J. N. Am. Benthol. Soc., 2008, 27(1):16–37
© 2008 by The North American Benthological Society
DOI: 10.1899/06-081.1
Published online: 13 November 2007

# Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages

**Daren M. Carlisle[1,4], Charles P. Hawkins[2,5], Michael R. Meador[1,6], Marina Potapova[3,7], AND James Falcone[1,8]**

[1] *National Water-Quality Assessment Program, US Geological Survey, 12201 Sunrise Valley Drive, MS 413, Reston, Virginia 20192 USA*
[2] *Western Center for Monitoring and Assessment of Freshwater Ecosystems, Department of Aquatic, Watershed, and Earth Resources, and the Ecology Center, Utah State University, Logan, Utah 84322-5210 USA*
[3] *Patrick Center for Environmental Research, The Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Philadelphia, Pennsylvania 19103 USA*

**Abstract.** We developed empirical models for fish, macroinvertebrate, and diatom assemblages to assess the biological condition of 268 streams sampled from 1993 to 2002 in 7 major river basins in the Appalachian region of the USA. These models estimate the expected taxonomic composition at each site based on observed variation in taxonomic composition at reference sites. The index, O/E, is the ratio of the number of predicted taxa that were observed (O) to that expected (E) to occur at a site and is a measure of taxonomic completeness. We compared how O/E for each assemblage varied among major landuse settings and whether impaired assemblages were associated with particular physicochemical conditions. We also examined concordance among assemblages in their response to stress. Biological, chemical, and physical data were collected following consistent protocols. We used land-cover criteria, published data, and topographic maps to classify sites by major landuse setting. Fish, macroinvertebrate, and diatom assemblages had been sampled at 73, 108, and 52, respectively, of the least disturbed sites used to establish reference conditions. The models accounted for a substantial portion of the natural variation in taxonomic composition across sites that was associated with biogeographic, climatic, and basin-scale factors and generally were unbiased across the range of environmental gradients observed in the region. Assessments at nonreference sites showed that impairment of fish and macroinvertebrate assemblages was most strongly associated with agriculture and urban land uses, whereas impairment of diatom assemblages was most strongly associated with mining in the basin. Concordance in assessments among assemblages was not strong. Assessments based on 2 assemblages differed in 28 to 57% of cases, and assessments were never concordant for cases where all 3 assemblages were sampled. Furthermore, only ½ of these cases would have been assessed as ecologically impaired had only 1 assemblage been sampled. Differences between observed and predicted frequencies of occurrence for individual taxa were generally consistent with known tolerances to environmental stressors and might aid in identifying causes of biological impairment.

**Key words:** biological assessment, RIVPACS, biological monitoring, biological indicators, predictive models, O/E, fish, macroinvertebrates, diatoms, land use, National Water-Quality Assessment.

Indicators of ecological integrity are needed to assess the condition of the USA's ecosystems (NRC 2000, Heinz Center 2002). However, at present, there are no regional or national criteria for assessing the biological condition of streams (Heinz Center 2002) largely because of a lack of indicators that can be aggregated consistently across regional and national spatial scales

[4] E-mail addresses: dcarlisle@usgs.gov
[5] chuck.hawkins@usu.edu
[6] mrmeador@usgs.gov
[7] potapova@acnatsci.org
[8] jfalcone@usgs.gov

(NRC 2000, USGAO 2000). Most states conduct biological assessments of wadeable streams, but use different approaches to field sampling, laboratory processing, and data analysis (Carter et al. 2006). As a result, inferences about how the conditions of stream ecosystems vary at regional and national scales are limited to qualitative summaries of these different assessments (e.g., USEPA 2002a). Until methods are developed to relate one type of assessment to another (Davies and Jackson 2006), sound inferences regarding the overall biological condition of the nation's waters are most likely to be derived from nationally consistent monitoring programs (e.g., USEPA 2006).

Most assessments of ecological condition are based on comparisons of indicators observed at a site with their expected values. Expected values generally are defined by the distributions of indicator values observed at reference sites within a region (Stoddard et al 2006). Biological assessments based on predictive models such as the River InVertebrate Prediction And Classification System (RIVPACS; Moss et al. 1987, Wright 2000) measure impairment in terms of the ratio of the number of predicted taxa that were observed (O) at a site to the number of taxa that were expected to occur (E), or O/E. Predictions are based on biota–environment relationships derived from data collected at reference sites. An important quality of this approach is that modeling accounts for at least a portion of the potentially confounding effects of natural environmental gradients, and thus, improves both accuracy and precision of assessments.

Ideally, assessments of biological condition would be based on multiple biological assemblages, thereby providing insight into the overall biological integrity of an ecosystem. As of 2001, 77% of stream monitoring entities in the USA had used >1 assemblage (fish, macroinvertebrates, or diatoms) to assess the ecological condition of wadeable streams (USEPA 2002b). Biological assessments based on multiple taxonomic groups should be more comprehensive than assessments based on single assemblages because taxa in different groups tend to vary in trophic position, life history, body size, and other ecological/biological factors. Therefore, based on general biological principles, we would not expect the response of a single taxonomic assemblage to represent the condition of all other groups completely. Moreover, if one of the goals of biological assessment is to identify biological alterations associated with any number of possible stressors before overall biological integrity is substantially compromised, then use of multiple assemblages should improve the sensitivity of such assessments. However, interpretation might be complicated if assessments are not based on consistent measures or indicators among assemblages. Recognition and implementation of multiassemblage assessments appears to be growing (USEPA 2005), but most assessment comparisons have been made after different approaches (e.g., suites of indicators) were applied to each assemblage, often without explicit recognition of the precision and accuracy of each method (e.g., Passy et al. 2004, Griffith et al. 2005). Comparisons of multiassemblage assessments might be enhanced if consistent ecological attributes (such as taxonomic completeness, sensu Hawkins 2006) and assessment thresholds were used for all assemblages.

We describe the application of 3 RIVPACS-type predictive models to assess the condition of fish, macroinvertebrate, and diatom assemblages in >200 Appalachian streams in contrasting landuse settings. We compare how assessments varied among assemblages and determine how physical and chemical habitat varied among classes of biological condition for each assemblage.

## Methods

### *Study area and site selection*

Chemical, physical, and biological data were collected between 1993 and 2002 in 7 river basins that are wholly or partly within the Appalachian Mountains (Ozark, Ouachita–Appalachian Forest Ecoregion). Major river basins included the Allegheny–Monongahela, Delaware, Kanawha, Lower Susquehanna, Mobile, Potomac, and Tennessee (Fig. 1). Site selection was based on a targeted intensive and synoptic design used by the US Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program (Gilliom et al. 1995). Most sites used in our study were selected as part of synoptic designs throughout each major river basin. Sites were targeted for sampling in streams that met 1 of 3 criteria: 1) wadeable streams that were potentially influenced by several anthropogenic stressors, 2) relatively small (e.g., $2^{nd}$–$3^{rd}$ order) streams that were representative of specific landuse conditions (e.g., agricultural land use), or 3) similar-size streams within similar natural environmental settings along gradients of land use for synoptic studies (Tate et al. 2005). Fish, macroinvertebrate, and diatom assemblages were sampled at 134, 210, and 124 sites, respectively.

### *Field sampling and data preparation*

Sampling generally was conducted during periods of baseflow. Sampling-reach lengths were 20× the mean wetted channel width, roughly equivalent to 1 meander wavelength (Fitzpatrick et al. 1998). A

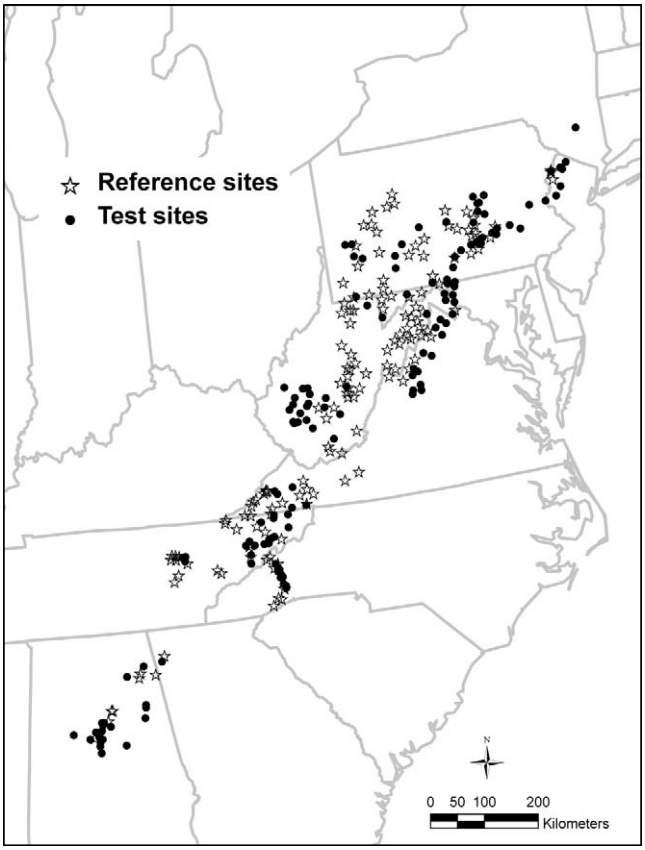



FIG. 1. Locations of reference and test sites in the Appalachian ecoregion sampled as part of the US Geological Survey National Water-Quality Assessment Program from 1993 to 2002.

sampling distance of at least 1 meander wavelength is likely to include at least 2 examples each of 2 different major habitat types (pools, riffles, runs; Leopold et al. 1964). Minimum reach length was 150 m, and maximum reach length was 300 m.

Fish were sampled with a single backpack electrofisher following the methods of Meador et al. (1993). Electrofishing crews usually consisted of 4 people, and electrofishing was conducted in an upstream direction. Block nets were not used to isolate the sampling area because block nets do not appear to improve estimates of species richness (Vadas and Orth 1993, Simonson and Lyons 1995, Edwards et al. 2003). Fish were removed from the water with dip nets, and upon completion of the 1$^{st}$ pass fish that could be identified in the field were counted and transported downstream from the sampling reach. A 2$^{nd}$ pass was done, and the data from both passes were combined. Two passes were used because Meador et al. (2003) showed that a single pass might underrepresent cyprinids and centrarchids. Fish that could not be identified in the field were retained for identification and enumeration in the laboratory (Walsh and Meador 1998).

Macroinvertebrates were sampled following Cuffney et al. (1993) and Moulton et al. (2002). Five discrete collections (each 0.25 m$^2$) were taken from riffles with a Slack sampler, passed through a 500-μm-mesh sieve, and combined into a single composite sample. Organisms were fixed in the field with 10% buffered formalin. Sample processing followed methods described in Moulton et al. (2000) and included removal of large organisms followed by a 300-organism fixed-count subsampling procedure. Most taxa were identified to genus, but some noninsect taxa were identified to family or higher levels of taxonomic resolution. Individuals with ambiguous taxonomic identifications were either assigned to a higher parent or excluded from analysis.

Diatoms were sampled following methods described by Porter et al. (1993). Quantitative samples were collected from hard substrates, mostly rocks, by scraping diatoms from substrates collected from riffles in the sampling reach. The size of the scraped area was determined from a mass–area relationship, where mass was measured with aluminum foil fitted to the area of rock sampled. Samples were fixed with 5% formalin. Diatoms were identified and enumerated from permanent slides under 1000× power at the Patrick Center for Environmental Research, (Academy of Natural Sciences, Philadelphia, Pennsylvania) following methods described in Charles et al. (2002).

Streamwater chemistry and physical habitat were measured at each site. Nutrients, suspended sediments, and major ions were measured from composite samples collected incrementally across the stream channel (Shelton 1994). Water temperature, dissolved $O_2$, pH, and specific conductance were measured directly in the field (Wilde and Radtke 1998). Physical habitat was measured following Fitzpatrick et al. (1998). Substrate particle size was estimated along 6 to 11 transects at each site. Percent fine/coarse substrates was estimated by calculating the proportion of quadrats along a transet that were primarily fine (<2 mm) or coarse (>16 mm) particle size.

A geographic information system was used to assemble basin-level hydrologic, topographic, geologic, climatic, land-cover, and soils data. Basin boundaries (USGS 2004a) were digitized at the 1:24,000 scale. Climatological data represent 30-y (1961–1990) averages (OCS 2004). Soils data were derived from the US Department of Agriculture State Soil Geographic Database (USDA 2004). Land use in each basin was derived from National Land Cover Data (NLCD) based on Landsat imagery for the period 1990 to 1992 (USGS 2004c). Stream segments were defined

based on the USGS National Hydrography Dataset 1:100,000 (USGS 2004b). Topographic information was derived from 100-m-resolution USGS National Elevation Data (USGS 2004a).

*Designation of reference sites*

Construction of predictive models requires a set of reference sites from which expected conditions at test sites are estimated. All basins in the Appalachians have been influenced by human activities to some extent. Therefore, least-impaired streams (sensu Stoddard et al. 2006) were regarded as reference streams in this region, and sites in these streams were assumed to represent the current biological potential of streams, given historic anthropogenic disturbance. Sites known to be affected by mining, dams/diversions, or point-source discharges were excluded when reference sites were selected. The judgment of local biologists, who selected sites to represent undeveloped or natural conditions, also was used. Last, land-cover data were used to identify least-disturbed basins. NLCD data were used to select watersheds with <5% urban and <50% agriculture land cover. Care was taken to exclude basins with urban land cover near the stream. The threshold for agriculture land use excluded basins with significant (>15%) row-crop farming. The remaining agriculture land cover included primarily pasture lands. This threshold was necessary to assess the low-elevation streams, and we acknowledge that using it probably resulted in lower-quality references sites than used for mountainous areas. Statistical comparisons of physicochemical conditions among landuse classes (described in *Basin landuse classifications of test sites*) were used to evaluate our reference selection criteria. Our selection criteria resulted in 73, 108, and 52 reference sites for fish, macroinvertebrate, and diatom assemblages, respectively. When >1 biological sample for a given assemblage had been collected at a reference site, only 1 sample, randomly selected, was used to construct predictive models.

The procedures described by Moss et al. (1987) and Clarke et al. (2003) were used to determine if the predictor variable values of each test site belonged to the same statistical population as the reference sites. The Mahalanobis squared distances between the locations of test-site values in predictor space and the centroids of each classification group are distributed as a $\chi^2$ function. Any test site was regarded as outside the range of the model if the smaller of these values exceeded the critical $\chi^2$ value for $\alpha = 0.01$ and $N$ degrees of freedom, where $N$ equals the number of discriminant functions.

*Model construction and calculation of O/E*

Detailed descriptions of RIVPACS modeling procedures are given elsewhere (Moss et al. 1987, Clarke 2000, Hawkins and Carlisle 2001), and are only briefly outlined here. Model construction and implementation followed these steps. First, reference sites were classified by cluster analysis based on assemblage composition after removing rare (occurring in <5% of sites) taxa. All pairwise Bray–Curtis dissimilarities were calculated on presence/absence data, and unweighted pair-group arithmetic averaging ($ß = -0.1$) was used to create a dendrogram from which groups of sites were derived. The number of groups selected depended on the number of reference sites available for each assemblage and the structure of the dendrogram. In general, an attempt was made to define groups such that within-group similarity was as high as possible, the level of between-group similarity used to cut the dendrograms was similar across assemblages, and group size was >5 sites. The all-subsets multiple discriminant analysis routines of Van Sickle et al. (2006) were used to select a set of predictor variables that were unrelated to human disturbance and that produced the most precise and accurate estimates of the probabilities of capture (Pc) for each taxon at each reference site. In RIVPACS-type models, taxon Pc values are estimated by weighting the frequencies of taxon occurrence within each reference-site group by a site's probability of group membership, as estimated from the discriminant model. This weighting essentially smoothes the frequencies of occurrence observed in the discrete groups along environmental continua. E was estimated at a site as the sum of all $Pc > 0.5$. O/E values were calculated as the number of predicted ($Pc > 0.5$) taxa collected (O) divided by E. A cutoff of $Pc > 0.5$ was used because O/E values based on $Pc < 0.5$ tend to be less precise than those based on $Pc > 0.5$ (Ostermiller and Hawkins 2004).

Two additional analyses that described how individual taxa varied across sites were used to aid in biological interpretation of predictions and assessments. First, an indicator species analysis was conducted following the method of Dufrêne and Legendre (1997) to identify those taxa that were most strongly associated with reference-site groups. Taxa with statistically significant ($p < 0.05$) indicator values (based on Monte Carlo simulations; McCune and Mefford 1999) were recorded. After applying the models to test sites, observed frequencies of occurrence (OF) within groups of test sites classified by major landuse class were compared with expected frequencies (EF) estimated by summing predicted Pcs across

sites (Knapp et al. 2005, deZwart et al. 2006). To facilitate comparison among taxa, the ratio OF/EF was used as a measure of taxon sensitivity to stress, where values >1 imply taxa are tolerant of alterations from reference conditions and values <1 imply taxa are intolerant of alterations.

*Model performance*

The usefulness of a predictive model in bioassessment is best judged by its ability to predict the expected assemblage accurately and precisely, which is a function of how well the model accounts for natural variability in E. When applied to reference-quality sites, good models should produce mean O/E values of ~1 (accuracy) and small standard deviations (SD) of O/E values (precision). Ideally, model performance would be assessed by applying models to an independent set of validation samples (typically 30–50 samples) and measuring model accuracy and precision (Van Sickle et al. 2006). Unfortunately, the small number of reference-site samples available for model-building precluded setting aside a set of validation samples and conducting a full validation. Therefore, internal validation, in which model predictions are compared with calibration samples, had to be used. In several previous analyses, calibration and validation data have often agreed well (Hawkins et al. 2000b, Hawkins 2006, Van Sickle et al. 2006).

Model performance also was evaluated by comparing the SD of reference-site O/E values to estimates of the upper and lower bounds of model precision. The lower bound of model precision is estimated from a null model in which E is predicted to be the same everywhere (i.e., Pc values for each taxon are assumed to be fixed across all reference sites). Such null models account for no natural variation in E (Van Sickle et al. 2005). Therefore, the SD of O/E from the null model is the lower limit of precision attainable by a predictive model. A theoretical upper limit of model precision was obtained following the method of Van Sickle et al. (2005). This method estimates the SD of O/E that is attributable only to replicate-sampling error, which would be the only expected source of error if the model accounted for all systematic effects of natural variability in E. This estimate of replicate-sampling error is theoretical, but is expected to approximate the error estimate from replicate sampling (which was unavailable).

*Basin landuse classifications of test sites*

To assess the potential effects of different landuse (i.e., prevalent human activity) practices on the 3 biological assemblages, each site was classified on the basis of 6 major types of land cover (from National Land Cover Dataset, USGS 2004c) in the contributing basin: urban, urban–forest, agriculture, agriculture–forest, mixed, and mining. Basins with >25% urban and <33% agriculture land cover were classified as either urban (forest land cover <33%; $n = 26$ basins) or urban–forest (forest land cover >33%; $n = 15$ basins). Basins with <25% urban and >33% agriculture land cover were classified as agriculture (forest land cover <33%; $n = 22$ basins) or agriculture–forest (forest land cover >33%; $n = 9$ basins). Basins with >10% urban and >33% agriculture land cover were classified as mixed ($n = 11$ basins). Basins in which mining was a prominent land use ($n = 21$ basins) were identified from published reports (Paybins et al. 2000, Chambers and Messinger 2001).

Following this initial classification, 7.5-min topographic maps of each basin were inspected visually. Human activities such as urbanization, dams, and diversions that were proximal to sampling reaches were noted, and several basins were reclassified. Five basins originally in the urban–forest class were reclassified as urban because of the short distance between the urban area and the sampling reach. This final classification obscured the apparent distinctions between some landuse classes but should have increased the likelihood that the classes represented general land modifications that might potentially influence test sites. Mean physicochemical conditions (temperature, conductance, pH, % fines, and concentrations of dissolved $O_2$, suspended sediment, $Cl^-$, $SO_4^{2-}$, and nutrients) and % riparian land cover classified as forest, agriculture, and urban were calculated within in each landuse class. Student's *t*-tests were used to determine if these values differed between test and reference sites ($p < 0.05$). Log($x$) transformations were used when needed to meet normality assumptions.

*Analyses of biological condition*

Three types of questions were addressed based on these biological assessments: 1) Did mean biological conditions differ among streams classified by basin land use? 2) What physicochemical attributes differed between biologically unimpaired and impaired streams? 3) What percentage of test-site samples was considered impaired based on the 3 different assemblages? Analysis of variance (ANOVA) was used to determine if variation in O/E values was associated with landuse class, and Bonferroni-adjusted post-hoc tests were applied to test for differences in mean O/E values between landuse classes. Questions 2 and 3 required that we set O/E thresholds for defining

biological impairment. Question 2 was best approached by setting common, biologically comparable thresholds for all 3 assemblages. Therefore, 2 classes of biological impairment were defined by the severity of assemblage alteration: O/E $> 0.8$ was considered unimpaired, O/E $\leq 0.8$ was considered impaired. This threshold was selected by assuming that 20% loss of expected taxa (O/E $= 0.80$) represents a reasonable balance between losses that are ecologically meaningful and losses that are statistically detectable. Student's *t*-tests were used to compare the mean values of physicochemical data (same as those compared among landuse classes) between impaired and unimpaired test sites ($p < 0.05$).

These fixed thresholds could not be used to compare the sensitivity of the different models in detecting impairment at individual sites (question 3) because the statistical test of whether a single sample is equivalent to reference depends on model precision, which varied among the 3 models. Therefore, statistically comparable thresholds were set as the $10^{th}$ and $90^{th}$ percentiles of the distribution of reference values. Any O/E value for a test site that fell outside of this range of values was considered impaired for these analyses. Similar thresholds have been used elsewhere for comparative purposes (Hawkins 2006). This analysis was based on test sites where $\geq 2$ assemblages were sampled concurrently. Invertebrates and diatoms were both sampled on 108 occasions. Invertebrates and fish assemblages were both sampled on 61 occasions. Diatom and fish assemblages were both sampled on 40 occasions. All 3 assemblages were sampled concurrently 40 times.

## Results

### *Environmental heterogeneity among sites*

Natural environmental characteristics varied considerably among basins, but differences between reference and test basins were subtle (Table 1). The size of reference basins varied by several orders of magnitude and, on average, reference basins were slightly larger than test basins. Site elevation varied 5- to 10-fold across reference sites (variation in reference-site elevation differed depending on the assemblage examined), and, on average, reference sites were slightly higher than test sites. Other environmental factors varied 2- to 35-fold among reference basins, but were similar between reference and test basins. Furthermore, $\chi^2$ tests showed that the predictor variables of most (>95%) test sites were within the range of predictor variables of reference sites. The small number of test sites that differed environmentally from reference sites were excluded from analysis. In general, the population of reference sites appeared adequate to characterize the range of naturally occurring conditions present at test sites.

### *Biological classification of sites and spatial heterogeneity in biotic structure*

Biological classification of reference sites differed depending on which assemblage was used to classify the sites. Regardless of the assemblage used, reference-site groups generally were characterized by specific indicator taxa (Appendices 1–3). The geographic distribution of reference-site groups showed strong spatial structure within the ecoregion (Fig. 2A–C), but the general pattern of the spatial distribution of reference-site groups differed depending on the assemblage used to classify sites.

Eleven reference-site groups were identified based on fish assemblages. Groups were distributed along a distinct latitudinal gradient (Fig. 2A). Three groups were mainly in the northern Appalachians, 6 groups were in the middle Appalachians, and 1 group was in the southern Appalachians (Fig. 2A). In 1 group, sites were broadly distributed along a latitudinal gradient from the middle to the northern Appalachians. In 7 groups, sites within groups were in close proximity to one other. Fish assemblages in the middle Appalachian region of eastern Virginia, eastern West Virginia, and western Tennessee were especially diverse.

Nine reference-site groups were identified based on macroinvertebrate assemblages. The spatial distribution of the groups was less strongly structured by latitude than observed in fish assemblages (Fig. 2B). Macroinvertebrate assemblages showed a high degree of local heterogeneity throughout the region, but sites in 4 groups were in close proximity to one another, indicating high local-scale similarity. The most diverse area within the region appeared to be in the northern Appalachians, but the apparent diversity of groups in this area might be an artifact of the large number of reference sites in this region.

Six reference-site groups were identified based on diatom assemblages. Groups showed some evidence of distribution along a latitudinal gradient in the southern and middle Appalachians (Fig. 2C). Diatom assemblages showed a high degree of heterogeneity in the northern Appalachians (Pennsylvania) where 5 of the 6 groups were present. One of these 5 groups was otherwise found only in the most southern most part of the study area.

### *Model predictors and performance*

The all-subsets analysis showed that fish assemblage composition (described by the 11 reference-site groups)

TABLE 1. Mean, minimum (min), and maximum (max) values of environmental variables in basins where fish, macroinvertebrate, and diatom assemblages were sampled in the Appalachian ecoregion between 1993 and 2002.

| | Fish | | | | | | Macroinvertebrates | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Reference ($n$ = 73) | | | Test ($n$ = 54) | | | Reference ($n$ = 108) | | | Test ($n$ = 100) | | |
| Variable | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max |
| Basin size ($\times$10 km$^2$) | 1 | 166 | 2973 | 0.5 | 112 | 3063 | 1 | 59 | 978 | 0.5 | 70 | 3063 |
| Elevation (m asl) | 165 | 416 | 679 | 119 | 300 | 729 | 89 | 360 | 993 | 56 | 225 | 732 |
| Slope (%) | 5 | 16 | 28 | 2 | 12 | 33 | 3 | 15 | 29 | 1 | 10 | 36 |
| Drainage density (km/km$^2$) | 0.6 | 0.9 | 1.8 | 0.4 | 0.8 | 1.6 | 0.5 | 1.0 | 2.3 | 0.1 | 0.8 | 1.9 |
| Segment gradient (m/km) | <1 | 5 | 24 | <1 | 4 | 17 | <1 | 7 | 66 | <1 | 6 | 33 |
| Clay (mean % in basin soils) | 16 | 26 | 51 | 16 | 30 | 53 | 14 | 24 | 51 | 8 | 29 | 59 |
| Silt (mean % in basin soils) | 34 | 45 | 57 | 32 | 45 | 56 | 34 | 49 | 58 | 29 | 48 | 58 |
| Sand (mean % in basin soils) | 15 | 29 | 43 | 12 | 25 | 41 | 12 | 27 | 41 | 10 | 23 | 54 |
| Mean soil permeability (cm/h) | 1 | 3 | 11 | 1 | 2 | 8 | 1 | 4 | 8 | 1 | 3 | 7 |
| Mean annual precipitation (cm) | 103 | 136 | 200 | 99 | 132 | 152 | 96 | 123 | 157 | 97 | 120 | 151 |

was best predicted by 6 variables: latitude, drainage area, site elevation, mean annual precipitation, soil permeability, and soil clay content. The precision of this model (SD = 0.15) was markedly better than the null model (SD = 0.42) and nearly equivalent to sampling error (SD = 0.14), showing that the full model accounted for essentially all of the systematic variation in fish species composition among reference sites.

Macroinvertebrate assemblage composition (9 reference-site groups) was best predicted by 5 variables: drainage area, mean annual temperature, latitude, longitude, and drainage density (stream km/km$^2$). The model was precise (SD = 0.11), but its precision was not much better than that of the null model (SD = 0.12). Furthermore, the null-model SD was only slightly higher than the estimate of sampling error (SD = 0.11).

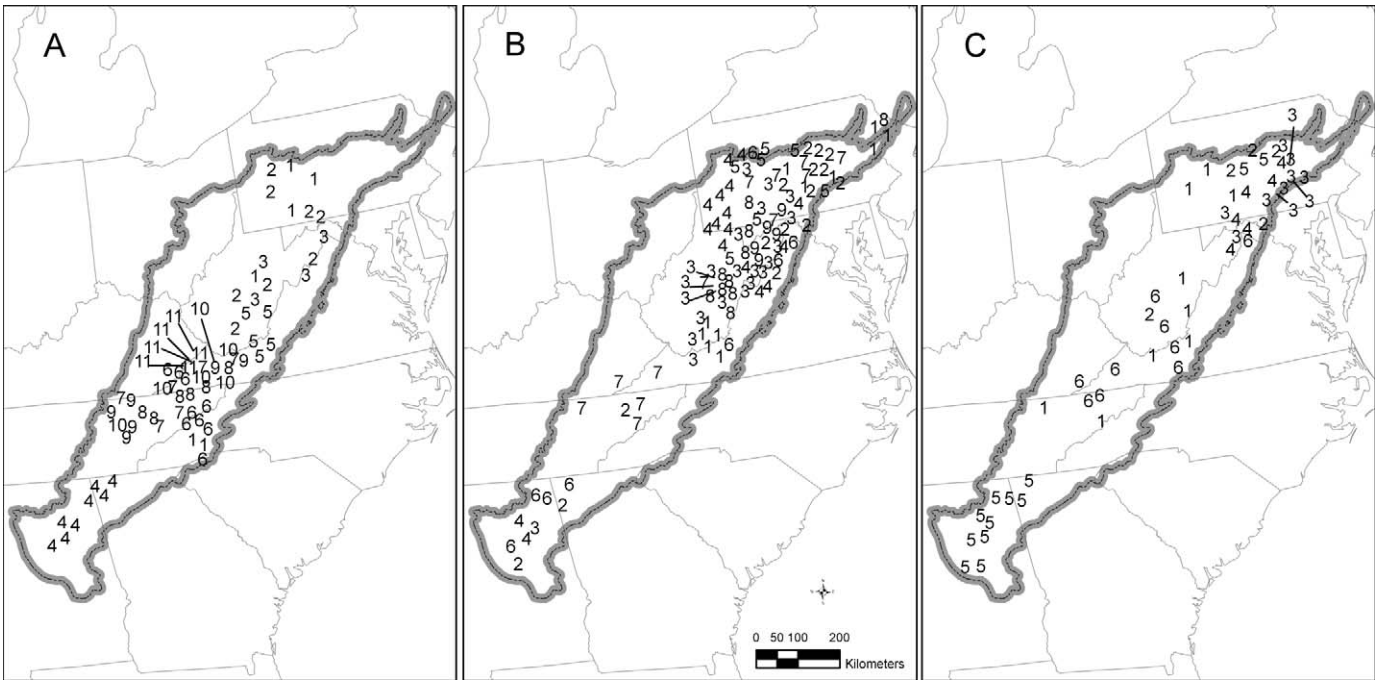



FIG. 2. Geographic distribution of reference sites classified by their biological similarity based on fish (A), macroinvertebrate (B), or diatom (C) assemblages. Site locations are shown as numbers that indicate groups of sites that were classified as biologically similar. Groups differ depending on the assemblage used to classify the sites. See Appendices 1 (fish), 2 (macroinvertebrates), or 3 (diatoms) for lists of indicator taxa for each group of sites. The outline shows the approximate boundary of the Ozark, Ouachita–Appalachian Forest Ecoregion.

TABLE 1. Extended.

| Diatoms | | | | | |
|---|---|---|---|---|---|
| Reference ($n$ = 52) | | | Test ($n$ = 65) | | |
| Min | Mean | Max | Min | Mean | Max |
| 1 | 250 | 6736 | 0.5 | 37 | 743 |
| 64 | 271 | 679 | 56 | 179 | 433 |
| 3 | 13 | 27 | 1 | 7 | 33 |
| 0.6 | 0.9 | 1.8 | 0.1 | 0.7 | 1.7 |
| <1 | 7 | 66 | <1 | 6 | 33 |
| 15 | 26 | 51 | 12 | 30 | 43 |
| 34 | 46 | 55 | 36 | 49 | 56 |
| 16 | 28 | 41 | 9 | 21 | 39 |
| 1 | 3 | 8 | 1 | 2 | 6 |
| 101 | 123 | 151 | 97 | 123 | 151 |

The full model was no better than the null model for macroinvertebrates, but the full model was used for 2 reasons: 1) it did predict differences in composition (Fig. 2A) between sites even though overall richness did not vary much between reference sites, and 2) the model enabled us to identify test sites that were outside the range of natural environmental conditions represented by the population of reference sites.

Diatom assemblage composition (6 reference-site groups) was best predicted by 5 variables: mean annual temperature, longitude, soil clay content, mean annual precipitation, and drainage area. This model (SD = 0.16) was substantially more precise than the null model (SD = 0.25), and its precision was close to sampling error (SD = 0.13).

*O/E at test sites*

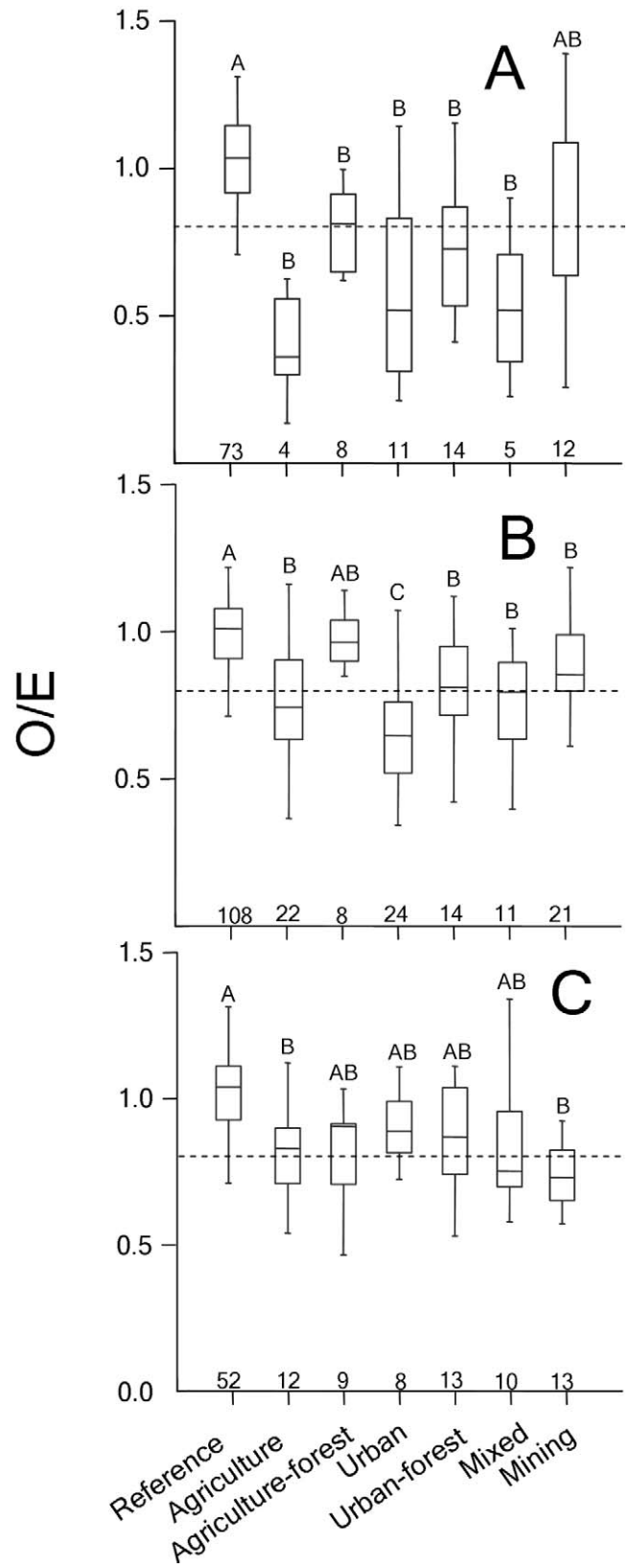
The degree to which test-site O/E values deviated from reference depended on the assemblage under consideration. In general, fish assemblages (mean O/E = 0.68, 95% confidence interval [CI] = 0.60–0.76) appeared to be more impaired than macroinvertebrate (mean O/E = 0.78, 95% CI = 0.73–0.82) and diatom assemblages (mean O/E = 0.83, 95% CI = 0.79–0.88). For fish, estimates of E at test sites varied from 4 to 22 taxa, and O varied from 0 to 25 taxa. For macroinvertebrates, estimates of E at test sites (16–22 taxa) did not vary substantially, whereas O varied from 1 to 22 taxa. For diatoms, estimates of E at test sites varied from 7 to 20 taxa, and O varied from 1 to 23 taxa.

*Physicochemical conditions in basin landuse classes*

Two or more physicochemical characteristics differed significantly between reference and test sites in all landuse classes (Table 2). Sites in agriculture basins had higher $Cl^-$, total P, total N, and % agriculture riparian land use and lower % forest riparian land use than sites in reference basins. Agriculture–forest basins had higher suspended sediment, total P, total N, and % agriculture riparian land use and lower % forest riparian land use than sites in reference basins. Sites in urban basins had higher suspended sediment, $Cl^-$, total N, and % urban riparian land use and lower % forest riparian land use than sites in reference basins.

TABLE 2. Mean physicochemical conditions of test sites in Appalachian streams within basin landuse classes. Landuse classes were: urban (>25% urban, <33% agriculture, <33% forest; $n = 26$), urban–forest (>25% urban, <33% agriculture, >33% forest; $n = 15$), agriculture (<25% urban, >33% agriculture, <33% forest; $n = 22$), agriculture–forest (<25% urban, >33% agriculture, >33% forest; $n = 9$), mixed (>10% urban, >33% agriculture; $n = 11$), and mining (mining activity within the basin; $n = 21$). Mean values in bold font were statistically ($p < 0.05$, Bonferroni adjusted) different from mean values at reference sites.

| Variable | Sample size | Landuse class | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Reference | Agriculture | Agriculture–forest | Urban | Urban–forest | Mixed | Mining |
| Temperature (°C) | 232 | 20 | 18 | 18 | 19 | 19 | 18 | 22 |
| Dissolved $O_2$ (mg/L) | 113 | 9 | 10 | 9 | 8 | 10 | 10 | 9 |
| pH | 140 | 7.7 | 7.9 | 8.0 | 7.6 | 7.5 | 7.9 | 7.5 |
| Suspended sediment (mg/L) | 103 | 7 | 27 | **102** | **62** | 11 | 29 | 6 |
| $SO_4^{2-}$ (mg/L) | 169 | 53 | 17 | 23 | 35 | 29 | 35 | **196** |
| $Cl^-$ (mg/L) | 169 | 8 | **11** | 11 | **27** | 21 | **53** | 8 |
| Specific conductance (μS/cm) | 117 | 252 | 384 | 339 | 414 | 183 | 528 | **510** |
| Total P (mg/L) | 122 | 0.02 | **0.06** | **0.26** | 0.05 | 0.05 | **0.11** | 0.02 |
| Total N (mg/L) | 122 | 0.59 | **4.81** | **5.23** | **2.57** | 0.76 | **2.99** | 0.62 |
| % fine substrate | 99 | 8 | 7 | 5 | 8 | 17 | 16 | 7 |
| % riparian forest | 268 | 75 | **22** | **43** | **33** | **58** | **43** | 83 |
| % riparian agriculture | 268 | 20 | **73** | **52** | 10 | 10 | **35** | **7** |
| % riparian urban | 268 | 2 | 2 | 2 | **25** | **25** | **15** | 5 |




Sites in urban–forest basins had higher % urban riparian land use and lower % forest riparian land use than sites in reference basins. Mixed landuse basins had higher $Cl^-$, total P, total N, % urban riparian land use, and % agriculture riparian land use and lower % forest riparian land use than sites in reference basins. Sites in mining basins had higher specific conductance and $SO_4^{2-}$and lower % agriculture riparian land use than sites in reference basins.

*Variation in O/E values with basin land use*

The land use most strongly associated with reductions in O/E at test sites differed depending on which assemblage was under consideration. Nearly ½ the variation in fish O/E values was associated with landuse class ($F = 18.16$, $p < 0.001$, $R^2 = 0.49$, $n = 127$; Fig. 3A). The lowest mean fish O/E values occurred in sites in agriculture (0.39), urban (0.57), urban–forest (0.71), and mixed (0.54) basins. However, these values were based on small sample sizes ($n < 10$) for some landuse classes. Mean fish O/E values from sites in agriculture–forest (0.75) and mining (0.82) basins were not statistically different from reference values. Forty-four % of the variation in macroinvertebrate O/E values was associated with landuse class ($F = 26.38$, $p < 0.001$, $R^2 = 0.44$, $n = 208$; Fig. 3B). The lowest mean macroinvertebrate O/E values occurred in sites in urban (0.62) basins. Intermediate mean macroinvertebrate O/E values occurred in sites in agriculture (0.76), mixed (0.76), urban–forest (0.81), and mining (0.86) basins. Mean macroinvertebrate O/E values from sites in agriculture–forest (0.97) basins were not statistically different from reference values. About 20% of the variation in diatom O/E values was associated with landuse class ($F = 4.68$, $p = 0.003$, $R^2 = 0.19$, $n = 117$; Fig. 3C), but mean diatom O/E values were statistically different from reference values only in sites in mining (0.74) and agriculture (0.79) basins. Mean O/E values from streams in agriculture–forest (0.81), mixed (0.84), urban (0.85), and urban–forest (0.85) basins were not statistically different from reference values.

←

FIG. 3. Box-and-whisker plots showing distributions of observed (O) to expected (E) ratios (O/E) for fish (A), macroinvertebrate (B), and diatom (C) assemblages among landuse classes. Boxes show quartiles, the bars in the boxes show medians, whiskers show values >1.5× the range of quartiles. Distributions with different letters were significantly different ($p < 0.05$, Bonferroni-adjusted). Sample sizes are indicated below each distribution. The long dashed line in each panel shows the impairment threshold of 0.80.

TABLE 3. Mean values of physicochemical variables in test sites with unimpaired and impaired fish, macroinvertebrate, or diatom assemblages. Sample size ($n$) of each group (unimpaired, impaired) varies because some variables were not sampled at all locations. $p < 0.05$ indicates a significant difference between values of a variable at sites classified as impaired or unimpaired based on the assemblage tested.

| **Assemblage tested**/variable | Unimpaired | Impaired | $n$ | $p$ |
|---|---|---|---|---|
| **Fish** | | | | |
| Temperature (°C) | 18 | 20 | 13, 24 | 0.158 |
| Dissolved $O_2$ (mg/L) | 9 | 9 | 13, 22 | 0.551 |
| pH | 7.8 | 7.9 | 13, 23 | 0.299 |
| Suspended sediment (mg/L) | 17 | 8 | 13, 24 | 0.845 |
| $SO_4^{2-}$ (mg/L) | 96 | 38 | 23, 24 | 0.637 |
| $Cl^-$ (mg/L) | 9 | 11 | 13, 24 | 0.750 |
| Specific conductance (μS/cm) | 255 | 336 | 13, 24 | 0.336 |
| Total P (mg/L) | 0.05 | 0.26 | 13, 24 | 0.559 |
| Total N (mg/L) | 1.20 | 1.60 | 13, 24 | 0.810 |
| % fine substrate | 15 | 16 | 10, 18 | 0.951 |
| % riparian forest | 59 | 53 | 18, 36 | 0.347 |
| % riparian agriculture | 19 | 21 | 18, 36 | 0.709 |
| % riparian urban | 19 | 23 | 18, 36 | 0.630 |
| **Macroinvertebrate** | | | | |
| Temperature (°C) | 20 | 20 | 50, 47 | 0.649 |
| Dissolved $O_2$ (mg/L) | 9 | 9 | 18, 23 | 0.415 |
| pH | 7.9 | 7.6 | 19, 27 | 0.034 |
| Suspended sediment (mg/L) | 42 | 45 | 13, 15 | 0.871 |
| $SO_4^{2-}$ (mg/L) | 18 | 20 | 31, 30 | 0.297 |
| $Cl^-$ (mg/L) | 18 | 20 | 31, 30 | 0.091 |
| Specific conductance (μS/cm) | 523 | 381 | 24, 12 | 0.389 |
| Total P (mg/L) | 0.09 | 0.06 | 17, 27 | 0.550 |
| Total N (mg/L) | 2.59 | 3.93 | 17, 27 | 0.097 |
| % fine substrate | 11 | 12 | 28, 24 | 0.786 |
| % riparian forest | 55 | 37 | 48, 52 | <0.001 |
| % riparian agriculture | 31 | 34 | 48, 52 | 0.622 |
| % riparian urban | 9 | 24 | 48, 52 | <0.001 |
| **Diatom** | | | | |
| Temperature (°C) | 19 | 19 | 39, 26 | 0.835 |
| Dissolved $O_2$ (mg/L) | 9 | 9 | 11, 15 | 0.529 |
| pH | 7.8 | 7.7 | 10, 16 | 0.529 |
| Suspended sediment (mg/L) | 52 | 33 | 14, 14 | 0.908 |
| $SO_4^{2-}$ (mg/L) | 27 | 64 | 22, 15 | 0.127 |
| $Cl^-$ (mg/L) | 15 | 17 | 22, 15 | 0.912 |
| Specific conductance (μS/cm) | 388 | 406 | 6, 8 | 0.733 |
| Total P (mg/L) | 0.06 | 0.10 | 22, 15 | 0.293 |
| Total N (mg/L) | 3.71 | 3.37 | 22, 15 | 0.801 |
| % fine substrate | 9 | 14 | 22, 27 | 0.331 |
| % riparian forest | 35 | 46 | 39, 26 | 0.057 |
| % riparian agriculture | 34 | 36 | 39, 26 | 0.743 |
| % riparian urban | 27 | 12 | 39, 26 | 0.008 |

*Physicochemical conditions associated with biological impairment classes*

Physicochemical conditions did not differ between test sites with impaired fish assemblages and test sites with unimpaired fish assemblages (Table 3). Test sites with impaired macroinvertebrate assemblages had higher % urban riparian land use and lower % forest riparian land use than test sites with unimpaired macroinvertebrate assemblages (Table 3). Test sites with impaired diatom assemblages tended to have conditions associated with less, rather than more stress. For example, % urban riparian land cover was lower at test sites with impaired diatom assemblages than at test sites with unimpaired assemblages (Table 3).

*Concordance in site-specific assessments among assemblages*

Assessments of test sites at which >1 assemblages were sampled contemporaneously frequently differed depending on which assemblage was used to judge

TABLE 4. Concordance of test-site assessments based on 2 assemblages. Impairment thresholds for each assemblage = 10th and 90th percentiles of reference-site observed/expected values. Each row is a possible outcome of comparing assessments of 2 assemblages. A = first assemblage listed in column heading, B = second assemblage listed in column heading, $n$ = number of test sites where both assemblages were sampled.

| Assessment outcome | Diatom/fish ($n$ = 40) | Diatom/macroinvertebrate ($n$ = 108) | Fish/macroinvertebrate ($n$ = 61) |
|---|---|---|---|
| Both unimpaired | 9 (22%) | 17 (16%) | 13 (21%) |
| Both impaired | 14 (35%) | 29 (27%) | 25 (41%) |
| A unimpaired, B impaired | 9 (22%) | 42 (39%) | 11 (18%) |
| A impaired, B unimpaired | 8 (21%) | 20 (18%) | 12 (20%) |

impairment (Table 4). Assessments based on diatoms and macroinvertebrates were most likely to differ (57% of all cases), and assessments based on macroinvertebrate assessments indicated impairment more frequently (39%) than assessments based on diatoms (18%). Assessments based on macroinvertebrates and fish were the least likely to differ (38% of all cases), and about the same percentage of test sites was assessed as impaired in assessments based on the 2 assemblages (macroinvertebrates: 18%, fish: 20%). Assessments based on diatoms and fish differed with an intermediate frequency (43%), and a similar percentage of samples was assessed as impaired in assessments based on the 2 assemblages (diatoms: 21%, fish: 22%). Forty sampling events included data collection for all 3 assemblages. At least 1 assessment differed from the others in all (100%) cases, i.e., all test sites would have been considered impaired based on at least 1 assemblage. In contrast, use of any single assemblage would have suggested impairment at only 22 to 23 (55–57%) test sites. The disparity between assessments based on different assemblages was a function of differences in the responsiveness of taxa in each assemblage to stress (fish > invertebrates > diatoms), the specific stressors to which taxa responded, and differences in precision of the models (macroinvertebrates >> fish > diatoms).

*Sensitive and tolerant taxa*

OF/EF varied widely among taxa (Table 5). Sample-size issues caused us to limit these analyses to 3 major landuse classes: urban (including urban–forest), agriculture (including agriculture–forest), and mining. In agriculture basins, 15 fish taxa, 8 macroinvertebrate taxa, and 4 diatom taxa were detected in <50% of the expected number of samples. In contrast, 16 fish taxa, 19 macroinvertebrate taxa, and 11 diatom taxa were detected >2× more frequently than expected. In urban basins, 5 fish taxa, 6 macroinvertebrate taxa, and 3 diatom taxa were detected in <50% of the expected number of samples. In contrast, 1 fish taxon, 8 macroinvertebrate taxa, and 15 diatom taxa were detected >2× more frequently than expected. In mining basins, 8 macroinvertebrate taxa and 3 diatom taxa were detected in <50% of the expected number of samples, and 3 diatom taxa were detected >2× more frequently than expected.

## Discussion

Interpretation of biological assessments must be grounded in a sound understanding of what an indicator measures and the statistical factors that affect its ability to distinguish a stress-induced change in value from random sampling error. O/E is essentially a measure of average change in taxon detectabilities, which implies a change in the composition (list of specific taxa) or structure (relative abundances) of an assemblage. Interpretation of O/E is independent of the type of biota to which it is applied because O/E has a reasonably direct biological interpretation. Moreover, it is well suited for assessment programs that span large, heterogeneous regions and that use multiple assemblage types because it is a relative measure of biological condition. However, its interpretation in the context of biological impairment requires that we understand the factors that can cause individual taxon detectabilities to change, the biological significance of those changes, and the statistical properties of the aggregate indicator.

*General performance and ecological interpretability of models*

Successful use of predictive models for bioassessment requires that models accurately predict the taxa that should have occurred at a site before any stress that might have affected the biota, i.e., the biota that would be present under reference conditions. These predictions also must be made with sufficient precision that ecologically meaningful departures from expected conditions can be detected. Therefore, any validation of a RIVPACS-type model must be based on how well it predicts the biota expected under reference conditions. If models predict well, then a departure from

predictions at a nonreference site must be the consequence of either sampling error or alteration of the biota by a stressor. Ideally, RIVPACS-type predictions are evaluated based on independent validation samples. Unfortunately, in many cases, sufficient validation data are not available, and internal validation must be used as a first approximation of how well the models perform.

The predictor variables used in the models can be interpreted in terms of causal mechanisms affecting taxon distributions. Thus, they can provide insight into the factors that most strongly influence natural communities and, hence, the factors that, if altered by human activity, should have the strongest effect on biological integrity. RIVPACS-type models that derive estimates of E from taxon-specific Pc also provide the opportunity to identify specific taxa that are either not detected but predicted or detected but not predicted. Such summaries can lead to insight regarding the relative sensitivity or tolerance of different taxa to environmental alteration (e.g., Knapp et al. 2005, deZwart et al. 2006). In general, each of the models developed in our study uses predictor variables that can be interpreted largely in terms of existing knowledge of the factors that influence stream biota.

The fish model was especially satisfying in terms of interpretability and performance. The predictors selected for use in this model imply that fish distributions are strongly controlled by temperature (latitude and elevation) and habitat availability associated with stream size (drainage area), with more subtle controls by hydrologic properties (precipitation, soil permeability, and amount of clay in soils). These results are generally consistent with a large body of literature that describes strong geographic patterns in stream fish distributions in the eastern USA (Mayden 1987, Angermeier and Winston 1999), and also agree with recent studies that show hydrologic properties can influence stream fish distributions (Lamouroux et al. 2002). The model was remarkable in that it accounted for essentially all of the systematic variation in species richness that occurred among reference sites within the region and was more precise than a model constructed for Ohio stream fish assemblages (Hawkins 2006, deZwart et al. 2006). The large SD of the null model O/E implies that either substantial species turnover occurs among sites within the region, total species richness varies considerably among sites, or both. The small estimate of sampling error implies that sampling was rigorous enough to predict consistently the common (i.e., $Pc > 0.5$) species present at each site. Given the equally small model error, the only way the precision of this model could be further improved is if sampling error were reduced by increasing sampling effort.

The macroinvertebrate model also was interpretable and precise. The dominant role of temperature and latitude as predictors in the model implies that stream macroinvertebrate assemblage composition is largely controlled by temperature gradients, with more subtle effects associated with factors that vary with stream size. These results also are also consistent with the literature describing factors that control stream macroinvertebrate distributions (Poff 1997, Vinson and Hawkins 1998, Hawkins et al. 2000b). Given the similar errors for the full model and for replicate sampling, improving the precision of this model would be difficult without greatly reducing sampling error by increasing sampling effort. This model is among the most precise macroinvertebrate models developed (cf. Hawkins 2006). On face value, the small null model error implies that we could use this simple null model as effectively as the full model without bothering with the statistical machinations involved with the full model. However, this result emerged because overall taxonomic richness at reference sites was similar even though composition varied. The full model accounted for variation in taxon identities, and this ability is needed to assess accurately what specific taxa are missing from individual sites with low O/E values and within a region as a whole.

The predictors used in the diatom model imply that diatom species distributions are largely controlled by the same climate and physicochemical factors associated with stream size (Stevenson 1997) that control fish and macroinvertebrate distributions. The use of soil type and clay content as predictors might also imply that diatom species composition is controlled partly by factors associated with the interaction of water and soils, e.g., hydrologic properties or water chemistry. The diatom model was less precise than either of the animal models. Therefore, it was not able to detect as small departures from reference conditions as the fish or (especially) the macroinvertebrate model. Full model error was measurably greater than the estimate of sampling error, and it might be possible to improve this model in the future by incorporating additional predictor variables that are specifically relevant to diatoms.

*Biological condition of streams in the Appalachian Region*

A major goal of the NAWQA program is to describe the status of the nation's streams within particular landuse settings. Use of O/E enabled us to address this goal by assessing biological condition in a standard way across sites in a large and diverse landscape and

TABLE 5. Taxa with ratios of observed frequencies (OF) to expected frequencies (EF) that indicated intolerance (OF/EF < 0.5) or tolerance (OF/EF > 2) of stressors at test sites in basins with different land uses in the Appalachian ecoregion. Only taxa that were predicted to occur or were detected at >10 sites were included. No fish taxon satisfied the above criterion in mining basins. $n$ = number of samples of each assemblage collected in the landuse class.

| **Assemblage**/land use | Intolerant | Tolerant |
|---|---|---|
| **Fish** | | |
| Agriculture ($n$ = 35) | *Campostoma anomalum* | *Catostomum commersoni* |
| | Cottidae | *Pimephales notatus* |
| | *Cyprinella galactura* | *Rhinichthys cataractae* |
| | *Etheostoma blennioides* | *Ameiurus natalis* |
| | *Etheostoma flabellare* | *Etheostoma olmstedi* |
| | *Etheostoma jessiae* | *Exoglossum maxillingua* |
| | *Etheostoma rufilineatum* | *Lepomis cyanellus* |
| | *Lepomis megalotis* | *Lepomis gibbosus* |
| | *Moxostoma duquesnei* | *Lepomis macrochirus* |
| | *Nocomis micropogon* | *Luxilus coccogenis* |
| | *Notropis amblops* | *Micropterus salmoides* |
| | *Notropis leuciodus* | *Notropis hudsonius* |
| | *Notropis telescopu* | *Noturus insignis* |
| | *Notropis volucellus* | *Rhinichthys atratulus* |
| | *Phenacobius uranops* | *Semotilus corporalis* |
| | | *Fundulus diaphanous* |
| Urban ($n$ = 31) | *Etheostoma coosae* | *Micropterus salmoides* |
| | *Luxilus chrysocephalus* | |
| | *Notropis chrosomus* | |
| | *Notropis xaenocephalus* | |
| | *Percina nigrofasciata* | |
| **Macroinvertebrate** | | |
| Agriculture ($n$ = 60) | Corydalidae | *Cardiocladius* |
| | Gomphidae | Asellidae |
| | Hemerodromiinae | *Cladotanytarsus* |
| | *Isonychia* | *Cricotopus* |
| | Perlidae | *Cryptochironomus* |
| | Philopotamidae | *Dubiraphia* |
| | *Psephenus* | Ephemeridae |
| | *Tanytarsus* | *Eukiefferiella* |
| | | *Gammarus* |
| | | Leptophlebiidae |
| | | *Pagastia* |
| | | *Parametriocnemus* |
| | | *Paratanytarsus* |
| | | Pleuroceridae |
| | | *Promoresia* |
| | | *Sialis* |
| | | Sphaeriidae |
| | | *Thienemanniella* |
| | | Turbellaria |
| Urban ($n$ = 46) | Corydalidae | Ancylidae |
| | Ephemerellidae | Asellidae |
| | Heptageniidae | *Cardiocladius* |
| | *Isonychia* | *Eukiefferiella* |
| | Perlidae | *Gammarus* |
| | Philopotamidae | Oligochaeta |
| | | *Paratanytarsus* |
| | | *Corbicula* |
| Mining ($n$ = 34) | Cambaridae | |
| | *Isonychia* | |
| | Perlidae | |
| | *Psephenus* | |
| | *Rheotanytarsus* | |
| | *Stenelmis* | |
| | *Tanytarsus* | |
| | *Thienemannimyia* | |

TABLE 5. Continued.

| **Assemblage**/land use | Intolerant | Tolerant |
|---|---|---|
| **Diatom** | | |
| Agriculture ($n = 45$) | *Achnanthes subhudsonis* var. *kraeuselii*<br>*Cymbella affinis*<br>*Fragilaria vaucheriae*<br>*Navicula germainii* | *Navicula lanceolata*<br>*Nitzschia inconspicua*<br>*Planothidium lanceolatum*<br>*Cyclotella meneghiniana*<br>*Gomphonema olivaceum*<br>*Nitzschia archibaldii*<br>*Navicula cryptocephala*<br>*Navicula trivialis*<br>*Staurosirella pinnata*<br>*Nitzschia palea*<br>*Sellaphora pupula* |
| Urban ($n = 44$) | *Achnanthes subhudsonis* var. *kraeuselii*<br>*Cymbella affinis*<br>*Melosira varians* | *Navicula lanceolata*<br>*Nitzschia inconspicua*<br>*Caloneis bacillum*<br>*Cyclotella meneghiniana*<br>*Navicula cryptocephala*<br>*Navicula gregaria*<br>*Navicula minima*<br>*Nitzschia frustulum*<br>*Staurosirella pinnata*<br>*Nitzschia palea*<br>*Nitzschia sinuata* var. *tabellaria*<br>*Gomphonema minutum*<br>*Navicula schroeteri* var. *escambia*<br>*Navicula subminuscula*<br>*Navicula rostellata* |
| Mining ($n = 16$) | *Cocconeis placentula*<br>*Navicula capitatoradiata*<br>*Planothidium lanceolatum* | *Gomphonema parvulum*<br>*Brachysira microcephala*<br>*Diatoma moniliformis* |

among different taxonomic groups. The most severely impaired biological assemblages occurred in streams draining basins with predominantly urban and agriculture land uses (Fig. 3), results that are generally consistent with the observations of others (e.g., Paul and Meyer 2001, Moore and Palmer 2005). Stream chemistry in basins dominated by urban and agriculture land use was predictably altered relative to reference basins (Table 2), but relationships between assemblage O/E values and physicochemical attributes provided limited insight into the types of stressors most likely to be affecting biota (Table 3). For example, O/E values for macroinvertebrate assemblages implied that alteration of the riparian corridor contributed to the degradation of stream biota, but other factors commonly observed or assumed to be associated with degradation of macroinvertebrate assemblages (e.g., temperature, bed substrates) showed little difference, on average, between impaired and unimpaired assemblages. Moreover, impairment of fish assemblages was not associated with any of the physicochemical stressors that we measured. Last, relationships between the condition of diatom assemblages and potential physicochemical stressors were counterintuitive.

We think that at least 3 related reasons explain the weak relationships between physicochemical conditions and impairment of fish and diatom assemblages. First, we did not consider all possible stressors acting on Appalachian streams, particularly non-native fish species and degradation of instream habitat, both of which are known to impair fish assemblages in the Appalachians and elsewhere (USEPA 2000, de Zwart et al. 2006). Second, stressors that are episodic (e.g., pulses of acidic pH) probably were not characterized adequately because our measurements of physicochemical conditions were limited to a single grab sample on or near the day of biological sampling. Third, nutrients associated with anthropogenic disturbances might increase diatom taxonomic richness in environments where nutrients naturally limit primary production (Potapova et al. 2005). Our apparently counterintuitive finding that diatom assemblage impairment was associated with less urbanization is supported by observations that algal taxonomic richness increased or remained unchanged along urban landuse gradients (Potapova et al. 2005). In

such cases, the relative abundance of specific autecological groups appears to change dramatically, a response that would probably escape detection by indicators based on the presence/absence of taxa (i.e., RIVPACS-type models).

In summary, our measures of potential physicochemical stressors provided little information about their relative importance in affecting biological condition. This result illustrates, as previously argued by others (e.g., Yoder and Rankin 1998, Karr and Yoder 2004), that direct measures of biological quality are needed to assess the condition of ecosystems. Physicochemical measures often characterize potential biological quality ineffectively because the stressors that most strongly affect biota at a site might not be measured or analytic methods can only crudely approximate the factors to which the biota respond. Therefore, sole reliance on chemical indicators might often lead to incorrect assessments, i.e., to the conclusion that impairments do not exist when biological communities are, in fact, impaired but have not been sampled. When biological impairment is detected, understanding the reasons for the impairment might be better addressed by considering the sensitivities of those taxa that affected O/E (e.g., Hemsley-Flint 2000, deZwart et al. 2006) than attempting to infer causation from limited stressor data.

O/E is a standardized measure. Therefore, assessments of biological condition theoretically can be compared directly among taxonomic groups and regions (Hawkins 2006). The macroinvertebrate O/E values at test sites in our study were slightly higher than values observed at test sites in North Carolina and the Mid-Atlantic Highlands (Hawkins 2006), similar to values observed in streams in Maine and Ohio (Hawkins 2006), and considerably lower than values observed in logged watersheds in California (Hawkins et al. 2000b). In general, these comparisons imply that the biological impairment of streams sampled by NAWQA in the Appalachian ecoregion was similar in magnitude to the impairment of streams in the midwestern and other parts of the eastern USA.

*Sensitive and tolerant taxa*

Predictive models generate site-specific estimates of the Pc for each taxon, and these estimates can enhance interpretations of biological assessments. Taxa that occur more or less frequently than expected across a set of test sites might reveal potential causes of impairment if tolerances to stressors are known for these taxa. We illustrate this idea with results from mining basins, where diatom and macroinvertebrate assemblages frequently were impaired. The loss of *Isonychia* (Ephemeroptera:Isonychiidae), Perlidae (Plecoptera), *Psephenus* (Coleoptera:Psephenidae), and *Tanytarsus* (Chironomidae:Diptera) from streams in mining basins is consistent with experimental and observational studies showing that these taxa are sensitive to dissolved metals and reduced pH associated with mine effluent (Clements et al. 1992, Chambers and Messinger 2001). Similarly, the alteration of diatom assemblages in mining basins was largely consistent with known tolerances of specific taxa and indicated that reduced pH and elevated metals were the primary stressors. *Cocconeis placentula* var. *euglypta* and *Navicula capitatoradiata* are well-known alkaliphilous taxa (van Dam et al. 1994), and both species were absent from most samples collected from mining-affected sites at which they were predicted to occur. *Cocconeis placentula* also was absent from mining-affected Ohio streams (Verb and Vis 2000). Two diatom species that occurred more often than predicted in mining basins, *Brachysira microcephala* (= *Brachysira vitrea*) and *Diatoma monliliformis*, have high $SO_4^{2-}$ optima (Potapova and Charles 2003). *Brachysira microcephala* also tolerates acidic conditions (Krammer and Lange-Bertalot 1986, Wolfe and Kling 2001) and has been reported from streams in which pH fluctuates between neutral and acidic (Verb and Vis 2000). *Diatoma monliliformis* prefers high concentrations of ions and is sometimes associated with mineral pollution (Bahls et al. 1984). Although the stressors created by mining activities are well known, this example demonstrates that patterns in assemblage composition derived from predictive models can be integrated with autecological and functional traits to diagnose potential causes of biological impairment.

The distributions of fish, macroinvertebrate, and diatom taxa across urban and agriculture basins also were consistent with generally recognized tolerance classifications. *Cottus* sp. were absent from most of the agriculture basins. Species-level identifications are difficult for this taxon, but the genus as a whole appears to be sensitive to a variety of anthropogenic stressors (Barbour et al. 1999). Longear sunfish (*Lepomis megalotis*) and several darter species (*Etheostoma*) were conspicuously absent from urban and agriculture basins, and these species are sensitive to most anthropogenic disturbances (Barbour et al. 1999, Scott 2006). Conversely, many fish species that are widely distributed or tolerant (Barbour et al. 1999, Scott 2006) were overrepresented in streams draining urban and agriculture basins. The macroinvertebrate taxa Asellidae, Oligochaeta, and *Corbicula* also occurred more frequently than expected in streams in urban basins, and these taxa are generally tolerant of disturbed conditions (Barbour et al. 1999, Voshell

2002). The Asian clam (*Corbicula*) possesses adaptations that enable it to persist in systems with highly variable flow and substrate conditions (McMahon 1999), characteristics that are typical of urban streams (Paul and Meyer 2001). The diatom taxa found more frequently than expected in streams in urban and agriculture basins (several species of *Navicula, Nitzschia,* and *Sellaphora*) are thought to be indicative of excess siltation and nutrient enrichment (Potapova and Charles 2006).

*Biological assessments based on multiple assemblages*

Use of multiple assemblages in biological assessments should ensure a more robust characterization of overall biological integrity than is possible with 1 assemblage. We should expect that indicators based on different taxa would respond differently to the same set of stressors because different taxonomic groups typically occupy different regions of overall niche space and have very different environmental requirements and tolerances. Our results are consistent with several studies that show that most taxonomic groups are not concordant in their response to either natural or stress gradients (Paavola et al. 2003, 2006, Passey et al. 2004, Griffith et al. 2005, Johnson et al. 2006). This lack of concordance has 2 main implications. First, an accurate inference regarding the overall biological integrity of an ecosystem will require examination of multiple assemblages. Second, we need ways to integrate the results of multiple assessments into a single inference regarding the overall biological condition of a water body. The second implication will require that we address 3 issues: 1) Should we weight results from different assemblages and, if so, how? 2) Should multiple assessments be integrated into a single score, or should they be applied independently? 3) What implications do differences in assessment error among groups have for setting thresholds defining impairment?

Addressing these issues will require considerable discussion and analysis and is largely outside of the scope of our paper. However, we offer the following comments as a means of starting the dialogue that policymakers will need as their agencies move toward more frequent use of multiple assemblages in bioassessments. If weighting is used, we suggest that it be based on the ecological importance of each group. However, characterizing the ecological importance of different groups will not be a trivial task and will require deep understanding of the direct ecological services provided by each set of taxa and of their importance in maintaining overall ecosystem integrity. Integration of assessments into a single score probably will have appeal for managers and policymakers because of its ease of communication. O/E is amenable to simple integration because its units of measure are the same across assemblages. However, simple averaging of O/E values could easily hide serious alterations in one assemblage if other assemblages were unaffected. For example, if diatoms, invertebrates, and fish had O/E values of 0.95, 0.98, and 0.5, then the average value would be 0.81, a value near the statistical threshold often used to infer from a single assemblage that a site is different from reference. However, in this case, the site would be missing 50% of its expected fish taxa, which might very well be an unacceptable condition. We think a more ecologically defensible approach would be to apply the assessments independently on multiple assemblages.

We can envision at least 2 ways to use independent assessments to determine if a water body is meeting its designated biological uses (i.e, similar to reference or some other agreed upon condition). First, define threshold values for each group based on the uncertainty in estimated O/E associated with each group. If the values for any group fall below that threshold value, conclude that the site is not in reference condition. A more sophisticated approach would be to use the joint distributions of the reference-site values for each group and determine whether the joint O/E values observed at a test site are outside this multivariate distribution. Bowman and Somers (2006) describe one such procedure for conducting this type of analysis.

Ideally, the error associated with estimating O/E would be similar for each group, so that the likelihood of committing type I and type II statistical errors would be similar among groups. In our study, the invertebrate model was more precise than the diatom and fish models (which were similar in precision). Thus, more subtle alterations could be detected in invertebrate assemblages than in fish or diatom assemblages. To put the 3 assemblages on similar footing, the performance of the diatom and fish models would have to be improved or the threshold used to infer alteration in the invertebrate assemblage would have to changed to be more similar to those used for diatoms and fish. A more sensible approach would be to determine on ecological grounds what level of alteration was important to detect in each assemblage and then refine the models until those levels of uncertainty were achieved.

Uncertainty in O/E estimates is associated primarily with the ability of models to describe real relationships between environmental gradients and assemblage composition, and this ability is strongly influenced by the number of observations used to calibrate

models. In some cases, model performance could be improved by adding reference sites. However, this approach also has its limits. Ultimately, model performance is constrained by sampling error and by the degree to which taxonomic composition is controlled by spatial gradients in measurable environmental features. Sampling error is a function of the heterogeneity in species distributions within reaches and sampling effort. In some cases, the sampling effort required to achieve a desired sampling error might be prohibitively large.

An equally important constraint on model performance is the fact that different types of assemblages might be structured by different processes, some of which operate at different spatial and temporal scales. For example, the fish model probably was less precise than the invertebrate model because the conditions within any specific reach do not represent the full set of factors required by most fish species. Many species use upstream or downstream sections to meet certain life-history needs. If those factors are lacking, the species probably will be absent from the sampled reach even if the conditions there are otherwise suitable. The diatom model was probably less precise than the invertebrate model for different reasons. In this case, because of their short life histories, diatom assemblages are probably more sensitive to seasonal variation in conditions within a reach than are either invertebrates or fish assemblages. Variation in sampled assemblage composition associated with these temporal signals will increase model error regardless of how well the assemblage in characterized at any given time.

*Use of predictive models for large-scale assessments*

Comparison of site assessments among regions or larger study areas can be complicated by many natural (McCormick et al. 2000) and anthropogenic (e.g., inconsistent field, laboratory, and analytical methods) factors (Hawkins et al. 2000a, Hawkins 2006). Although imperfect for the reasons discussed above, the predictive-model framework enabled us to correct for potentially confounding effects of natural gradients, use a consistent measure of biological condition, and apply a generally uniform assessment criterion across several distinct river basins within the Appalachian ecoregion. Such a standardized approach should be of value to water-resources scientists who need to compare and interpret biological data across geographically, geologically, climatically, and hence, ecologically diverse basins (Heinz Center 2002, Hawkins 2006, USEPA 2006). This capability is particularly important for the USGS NAWQA program, in which the ecological effects of human activities (e.g., urbanization and nutrient applications) on the land are studied in disparate river basins across the nation.

## Acknowledgements

We thank the many USGS personnel who diligently applied sampling protocols. Colleague reviews by Terry Maret, Chris Mebane, and Sandy Cooper greatly improved the clarity of the manuscript. We also appreciate the suggestions of Leska Fore, Pamela Silver, and 2 anonymous referees. CPH was supported, in part, by EPA Science To Achieve Results (STAR) grant R-82863701.

## Literature Cited

ANGERMEIER, P. L., AND M. R. WINSTON. 1999. Characterizing fish community diversity across Virginia landscapes: prerequisite for conservation. Ecological Applications 9: 335–349.

BAHLS, L. L., E. E. WEBER, AND J. O. JARVIE. 1984. Ecology and distribution of major diatom ecotypes in the southern Fort Union coal region of Montana. U.S. Geological Survey Professional Paper 1289. US Government Printing Office, Washington, DC.

BARBOUR, M. T., J. GERRITSEN, B. D. SNYDER, AND J. B. STRIBLING. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. 2nd edition. EPA 841-B-99–002. Office of Water, US Environmental Protection Agency, Washington, DC.

BOWMAN, M. F., AND K. M. SOMERS. 2006. Evaluating a novel test site analysis (TSA) bioassessment approach. Journal of the North American Benthological Society 25:712–727.

CARTER, J. L., V. H. RESH, D. M. ROSENBERG, AND T. B. REYNOLDSON. 2006. Biomonitoring in North American rivers: a comparison of methods used for benthic macroinvertebrates in Canada and the United States. Pages 203–228 *in* G. Ziglio, M. Siligardi, and G. Flaim (editors). Biological monitoring of rivers. John Wiley and Sons, West Sussex, UK.

CHAMBERS, D. B., AND T. MESSINGER. 2001. Benthic macroinvertebrate communities and their responses to selected environmental factors in Kanawha River Basin, West Virginia, Virginia, and North Carolina. U.S. Geological Survey Water Resources Investigation Report 01–4021. US Geological Survey, Charleston, West Virginia.

CHARLES, D. F., C. KNOWLES, AND R. S. DAVIS. 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. Patrick Center for Environmental Research, Academy of Natural Sciences, Philadelphia, Pennsylvania.

CLARKE, R. 2000. Uncertainty in estimates of biological quality based on RIVPACS. Pages 39–54 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters. Freshwater Biological Association, Ambleside, UK.

CLARKE, R. T., J. R. WRIGHT, AND M. T. FURSE. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160:219–233.

CLEMENTS, W. H., D. S. CHERRY, AND J. H. VAN HASSEL. 1992. Assessment of the impact of heavy metals on benthic communities at the Clinch River (Virginia): evaluation of an index of community sensitivity. Canadian Journal of Fisheries and Aquatic Sciences 49:1686–1694.

CUFFNEY, T. F., M. E. GURTZ, AND M. R. MEADOR. 1993. Methods for collecting benthic macroinvertebrate samples as part of the National Water-Quality Assessment Program. U.S. Geological Survey Open File Report 93–406. US Geological Survey, Reston, Virginia.

DAVIES, S. P., AND S. K. JACKSON. 2006. The Biological Condition Gradient: a conceptual model for interpreting detrimental change in aquatic ecosystems. Ecological Applications 16:1251–1266.

DEZWART, D., S. D. DYER, L. POSTHUMA, AND C. P. HAWKINS. 2006. Use of predictive models to attribute potential effects of mixture toxicity and habitat alteration on the biological condition of fish assemblages. Ecological Applications 16:1295–1310.

DUFRÊNE, M., AND P. LEGENDRE. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67:345–366.

EDWARDS, M. R., D. L. COMBS, AND S. B. COOK. 2003. Comparison of single-pass electrofishing to depletion sampling for surveying fish assemblages in small warm-water streams. Journal of Freshwater Ecology 18:625–634.

FITZPATRICK, F. A., I. R. WAITE, P. J. D'ARCONTE, M. R. MEADOR, M. A. MAUPIN, AND M. E. GURTZ. 1998. Revised methods for characterizing stream habitat—National Water-Quality Assessment Program. U.S. Geological Survey Water-Resources Investigations Report 98–4052. US Geological Survey, Raleigh, North Carolina.

GILLIOM, R. J., W. M. ALLEY, AND M. E. GURTZ. 1995. Design of the national water-quality assessment program: occurrencedistribution of water-quality conditions. Geological Survey Circular U.S. 1112. US Geological Survey, Sacramento, California.

GRIFFITH, M. B., B. H. HILL, F. H. MCCORMICK, P. R. KAUFMANN, A. T. HERLIHY, AND A. R. SELLE. 2005. Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. Ecological Indicators 5:117–136.

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., AND D. M. CARLISLE. 2001. Use of predictive models for assessing the biological integrity of wetlands and other aquatic habitats. Pages 59–84 *in* R. B. Rader, D. P. Batzer, and S. A. Wissinger (editors). Biomonitoring and management of North American freshwater wetlands. John Wiley and Sons, New York.

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 19:541–556.

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 10:1456–1477.

HEINZ CENTER. 2002. The state of the nation's ecosystems. Cambridge University Press, New York, New York.

HEMSLEY-FLINT, B. 2000. Classification of the biological quality of rivers in England and Wales. Pages 55–70 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters. Freshwater Biological Association, Ambleside, UK.

JOHNSON, R. K., D. HERING, M. T. FURSE, AND P. F. M. VERDONSCHOT. 2006. Indicators of ecological change: comparisons of the early response of four organism groups to stress gradients. Hydrobiologia 566:139–152.

KARR, J. R., AND C. O. YODER. 2004. Biological assessment and criteria improve TMDL planning and decision-making. Journal of Environmental Engineering 130:594–604.

KNAPP, R. A., C. P. HAWKINS, J. LADAU, AND J. G. MCCLORY. 2005. Fauna of Yosemite National Park lakes has low resistance but high resilience to fish introductions. Ecological Applications 15:835–847.

KRAMMER, K., AND H. LANGE-BERTALOT. 1986. Bacillariophyceae. 1. Teil: Naviculaceae. Die Süßwasserflora Mitteleuropa 2/1. J. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer (editors). G. Fisher, Stuttgart, Federal Republic of Germany.

LAMOUROUX, N., N. L. POFF, AND P. L. ANGERMEIER. 2002. Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients. Ecology 83:1792–1807.

LEOPOLD, L. B., M. G. WOLMAN, AND J. P. MILLER. 1964. Fluvial processes in geomorphology. W. H. Freeman, San Francisco, California.

MAYDEN, R. L. 1987. Historical ecology and North American highland fishes: a research program in community ecology. Pages 203–222 *in* W. J. Mathews and D. C. Heins (editors). Community and evolutionary ecology of North American stream fishes. University of Oklahoma Press, Norman, Oklahoma.

MCCORMICK, F. H., D. V. PECK, AND D. P. LARSEN. 2000. Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. Journal of the North American Benthological Society 19:385–404.

MCCUNE, B., AND M. J. MEFFORD. 1999. PC-ORD. Multivariate analysis of ecological data, version 4. MjM Software Design, Gleneden Beach, Oregon.

MCMAHON, R. F. 1999. Invasive characteristics of the freshwater bivalve, *Corbicula fluminea*. Pages 315–343 *in* R. Claudi and J. H. Leach (editors). Nonindigenous freshwater organisms: vectors, biology and impacts. Lewis Publishers, Boca Raton, Florida.

MEADOR, M. R., T. F. CUFFNEY, AND M. E. GURTZ. 1993. Methods for sampling fish communities as part of the National

Water-Quality Assessment Program. U.S. Geological Survey Open File Report 93–104. US Geological Survey, Raleigh, North Carolina.

MEADOR, M. R., J. P. MCINTYRE, AND K. H. POLLOCK. 2003. Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure. Transactions of the American Fisheries Society 132:39–46.

MOORE, A. A., AND M. A. PALMER. 2005. Macroinvertebrate biodiversity in agricultural and urban headwater streams: implications for conservation and management. Ecological Applications 15:1169–1177.

MOSS, D., M. T. FURSE, J. F. WRIGHT, AND P. D. ARMITAGE. 1987. The prediction of the macro-macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41–52.

MOULTON, S. R., J. L. CARTER, S. A. GROTHEER, T. F. CUFFNEY, AND T. M. SHORT. 2000. Methods of analysis by the US Geological Survey National Water-Quality Laboratory: processing, taxonomy, and quality control of benthic macroinvertebrate samples. U.S. Geological Survey Open File Report 00–212. US Geological Survey, Reston, Virginia.

MOULTON, S. R., J. G. KENNEN, R. M. GOLDSTEIN, AND J. A. HAMBROOK. 2002. Revised protocols for sampling algal, macroinvertebrate, and fish communities as part of the National Water-Quality Assessment program. U.S. Geological Survey Open-file Report 02–150. US Geological Survey, Reston, Virginia.

NRC (NATIONAL RESEARCH COUNCIL) 2000. Ecological indicators for the nation. National Academy Press, Washington, DC.

OCS (OREGON CLIMATE SERVICE). 2004. PRISM data set. Oregon State University, Corvallis, Oregon. (Available from: http://prism.oregonstate.edu/)

OSTERMILLER, J. D., AND C. P. HAWKINS. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23:363–382.

PAAVOLA, R., T. MUOTKA, R. VIRTANEN, J. HEINO, D. JACKSON, AND A. MÄKI-PETÄYS. 2006. Spatial scales effects community concordance among fishes, benthic macroinvertebrates, and bryophytes in streams. Ecological Applications 16: 368–379.

PAAVOLA, R., T. MUOTKA, R. VIRTANEN, J. HEINO, AND P. KREVIVI. 2003. Are biological classifications of headwater streams concordant among multiple taxonomic groups? Freshwater Biology 48:1912–1923.

PASSY, S. I., R. W. BODE, D. M. CARLSON, AND M. A. NOVAK. 2004. Comparative environmental assessment in the studies of benthic diatom, macroinvertebrate, and fish communities. International Review of Hydrobiology 89: 121–138.

PAUL, M. J., AND J. L. MEYER. 2001. Streams in the urban landscape. Annual Review of Ecology, Evolution, and Systematics 32:333–365.

PAYBINS, K. S., T. MESSINGER, J. H. EYCHANER, D. B. CHAMBERS, AND M. D. KOZAR. 2000. Water quality in the Kanawha-New River basin, West Virginia, Virginia, and North Carolina, 1996–98. U.S. Geological Survey Circular 1204. US Geological Survey, Reston, Virginia.

POFF, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction stream ecology. Journal of the North American Benthological Society 16:391–409.

PORTER, S. D., T. F. CUFFNEY, M. E. GURTZ, AND M. R. MEADOR. 1993. Methods for collecting algal samples as part of the National Water-Quality Assessment Program. U.S. Geological Survey Open-file Report 93–049. US Geological Survey, Reston, Virginia.

POTAPOVA, M., AND D. F. CHARLES. 2003. Distribution of diatoms in U.S. rivers in relation to conductivity and ionic composition. Freshwater Biology 48:1311–1328.

POTAPOVA, M., AND D. F. CHARLES. 2006. Diatom metrics for measuring eutrophication in rivers of the United States. Ecological Indicators 7:48–70.

POTAPOVA, M., J. F. COLES, E. M. GIDDINGS, AND H. ZAPPIA. 2005. A comparison of the influences of urbanization in contrasting environmental settings on stream benthic algal assemblages. Pages 333–359 *in* L. R. Brown, R. H. Gray, R. H. Hughs, and M. R. Meador (editors). Effects of urbanization on stream ecosystems. American Fisheries Society Symposium 47. American Fisheries Society, Bethesda, Maryland.

SCOTT, M. C. 2006. Winners and losers among stream fishes in relation to land use legacies and urban development in the southeastern US. Biological Conservation 127:301–309.

SHELTON, L. R. 1994. Field guide for collecting and processing stream-water samples for the national water-quality assessment program. U.S. Geological Survey Open File Report 94–455. US Geological Survey, Sacramento, California.

SIMONSON, T. D., AND J. LYONS. 1995. Comparison of catch per effort and removal procedures for sampling steam fish assemblages. North American Journal of Fisheries Management 15:419–427.

STEVENSON, R. J. 1997. Scale-dependent determinants and consequences of benthic algal heterogeneity. Journal of the North American Benthological Society 16:248–262.

STODDARD, J. L., D. P. LARSEN, C. P. HAWKINS, R. K. JOHNSON, AND R. H. NORRIS. 2006. Setting expectations for the ecological condition of running waters: the concept of reference condition. Ecological Applications 16:1267–1276.

TATE, C. M., T. F. CUFFNEY, G. MCMAHON, E. M. P. GIDDINGS, J. F. COLES, AND H. ZAPPIA. 2005. Use of an urban intensity index to assess urban effects on streams in three contrasting environmental settings. Pages 219–315 *in* L. R. Brown, R. H. Gray, R. H. Hughs, and M. R. Meador (editors). Effects of urbanization on stream ecosystems. American Fisheries Society Symposium 47. American Fisheries Society, Bethesda, Maryland.

USDA (US DEPARTMENT OF AGRICULTURE). 2004. State soil geographic database. US Department of Agriculture, Washington, DC. (Available from: http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/)

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2000. Mid-

Atlantic highlands stream assessment. EPA/903/R-00/015. US Environmental Protection Agency, Philadelphia, Pennsylvania.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2002a. National water quality inventory: 2000 report. EPA 841-R-02–001. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2002b. Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: streams and wadeable rivers. EPA 822-R-02–048. Office of Environmental Information and Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2005. Use of biological information to better define designated aquatic life uses in state and tribal water quality standards. EPA 822-R-05–001. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2006. Draft wadeable streams assessment: a collaborative survey of the Nation's streams. EPA 841-B-06–002. Office of Water, US Environmental Protection Agency, Washington, DC.

USGAO (US GENERAL ACCOUNTING OFFICE). 2000. Water quality: key EPA and state decisions limited by inconsistent and incomplete data. Report to Chairman, Subcommittee on Water Resources and the Environment, Committee on Transportation and Infrastructure, House of Representatives. GAO/RCED-00–54. US General Accounting Office, Washington, DC.

USGS (US GEOLOGICAL SURVEY). 2004a. National elevation database. US Geological Survey, Reston, Virginia. (Available from: http://gisdata.usgs.gov/ned/)

USGS (US GEOLOGICAL SURVEY). 2004b. National hydrography dataset. US Geological Survey, Reston, Virginia. (Available from: http://nhd.usgs.gov/)

USGS (US GEOLOGICAL SURVEY). 2004c. National land cover characterization. US Geological Survey, Reston, Virginia. (Available from: http://landcover.usgs.gov/natllandcover.asp)

VADAS, R. L., AND D. J. ORTH. 1993. A new technique for estimating the abundance and habitat use of stream fishes. Journal of Freshwater Ecology 8:305–317.

VAN DAM, H., A. MERTENS, AND J. SINKELDAM. 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. Netherlands Journal of Aquatic Ecology 28:117–133.

VAN SICKLE, J., C. P. HAWKINS, D. P. LARSEN, AND A. T. HERLIHY. 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24:178–191.

VAN SICKLE, D. H. HUFF, AND C. P. HAWKINS. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology 51:359–372.

VERB, R. G., AND M. L. VIS. 2000. Comparison of benthic diatom assemblages from streams draining abandoned and reclaimed coal mines and nonimpacted sites. Journal of the North American Benthological Society 19:274–288.

VINSON, M. R., AND C. P. HAWKINS. 1998. Biodiversity of stream insects: variation at local, basin, and regional scales. Annual Review of Entomology 43:271–293.

VOSHELL, J. R. 2002. A guide to common freshwater invertebrates of North America. McDonald and Woodward Publishing, Blacksburg, Virginia.

WALSH, S. J., AND M. R. MEADOR. 1998. Guidelines for quality assurance and quality control of fish taxonomic data collected as part of the National Water-Quality Assessment Program. U.S. Geological Survey Water Resources Investigations Report 98–4239. US Geological Survey, Raleigh, North Carolina.

WILDE, F. D., AND D. B. RADTKE (EDITORS). 1998. National field manual for the collection of water-quality data. U.S. Geological Survey techniques of water-resources investigations. Book 9. Handbooks for water-resources investigations. Chapter A6, Field measurements. US Government Printing Office, Washington, DC.

WOLFE, A. P., AND H. J. KLING. 2001. A consideration of some North American soft-water *Brachisyra* taxa and description of *B. arctoborealis* sp. Nov. Pages 243–264 *in* R. Jahn, J. P. Kociolek, A. Witkowski, and P. Compère (editors). Lange-Bertalot-Festschrift: Studies on diatoms. Koeltz, Koenigstein, Germany.

WRIGHT, J. F. 2000. An introduction to RIVPACS. Pages 1–24 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters. Freshwater Biological Association, Ambleside, UK.

YODER, C. O., AND E. T. RANKIN. 1998. The role of biological indicators in a state water quality management process. Environmental Monitoring and Assessment 51:61–88.

*Received: 9 August 2006*
*Accepted: 20 September 2007*

APPENDIX 1. Fish indicator species (sensu Dufrêne and Legendre 1997) for each of 10 groups of reference sites in the Appalachian ecoregion. Group numbers match those in Fig. 2A. One group (11) had no indicator species.

| Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
|---|---|---|---|---|
| Cottidae | *Luxilus cornutus* | *Rhinichthys atratulus* | *Notropis chrosomus* | *Noturus insignis* |
| *Ichthyomyzon greeleyi* | *Stizostedion canadense* | *Salvelinus fontinalis* | *Etheostoma coosae* | *Nocomis leptocephalus* |
| *Notropis spectrunculus* | *Notropis atherinoides* | *Semotilus atromaculatus* | *Percina nigrofasciata* | *Luxilus albeolus* |
| *Catostomus commersoni* | *Semotilus corporalis* | *Etheostoma flabellare* | *Notropis xaenocephalus* | *Notropis hudsonius* |
| | | *Clinostomus funduloides* | *Hypentelium etowanum* | *Micropterus dolomieu* |
| | | *Moxostoma rhothoecum* | *Micropterus coosae* | *Ambloplites rupestris* |
| | | | *Campostoma oligolepis* | *Notropis volucellus* |
| | | | *Lepomis cyanellus* | *Hypentelium nigricans* |
| | | | *Etheostoma stigmaeum* | *Percina roanoka* |
| | | | *Cyprinella venusta* | *Etheostoma blennioides* |
| | | | *Lepomis microlophus* | *Notropis rubellus* |
| | | | *Phenacobius catostomus* | *Rhinichthys cataractae* |
| | | | *Notropis stilbius* | *Percina oxyrhynchus* |
| | | | *Moxostoma poecilurum* | *Campostoma anomalum* |
| | | | *Lepomis miniatus* | *Etheostoma caeruleum* |
| | | | *Percina kathae* | *Pylodictis olivaris* |
| | | | *Cyprinella callistia* | |

| Group 6 | Group 7 | Group 8 | Group 9 | Group 10 |
|---|---|---|---|---|
| *Notropis leuciodus* | *Cyprinella galactura* | *Cyprinella spiloptera* | *Percina aurantiaca* | *Lythrurus lirus* |
| *Luxilus coccogenis* | *Etheostoma rufilineatum* | *Etheostoma zonale* | | *Phenacobius uranops* |
| *Nocomis micropogon* | *Notropis photogenis* | *Ictalurus punctatus* | | *Notropis amblops* |
| *Etheostoma swannanoa* | *Moxostoma erythrurum* | *Moxostoma duquesnei* | | *Notropis telescopus* |
| *Etheostoma chlorobranchium* | *Dorosoma cepedianum* | *Phenacobius crassilabrum* | | *Etheostoma jessiae* |
| | *Etheostoma simoterum* | *Moxostoma carinatum* | | *Luxilus chrysocephalus* |
| | *Lepomis auritus* | *Etheostoma acuticeps* | | *Erimystax dissimilis* |
| | *Noturus eleutherus* | *Carpiodes carpio* | | *Pimephales notatus* |
| | *Etheostoma camurum* | | | *Lepomis megalotis* |
| | *Pomoxis annularis* | | | |

APPENDIX 2. Macroinvertebrate indicator taxa (sensu Dufrêne and Legendre 1997) for each of 9 groups of reference sites in the Appalachian ecoregion. Group numbers match those in Fig. 2B.

| Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | Group 7 | Group 8 | Group 9 |
|---|---|---|---|---|---|---|---|---|
| Brachycentridae | *Stenelmis* | *Orthocladius* | *Phaenopsectra* | *Cricotopus/ Orthocladius* | Philopotamidae | *Cardiocladius* | Chloroperlidae | Gomphidae |
| Ephemerellidae | *Isonychia* | Perlidae | *Microtendipes* | *Krenosmittia* | *Rheocricotopus* | | Leptophlebiidae | *Psephenus* |
| Turbellaria | *Caenis* | *Thienemanniella* | Oligochaeta | *Cricotopus* | Simuliidae | | *Micropsectra* | *Thienemannimyia* |
| *Neophylax* | Hydroptilidae | | *Cryptochironomus* | *Paratanytarsus* | Pleuroceridae | | *Rhyacophila* | *Larsia* |
| | Psychomyiidae | | *Stempellinella* | *Nanocladius* | Asellidae | | Leuctridae | Dolichopodidae |
| | *Ephoron* | | *Nilothauma* | *Pagastia* | | | Baetidae | Polycentropodidae |
| | *Tricorythodes* | | Physidae | *Antocha* | | | *Oulimnius* | Tabanidae |
| | Sphaeriidae | | *Tanytarsus* | *Sublettea* | | | Peltoperlidae | *Demicryptochironomus* |
| | *Microcylloepus* | | *Dicrotendipes* | Acari | | | *Parachaetocladius* | *Hydraena* |
| | Hydropsychidae | | Ephemeridae | *Parametriocnemus* | | | *Micropsectra/ Tanytarsus* | Muscidae |
| | Heptageniidae | | | | | | Glossosomatidae | *Paramerina* |
| | | | | | | | *Hexatoma* | *Odontomyia* |
| | | | | | | | Capniidae | |

APPENDIX 3. Diatom indicator taxa (sensu Dufrêne and Legendre 1997) for each of 6 groups of reference sites in the Appalachian ecoregion. Group numbers match those in Fig. 2C.

| Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 |
|---|---|---|---|---|---|
| *Encyonema silesiacum* | *Caloneis bacillum* | *Navicula cryptotenella* | *Gomphonema apuncto* | *Nitzschia sociabilis* | *Sellaphora seminulum* |
| *Synedra ulna* | *Navicula tripunctata* | *Fragilaria capucina* var. *gracilis* | *Nitzschia sinuata* var. *tabellaria* | *Gomphonema minutum* | *Navicula antonii* |
| *Cymbella cistula* | *Cocconeis pediculus* | *Cymbella tumida* | *Cymbella delicatula* | *Navicula reichardtiana* | *Navicula caterva* |
| *Fragilaria capucina* | *Pleurosira laevis* | *Navicula germainii* | *Gomphonema sphaerophorum* | *Nitzschia palea* var. *debilis* | *Nitzschia recta* |
| *Encyonema minutum* | *Rhoicosphenia abbreviata* | *Synedra ulna* var. *contracta* | *Denticula elegans* | *Karayevia clevei* | *Navicula trivialis* |
| *Nitzschia clausii* | | *Gomphonema kobayasii* | *Fragilaria tenera* | *Achnanthidium pyrenaicum* | *Nitzschia amphibia* |
| *Navicula angusta* | | *Gomphonema parvulum* | *Encyonopsis microcephala* | *Nitzschia dissipata* | *Navicula subminuscula* |
| *Geissleria schoenfeldii* | | *Gomphoneis eriense* var. *variabilis* | | *Navicula subrotundata* | *Navicula rostellata* |
| *Gomphonema angustatum* | | *Achnanthidium minutissimum* | | *Diatoma vulgaris* | *Amphora inariensis* |
| *Nitzschia acicularis* | | *Navicula symmetrica* | | *Navicula minima* | *Planothidium frequentissimum* |
| *Gomphonema olivaceoides* | | *Gomphonema angustatum* | | | *Nitzschia agnita* |
| | | *Nitzschia archibaldii* | | | *Gomphonema affine* |
| | | *Fragilaria capucina* var. *rumpens* | | | |
| | | *Nitzschia frustulum* | | | |
| | | *Navicula menisculus* | | | |
| | | *Staurosirella pinnata* | | | |