Use of Benthic Macroinvertebrate Indices to Assess Aquatic Health in a Mixed-Landuse Watershed

Christopher Riley
Great Lakes Center, Buffalo State College
Buffalo, New York 14222 USA

Shreeram Inamdar[a]
Bioresources Engineering, University of Delaware
Newark, Delaware 19716 USA

and

Christopher Pennuto
Biology, Buffalo State College
Buffalo, New York 14222 USA

ABSTRACT

Seven family-level metrics were determined from benthic macroinvertebrate samples collected monthly across four different landuse sites in the Cazenovia Creek watershed in western New York. We evaluated which metrics were most appropriate for, and effective in, discerning differences among the sites. Dramatic and consistent seasonal differences in metric scores were observed across all sites. All metric scores indicated markedly better biological health during winter months than during summer months. The biological impairment designation was highest in the summer at the suburban and agricultural sites. Coefficients of variation were considerably greater at the lower stream order sites. The metrics that were most appropriate and effective in assessing benthic assemblage health in this study were: richness, percent model affinity, family-level biotic index, and Ephemeroptera-Plecoptera-Trichoptera index. These indices correlated with a low amount of redundancy in a Pearson matrix, had significant discriminatory power in assessing biological impairments across sites, and had low variation within sites and seasons. This suggests these metrics should be selected for bioassessments in similar Great Lakes watersheds.

INTRODUCTION

Numerous studies have shown that the use of benthic macroinvertebrates is an effective monitoring tool for measuring continuous and chronic effects from pollution (Karr et al. 1985, Plafkin et al. 1989, Kashian and Burton 2000), evaluating stream degradation from storm water runoff and point source discharges (Plafkin et al. 1989), and indicating stream recovery (Yandora 1998). Benthic communities also integrate the effects of different pollutant stressors over time and provide an ecological measure of fluctuating environmental conditions. These attributes have led to the use of benthic indices for monitoring nonpoint source pollution and for watershed assessments (Karr and Yoder 2004, National Research Council 2001).

Past research has indicated that watershed landuse can affect aquatic communities in streams (Lenat and Crawford 1994, Quinn 2000, Roy et al. 2003). Urbanization frequently impacts aquatic ecosystems through increases in nonpoint source pollutants such as heavy metals, oil, pesticides, road salt, organic materials, nutrients, and sediments (Lenat and Crawford 1994, Rosenberry et al. 1999, Fitzpatrick et al. 2004). Alterations in water quality, hydrology, and channel morphology have been shown to negatively impact the resident benthic communities, reducing their richness, diversity, and overall biological health (Stepenuck et al. 2002). The influence of season on benthic communities and bioassessments, however, has not been extensively explored.

[a] Corresponding author; E-mail: inamdar@udel.edu

There can be large temporal variations in benthic community structure between seasons for both impaired and non-impaired streams (Linke et al. 1999), which can have important implications for bioassessments. Investigating how certain metrics respond across seasons may allow the researcher to compensate for these changes, select metrics that are seasonally less variable, and improve the overall accuracy of bioassessments using benthic communities.

Multimetric biological indices have proven effective in the assessment of ecological conditions in a variety of management settings and in diverse ecoregions and are currently employed in 48 states (Kashian and Burton 2000). Since certain metrics are sensitive to rare taxa while others are sensitive to abundant taxa different indices measure different aspects of macroinvertebrate assemblages and can lead to different results (Kashian and Burton 2000). Thus, by using several different metrics that analyze multiple aspects of the benthic community a more accurate measure of biological condition may be obtained (Rosenberg and Resh 1993, Bennett et al. 2004).

The key is to select metrics that are most appropriate for assessing the overall health of the benthic assemblage while limiting metric redundancy (e.g., Bennett et al. 2004). We evaluated the performance of seven metrics to examine which metrics were most appropriate for, and effective in, discerning differences among sites. Monthly benthic sampling was performed over a year across four sites representing various landuses. Water quality monitoring was also performed at the four sites. In addition, a habitat survey was done at each sampling location to assess the similarity of benthic habitats across the sites and validate the legitimacy of comparing benthic metric scores among the four sites. Specific questions investigated in this study were: (1) How do benthic indices vary across the four sites representing different landuses and water quality? (2) What are the seasonal variations in benthic indices and what are the implications for biomonitoring? (3) What selected benthic indices are most appropriate for the assessment of biological impairment across sites and seasons?

METHODS AND MATERIALS

Study watershed and sampling sites

This study was conducted in the 350 km^2 Cazenovia Creek watershed located in western New York. Cazenovia Creek watershed is a subwatershed of the Buffalo River drainage basin (1,098 km^2). Landuse in the watershed ranges from urban/suburban near the outlet of the watershed to agricultural and forested in the headwater areas. Average temperature and precipitation in the area are 8.8 °C and 95 cm/yr, respectively (USDA 1986). Four sites were selected to include all the major types of landuse that influence streams in this watershed. All sampling reaches were 100 m in length.

The suburban landuse site (SUB) was located on Cazenovia Creek in the town of West Seneca, where the creek was a fourth-order stream approximately 18 m wide and 0.5 - 1 m deep. The rural landuse site (RUR) was located on the third-order East Branch of Cazenovia Creek in the town of Holland; the channel was approximately 10 m wide and 0.5 – 1 meter deep. The forested landuse site (FOR) was on Sprague Brook, a second-order stream in Glenwood with a channel width of 5 m and depth of 0.1 – 0.8 m. NY. The site designated as the agricultural landuse site (AGR) was located in Glenwood on Spencer Brook, a second-order stream that averages 5 m in width and 0.1 to 0.75 m in depth at the sampling site.

Habitat assessment

In July 2005 a visual habitat assessment was conducted over a 100 m reach at each site (Table 1) using ten parameters of the USEPA Rapid Bioassessment Protocol (RBP; Barbour et al. 1999). Current velocity was measured during July 2003 using a

Marsh-McBirney electronic flowmeter. Except for the riparian zone width parameter, no habitat measure varied more that 25% from conditions at the forested site. In addition, substrate particle size, current velocity, and canopy cover were assessed relative to the New York State Department of Environmental Conservation (NYSDEC) habitat comparability criteria (Bode et al. 1990). The ratio of scores between a reference site and the other stations provides a percent comparability measure for each site. All four sites had scores within acceptable ranges to permit comparison among sites.

Water quality monitoring

From July 2003 to July 2004, YSI P6600 data sondes (YSI Inc.) were used to measure temperature, specific conductivity, dissolved oxygen, pH, and turbidity in 15-min intervals at each site during spring, summer, and fall and at 1-h intervals during the winter. The data sondes were housed in perforated PVC tubes, which provided protection but allowed for free flow of water. Data were downloaded bi-monthly except between December 2003 and March 2004 when no downloading or maintenance took place due to ice cover. Several months of data were rejected due to the failure of one or more sondes. Gaps in data occurred during July, October, November, December through April, and June. Still, simultaneous operation of all sondes resulted in greater than 7,000 concurrent measurements per site for each water quality parameter, except dissolved oxygen for which there were approximately 2,800 concurrent measurements per site.

Benthic macroinvertebrate sampling

Benthic sampling took place monthly from July 2003 to June 2004, with the exception of January and February due to ice cover. Samples were collected from all sites on the same day near the middle of the month, preferably during baseflow conditions, and never within seven days of a storm event. Procedures for sampling benthic macroinvertebrates were modified from Barbour et al. 1999 and samples were collected in riffle sections of the stream. A D-frame net (40-cm width) with a 0.5-mm mesh was used in the collection of benthos. Sampling began at the downstream end of

Table 1. Habitat assessment summary for four sites in the Cazenovia Creek watershed based on habitat parameters from the EPA RBP and the NYSDEC biomonitoring.

Habitat Parameter	Suburban (SUB)	Rural (RUR)	Forested (FOR)	Agricultural (AGR)
USEPA RBP				
Available substrate	Optimal	Optimal	Optimal	Optimal
Embeddedness	Optimal	Optimal	Optimal	Optimal
Velocity/depth regime	Optimal	Optimal	Optimal	Optimal
Sediment deposition	Optimal	Optimal	Optimal	Optimal
Channel flow status	Optimal	Optimal	Suboptimal	Suboptimal
Channel alteration	Suboptimal	Optimal	Optimal	Suboptimal
Frequency of riffles	Optimal	Optimal	Optimal	Optimal
Bank stability	Optimal	Optimal	Suboptimal	Suboptimal
Vegetative protection	Suboptimal	Optimal	Optimal	Suboptimal
Riparian zone width (m)	5-10	>100	>100	10-100
NYSDEC criteria				
Velocity (cm/s)	49	38	30	34
Substrate phi	-3.9	-3.4	-4.4	-3.9
Canopy cover (%)	0	20	10	5

the reach and proceeded upstream. A total of 20 kick samples of 20 seconds each was taken over the length of the reach. The 20 kick samples were consolidated into a single homogeneous sample, preserved with ethanol and processed in the laboratory following slightly modified procedures of Bode et al. (1996). All macroinvertebrates were picked from randomly selected grid squares until 100 organisms had been counted. Identification was to the family level.

Computation of benthic metrics

The seven metrics computed were: family-level taxa richness (RICH); Shannon-Weiner diversity index (SDI): percent model affinity (PMA); family-level biotic index (FBI); Ephemeroptera-Plecoptera-Trichoptera index (EPTI); percent Chironomidae (PCHIR); and percent dominant family (PDOM). Metrics were selected based upon established research in this ecoregion (Bode et al. 1990), project goals, and resources available. Benthic metric scores were compiled for each site from the 10 months of sampling and annual and seasonal (winter versus summer) means were calculated. Winter means were developed from November, December, March, and April data while summer means were compiled from June, July, August, and September data. A repeated measures analysis of variance (ANOVA) was used to test for significant differences among sites (SUB, RUR, FOR and AGR; d.f. = 3,39) for each of the seven metrics and between summer and winter months (d.f. = 1,30). All metric scores were 4^{th}-root transformed prior to testing.

The level of biological impairment at each site was determined by comparing average summer scores on five of the seven metrics against the USEPA RBP II biological impairment criteria (Plafkin et al. 1989) and the NYSDEC Index of Expected Values for Flowing Waters in NYS (Bode et al., 1990). The USEPA RBP II criterion uses percent similarity to a reference site to determine biological impairment (Plafkin et al. 1989). Here, the particular site with the most favorable metric score was used as the reference site. Two of the metrics (SDI and PCHIR) were not applicable to either criterion. Coefficients of variation were determined from average summer and winter month data for each site on all seven metrics to investigate how metric scores were changing across sites and seasons. To test how well the seven metrics correlated

Table 2. Annual descriptors for the physicochemical data collected from four sites within the Cazenovia Creek watershed.

Site	Temp (°C)	SpCon (ms/cm)	DO (mg L^{-1})	pH	Turbidity (NTU)
Suburban					
Mean	21.3	0.391	8.8	8.2	40
Maximum	32.6	0.502	14.6	9.2	1456
Minimum	5.6	0.143	5.8	5.6	0
Rural					
Mean	17.1	0.393	10.2	7.9	26
Maximum	25.1	0.472	17.4	9.0	1308
Minimum	4.9	0.092	7.1	7.2	0
Forested					
Mean	17.5	0.284	9.4	8.4	3
Maximum	28.5	0.346	15.7	9.0	826
Minimum	5.0	0.080	6.9	7.6	0
Agricultural					
Mean	15.9	0.357	9.4	8.2	21
Maximum	25.6	0.452	13.9	8.9	951
Minimum	4.9	0.091	7.2	7.6	0

with one another and assess the redundancy of each measure, monthly metric scores from each site were correlated and inserted into a Pearson correlation matrix. Forty RICH scores -- one score from each of four sites over 10 months of sampling were correlated with 40 EPTI scores (d.f. = 38; r critical = 0.325).

RESULTS

Water quality patterns

Stream water quality data indicated that all four streams had healthy median dissolved oxygen (DO) levels and were well buffered but differed considerably in their dissolved constituents (SpCon), turbidity (NTU), and temperature (T) regimes (Table 2). The SUB had the highest maximum T, SpCon, and NTU and the highest median T. At the other end of the spectrum, FOR had the lowest maxima and medians for SpCon and NTU. Water quality values (maxima and medians) for the RUR and AGR sites were between those from the FOR and SUB sites.

Trends in benthic metrics

All metrics indicated better benthic community health at the RUR and FOR sites based on year-averaged scores from the 10 months of sampling (Table 3). Whereas four of the metrics (RICH, SDI, PMA and PDOM) indicated that RUR had the healthiest macroinvertebrate community, scores from the three other metrics (FBI, EPTI, and PCHIR), indicated that FOR had the healthiest macroinvertebrate community. Conversely, five of the seven metrics (RICH, SDI, PMA, PCHIR, and PDOM) indicated that AGR had the most impaired macroinvertebrate community, while the remaining two metrics (FBI and EPTI) indicated that SUB had the most impaired macroinvertebrate community. The repeated measures ANOVA indicated significant ($P < 0.05$) differences among sites on all metrics except SDI. The greatest differences in metric scores among the four sites were found with PMA, FBI, and EPTI.

The RUR and FOR were assessed as non-impaired on all four applicable metrics (RICH, FBI, EPTI, and PDOM) when their summer-averaged scores were compared to the framework of USEPA RBP II biological impairment criteria (Table 4), indicating that these two sites were healthiest in terms of their aquatic biota. Although most year-averaged metrics ranked SUB above AGR on benthic health, comparisons to the RBP II bio-impairment criteria indicated that the summer-averaged scores at these

Table 3. Repeated measures ANOVA for testing differences in the benthic metric scores among sites (yearly averages; d.f. = 3,39) and between seasons (summer versus winter; d.f. = 1,30). RICH = family taxa richness, SDI = Shannon-Weiner diversity index, PMA = percent model affinity, FBI = family-level biotic index, EPTI = Epheroptera-Plecoptera-Trichoptera index, PCHIR = % Chironomidae, RDOM = % dominant family.

Factor	RICH	SDI	PMA	FBI	EPTI	PCHIR	PDOM
Among sites							
Suburban	15.9	2.1	71.1	4.9	38.5	28.1	27.3
Rural	16.4	2.1	73.2	4.2	55.6	25.3	24.2
Forested	15.4	2.1	64.1	3.9	62.9	21.3	27.0
Agricultural	13.4	1.9	62.3	4.4	48.4	35.0	33.4
P-value	0.0112	0.1691	0.0016	0.0023	0.0020	0.0163	0.0401
Between seasons							
Summer	13.1	1.8	62.9	5.0	32.6	44.1	30.0
Winter	16.7	2.2	71.9	3.8	67.8	17.0	27.3
P-value	0.0001	0.0003	0.0096	<0.00001	<0.00001	<0.00001	0.6380

sites provided the same assessments. Both sites were non-impaired on FBI and RICH, moderately-impaired on the EPTI, and slightly-impaired on the PDOM metric. NYSDEC Biological Impairment Criteria of expected index values for flowing waters in New York contrasted with the RBP II results. All four sites were assessed as slightly-impaired on the FBI metric (albeit at the lower end of the impairment scale), while the forested and agricultural sites were assessed as slightly-impaired on PMA.

Seasonal variation of metrics

During summer, all seven metrics showed RUR as having the best benthic health, and AGR as having the worst benthic health (Table 5). However, during the winter months, only three metrics ranked RUR as having the best benthic health, while FOR had the most favorable scores on five of the seven metrics. Similar to summer month averages, metrics showed AGR had the worst benthic health on five of the seven metrics during the winter months. Therefore, as was the case with year-averaged metric scores, metrics indicated that RUR and FOR had healthier benthic communities than SUB and AGR during summer and winter month comparisons.

Benthic metric scores displayed considerable differences between summer and winter. There was significant ($P < 0.05$) seasonal difference in all metrics except PDOM. The community composition metrics showed the largest seasonal variation, with mean EPTI scores increasing 108% from summer to winter, and mean PCHIR scores decreasing 159% from summer to winter. Results from the other metrics showed, as compared to summer months, winter month scores were: 27% higher for RICH; 25% higher for SDI; 14% higher for PMA; 31% lower for FBI; and 10% lower for PDOM.

All metrics showed greater percent change between seasons than among sites. That is, scores changed more within sites over the one-year sampling period than among sites during any one month. Overall, the percent change of RICH averaged 28% between the four sites and 41% within-sites over the one-year sampling period.

Table 4. Comparison of summer-averaged benthic macroinvertebrate metric from Cazenovia Creek watershed to USEPA bioimpairment criteria (RBP II) showing % similarity to high RBP II scores. Also shown are NYSDEC biological assessment scores. Interpretations (parenthetically): N = non-impaired; M = moderately impaired; S = slightly impaired.

Metric	Suburban (SUB)	Rural (RUR)	Forested (FOR)	Agricultural (AGR)
USEPA				
RICH Score	14.0	14.3	12.0	12.0
	(N)	(N)	(N)	(N)
FBI Score	5.14	4.8	4.9	5.2
	(N)	(N)	(N)	(N)
EPTI Score	27.8	41.0	38.3	23.3
	(M)	(N)	(N)	(M)
PDOM Score	31.3	27.3	28.0	33.5
	(S)	(N)	(N)	(S)
NYSDEC				
PMA Score	67.3	72.0	58.8	53.5
	(N)	(N)	(S)	(S)
FBI Score	5.1	4.8	4.9	5.2
	(S)	(S)	(S)	(S)

Average between-site percent change for SDI across the four sites over the year-long study period was 23%; within-site percent change averaged 44% for this metric. For PMA between-site percent change averaged just 21%, while within-site percent change averaged 33%. Similarly, percent change averaged 25% between-sites and 40% within-sites for FBI, 50% between-sites and 74% within-sites for EPTI, and 44% between-sites and 64% within-sites for the PDOM metric. PCHIR displayed the most difference, in terms of both between-site and within-site percent change. Between-site percent change averaged 69%, while within-site percent change averaged 91%.

Coefficients of variation for each metric and site (Table 5) indicated that variation in metric scores increased with stream order. For all seven metrics, the second order sites (FOR and AGR) had mean variations of 29% and 28% respectively, the third order site (RUR) had a mean variation of 24%, and the fourth order site (SUB) had a mean variation of 19%. There were no significant ($P > 0.05$) differences between summer and winter variation in metric scores (i.e., metrics with low variation had low variation across both summer and winter). The composition metrics (EPTI and PCHIR) had higher variation both within and between the two seasons than did the other metric types.

Metric redundancy

Using a Pearson correlation matrix we investigated the redundancy of the seven metrics by metric type (richness, similarity, tolerance, and composition). In terms of metric type, the Pearson correlation matrix indicated that the richness and composition metrics had the highest redundancy (Table 6). The high redundancy is a result of both RICH and SDI measuring very similar aspects of the benthic assemblage. The only similarity metric examined (PMA) was not strongly redundant with any other metrics. In the matrix, PMA shared its highest r value with SDI ($r = 0.535$). Between the two tolerance metrics (FBI and PDOM), there was no significant redundancy, as the two metrics shared an r value of 0.300. The highest degree of redundancy occurred between the composition metrics (EPTI and PCHIR). These two metrics correlated with an r value of -0.830 in the Pearson matrix, the negative sign indicating the strong inverse relationship between these two community descriptors.

Table 5. Mean benthic scores for summer and winter months from four sites in the Cazenovia Creek watershed. SUB = suburban, RUR = rural, FOR = forested, AGR = agricultural. Coefficients of variation are in parenthesis.

	Summer				Winter			
	SUB	RUR	FOR	AGR	SUB	RUR	FOR	AGR
RICH	14.0	14.3	12.0	12.0	16.0	17.5	18.0	15.3
	(8.2)	(6.7)	(18.0)	(23.6)	(10.2)	(14.4)	(21.3)	(9.8)
SDI	1.88	1.9	1.7	1.6	2.2	2.3	2.3	2.1
	(11.1)	(18.9)	(25.1)	(26.5)	(8.9)	(9.8)	(12.6)	(9.5)
PMA	67.3	72.0	58.8	53.5	73.5	79.3	66.3	68.5
	(11.9)	(13.0)	(3.8)	(11.2)	(9.6)	(8.8)	(15.7)	(9.4)
FBI	5.1	4.8	4.9	5.2	4.5	3.7	3.4	3.7
	(4.9)	(7.8)	(14.9)	(4.7)	(6.1)	(11.5)	(13.3)	(25.8)
EPTI	27.8	41.0	38.3	23.3	50.3	65.8	84.8	70.5
	(31.7)	(29.9)	(29.8)	(28.2)	(29.5)	(16.0)	(11.1)	(31.8)
PCHIR	42.3	38.5	44.5	51.3	20.3	19.3	5.5	23.0
	(26.2)	(40.4)	(44.6)	(35.5)	(38.8)	(68.8)	(119.2)	(98.1)
PDOM	31.3	27.3	28.0	33.5	27.0	22.5	27.0	32.5
	(30.4)	(60.6)	(50.3)	(46.4)	(32.0)	(21.9)	(30.1)	(24.2)

Table 6. Pearson correlation matrix of r values comparing seven family-level metrics with one another with data from four sites in the Cazenovia Creek watershed (n = 7, df = 38, r critical = 0.325).

	RICH	SDI	PMA	PDOM	FBI	EPTI
RICH						
SDI	0.818					
PMA	0.362	0.535				
PDOM	-0.484	-0.643	-0.290			
FBI	-0.627	-0.636	-0.199	0.300		
EPTI	0.558	0.586	0.237	-0.238	-0.930	
PCHIR	-0.761	-0.854	-0.366	0.510	0.843	-0.830

DISCUSSION

How do benthic indices vary across the four sites representing different landuses and water quality?

All seven metrics indicated that RUR and FOR had the best benthic community health on year-averaged, summer-averaged and winter-averaged scores. Conversely, SUB and AGR had the poorest benthic health scores on all year-averaged metric scores. The progression of biological impairment among sites according to USEPA biocriteria using summer-averaged scores was: RUR and FOR were non-impaired on all four metrics; and SUB and AGR were slightly-impaired on the PDOM and moderately-impaired on the EPTI. Differences in how the assessed level of biological impairment can change based on methodology can be seen in the comparisons of metric scores using the USEPA RBP II biological impairment criteria and the NYSDEC expected index values for flowing waters in New York. According to the NYSDEC expected index values, FOR and AGR were slightly-impaired on the PMA metric, and all four sites were assessed as slightly-impaired on the FBI metric, although this assessment was borderline according to the NYSDEC criteria. Stream order most likely played a role in the level of assessed biological impairment for the PMA metric, as scores on this measure were consistently lower at the two second-order sites than the third- and fourth-order sites.

Greater biological impairment was expected at SUB as this site is surrounded by a high percent of impervious surfaces and has narrow riparian buffers (often <10 m) and reduced canopy cover. These characteristics allow runoff to reach the stream without being significantly filtered and likely contributed to increased stream temperatures and decreased dissolved oxygen measured at this site. Paul and Meyer (2001) found that although urban and suburban land tends to be a small percentage of total catchment area, it exerts a disproportionately large influence on water quality both proximately and over distances. Marked declines in benthic indices have been found when percent impervious area (used as a surrogate of percent urban area) is greater than 8% - 15% (Stepenuck et al. 2002, Roy et al. 2003). Past research has also indicated that agricultural landuse negatively impacts the benthic assemblage (Davis et al. 2003) but to a lesser degree than urban landuse (Stepenuck et al. 2002). Fitzpatrick et al. (2004)

found that the Illinois fish Alternative Index of Biotic Integrity and the macroinvertebrate index ranged from poor to excellent in agricultural/rural streams but varied between poor and fair for streams with more than 10% urban land. Our results suggested greater biological impairment at AGR than at SUB but lower water quality at SUB.

The RUR had non-impaired conditions on all metrics under the USEPA RBP II criteria. The percent of active crop land surrounding the rural site is relatively low compared to the primarily agricultural site. Previous studies have shown that streams in agricultural catchments usually remain in good condition until the extent of agriculture increases to 30-50% of total catchment area (Quinn and Hickey 1990, Quinn 2000, Bennett et al. 2004). In fact, Quinn (2000) observed that small increases in agriculture actually increased pollution intolerant benthic macroinvertebrate abundance, but increases greater than 30% resulted in declines of intolerant taxa and increases of pollution tolerant taxa. Bennett et al. (2004) found elevated macroinvertebrate index scores for streams subject to nutrient enrichment from cattle runoff. The prevailing thought is that small increases in nutrients aid periphyton growth and assist benthic assemblages, but when a certain threshold is exceeded, eutrophication increases biochemical oxygen demand, which depletes dissolved oxygen levels and stresses resident benthos (Nelson and Booth 2002). Such a stress relationship is suggested in our research in relation to RUR and AGR. The RUR, with a low percent of active cropland had metric scores that exceeded those of even FOR, while the AGR with a high percent of active cropland, displayed metric scores that were consistently among the least favorable of all sites.

The variability of coefficients of variation that we recorded suggests that metric variation decreases with increasing stream order. The findings of greater variability of benthic metric scores from low order streams is supported by Linke et al. (1999) who found metric variation to be higher on low order streams. Compared to higher order streams (third- through sixth-order) there is generally much less total habitat in low order streams simply due to the lower amount of substrates available for colonization. Also, whereas higher order streams tend to have greater equilibrium in flow, temperature, and dissolved oxygen regimes, lower order streams tend to be subject to comparatively larger fluctuations, with quicker changes in stream conditions (Vannote et al. 1980, Allan 1995). The effects of macroinvertebrate life histories may also influence the variability of benthic metrics across stream orders. When a community is dominated by few families, the life history patterns of those families may result in large metric variations between seasons. The influence of stream order may have particularly affected metric outcomes in regards to the richness (RICH and SDI) and composition (EPTI and PCHIR) measures.

What are the seasonal variations in benthic metrics and what are the implications for biomonitoring?

The benthic assemblage health, as measured by the seven metrics, was much better during the winter months than during the summer months. The differences in scores were such that sites that were assessed as slightly or moderately impaired under the NYSDEC biological impairment criteria during the summer would have been assessed as non-impaired under the same criteria during the winter (e.g., FOR and AGR on the PMA metric). For the FBI, all four sites were assessed as slightly-impaired during the summer months, but all four would have been assessed as non-impaired during the winter months. These findings are consistent with those of Hilsenhoff (1988), who found lower FBI values (indicating improved benthic community health) during the winter months. Although the trend of poorer metric scores in the summer and improved scores in the winter was true for all metrics, it was especially pronounced for the richness (RICH and SDI) and composition (EPTI and PCHIR) metrics. PMA and PDOM displayed the least change between summer and winter. This was similar to Linke et al. (1999) who found PDOM showed no significant change between summer and winter, while RICH and FBI showed much larger seasonal variations.

Certain metrics (EPTI and PCHIR) had much greater seasonal variation than others. This was most likely due to the narrower range of families these metrics examined and the life history effects of those families. Life histories of families within a community have been shown to alter the observed composition of the benthic assemblage throughout the year (Linke et al. 1999). The degree to which life histories affect benthic indices depends on the families examined (Rosillon 1987). For example, *Capnia*, a Plecopteran which is widely used as an indicator of good water quality and whose abundance is assessed in the EPTI, is much more likely to appear in winter samples than summer samples (Minshall 1981). Similar to Minshall (1981), we found consistently fewer Ephemeropterans and Plecopterans during the summer months as compared to the winter months. Conversely, chironomid abundance was markedly higher at all sites during the summer months. Rosillon (1987) also observed a steep increase of ephemeroptera and plecoptera from summer to winter, and an opposite tendency for chironomids.

Within seasons, PMA, RICH, FBI and SDI showed markedly lower variation than EPTI, PCHIR and PDOM. This finding is supported by results of Novak and Bode (1992) who observed lower variation for PMA, RICH, and FBI than other metrics including EPTI and PDOM. Lenat (1987) and Linke et al. (1999) also observed lower variation in RICH and SDI compared to EPTI and PDOM. Novak and Bode (1992) sampled two unpolluted streams once per month for a year, and found the mean coefficients of variation to be 14% for PMA and 22% for RICH. Over the one-year sampling period, our results showed mean coefficients of variation to be very similar -- 12% for PMA, and 16% for RICH.

In every case benthic indices varied more within a site over the course of the year-long sampling period than they did among sites in any one month. This suggests that season influenced benthic metric scores more than sampling location. In their investigation of seasonal variability of benthic indices, Linke et al. (1999) used cluster analysis of sites and found that sites sampled at the same time of the year had more chance of ending up in a cluster together than the same site sampled at differing times of the year. Linke et al. (1999) also showed that the clustering of sites was different for summer and winter sampling.

The implications for these seasonal variations in benthic metric values are significant, and seasonality is particularly important when only a few sites are being sampled (Plafkin et al. 1989). Most biomonitoring programs focus on summer sampling to maximize the accessibility to sites and benthic communities, limit equipment restrictions, minimize year-to-year variability (Barbour et al. 1999), maximize sampling safety, and/or obtain a "worst-case scenario" of the benthic assemblage health. Summer tends to be the season of greatest natural stress on the benthic community (due to high temperatures and low dissolved oxygen), and results from summer sampling may indicate biological impairment when there is only natural variation (Poff et al. 1997, Poole et al. 2004). When the goal of a biomonitoring program is to assess the effectiveness of stream restoration or existing best management practices, summer month sampling may not provide an accurate picture of benthic community health. Sampling during every season would be recommended for this situation, so that natural seasonal stress and variation may be noticed and accounted for.

What selected benthic indices are most appropriate for the assessment of biological impairment across sites and seasons?

The abilities of a metric to discriminate significant differences across sites, be applicable to state and federal biomonitoring programs, maintain low within-site and within-season variation, and have low redundancy within its metric type (richness, similarity, tolerance, or composition) were used as criteria to assess the appropriateness of the seven metrics. As RICH and SDI measure very similar aspects of the benthic community, it was not surprising that they were highly correlated (r = 0.818). Between these two measures, RICH is preferred because of its low variation, its ability to discriminate significant differences between sites, and its widespread use in state and federal biomonitoring programs. A caveat to using RICH is that this metric may be naturally lower in low order streams (first and second order), and higher in mid-order streams (third through fifth order) as a result of stream width, daily and annual temperature fluctuations, and flow continuity providing a greater area of stable habitat in mid-order streams (Vannote et al. 1980). Such natural variations may have been responsible for the lower RICH scores we found at the two second order streams.

PMA, the only similarity metric examined, was not strongly redundant with any other metrics. PMA had the lowest coefficient of variation both within sites and within seasons, was able to identify significant among-site differences, and had the second lowest percent change between seasons (only PDOM was lower).

Between the two tolerance metrics (FBI and PDOM) there was no significant redundancy (r = 0.300). FBI exhibited low coefficients of variation within sites and within seasons, had robust discriminatory power between the four sites, is widely-used in state and federally biomonitoring programs, and has been shown to be a reliable metric over a wide geographic range (Hilsenhoff 1988). PDOM identified significant differences among sites, is widely used by government biomonitoring programs, but had inconsistent variation within sites and seasons. Considering this, FBI was the preferred tolerance metric. However, PDOM was the only metric that did not change significantly between summer and winter months. Therefore, if the goal of a biomonitoring program is to evaluate the effectiveness of stream restoration and sampling is to be done through multiple seasons, our results suggest that PDOM may be a useful metric due to its lack of change between seasons. Of course, by being aware of the seasonal changes in selected metrics and comparing samples only within seasons, errors related to seasonal variation may be reduced.

The highest degree of redundancy occurred between the composition metrics (EPTI and PCHIR). Consistently, as EPTI increased, PCHIR decreased. EPTI had considerably lower coefficients of variation both within sites and within seasons. Although both metrics were able to discriminate significant differences among sites, PCHIR varied much more between summer and winter. Doberstien et al. (2000) also observed that EPTI and PCHIR displayed high and often unstable seasonal variability among sampling sites. High variability in a metric over the course of the year does have significant implications for bioassessments that sample only once a year or those that attempt to compare samples taken from different times of the year.

Results from the above criteria suggest RICH, PMA, FBI, and EPTI should be used in future assessments of similar Great Lakes watersheds. Supporting these findings is past research that has found FBI and EPTI to be measures that accurately account for the tolerance range of taxa to organic pollution (Hilsenhoff 1988, Rosenberg and Resh 1993). For instance, many EPTI taxa may be eliminated from a habitat that experiences even small increases of organic pollution (Hilsenhoff 1988). Previous work by Roy et al. (2003) indicated that RICH, EPTI, and FBI were the metrics most sensitive to environmental stress.

ACKNOWLEDGEMENTS

Funding for this research was provided by the Great Lakes Basin Program of the Great Lakes Commission, Ann Arbor, Michigan. The Great Lakes Center at Buffalo State College provided a tuition scholarship to C. Riley. Logistical support provided by John Whitney, USDA-NRCS, is much appreciated. We also thank Josh Brown for allowing access to his property.

LITERATURE CITED

Allan, J. D. 1995. Stream Ecology: structure and function of running waters. Chapman and Hall, London, UK.

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, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Bennett. H. H., M.W. Mullen. P.M. Stewart. J.A. Sawyer. and E.C. Webber. 2004. Development of an invertebrate community index for an Alabama coastal plain watershed. Journal of American Water Resources Association 40: 43-51.

Bode, R.W., M.A. Novak, and L.E. Abele. 1996. Quality assurance work plan for biological stream monitoring in New York State. New York State Department of Environmental Conservation, New York. 89p.

Bode, R. W., M.A. Novak, and L.E. Abele. 1990. Biological Impairment Criteria for Flowing Waters in New York State. New York State Department of Environmental Conservation, Albany, New York. 110p.

Davis, S., S. Golladay, G. Vellidis, and C. Pringle. 2003. Macroinvertebrate biomonitoring in intermittent coastal plain streams impacted by animal agriculture. Journal of Environmental Quality 32: 1036-1043.

Doberstein, C.P., J.R. Karr, and L.D. Conquest. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44: 355-371.

Fitzpatrick, F. A., M.A. Harris, T.L. Arnold, and K.A. Richards. 2004. Urbanization influences on aquatic communities in northeastern Illinois streams. Journal of American Water Resources Association 40:461-475.

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. Journal of North American Benthological Society 7:65-68.

Karr, J. R. and C.O. Yoder. 2004. Biological assessment and criteria improve TMDL decision making. Journal of Environmental Engineering 130:594-604.

Karr, J. R., R.C. Heidinger, and E.H. Helmer. 1985. Sensitivity of the index of biotic integrity to changes in chlorine and ammonia levels from wastewater treatment facilities. Journal of Water Pollution Control Federation 57: 912-915.

Kashian, D.R., and T.M. Burton. 2000. A comparison of macroinvertebrates of two Great Lakes coastal wetlands: testing potential metrics for an index of ecological integrity. Journal of Great Lakes Research 26:460-481.

Lenat, D.R. 1987. The macroinvertebrate fauna of the Little River, North Carolina: taxa list and seasonal trends. Archives of Hydrobiology 110: 19-43.

Lenat, D.R. and J.K. Crawford. 1994. Effects of land use on water quality and aquatic biota of three North Carolina piedmont streams. Hydrobiologia 294:185-199.

Linke, S., R.C. Baily, and J. Schwindt. 1999. Temporal variability of stream bioassessments using benthic macroinvertebrates. Freshwater Biology 42: 575-584.

Minshall, G.W. 1981. Structure and temporal variations of the benthic macroinvertebrate community inhabiting Mink Creek, Idaho, U.S.A., a 3rd order Rocky Mountain stream. Freshwater Biology 1: 13-26.

National Research Council. 2001. Assessing the TMDL approach to water quality management. National Academy Press, Washington, D.C.
Nelson, E. J. and D.D. Booth. 2002. Sediment sources in an urbanizing, mixed land-use watershed. Journal of Hydrology 264, 51–68.
Novak, M.A. and R.W Bode. 1992. Percent model affinity: a new measure of macroinvertebrate community composition. Journal of North American Benthological Society 11:80-85.
Paul, M.J. and J.L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecological Systems 32:333-65.
Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid bioassessment protocols for use in stream and rivers: benthic macroinvertebrates and fish. EPA 440/4-89/001. Office of Water, U.S. Environmental Protection Agency, Washington, DC.
Poff, N.L., D.J. Allen, M.B. Bain, J.R. Karr, and K.L. Prestegaard. 1997. The natural flow regime: a paradigm for conservation and restoration of riverine ecosystems. BioScience 47:769-84.
Poole, G.C., J.B. Dunham, D.M. Keenan, S.T. Sauter, D.A. McCullough, C. Mebane, J.C. Lockwood, D.A. Essig, M.P. Hicks, D.J. Sturdevant, E.J. Materna, S.A. Spalding, J. Risley, and M. Deppman. 2004. The case for regime-based water quality standards. BioScience 54: 155-61.
Quinn, J.M. 2000. Effects of pastoral development. Pages 208-229 *In*: Collier, K.J. and M.J. Winterbourn (eds.), New Zealand stream invertebrates: ecology and implications for management. Caxton, Christchurch, New Zealand.
Quinn, J.M. and C.W. Hickey. 1990. The magnitude of the effects of substrate particle size, recent flooding, and catchment development on benthic macroinvertebrates in 88 New Zealand rivers. New Zealand Journal of Maritime Freshwater Research 24: 411-27.
Rosenberg, D.M. and V.H. Resh. 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York. 488pp.
Rosenberry, D.O., P.A. Bukaveckas, D.C. Buso, G.E. Likens, A.M. Shapiro, and T.C. Winter. 1999. Movement of road salt to a small New Hampshire lake. Water Air and Soil Pollution 109:179-206
Rosillon, D. 1987. Seasonal variation in the benthos of a chalk trout stream, the River Samson, Belgium. Hydrobiologia 126: 253-262.
Roy, A.H., A.D. Rosemond, D.S. Leigh, M.J. Paul, and J.B. Wallace. 2003. Habitat-specific responses of stream insects to land cover disturbance: biological consequences and monitoring implications. Journal of North American Benthological Society 22:297-307.
Stepenuck, K., R. Crunkilton, and L. Wang. 2002. Impacts of urban landuse on macroinvertebrate communities in southeastern Wisconsin streams. Journal of American Water Resources Association 38: 1041-1051.
United States Department of Agriculture (USDA). 1986. Soil survey of Erie County, New York. National Cooperative Soil Survey. USDA-NRCS, Erie County Office, East Aurora, NY.
Vannote, R. L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Science 37: 130-137.
Yandora, K. 1998. Rapid bioassessment of benthic macroinvertebrates illustrates water quality in small order urban streams in a North Carolina Piedmont city. City of Greensboro, Storm Water Services, 401 Patton Avenue, Greensboro, North Carolina 27406.

Received: 12 March 2007 Accepted: 8 June 2007