebrate indicators of biological condition. *Freshwater Science* 31(1): 182–190. doi: 10.1899/11-040.1.

This page intentionally left blank.