Recalibration of the Illinois Fish IBI Using a Continuous Scoring Methodology

Prepared for:

U.S. Environmental Protection Agency, Region 5
77 West Jackson Boulevard
Chicago, IL 60604-3507

and

Illinois Environmental Protection Agency
1021 North Grand Ave. East
Springfield, IL 62794-9276

Prepared by:

Edward T. Rankin
Midwest Biodiversity Institute
4673 Northwest Parkway
Hilliard, OH

and

Ben Jessup
Tetra Tech, Inc.
400 Red Brook Boulevard, Suite 200
Owings Mills, MD 21117

Final Report
February 8, 2016

Table of Contents

List of Tables

List of Figures

1.0 Introduction

This report summarizes the recalibration of the fish IBI (fIBI) for the State of Illinois Environmental Protection Agency (IEPA). This effort did not completely "start-from-scratch", but rather built upon the existing fIBI and took advantage of a larger dataset that has accrued since the original fIBI was developed. The recalibration also adds some features that are common components of more recent forms of fish IBIs. These advances include a continuous scoring methodology on a 0-100 continuous scale as opposed to the ordinal 0-6 scoring methodology used in the original Illinois fIBI (Smogor 2000, 2005). This recalibration also incorporates more recent data (up to and including 2010) and a refined identification of reference sites.

For all of the more recent data, reference sites were identified using criteria related to land uses and activities in the watershed of each site. The reference status of each site was identified by first comparing watershed conditions to quantitative criteria thresholds for 11 stressor variables. Sites that had relatively low stress for all criteria were candidate reference sites. The reference designations of sites was reviewed by IEPA and modified based on familiar knowledge of the sites. The process is described in greater detail in the report on Illinois Stream Macroinvertebrate Multimetric Index Development (Tetra Tech, 2016). All reference samples identified by IEPA for original calibration of the fIBI were assumed to represent reference sites regardless of current analytical results.

The original fIBI is applied differently in each of 13 distinct fish classification regions. Regional variations in IBI expectations along a geographic classification (e.g., IBI regions) are now accounted for by setting impairment thresholds based on fIBI reference values from each classification stratum. A classification analysis conducted with the recent and larger data set suggested that a simpler classification approach could be used (Appendix A). In the simpler approach most metrics could be calibrated statewide or across north-south latitude-based classification regions and size classifications based on both watershed size and stream width (Table 1). This simple classification scheme was used in preliminary assessments of metric responsiveness. Because the Larger classification is recognized with either condition of watershed size or wetted width, the points in following graphs can show overlap on each individual size scale.

Table 1. Site classes suggested in the classification analysis (Appendix A).

Class	**Description**
Smaller north- central (SNC)	≤ 40 foot wetted width and ≤ 200 mi^2 watershed, not in IBI regions 12 or 13 (approximately 37.7 N latitude)
Smaller south (SS)	≤ 40 foot wetted width and ≤ 200 mi^2 watershed, IBI regions 12 or 13
Larger (LNC or LS)	> 40 foot wetted width or > 200 mi^2 watershed, any location

When metric characteristics relative to environmental conditions were further studied during the course of metric recalibration, the classification strata were not always applicable. Either metrics were unresponsive to the north-south gradient, application of the categorical size classes was coarse compared to continuous gradient options, or stream slope was more influential than location or stream size. Classification strata or adjustments were applied specifically for each metric, depending on the greatest apparent influences on the reference metric values. Therefore, adjustments to individual metrics were made for all conditions of watershed size or stream slope. Only two metrics were adjusted for north-south location.

Stream Size vs. Stream Width

The original Illinois fIBI was calibrated using stream width (ft.) to account for increasing species richness with stream size (Smogor 2000). The classification analyses (Appendix A) confirmed that measures of stream size are important classification variables for Illinois stream fish assemblages. Plots of stream width vs. species richness metrics (e.g., native species, intolerant species) confirm that the upper threshold of these metrics is different between the Northern and Southern classification regions (Figure 1, top). Using drainage area as a replacement for stream width better resolved the North vs. South classification (e.g., Figure 1, bottom) for most of the fIBI metrics (excepting the minnow species metric). The threshold difference for native species, intolerant species, and several other species richness metrics seems related to the fact that Southern region streams are generally wider for the same drainage area than Northern streams. If the fish assemblage at a given stream size is related to the volume of habitat during summer-fall flow periods then stream width may overestimate species richness potential compared to drainage area. Drainage area is also a more repeatable measure of stream size, eliminating the influence of annual and intra-seasonal fluctuations in stream width and is thus a more stable measure for calibrating fIBI metrics (Appendix B). We decided to use the continuous drainage area adjustment in favor of stream width or the categorical north-south classes because that allowed for statewide calibration and application of metric scoring.

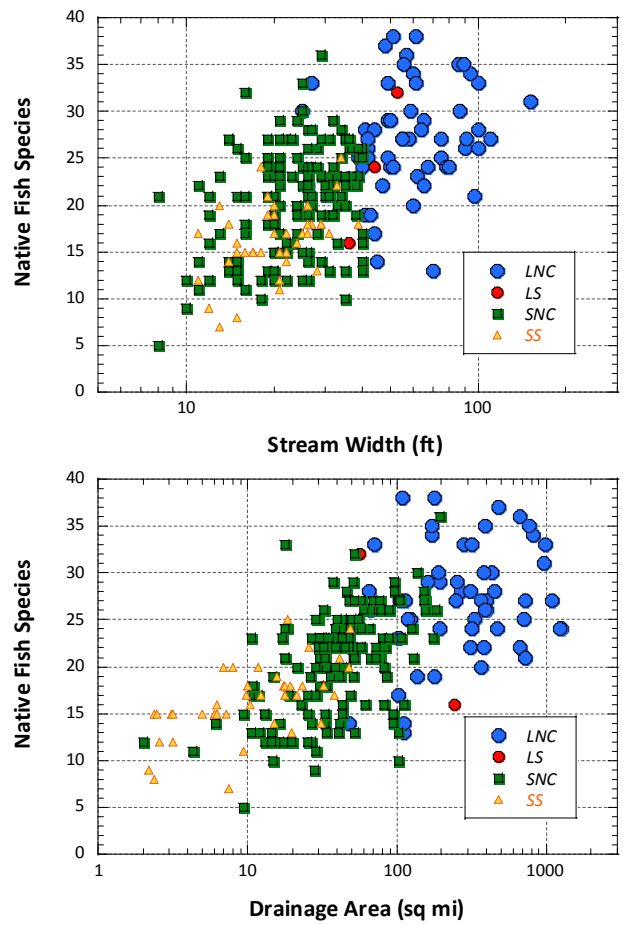


Figure 1. Plots of native fish species vs. stream width (top) and drainage area (bottom) for Illinois reference sites. Points coded as Large North-Central (LNC), Large Southern (LS), Small North-Central (SNC) or Small Southern (SS) sites

Calibration of Metrics by Stream Size for Continuous Scoring

A variety of approaches have been used to score metrics using a continuous approach and include: 1) different ways to set ceiling and floor values for metrics, 2) different methods to account for stream size or other continuous classification variables (e.g., elevation, gradient), and 3) calculating metric scores using linear interpolation vs. curve-fitting to data using a cumulative distribution function (CDF) of metric values). Some of the continuous calibration approaches for multimetric indices are provided in Table 2.

Table 2. Selected approaches to calibration and scoring of metrics for IBI type indices using a continuous rather than discrete scoring approach.

Author	Calibration with Stream Size	Scoring Method	Index Scoring Range	Max/Min. Score Set Points	ACRONYM
Mebane et al. (2003) - Fish	No	Curves of CDF of Metric Value	0-100	Min at 0; Max at 95th %tile ALL sites	**PacNW IBI**
EMAP – Fish	Yes, used residuals standardized to 100 km^2	Linear Interpolation	0-100	Min at 10th %tile ALL sites; Max at median of REF Sites	**EMAP FIBI**
Aparicio et al. (2011)	Yes, used quantile regression with drainage area	Linear Interpolation	0-100	Min at 0 or 5th %tile dep. on metric; Max at max. value or 95th %tile (e.g., CPUE)	**IBI-Jucar**
Klemm et al. (2003) - Macros	No	Linear Interpolation	0-100	Min – 25th %tile of Impaired Sites; Max – 75th %tile of Reference Sites	**MBII**
Rankin (2010) – Ohio Fish (CIBI) and Macros (CICI)	Yes, used quantile regression with drainage area	Linear Interpolation	0-100	Min – Minimum metric value; Max – 99th %tile of Reference Sites	**CIBI; CICI**

Because stream size is such an important calibration variable in Midwest warm-water streams, a quantile regression method was used in the Illinois data set to derive an upper threshold of species richness relative to drainage area. All species richness metrics, plus the proportion of species that are tolerant were associated with drainage area. No abundance metrics were associated with drainage area, but for the metrics percent individuals as specialist benthic invertivores or as mineral spawners, percent slope was a better predictive classification variable for expected reference values than drainage area.

2.0 Selection of a Scoring Approach

Different scoring approaches distribute scores along the gradient of condition in different ways. Figure 2 illustrates how the various continuous scoring methods would result in a different distribution of scores for the native species metric. A linear interpolation from zero to the 95th percentile (95 on the y-axis) provides an even distribution of scores along the entire metric range (green dashed line in Figure 2).

A linear interpolation between some percentile above zero (e.g., 5th percentile, red dashed line in Figure 2) or a curve defined by the actual distribution of data (red dots) or reference sites (blue dots) will less readily distinguish between sites at the lower ranges of species richness, but will have the effect of separating site scores more between 50 to 100 percent of maximum expected richness. In all the options, the 5% of values greater than the 95th percentile will not be distinguished from values at the 95th percentile. There is precedence in using the 95th percentile to eliminate the effects of high outliers and establish a basis for scoring (Barbour et al. 1999, Blocksom 2003, Mebane et al. 2003, Aparicio et al. 2011).

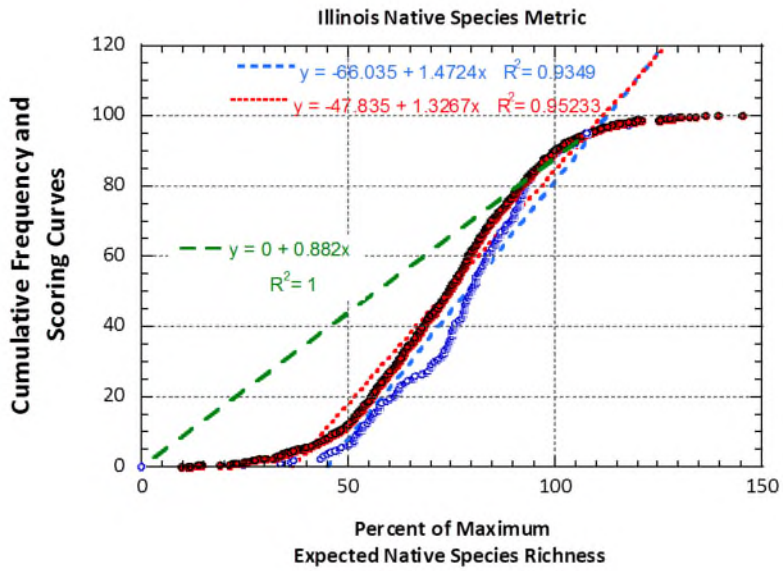


Figure 2. Cumulative distribution plots of the percent of maximum expected native fish species in Illinois streams ≤ 200 sq mi) for all data (red dots) and reference sites (blue dots). Dash lines represent three linear interpolation methods to generate metric scores.

Figure 3 illustrates distributions of the percent of expected native species richness values in categories of site disturbance (all, reference, stressed, and highly stressed). It also illustrates ranges of values used to set minimum (0) and maximum (10) metric scores, as summarized in Table 2. It is clear that the method of scoring can have important consequences on the range of metric values that will be within the sensitive scoring range. For fIBI metrics that we are calibrating in this report,

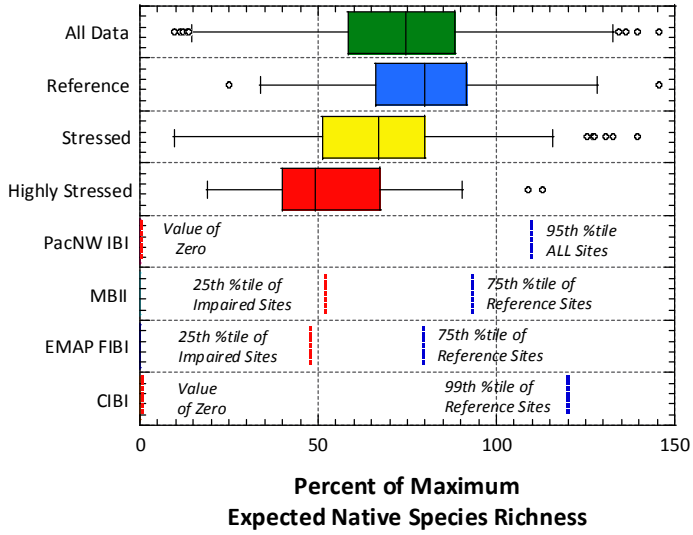


Figure 3. Box and whisker plots of percent of expected native species richness values for Illinois fish data (< 200 sq mi) at all sites, reference sites only, all stressed sites and highly stressed sites. Dash lines below represent various approaches to setting

scoring is most similar to the PACNW IBI approach, in which the minimum score is set at 0 and the maximum is the 95th percentile of ALL data. Exceptions include the Proportion of Individuals that are Generalist Feeders and the Proportion of Tolerant Species, which are scored with lower values having higher scores.

To compare the effect of these scoring approaches to the original Illinois fIBI, the native species metric using three of these approaches was scored (Figure 4) and the distribution of each newly scored metric was compared to the discrete scores (0-6) of the original Illinois fish IBI. Figure 4 (top) is a box-and-whisker plot using the 5th and 95th percentile of references sites as the floor and ceiling values and Figure 4 (middle) is a similar plot using ALL Illinois data. Figure 4 (bottom) uses zero as the floor value. All scores represented in the box plots are based on a linear interpolation between floor and ceiling values.

We wanted to keep as much continuity with the original fIBI as possible, so as in the fIBI, we equated the minimum score at a zero value for all metrics except the Proportion of Individuals that are Generalist Feeders and the Proportion of Tolerant Species. With this scoring approach, the intraquartile ranges of scored values (the boxes in relation to the x-axis) correspond with the original Illinois fIBI scores on the y-axis (Figure 4, bottom). In the approaches that use the 5th percentile as the scoring floor, scores that were 0 and 1 in the original Illinois fIBI scores (and some scores of 3) are all 0 on the new scoring scale. Distinction at the poor end of the scoring scale is only accommodated by the 0-95th percentile scoring scale. At the high end of the scale, only original Illinois fIBI scores of 6 have a non-outlier maximum greater than 10 and scores of 4, 5, and 6 have distinct intra-quartile ranges. This suggests that the upper end of the scale accounts for discernable differences in condition when using the 95th percentile as the upper scoring limit.

Metric Scoring Calculation

Because of the strong relationship between metrics and drainage area (or percent slope), we accounted for these classification effects by calculating a “maximum species richness line” using a 95th quantile regression using reference sites (unless the threshold pattern was unclear based on reference sites alone) for the part of the relationship that was clearly influenced by size or slope. For most metrics we used reference sites to generate 95th quantile regressions by drainage area or percent slope. For each metric we present the calibration curves plotted with both the reference data and separately for ALL data. For certain metrics, the number and range of reference sites (e.g., streams with lower slopes) was insufficient to define the maximum or minimum expectations, so we used all of the available data to derive metric expectations. This is discussed under each metric summary.

The break points between an effective slope and a plateau in metric values were determined by a visual inspection of the data in bi-plots. Evidence was also derived from ordinations (Appendix A). Each metric that decreased with increasing disturbance was converted to a value reflecting the “percent of expected maximum” for each site:

$$\text{Metric Score} = 10 * (\text{observation})/(\text{expected value})$$

Expected values that increase with drainage area or percent slope for each metric appear in Table 3. For all but the Proportion of Individuals that are Generalist Feeders and the Proportion of Tolerant Species metrics, the minimum value was the theoretic minimum (e.g., zero for each richness metric). The alternative scoring procedure for the "floor" of the exceptional metrics is detailed in the sections describing those metrics.

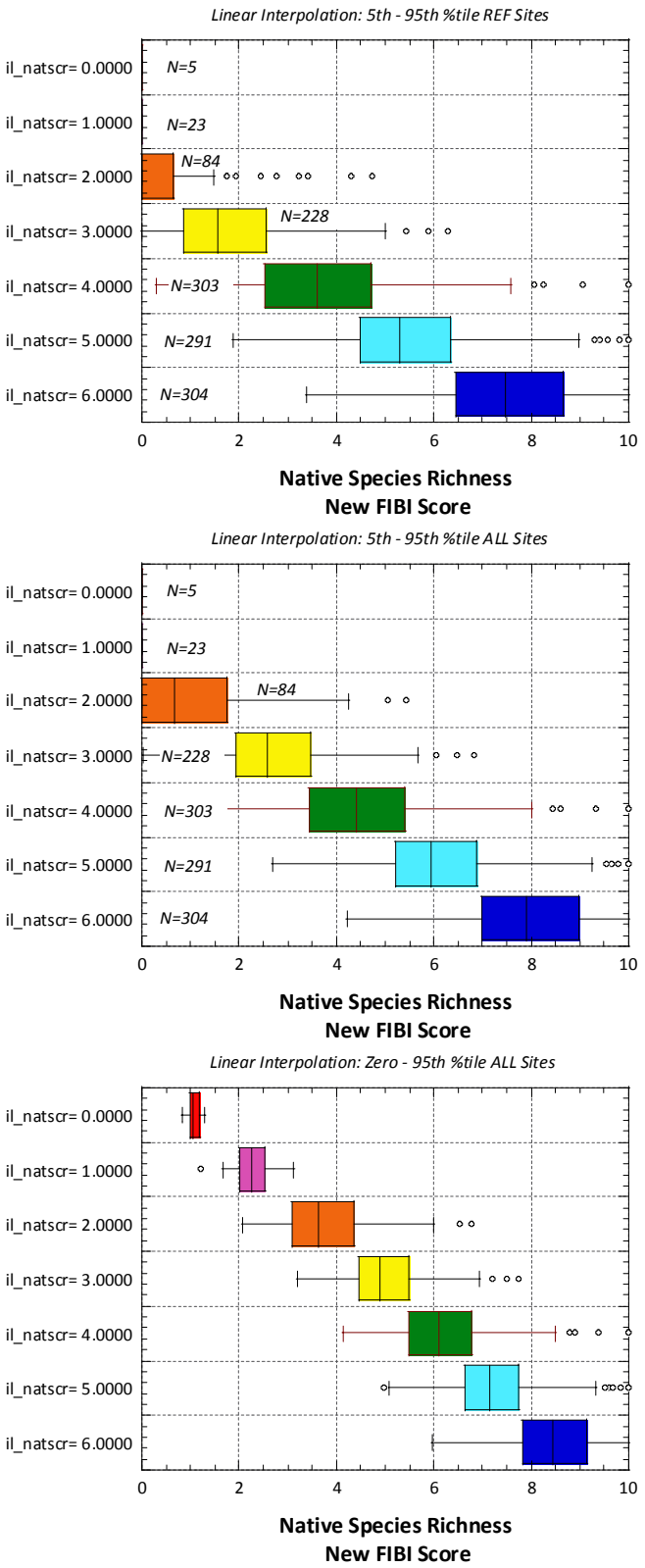


Figure 4. Box and whisker plots of native species richness metric values for Illinois fish data (≤ 200 sq mi) plotted vs. the existing discrete Illinois metric scores of zero to six. Top plot used the a 5th percentile of reference sites at the floor, middle plot used the 5th percentile of all data as the floor and the bottom plot used zero as the floor value.

Table 3. Scoring curves and ranges for each metric of the Illinois Continuous Fish IBI that decreases with increasing disturbance. These equations or values can be used to determine the metric scoring which is calculated as 10(Measured/Expected Value).*

			Drainage Area (DA) or slope regression		**No regression**	
Metric	**Metric Trend with Stress**	**Data Used for Calibra-tion**	**Max Value (Score 10)**	**Min. Value (Score 0)**	**Max Value (No Slope) (Score 10)**	**Min. Value (Score 0)**
Native Species Richness	-	Ref	10.06+12.41*Log10(DA) (≤ 200 sq mi DA)	Zero	38.62 (>200 sq mi DA)	Zero
Number of Native Sucker Species	-	Ref	0.497+3.78*Log10(DA) (≤ 200 sq mi DA)	Zero	9.19 (>200 sq mi DA)	Zero
Number of Native Sunfish Species	-	Ref	3.061+2.3596* Log10(DA) (≤ 40 sq mi DA)	Zero	6.84 (>40 sq mi DA)	Zero
Number of Native Intolerant Species	-	Ref	(4.03*Log10(DA))-0.214 (≤ 200 sq mi DA)	Zero	9.07 (>200 sq mi DA)	Zero
Number of Native Minnow Species	North-Central Streams					
	-	ALL	5.784+4.172*Log10(DA) (≤ 40 sq mi DA)	Zero	12.47 (>40 sq mi DA)	Zero
	Southern Streams					
	-	ALL	3.962+2.43*Log10(DA) (≤ 40 sq mi DA)	Zero	7.85 (>40 sq mi DA)	Zero
Number of Native Benthic Invertivore Species	-	Ref	0.0045+6.629*Log10(DA) (≤ 200 sq mi DA)	Zero	15.26 (>200 sq mi DA)	Zero
Proportion of Individuals that are Specialist Benthic Invertivores	-	ALL	30.536+2.291*Log10(PS) (≤ 0.1 Percent Slope)	Zero	28.24 (> 0.1 Percent Slope)	Zero
Proportion of Individuals that are Obligate Coarse-Mineral Substrate Spawners (and not Tolerant)	-	ALL	104.65+35.176*Log10(PS)	Zero	69.47% (> 0.1 Percent Slope)	Zero

Table 3 (continued). Scoring curves and ranges for each metric of the Illinois Continuous Fish IBI that increases with increasing disturbance. These equations or values can be used to determine the metric scoring which is calculated as 10(Measured/Expected Value).*

Metric	**Metric Trend with Stress**	**Data Used for Calibra-tion**	**Scoring Floor (Score 10) and Ceiling (Score 0)**
Proportion of Individuals that are Generalist Feeders	+	Ref	Ceiling: 100% Floor (≤ 0.1 Percent Slope): 45.5% Floor (≥ 0.1 Percent Slope): 16.94
Proportion of Tolerant Species	North-Central Streams		
	+	ALL	Floor: 24.516-6.00*Log10(DA)
			Ceiling: 77.888-17.603*Log10(DA)
	Southern Streams		
	+	ALL	Floor: 21.252-6.893*Log10(DA)
			Ceiling: 48.037-9.567*Log10(DA)

3.0 Metric Descriptions

The following sections explain and illustrate the scoring approach used for each metric and provides plots of reference sites (only) and ALL data.

Metric 1: Total Native Species Richness

The total native species richness metric includes any species that is native and excludes exotic species and hybrids. The curves on Figure 5 (both top and bottom graphs) represent the 95th quantile regression from reference sites up to 200 sq mi. The maximum expected value at 200 sq mi (38.62 species) was used as the maximum expectation for sites greater than 200 sq mi in catchment size. The reference site derived maximum (95th quantile) species line provides a good fit to reference data (Figure 5, top) and ALL available data (Figure 5, bottom).

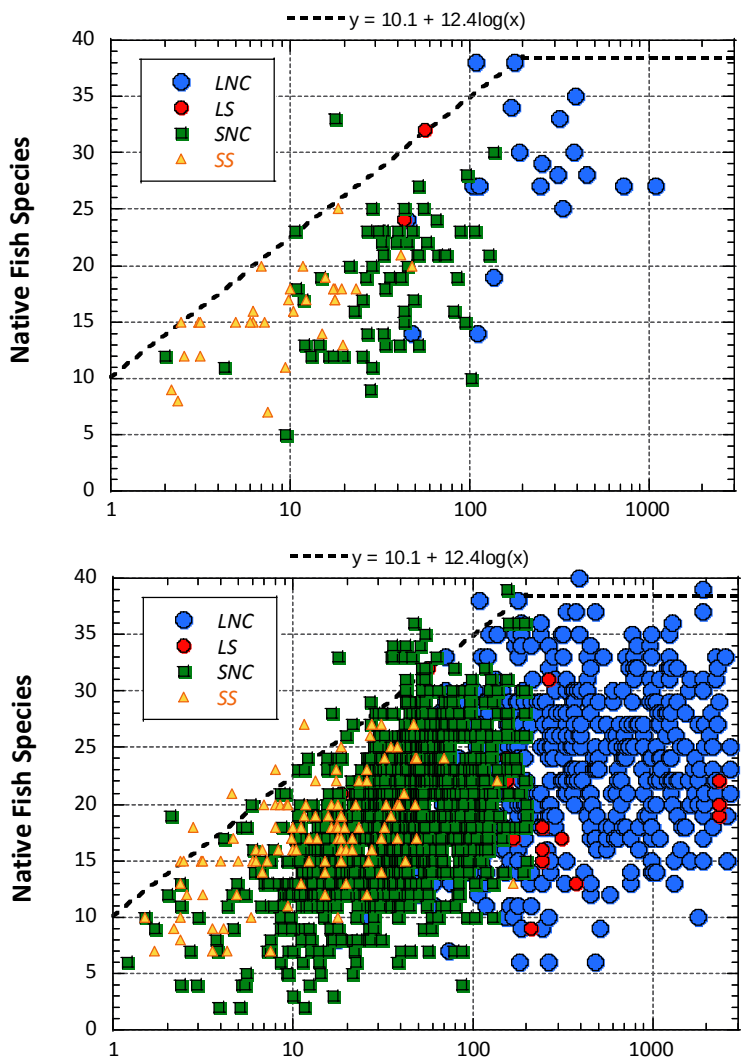


Figure 5. Plot of native fish species richness vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of reference sites up to 200 sq mi drainage; 95th quantile value was then extended straight across for sites greater than 200 sq mi.

Metric 2: Number of Native Sucker Species (Family Catostomidae)

This metric includes any sucker species that is native and in the Family Catostomidae. The curves on Figure 6 (both top and bottom plots) represent a 95th quantile regression from reference sites up to 200 sq mi. The maximum expected value at 200 sq mi (9.19) was used as the maximum expectation for sites greater than 200 sq mi in catchment size. The reference site derived maximum (95th quantile) species lines are plotted again versus ALL available data (Figure 6, bottom) to illustrate the fit to the entire data set.

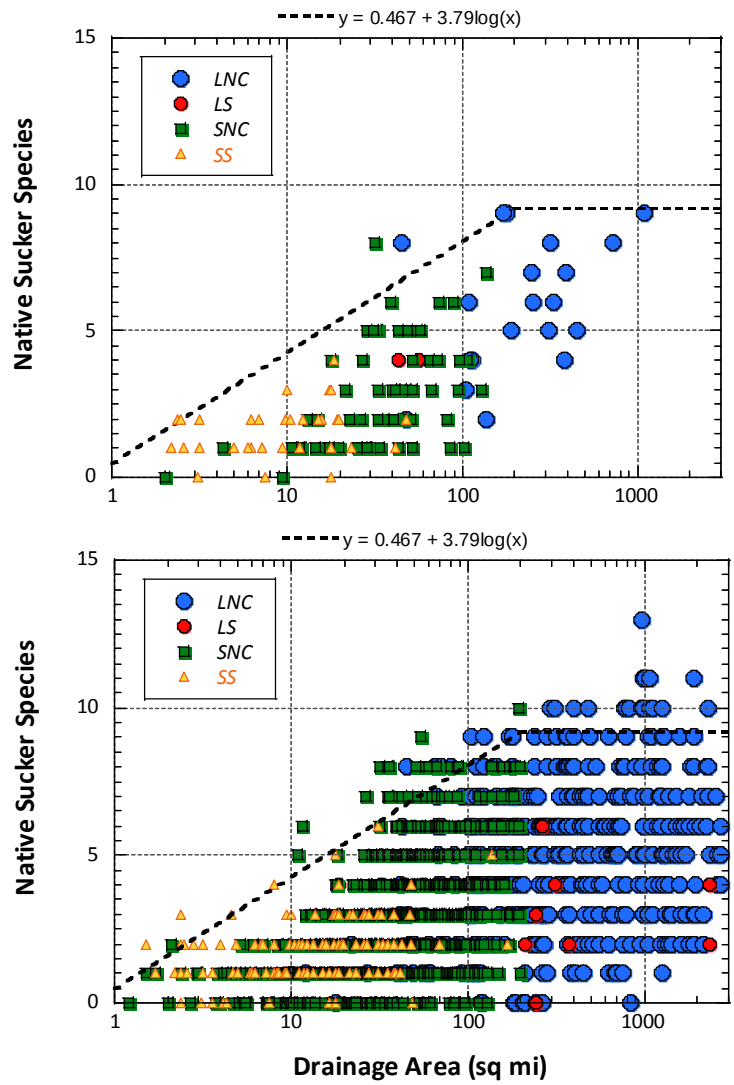


Figure 6. Plot of native sucker species richness vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of reference sites up to 200 sq mi drainage area.

Metric 3: Number of Native Sunfish Species (family Centrarchidae)

This metric includes any sunfish species that is native and in the family Centrarchidae. The curves on Figure 7 (both top and bottom plots) represent a 95th quantile regression from reference sites only up to 40 sq mi. The maximum expected value at 40 sq mi (6.84) was used as the maximum expectation for sites greater than 40 sq mi in catchment size. The reference site derived maximum (95th quantile) species lines are plotted again vs. ALL available data (Figure 7, bottom) to illustrate the fit to entire data set. We initially considered, based on the reference points, separate curves for the North-Central and Southern small streams. On the plot of ALL data the <40 sq mi curve seems to be a reasonable expectation for both regions.

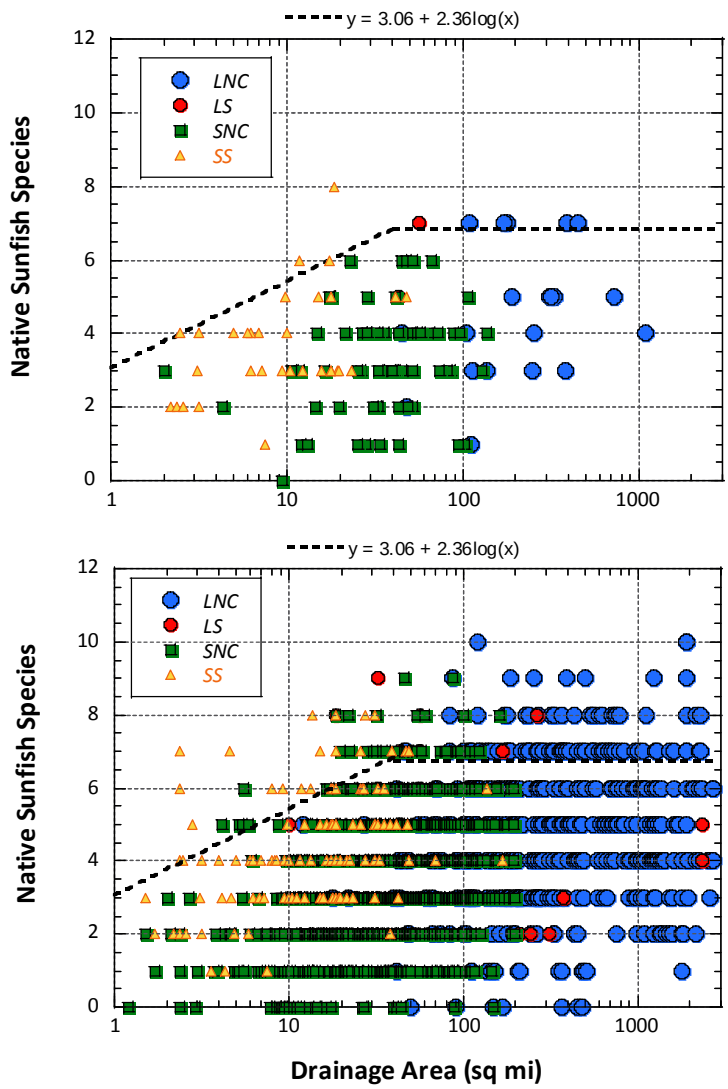


Figure 7. Plot of native sunfish species richness vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of reference sites up to 40 sq mi drainage; 95th quantile value was then extended straight across for sites greater than 40 sq mi.

Metric 4: Number of Native Intolerant Species

This metric includes any species that is native and considered intolerant. The curves on Figure 8 (both top and bottom plots) represent a 95th quantile regression line based on reference sites only and up to 200 sq mi. The maximum expected value at 200 sq mi (9.07) was used as the maximum expectation for sites greater than 200 sq mi in catchment size. The reference site derived maximum (95th quantile) species lines are plotted again versus ALL available data (Figure 8, bottom) to illustrate the fit to entire data set

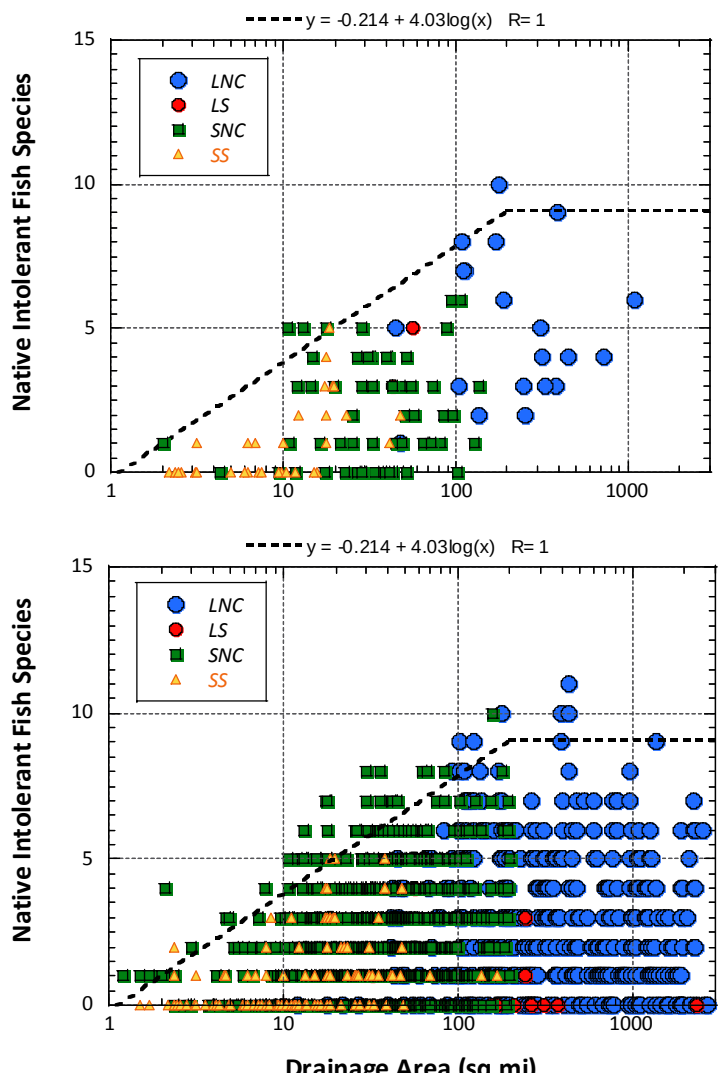


Figure 8. Plot of native sunfish species richness vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of reference sites up to 200 sq mi drainage; 95th quantile value was then extended straight across for sites greater than 200 sq mi.

Metric 5: Number of Native Minnow Species (family Cyprinidae)

This metric includes any species that is a native minnow in the family Cyprinidae. Figure 9 (top) is all reference sites. Figure 9 (middle) has ALL sites from southern streams and Figure 9 (bottom) has ALL sites from northern streams. The curves in Figure 9 represent 95th quantile regression lines using ALL data, one for North-Central and one for Southern classifications using ALL sites up to 40 sq mi. The maximum expected value at 40 sq mi (7.85, South; 12.47, North-Central) was used as the maximum expectation for sites greater than 40 sq mi in catchment size. We used ALL sites rather than reference sites alone because there were too few reference sites in smaller drainage area sites in the North-Central region and too few reference sites with large drainage areas in the South for this metric. It was thought that the maximum response across ALL sites provided a more accurate fit with drainage area because sample size was diluted by dividing the data into North-Central and Southern classification strata.

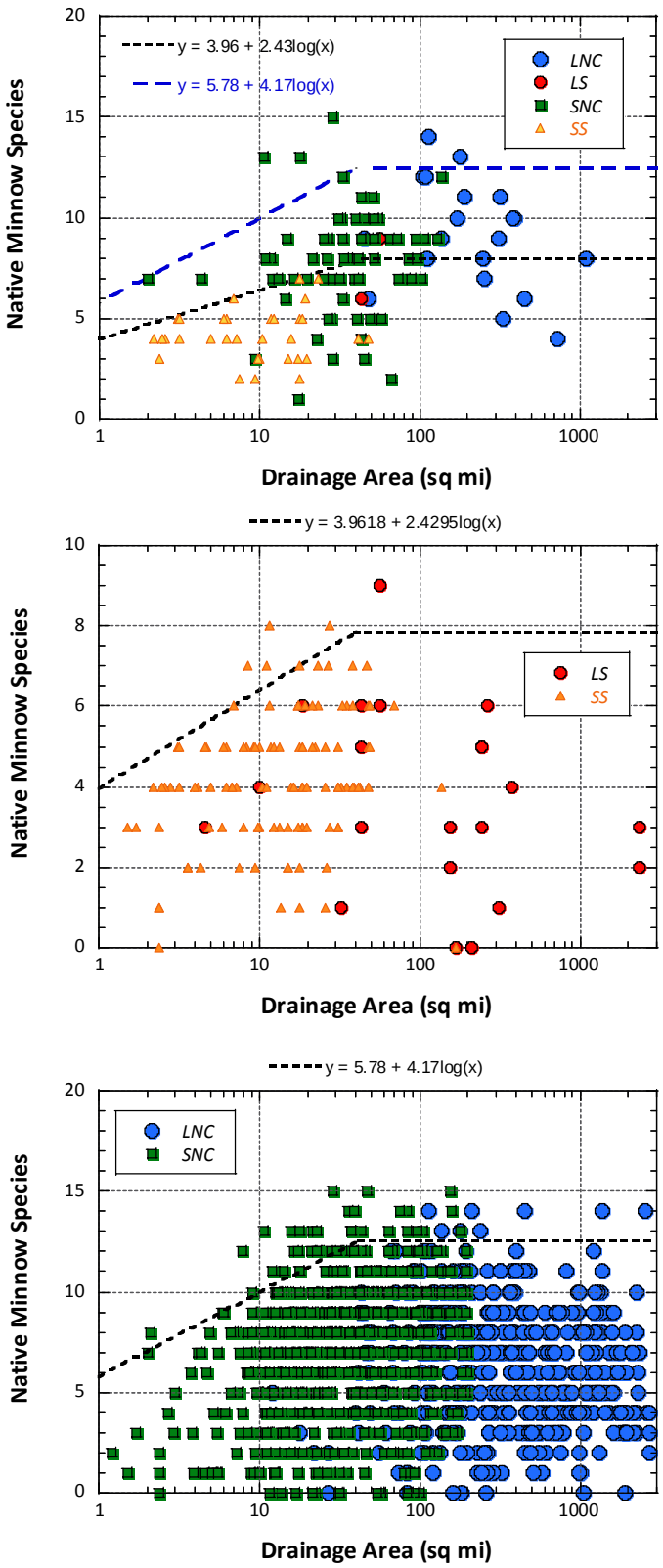


Figure 9. Plots of native minnow species vs. drainage area at reference sites (top) and separately for SS reference sites (bottom, left) and SNC sites (bottom, right). Points coded as Large North-Central (LNC), Large Southern (LS), Small North-Central (SNC) or Small Southern (SS) sites. Calibration curves were generated by using a 95th percentile quantile regression of reference sites up to 40 sq mi drainage; 95th quantile value was then extended straight across for sites greater than 40 sq mi.

Metric 6. Number of Native Benthic Invertivore Species

This metric includes any species that is native and is a benthic invertivore. It replaces the number of darter species metric because it is more inclusive of this important ecological attribute. The curves on Figure 10 (both top and bottom plots) represent a 95th quantile regression line based on reference sites only and up to 200 sq mi. The maximum expected value at 200 sq mi (15.26) was used as the maximum expectation for sites greater than 200 sq mi in catchment size. The reference site derived maximum (95th quantile) species lines are plotted again versus ALL available data (Figure 10, bottom) to illustrate the fit to entire data set

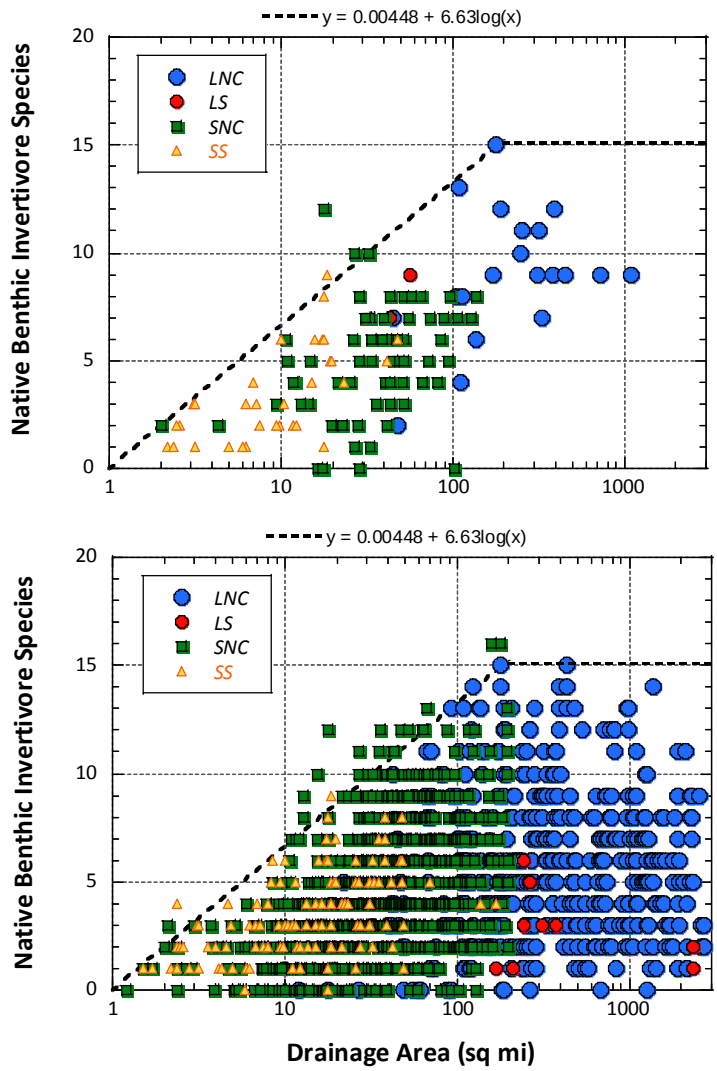


Figure 10. Plot of native benthic invertivore species vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of reference sites up to 200 sq mi drainage; 95th quantile value was then extended straight across for sites greater than 200 sq mi.

Metric 7: Proportion of Individuals that are Specialist Benthic Invertivores

This metric is the proportion of individuals that are specialist benthic invertivores. It does not include species such as creek chub that are generalist and not specialized in their feeding habits. The reference site database did not include very many sites of low gradient and none below a slope of 0.01 (Figure 11, top). The entire database represented a wide range of stream slopes and we use ALL data to derive the expectations (Figure 11, bottom, quantile regression orange dashed line). The slope of the curve using ALL data would be steeper if it was not strongly influenced by low gradient streams with higher proportions of specialist benthic invertivores. The reference site quantile regression is the black dashed line, which was not used in scoring. Because low gradient streams tend be depositional for fines, the apparent changing slope expectation below 0.1 percent slope may be related to widespread sedimentation impact across the lower Midwest rather than a real least impacted condition. In fact, most low gradient streams may not meet reference site definitions.

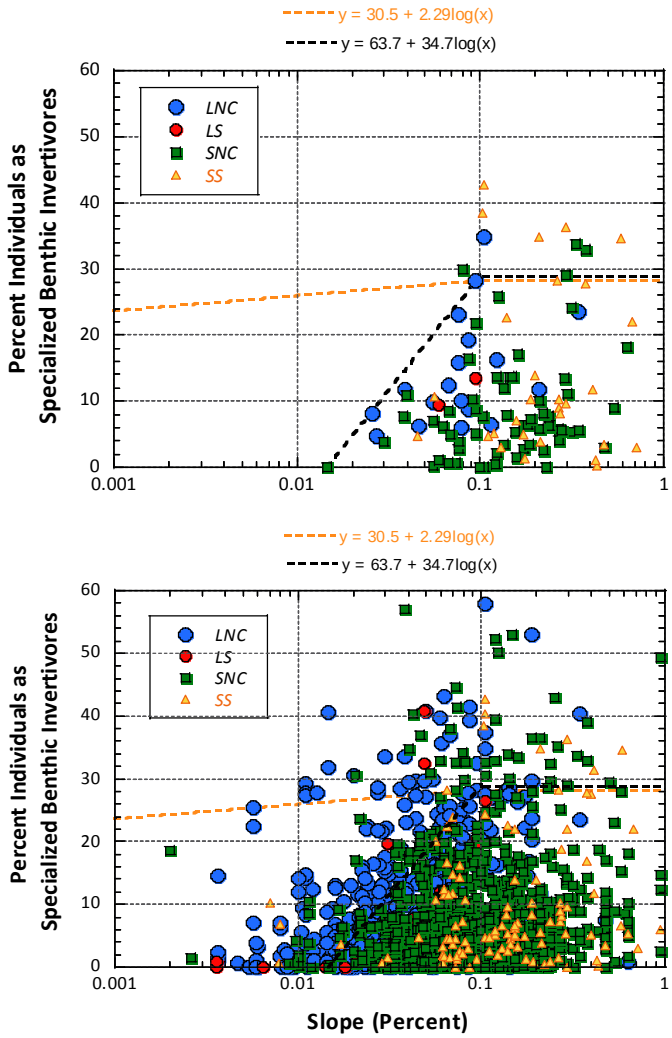


Figure 11. Plot of percent specialized benthic invertivores vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of ALL sites ≤ 0.1 percent slope; 95th quantile value was then extended straight across for sites greater than 0.1 percent slope.

Metric 8: Proportion of Individuals that are Generalist Feeders

This metric includes any species that is a generalist feeder. Smaller proportions of generalist feeders signify biological integrity so this metric is scored higher for lower metric values. The reference sites represented a restricted range of slopes compared to ALL sites in the database. A quantile regression line for the a 5th percentile (potentially the ceiling of the scoring scale) did not provide a meaningful regression line because of the scatter in the relationship. Therefore, we used a "flat" (zero slope) scoring threshold for sites $\leq$ 0.1 percent slope and those > 0.1 using the 5th percentile of the reference sites (Figure 12, top and bottom).

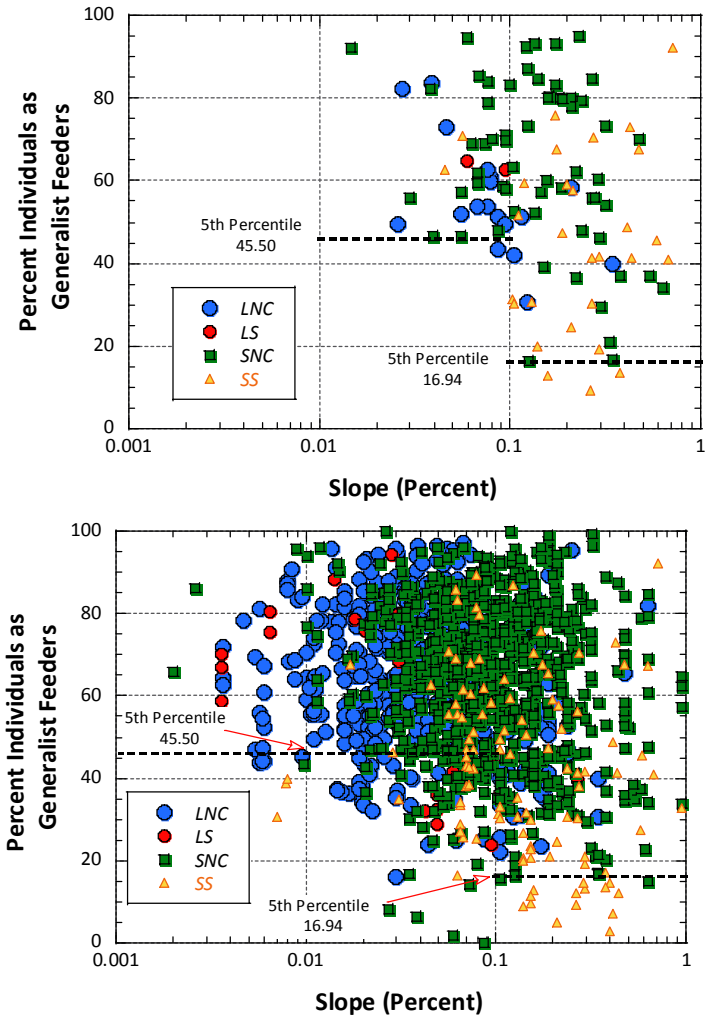


Figure 12. Plot of percent generalist feeders vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration lines were generated by using the 5th percentile of reference sites $\leq$ 0.1 percent slope, and separately, the 5th percentile of reference sites for greater than 0.1 percent slope.

Metric 9: Proportion of Individuals that are Obligate Coarse-Mineral Substrate Spawners (and not Tolerant)

This metric includes any species that is an obligate mineral-substrate (e.g., gravel) spawner and is not considered tolerant (e.g., creek chub). The reference sites represented a restricted range of slopes compared to ALL sites in the database (Figure 13, top and bottom). We calculated quantile regression for reference sites (black dashed line) and for ALL sites (blue dashed line) along percent slope values. We used the regression from ALL sites because it better reflected the expectations for low gradient values (Figure 13, top and bottom). If we used the reference regression, all sites with a stream slope less than 0.01 percent would get a score of 10 for this metric, which would inflate index scores for low gradient sites regardless of metric values.

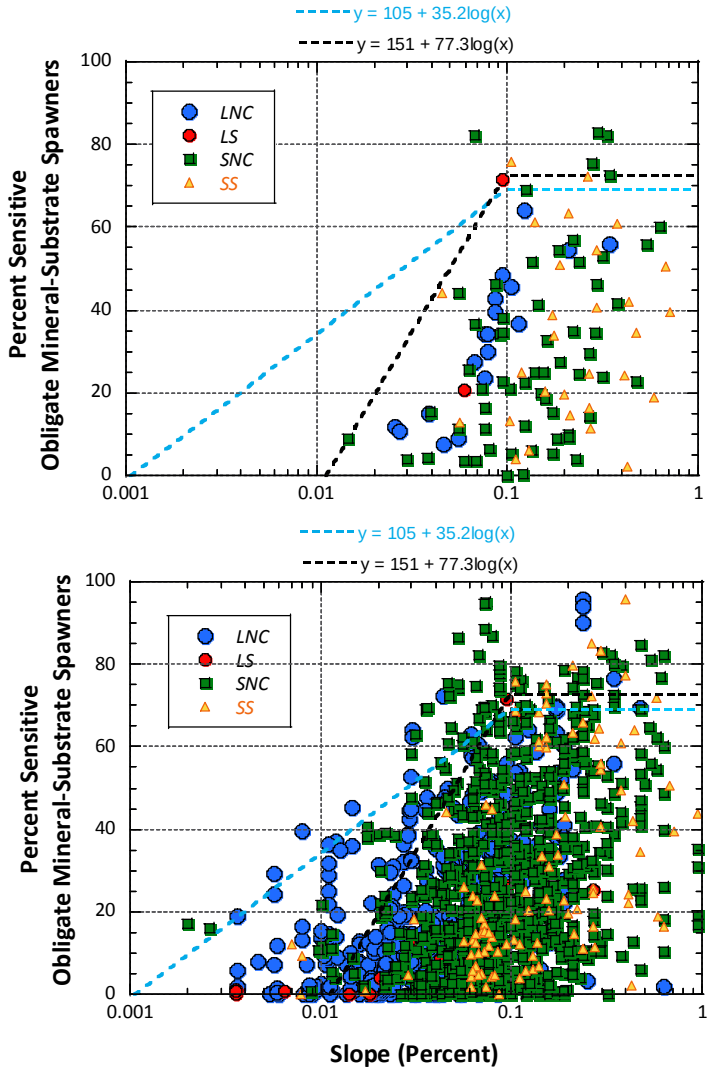


Figure 13. Plot of percent obligate mineral substrate spawners vs. drainage area (sq mi) for reference sites (top) and all available data (bottom). Calibration curve was generated by using a 95th percentile quantile regression of ALL sites $\leq$ 0.1 percent slope; 95th quantile value was then extended straight across for sites greater than 0.1 percent slope.

Metric 10: Proportion of Tolerant Species

This metric is the percent of total species (not individuals) that are considered tolerant of pollution. Although the floor of the reference sites appears similar between the North-Central and Southern classification regions, the ceilings appeared to differ, thus we decided to calculate different ceiling lines for calculating minimum scores (rather than using 100 percent as tolerant) (Figure 14). Because the ceiling values would be best measured with stressed sites we used ALL data for generating these lines. Because the ceiling (blue dashed lines) and floor (black dashed lines) expectations varied continuously with stream size we did not use a drainage area cutoff (e.g., 200 sq mi) as we did for other metrics (Figure 14, middle and bottom). For this metric, the score is calculated as:

$$10 * \frac{(\text{predicted ceiling} - \text{observed metric})}{(\text{predicted ceiling} - \text{predicted floor})}$$

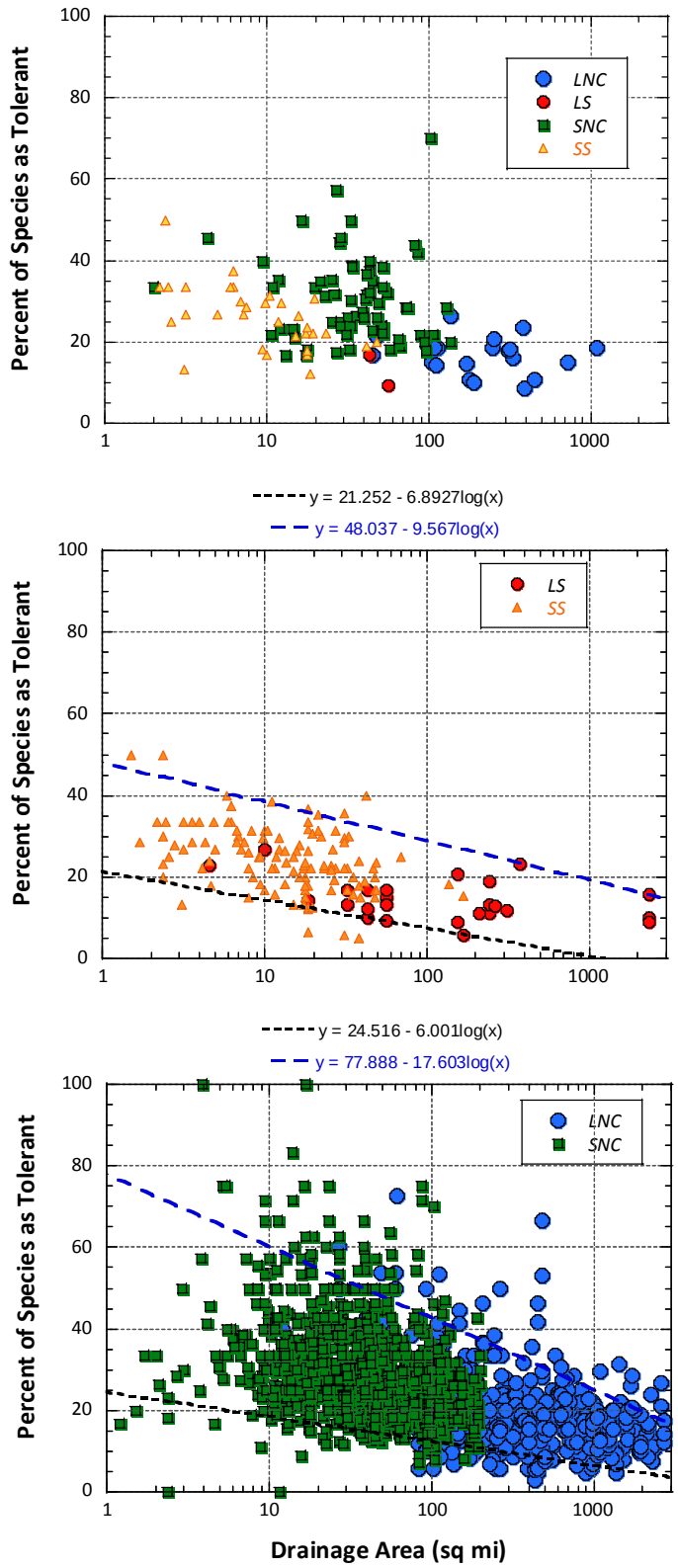


Figure 14. Plots of percent obligate mineral substrate spawners vs. drainage area at all sites (top) and separately for SS reference sites (bottom, left) and SNC sites (bottom, right). Points coded as Large North-Central (LNC), Large Southern (LS), Small North-Central (SNC) or Small Southern (SS) sites. Calibration curves were generated by using a 5th percentile floor quantile regression and a 95th percentile ceiling quantile regression of ALL sites.

Calculation of the IL Continuous IBI

Each metric of the new IBI has a potential score of 0-10 and there are ten metrics in the index. Index scores are calculated as outlined in the metric descriptions using IEPA taxa attributes and scoring formulas in Table 3. Metric scores with calculated values greater than 10 or less than 0 are re-set to the 10 and 0 values (respectively) before they are added in the multimetric index. The final continuous fish IBI score is calculated by adding scores of all 10 metrics, resulting in an index score that can range from zero to 100 points.

4.0 Comparison of the Existing IL Fish IBI with the Continuous Fish IBI

Statewide, the continuous fIBI was well correlated with original fIBI (Figure 15). We coded individual fish regions to visualize and tabulate which regions may contribute most to the variation between the original and continuous fIBI (Figure 16, Table 4). Along the existing fIBI scale, regions 2-6 in northern Illinois had higher continuous fIBI scores compared to continuous fIBI scores in the southern regions 7-11. Region 12 in the southern part of Illinois contributes the most variation in the relationship, as can be seen in a relatively broad scatter of points (Figure 16) and a relatively low r^2 value (Table 4). It also had higher continuous fIBI scores relative to other southern regions along the existing fIBI scale.

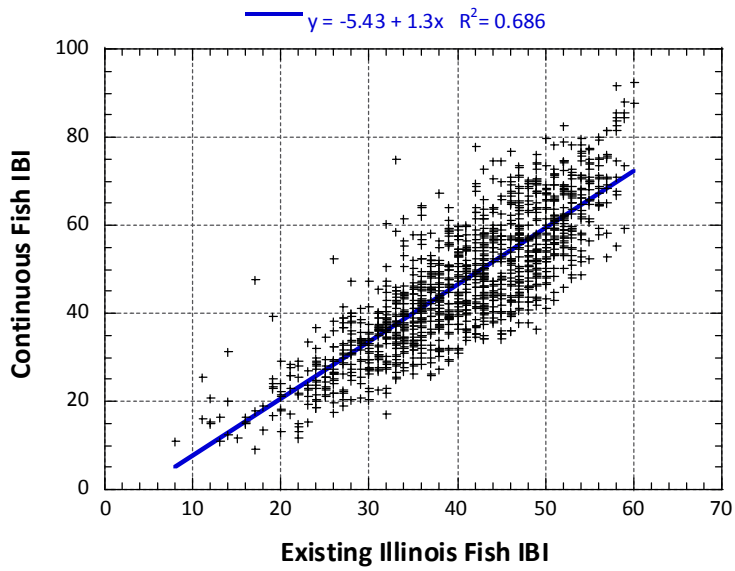


Figure 15. Plots of the existing Illinois Fish IBI versus the continuous Fish IBI Statewide.

The continuous fIBI potentially ranges between zero and 100 points, but in this data set the observed range is from 5 to 85. In contrast, the original fIBI potentially ranges from zero to 60 points and the observed range is 8 to 60. When applied on a statewide scale, a score of 20 is comparable for both index scales, as can be seen in the linear regression line in Figure 15. At the upper end of the scales, a score of 50 for the existing fIBI is comparable to a score of 60 for the continuous fIBI.

Table 4. Table of r^2 values between the existing Illinois Fish IBI and the continuous Fish IBI.

Region	2	3	4	5	6	7	8	9	10	11	12
r^2	0.88	0.95	0.84	0.78	0.83	0.89	0.87	0.81	0.67	0.79	0.63

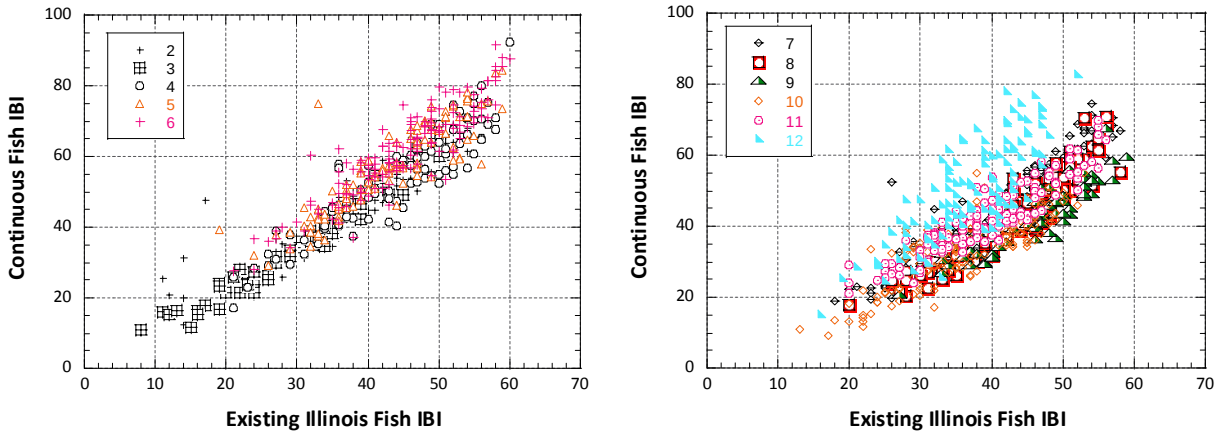


Figure 16. Plots of the existing Illinois Fish IBI vs the Continuous Fish IBI by region, for fish regions 2-6 (left) and fish regions 7-12 (right).

5.0 Conclusions

Our goal in this effort was to recalibrate the Illinois fish IBI using the same metrics that were used in the existing fIBI, but using a continuous scoring methodology and a simpler classification scheme. Recent analyses of discrete vs. continuous scoring approaches found a reduction of bias in IBI scores and in variability of identifying aquatic life impairment decisions when continuous approaches were used (Fore et al. 1994; Blocksom 2003; Dolph et al. 2010). The form of bias that can result from a discrete vs. a continuous score is the possibility of scores in a discrete approach that underestimate the quality of good sites and overestimate the quality of poor sites.

The recalibration depended on a scoring scale that was anchored in the metric values observed either in the least disturbed reference sites or in all sites. The 95th percentile of observed values were used to describe the best metric values possible in each site relative to watershed size or stream slope. The 95th percentile discounts extreme values that might be unrepresentative or might skew the scoring scale towards an unrealistically high expectation. The 95th percentile was used to define the scoring scale instead of a reference mean value because it allows for better distinction of conditions at the higher end of the assessment scale. If the mean of reference was used for defining the upper range of the scoring scale, there would be greater chance that the most excellent conditions would not be recognized because of a limited upper end of the scale.

The existing fIBI uses discrete scoring thresholds based on central tendency of the metric values of least disturbed fish samples. Metric scores of 5 and 6 are just below and above the central statistic of the metric. This allows for a conceptualization of metric and index scores relative to reference central tendency. The continuous scoring approach does not have the same conceptual framework, but instead is based on the highest representative potential metric values, which are mostly derived from reference sites but could include non-reference sites with excellent metric values.

The highest representative potential metric values in reference sites were expected to be similar to the highest values observed in all sites and this expectations was confirmed in comparative plots of metrics in all sites and reference sites. The 95th percentile from all sites was only used when the reference sites were not represented for all site types (such as low gradient sites) or when the North-Central and Southern regions were calibrated separately and the reference sites were sparse within each group. For example, all sites (not just reference) were used in scoring the Proportion of Individuals that are Specialist Benthic Invertivores. It is clear from Figure 11 that if we used reference sites to scale the scoring, any site with a low slope (<0.02) would score a 10. Because we did not want to score all low gradient sites with high metric scores, we used scoring expectations derived from all sites, which show different potential for this metric in low gradient sites.

The result of the recalibration shows a strong similarity to the previous index, but hopefully a reduction in bias that may be inherent to a discrete scoring approach. The statewide approach to classification (except for a North-South distinction for minnow species) results in some slight differences in index value distributions among fish regions that were previously used as primary classification strata (Figure 16). Northern sites (regions 2-6) appear to score somewhat higher using the continuous fIBI relative to southern sites (regions 7-11), as can be seen in Figure 16.

Most regions had high correlations between the continuous and existing fIBIs (Table 4). The greatest variability among fIBI scores was in fish region 12, which is essentially the Interior Plateau (north and south Shawnee Hills) ecoregion in Illinois. Sites in this region score substantially higher in the continuous fIBI compared to sites in other southern regions (Figure 16, right). This scoring difference may be reasonable given that this area of Illinois is generally the least disturbed region of the state based on analyses done by the Illinois Natural History Survey (Figure 17).

Because this may well result in a change in the scoring of a site with the continuous fIBI compared to the original fIBI, analyses of results should be compared in a series of specific situations, along the human disturbance gradient (poor to excellent conditions) to ensure that the ratings of the continuous IBI match the expectations of the field biologist. The hope is that even a small reduction in bias can resolve situations where sites are either over- or under-rated as to their biologic integrity. This in turn would be expected to improve the index ability to discriminate among key stressors in statistical analyses.

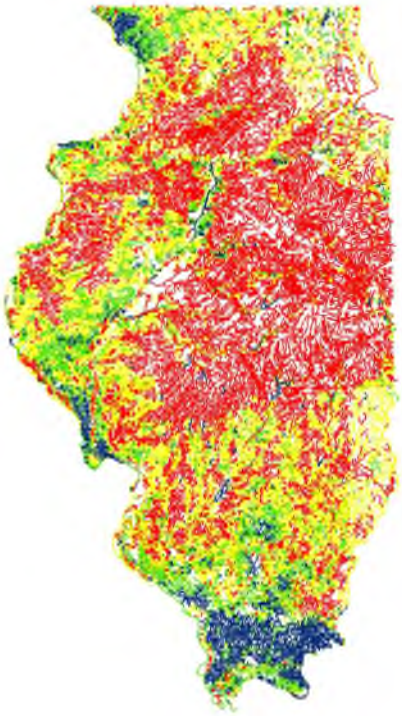


Figure 17. Fig 2 of Sass et al. (2002). Distribution map of disturbance ratings for proportions of disturbed land in the whole watersheds of Illinois streams. Stream arcs are color coded by disturbance rating: 1-5 = blue, 6-10 = green, 11-15 = yellow, 16-20 = red. Smaller ratings indicate less disturbance.

6.0 References

Aparicio, E., G. Carmona-Catot, P. B. Moyle and E. García-Berthou. 2011. Development and evaluation of a fish-based index to assess biological integrity of Mediterranean streams. Aquatic Conservation: Marine and Freshwater Ecosystems 21: 324–337.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers. EPA 841-B-99-002. USEPA, Washington.

Blocksom, K.A., 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic highland streams. Environmental Management 31, 670–682

Dolph, C.L., A.Y. Sheshukov, C.J. Chizinski, B. Vondracek and B. Wilson. 2010. The Index of Biological Integrity and the bootstrap: Can random sampling error affect stream impairment decisions? Ecological Indicators 10: 527–537.

Fore, L.S., Karr, J.R., Conquest, L.L., 1994. Statistical properties of an index of biological integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences 51, 1077–1087.

Sass, L., L.C. Hinz Jr., J. Epifanio and A.M. Holtrop. 2010. Developing a Multimetric Habitat Index for Wadeable Streams in Illinois. Illinois Natural History Survey, Institute of Natural Resource Sustainability, University of Illinois Champaign-Urbana, Illinois Department of Natural Resources, Watershed Protection Section. Prepared for: Illinois Natural History Survey. Grant / Project Number: (T-25-P-001).

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

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey, Urbana, Illinois. Special Publication no. 5.

Klemm, D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, W. T., Thoeny, M. B. Griffith, and W. S. Davis. 2003. Development and Evaluation of a Macroinvertebrate Biotic Integrity Index (MBII) for Regionally Assessing Mid-Atlantic Highlands Streams. Environmental Management 31:656–669.

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

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

Mebane, C.A., T. R. Maret and R. M. Hughes. 2003. An Index of Biological Integrity (IBI) for Pacific Northwest Rivers. Transactions of the American Fisheries Society 132: 239–261.

Rankin, E.T. 2010. Calibration of the Ohio IBI and ICI using continuous scoring methods. Environmental Management Program, Voinovich School of Leadership and Public Affairs, Building 22, The Ridges, Ohio University, Athens, OH 45701

Smogor, R. 2000. Draft Manual for Calculating Index of Biotic Integrity Scores for Streams in Illinois, August 2000. Prepared for: Illinois Environmental Protection Agency and Illinois Department of Natural Resources. 23 pp.

Smogor, R. 2005. Draft manual for Interpreting Illinois Fish-IBI Scores. Prepared for: Illinois Environmental Protection Agency. 26 pp.

Appendix A Site Classification for Fish in Illinois

Our interest in classifying streams for fish is twofold. First and foremost, we would like to find meaningful and efficient groupings of sites that will allow for development of a properly calibrated index of biological integrity (IBI). Second, we would like to find common ground between any new site classes and other classification schemes, especially the classes used in the current Illinois fish IBI and classes discussed in ongoing biological condition gradient (BCG) exercises. The IEPA has given valuable insights on classification issues during ongoing BCG conferences. Those insights have been explored through these analyses. The variables that have been discussed include stream size, which is related to sampling gear and methods. There are also some unique portions of the state, such as the Driftless Area and the Shawnee National Forest.

Classification is based on fish samples from best reference and reference sites, metrics of the existing IBI, and characteristics of the sites that are not subject to human influence (classification variables). We used two methods to relate the fish metrics to site characteristics. First, we looked at existing classification schemes to calculate classification strength for each scheme and variations with fewer groupings. Metric differences within and between groups of sites showed how the classification optimized precision within groups and distinction between them.

The second method used principal components analysis (PCA) to organize the ten metrics of the existing fish IBI into a few axes based on metric correlations among sites. We could then focus on the few axes to find classification variables that were correlated to the important and common assemblage variability. We could assign thresholds to the correlated classification variables that separated sites in the PCA ordination diagram into cohesive groupings, suggesting new site classes.

Regional classifications that were tested in the classification analysis included the existing fish IBI regions, level 3 and 4 ecoregions, and ecological drainage units (EDU). Other classification variables included sampling method, stream size (watershed size and stream width), latitude, longitude, gradient, elevation, depth, Julian day of sample, riffle and pool predominance, stream velocity, canopy cover and substrates. Samples from reference sites (prior to the final review by IEPA) were used for these analyses (n=115) (Figure 1).

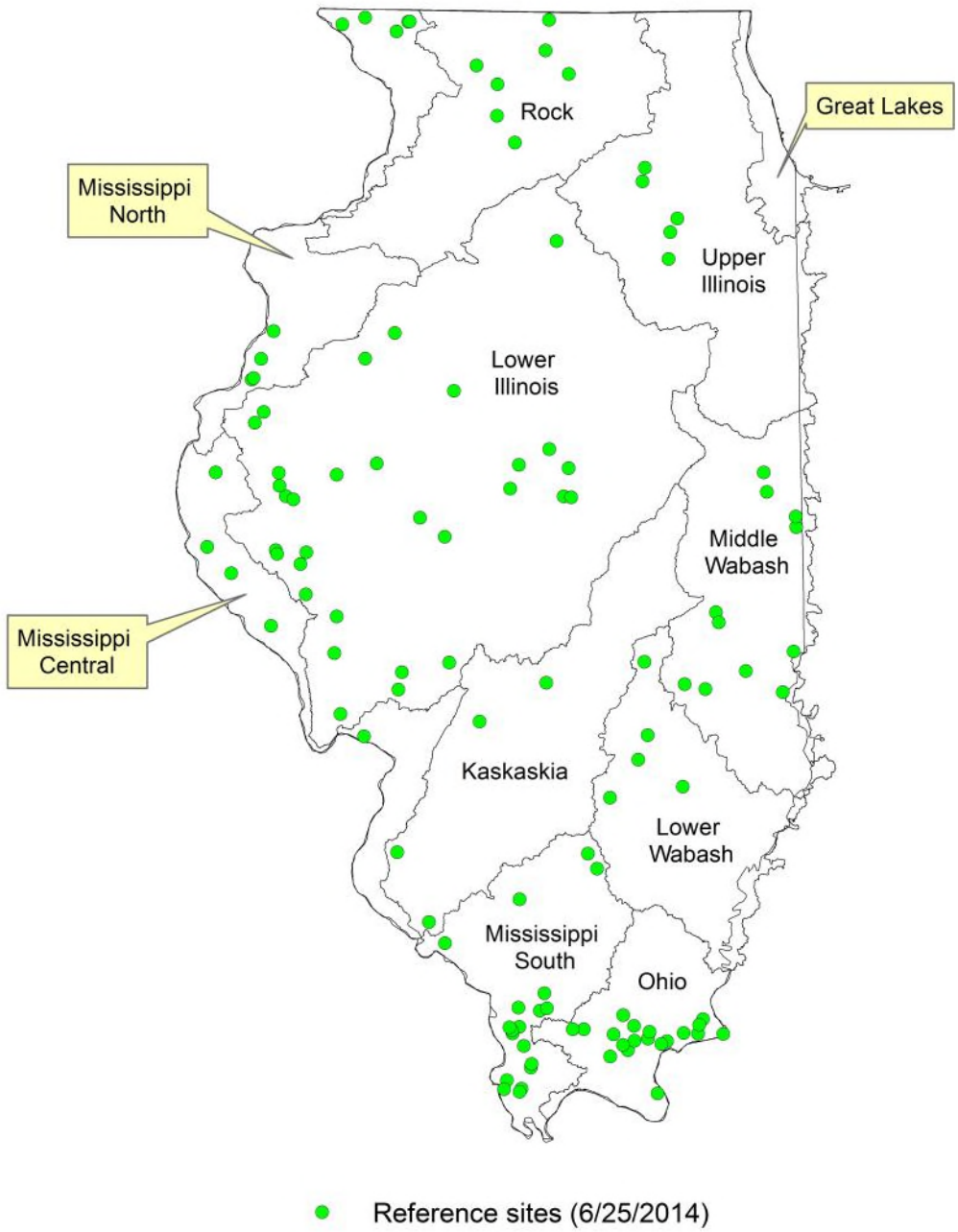


Figure 1. Samples from reference sites that were used in the classification analyses located within ecological drainage units.

We examined similarities among samples by calculating a Bray-Curtis (BC) similarity index for each reference sample pair, as demonstrated by Van Sickle (Van Sickle 1997, Van Sickle and Hughes 2000). We compared the 10 metrics in the existing fish IBI after transforming each to a percentage score of the observed range of reference values (Van Sickle and Hughes 2000). The samples were grouped according to the classification scheme to be tested. The average BC value within groups, weighted by sample size (W), was compared to the average BC value between

groups (B). Distinct classification was indicated by high similarity within groups compared to between groups, measured as the ratio of B/W and the difference of W-B.

The existing regional classification schemes had several categories which we attempted to simplify by pooling similar groups. The similarity between groups was estimated as the B for the two groups being considered for pooling. When B was similar to W for each group, pooling was considered.

We also looked at principal components analysis (PCA) based on metrics (converted to the percentage scale) to explore relevant relationships with the metrics and both categorical and continuous classification variables. The PCA condenses all the metrics into fewer axes with sets of metrics related to each axis. Classification variables can be related to the axes after they have been structured around the metrics. This allows interpretation of the environmental characteristics that are related to major variations in metrics.

Results

All Illinois classifications tested were weak compared to classifications reported by Van Sickle and Hughes (2000) and Hawkins et al. (2000). Nevertheless, the existing IBI classification scheme with 13 groups had better performance than the other groupings (Table 1). Groups with high B similarity were combined. When grouping as follows, 1+2, 4+5+6, 7+8+9, 10+11+13, and 12 (Figure 2); the classification strength for the 5 combined groups was less than the classification strength for the original 13 groups (Table 1). The strength of classification based on ecological drainage units (EDUs) also was tested both in the 10 original groups and in 5 combined groups, as follows, 5+8, 1+4+9, 2+10, 3, and 6+7 (Figure 3). The 10 EDU grouping (no reference sites were found in the Great Lakes unit) and the compressed grouping had slightly weaker classification strength compared to the existing IBI groups (Table 1). Level 3 ecoregions had classification strength similar to the compressed EDU.

Table 1. Classification strengths of alternative groupings, showing statistics derived from within (W) and between (B) group Bray-Curtis similarity values. Low B/W and high W-B show stronger classifications.

Classification scheme	**B/W**	**W-B**
Existing IBI regions (13)	0.90	0.08
Compressed IBI regions (5)	0.93	0.05
EDU regions (10)	0.92	0.06
Compressed EDU regions (5)	0.95	0.04
Level 3 ecoregion (5)	0.95	0.04
Methods (3 groups)	0.93	0.05
Size and location (3 groups)	0.93	0.05

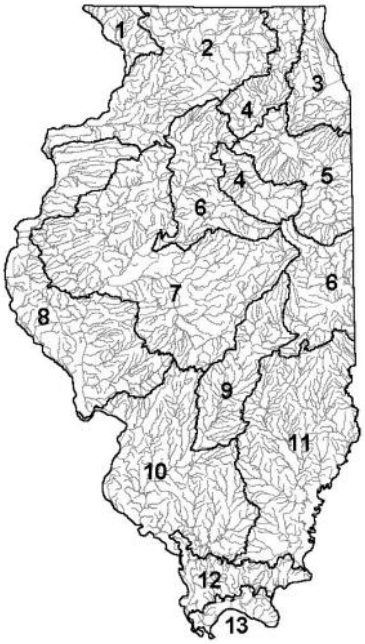


Figure 2 IBI regions in Illinois. Region 4 comprises two noncontiguous areas. We used the following groupings in our analysis: 1+2, 4+5+6, 7+8+9, 10+11+13, and 12.

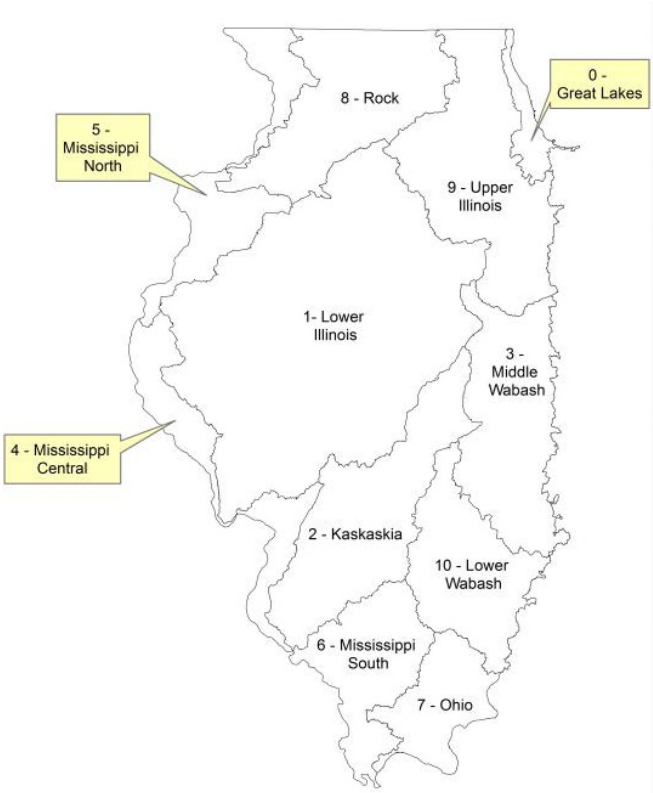


Figure 3.Ecological Drainage Units (EDUs) in Illinois. We used the following groupings in our analysis: 5+8, 1+4+9, 2+10, 3, and 6+7.

Based on distributions of total taxa metric values by sampling method (Figure 4), we grouped methods and tested the groupings as a possible classification scheme. We grouped 1) boatable only and electric seine, 2) boatable + minnow seine and other atypical methods, and 3) backpack shocker or minnow seine only. In the reference data set, the samples were mostly collected with electric seines (N = 92), followed by boat + minnow seine (N = 10), minnow seine (N = 5), and boat only (N = 4). An additional 2 samples were collected with a backpack shocker and 2 more were sampled with atypical methods. The BC comparison showed that this methods-based classification scheme was as strong as others (Table 1).

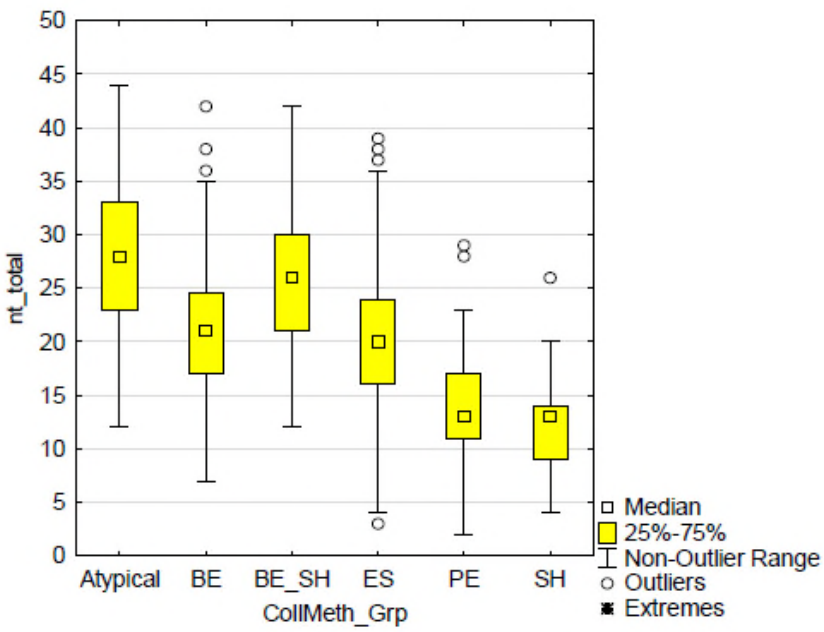


Figure 4. Total taxa by sample collection method, showing distributions for all sites (not only reference). BE = boat methods only, BE_SH = boat + minnow seine, ES = electric seine, PE = backpack shocker, and SH = minnow seine.

We used the PCA ordination to discern patterns with metrics and the categorical and continuous variables. The continuous variables were correlated to the PCA axes to find those that were related to the biological variability (Table 2). The first three axes had eigenvalues > 1.0 and cumulatively explained 76% of the variability. Stream size was correlated to the first PCA axis as measured by wetted width, channel width, and watershed size. No environmental variables were correlated to the second axis at r^2 values greater than 0.20. The strongest correlations on the second axis were with % sand and gradient. On the third axis, latitude and elevation were correlated. On the first axis, the metrics that were related to stream size included benthic invertivore species (bentinve), intolerant species (Intolspe), native species (Natfishs), sucker species (Natsucke), and proportion of tolerant species (Proptols) (also see Attachment 1). Proportions of generalist feeders (Propgenf) and mineral substrate spawners (Propmins) were

related to the second axis. On the third axis, the metrics related to latitude and elevation were minnow species (Natminns) and sunfish species (Natsunfi).

Table 2. Regression coefficients (r^2) for classification variables and metrics on the first 3 axes of the metric PCA, showing variables with regression coefficients > 0.20 on any single axis. Metric codes are explained in the text.

Axis:	1	2	3
Wetted Width (log feet)	**0.530**	0.049	0.011
Channel Width (log feet)	**0.369**	0.024	0.002
Watershed area (log mi^2)	**0.431**	0.123	0.022
Sand (proportion)	0.010	**0.198**	0.072
Gradient (log)	0.040	**0.197**	0.001
Latitude (degrees)	0.139	0.001	**0.306**
Elevation (feet)	0.108	0	**0.299**
Bentinve	**0.795**	0.029	0.023
Intolspe	**0.701**	0.042	0.030
Natfishs	**0.828**	0.108	0
Natsucke	**0.602**	0.126	0.002
Proptols	**0.625**	0.035	0.115
Propgenf	0.061	**0.693**	0.088
Propmins	0.091	**0.621**	0.042
Natminns	0.384	0	**0.457**
Natsunfi	0.268	0.119	**0.385**

When we looked at the stream size variables in relation to the ordination diagram, we could estimate (by eye on the first axis) thresholds for stream classes at wetted width of 40 feet (log40 = 1.6) and watershed size of 200 square miles (log200 = 2.3) (Figure 5). These estimates are subjective, debatable, and could include ranges. We compared the thresholds to changes observed in sampling gear. The wetted width of 40 (+/- 10) feet and watershed size of 200 (+/- 100) square miles correspond fairly closely with the transition from primarily electric seine samples to boat and boat plus minnow seine samples (Figure 6).

Thresholds for latitude and elevation on the third axis were estimated at 38N and 130 feet, respectively (Figure 7). Because we had previously observed in our data and heard from IEPA that the southern regions were biologically distinct, we used IBI regions 12 and 13 as surrogates for latitude. These corresponded to the southern ecoregions including the Shawnee Hills (ecoregion 71) and the Southern Ozarkian River Bluffs (72g). The relationship with elevation was somewhat redundant with latitude (Figure 8), so an elevation threshold was not used.

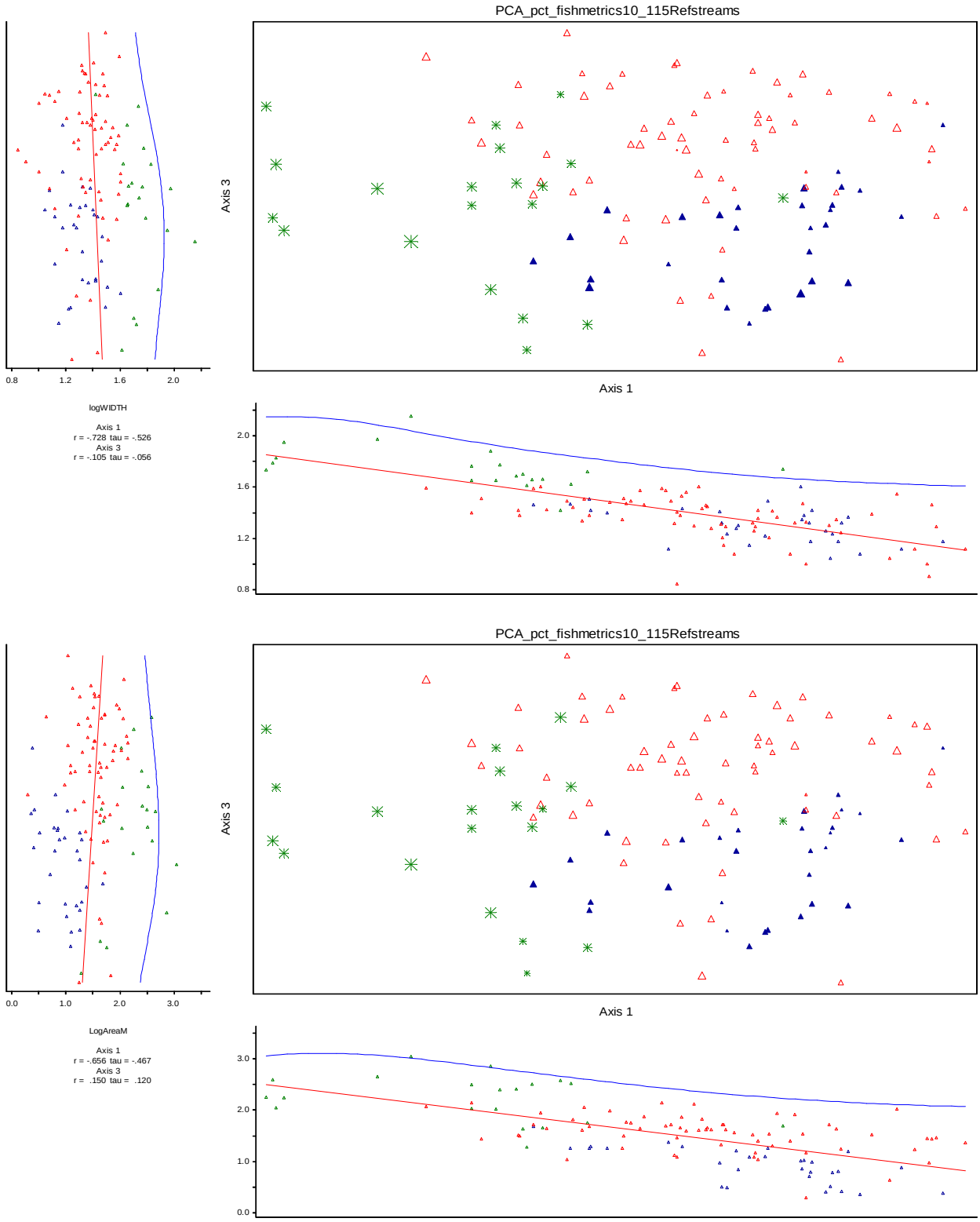


Figure 5. Wetted width (upper) and watershed size (lower) (both log transformed) in relation to the PCA axes 1 and 3. Marker size is related to the magnitude of each variable.

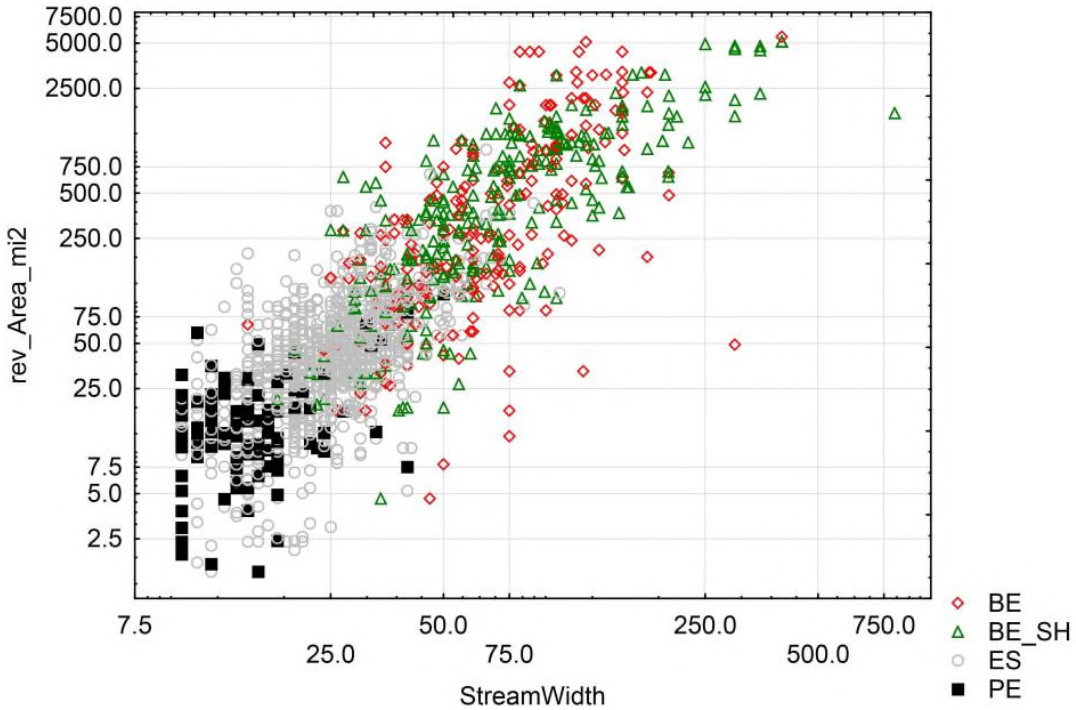


Figure 6. Relationship between drainage area (mi2) and stream width (ft). Axes are log-transformed. BE=boat only; BE_SH = boat plus minnow seine; ES = electric seine; PE = backpack shocker.

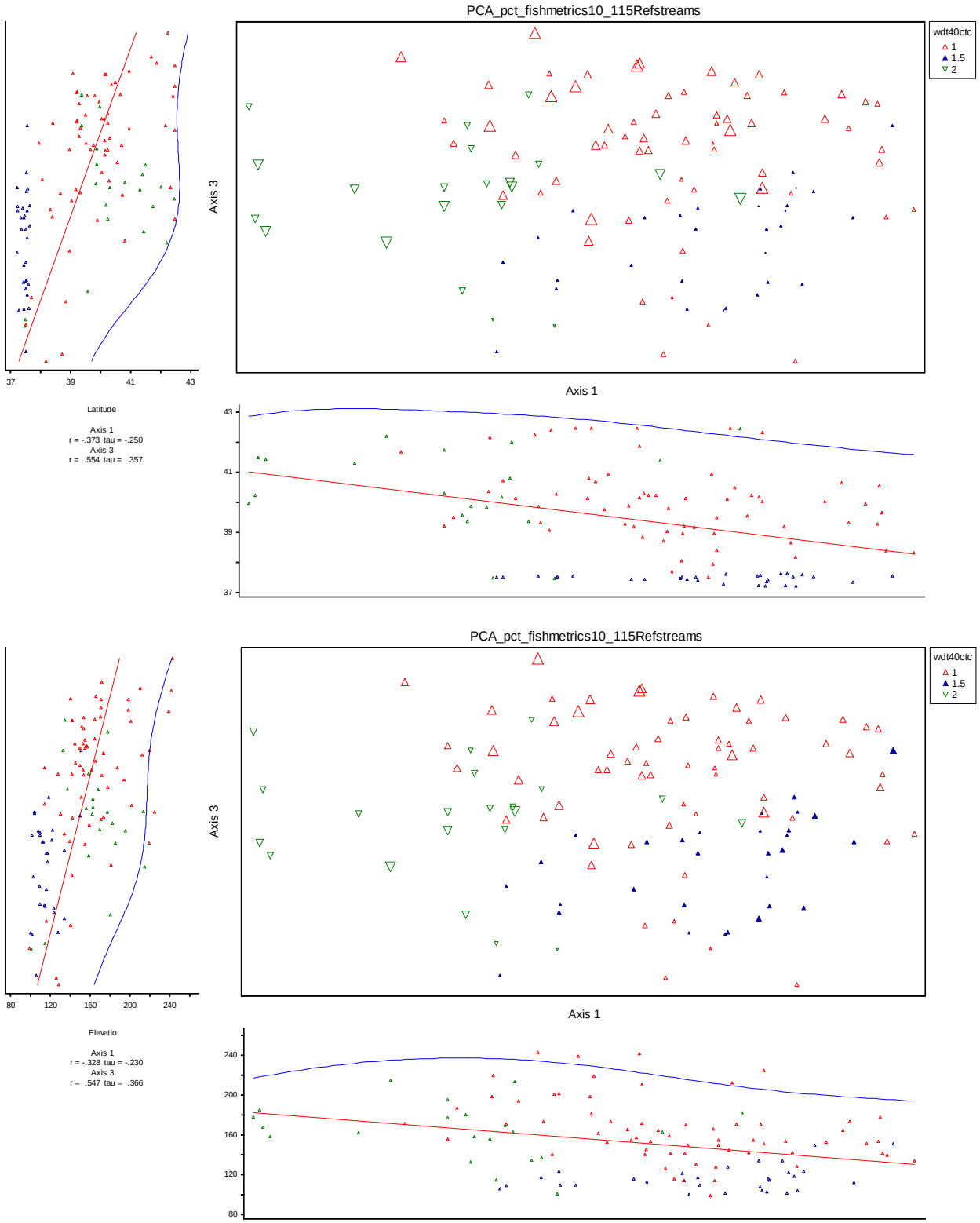


Figure 7. Latitude (upper) and elevation (lower) in relation to the PCA axes 1 and 3. Marker size is related to magnitude.

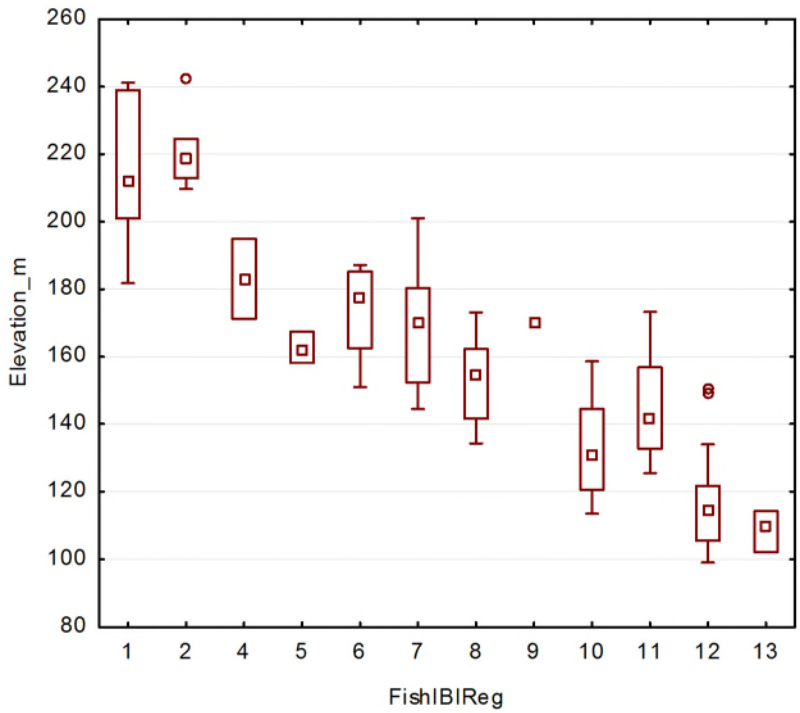


Figure 8. Elevations in the fish IBI regions.

We established three site classes based on size and location: smaller north-central sites, smaller southern sites, and larger sites (Table 3). A map of the southern region is shown in Figure 9. There are so few larger southern reference sites that the larger group was not divided by latitude. Of the 5 samples collected with boat-only methods, 3 were in the larger group. When plotting the potential site classes in ordination space (Figure 10), there was reasonable separation of the groups though with some overlap. The BC assessment showed that these groups were as effective as other potential groupings (Table 1).

Table 3. Potential site classes based on fish metrics in Illinois.

	Class	**Description**	**Reference N**
1	Smaller north-central	≤ 40 foot wetted width and ≤ 200 mi^2 watershed, not in IBI regions 12 or 13 (approximately 37.7 N latitude)	64
2	Smaller south	≤ 40 foot wetted width and ≤ 200 mi^2 watershed, IBI regions 12 or 13	30
3	Larger	> 40 foot wetted width or > 200 mi^2 watershed, any location	21

Metric distributions among stream size, method, and site classes were explored in a series of scatter plots (Attachment A). Native fish species increased with both watershed area and stream width. Relationships among metrics and stream size variables were also addressed in Appendix B. Reference quality streams were not found in the largest streams in the whole data set. The relationships we see in reference streams ranging up to 800 sq mi and 140 feet wide suggest log-linear or linear relationships. We suspect that metric values would level off to a plateau if the size range was increased. Plateaus are apparent for some metrics as width increases above 60 feet. A plateau is less obvious on the log-transformed watershed area scale.

Stream sizes cannot be predicted from sample collection methods. Only the less common methods were related to stream size, with boat methods in larger streams and backpack and minnow seine methods in smaller streams. Electric seine methods were used to collect 80% of the reference samples. Electric seine methods were used in all but the largest streams (>400 sq mi and >65 feet wide).

Relationships between metrics and stream width appear to be strongest for richness of native species, benthic invertivore species, sunfish species, sucker species, minnow species, and intolerant species. The proportion of tolerant species decreases with stream width. Proportions of individuals (benthic invertivores, generalist feeders, and mineral substrate spawners) were not strongly related to stream width. These relationships could be seen over the whole dataset and within each site class, though the plateau effect reduces the stream size effect in the larger stream site class.

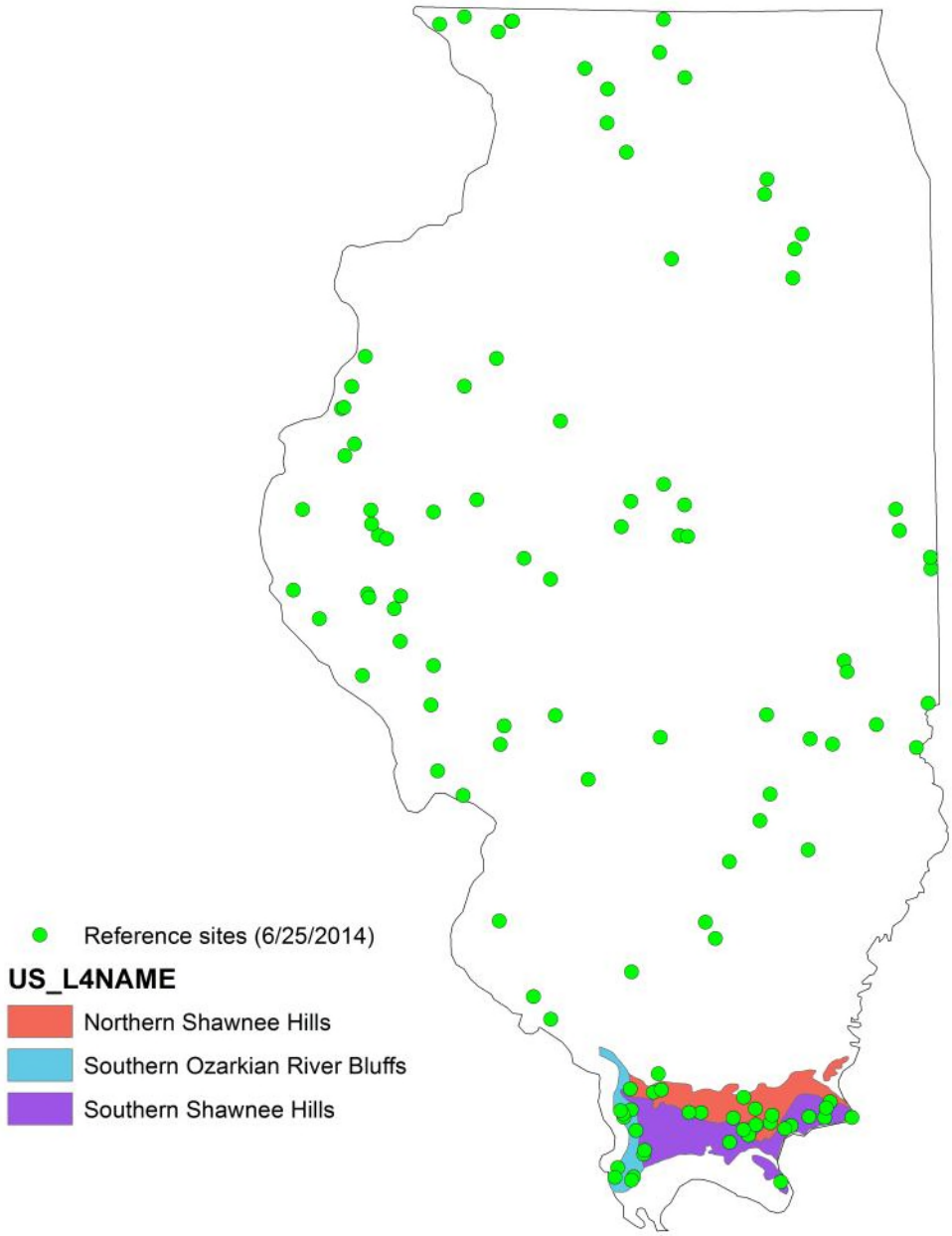


Figure 9. Level 4 ecoregions comprising the 'southern' region.

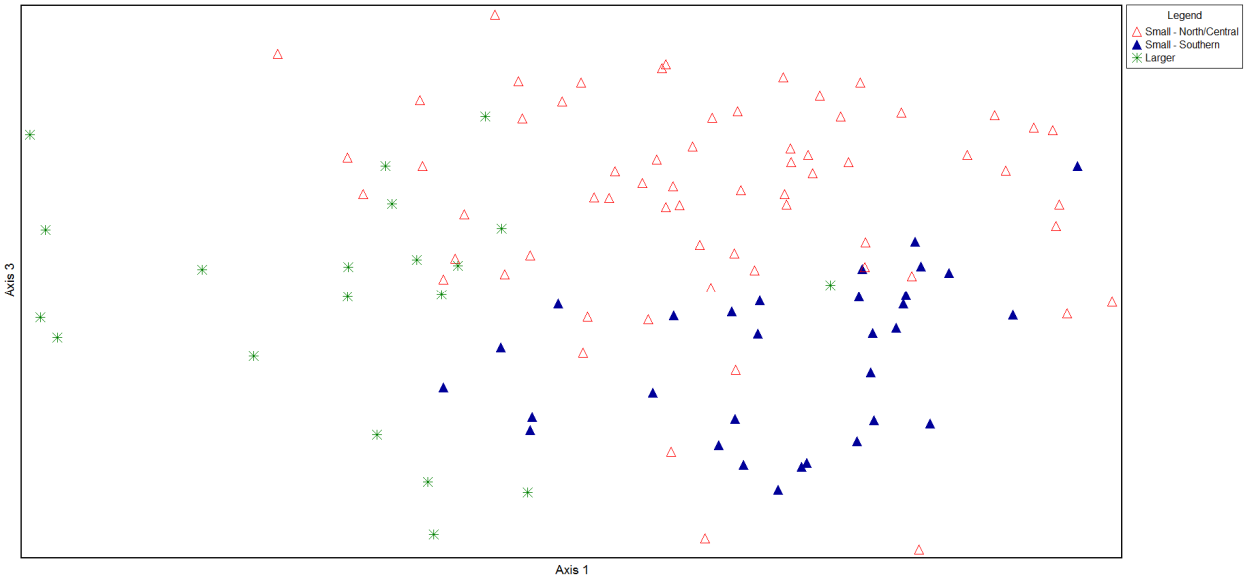


Figure 10. PCA ordination diagram showing sites in potential site classes: smaller north-central (red open triangles), smaller – south (blue closed triangles), and larger (stars).

Discussion

Van Sickle and Hughes (2000) found B/W and W-B values in the ranges 0.91 – 0.94 and 0.03 – 0.05, respectively, for comparing metrics among ecoregions or basins in western Oregon streams. These are comparable to our values in Table 1. Similar or better geographical classifications have been found in other studies (Hawkins et al. 2000). These classification strengths (W-B<0.20) are considered to be weak (Hawkins et al. 2000). We can conclude as others have, that choosing between similarly ordered geographic partitions (e.g., ecoregions versus watersheds) appears to be a fairly minor issue, if one's goal is to find an optimal way to classify stream ecosystems. Greater classification strength can be gained by greater partitioning, but that would also require more data from each class for development of robust assessment tools. Our expert knowledge of the biological meaning of plausible, competing models might be more informative than tests for small statistical differences between the models. Yet, if all of the proposed models show weak classification strength, then we can be convinced that classification is unnecessary.

There were some potential site classes that were poorly represented in the reference data set. These included sites where boat-only methods were used and southern sites outside of the Shawnee Hills or Southern Ozarkian River Bluffs. The lack of boatable reference sites is not surprising, because as streams enlarge to boatable systems, the surrounding landscape is generally more conducive to development. There is a possibility that the difference in samples collected by boat-only methods could be an important differentiation and might warrant classification. However, the distributions of metrics suggest that these samples are on the whole comparable to electric seine methods in reference sites. With these considerations and with

evidence from the PCA and metric distributions that the boat samples were not unique, we lumped the boat samples with other classes but recognize that there may be a need to separate them out, especially in degraded sites with much larger watersheds than we see in the reference data set. We might have seen more evidence for splitting this group (or others) if we used taxa instead of metrics in the analyses.

The comparison of within and between Bray-Curtis similarities showed that there were slight reductions in variability to be gained by classifying sites in Illinois. The 3 classes defined by stream size and location make a reasonable scheme to consider during index development. Alternatively, the 13 region IBI classification scheme showed marginally greater classification strength and could be used, though calibrating and testing metric responses in a dataset so finely divided would not be robust. Instead, the index could be developed statewide and then applied in the regions or classes when determining impairment thresholds.

The differences in classification strengths for the classification schemes is apparently related to the number of classes in each scheme. Table 1 suggests that the more classes, the better the classification strength. That makes sense for spatial/geographic classifications because within-similarity will be higher for small groups of sites very close together due to simple spatial autocorrelation. While a classification scheme with more categories is likely to have greater classification strength, the likelihood of overfitting the model is also greater. If we use a general rule of thumb that we should use 30-40 observations for each parameter to get a good model, then we might want to limit the number of recognizable classes to 3-4 for our 115 site data set.

Our recommendation is that development or recalibration of the IBI in Illinois should be conducted with some consideration of stream size, either through classification of width, watershed size or by boat-dependent methods. Classification could be categorical (more than the 2 classes we describe) or continuous, with metric adjustments when strong relationships are observed (Fausch et al. 1984). In the PCA, the metrics related to the first axis (Table 2) were not very responsive to other axes. That suggests that metric adjustment to size could be more important than regional classification. An alternative might be to express all metrics as a percentage of species or individuals in the total collection to avoid having to adjust assemblage size for watershed area.

Metric differences and Bray-Curtis similarities between sites are likely to decrease among potential site classes after accounting for stream size. One approach for developing the fish index would be to adjust the metrics, find the most responsive candidate index metrics, and repeat the similarity analysis based on the adjusted metrics in potential site classes. If there are high average similarities between potential classes (within the range of ‘within similarities’), that will be evidence for lumping classes. Thresholds, and perhaps metric selection, could be established for the lumped classes, assuming there will be 3-5, each with sufficient numbers of sites to characterize the reference index scores. If the classification strength of these lumped class is less than our current classification strengths (which are already weak), there will be little evidence that classification necessary.

The metrics related to latitude were native minnows and native sunfish. Sunfish were more diverse in the southern region and minnows were more diverse in the north-central region. The reference diversity of minnows in the south and sunfish in the north-central might be too low to expect a meaningful response to stress in the respective regions. Though the metrics had strongest regression coefficients with the third axis, they showed some response to the first axis also, suggesting that adjustment to stream size may be necessary as well as latitude.

The proportion of generalist feeders and of minnows were related to the second axis only. No other metrics were related to the second axis. The environmental variables most correlated with the second axis were percent sand and gradient. Sandy, low gradient sites had more generalist feeders and fewer minnows.

Recommendations

We will adjust metrics to stream size, probably using watershed size for adjustments because in contrast to stream width, watershed size is not subject to stream channel disturbance, measurement bias and variability, nor is it variable over time.

We will develop an index (or indices) within the three classes defined by stream size and location (Table 3). Index development might include statewide metric scoring for metrics that have similar response patterns among site classes and class-specific metric scoring when that is appropriate.

We will derive index thresholds within the three classes as appropriate. If statewide index calibration is feasible, we can test classification strengths for alternative schemes using the adjusted metrics and the new IBI. The alternatives will include considerations of existing site classification schemes. This might result in index thresholds for more or less than three site classes.

References Cited

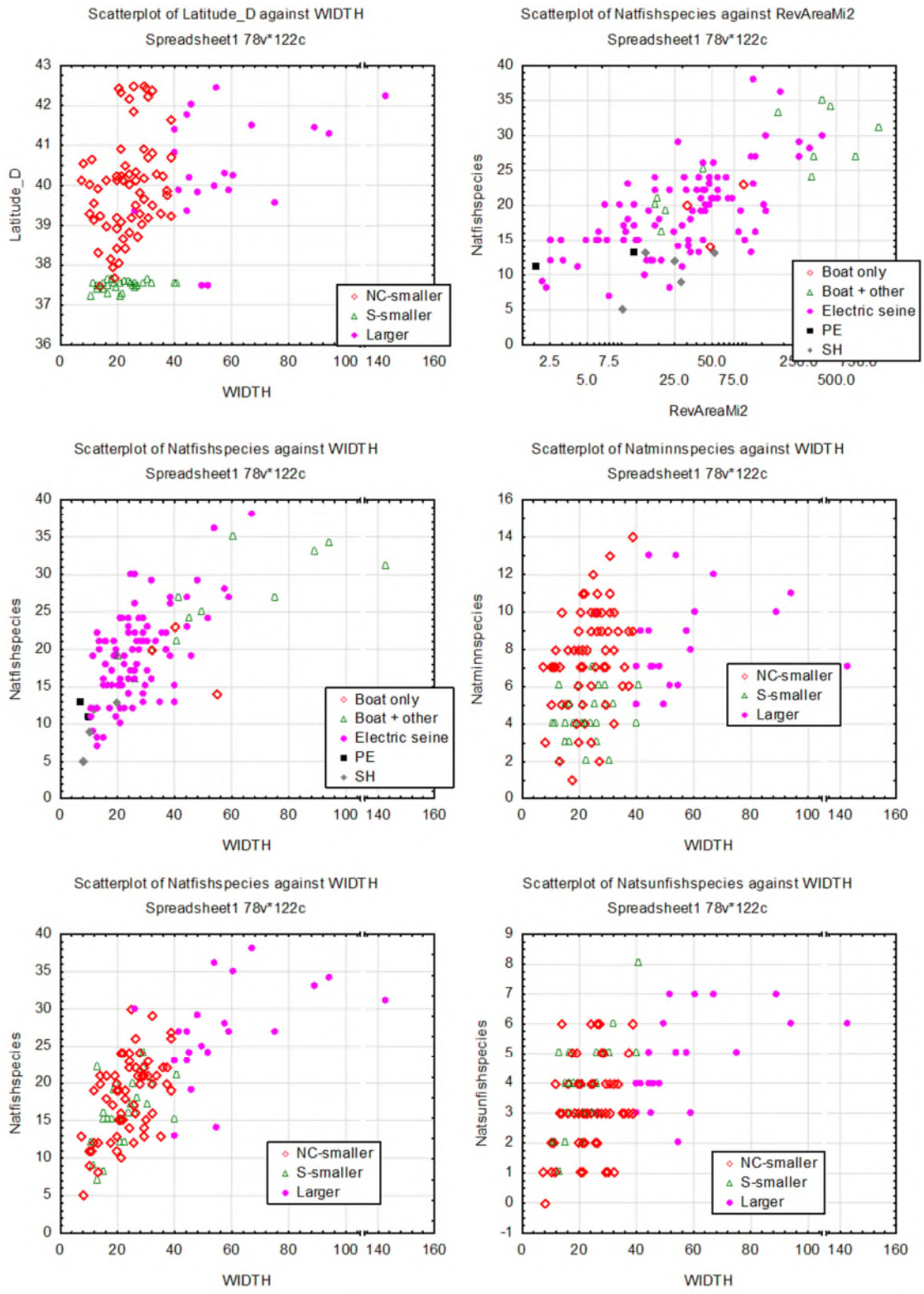
Fausch, K. D., J. R. Karr, and P. R. Yant. 1984. Regional application of an index of biotic integrity based on stream fish communities. Transactions of the American Fisheries Society 113:39-55.

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

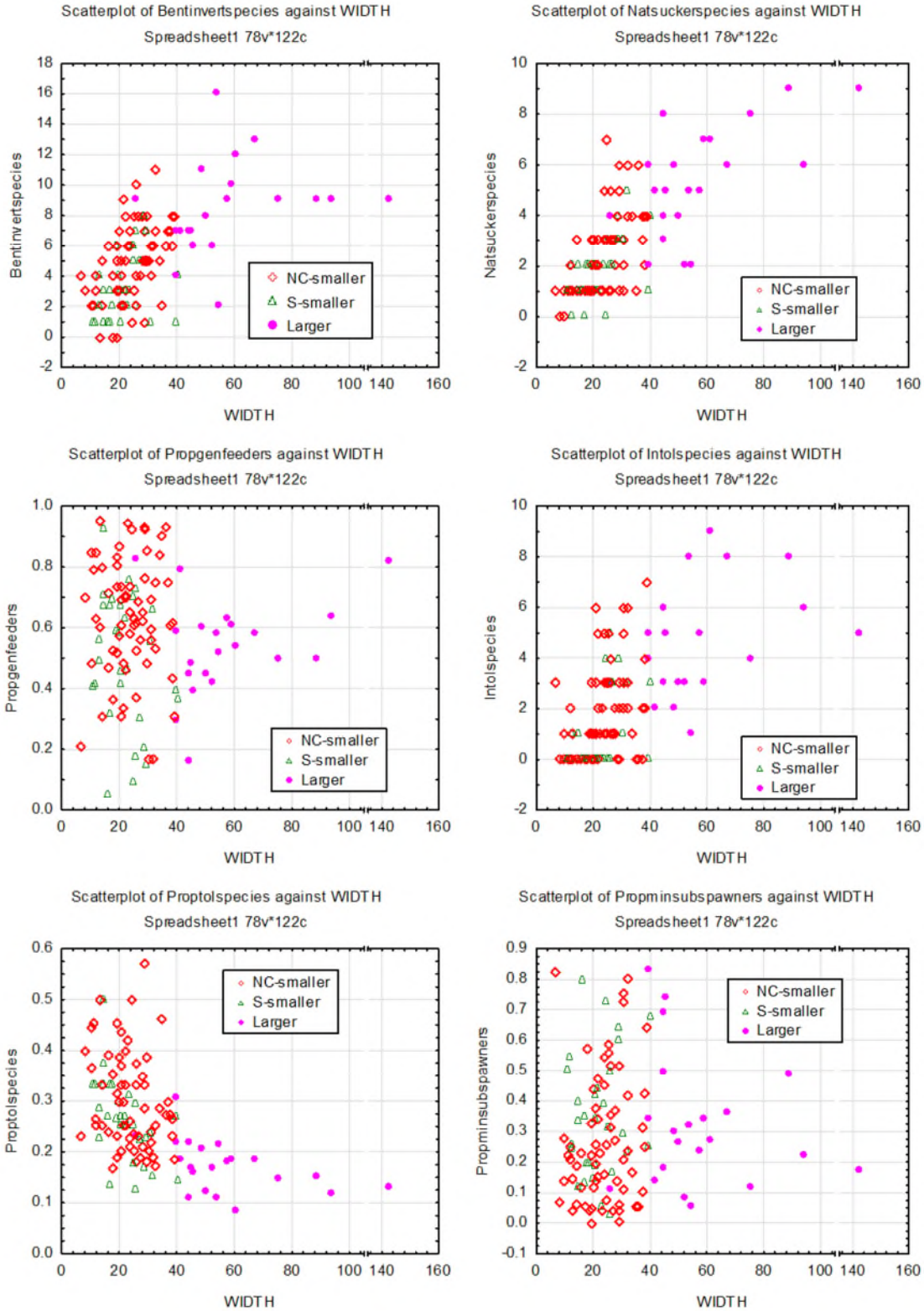
Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. Journal of Agricultural, Biological, and Environmental Statistics, 370-388.

Van Sickle, J., & Hughes, R. M. 2000. Classification strengths of ecoregions, watersheds, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society, 19(3), 370-384.

Attachment 1. Metric distributions among stream size, method, and site classes.



Attachment 1 (continued). Metric distributions among stream size, method, and site classes.



Appendix B.

A Comparison of Catchment Size and Stream Width as Predictors of Fish Species Richness

One of the earliest patterns recognized in fish species distribution and abundance in flowing waters was the pattern of increasing species richness with increasing stream size (Shelford 1911; Sheldon 1968; Allan and Castillo 2007). Small headwater streams, controlling for biogeography, are generally composed of fewer species per unit distance of stream compared to larger waters. This pattern is also an important component of the River Continuum concept (Vannote et al. 1980) which predicts that biological aspects of a stream respond to the gradual change of physical environmental conditions such as the width, depth, water, flow characteristics, temperature, and physical complexity. Derivation of multimetric indices such as the Index of Biotic Integrity (Karr et al. 1986) rely on calibrations of key physical factors such as stream size to develop species richness expectations for reference conditions so that deviations from such conditions that will reflect anthropogenic disturbance.

Examples of stream size variables used in IBI calibrations include stream order, drainage area (*i.e.*, catchment area), habitat volume, and stream width. Hughes and Omernik (1983) considered stream order to be too inaccurate as an estimator of stream size to calibrate species richness expectations. Many multimetric indices use drainage area for calibration because of its ease and consistency of calculation (e.g., Ohio EPA 1989; DeShon 1995; McCormick et al. 2001; Compton et al. 2003; Walters 2006; MPCA 2014). Pont et al. (2009) used "potential stream volume" in the Western US because of great variation in flows for wet vs. xeric areas; where stream volume is calculated as the product of catchment area and runoff. Others have used measures of stream width to calibrate indices (e.g., Whitter et al. 2007), although Pont et al (2009) thought such a direct measure of local stream size too likely to be affected by flow and channel alterations in the western USA. Here we consider the advantages and disadvantage of using drainage area vs. stream width measures to calibrate fish IBIs for Midwest streams with a particular focus on Illinois streams.

Stream Width and Drainage Area

Drainage area and bankfull width have been found to be well correlated across the US with factors such as discharge/runoff, riparian vegetation, sediment size, etc. (Faustini et al. 2009). Wetted width is a more variable parameter than bankfull width because it reflects local short-term variation in stream flows and runoff. Using stream data from Illinois where we had two or more sampling events, drainage area (sq mi) and wetted stream width (ft) are positively correlated (Figure 1, r^2=0.69) although there is enough scatter or deviation in the relationship that

choice of stream size measures could have a moderate influence on the calibration and scoring of stream size-dependent metrics in an IBI. Drainage area size is an invariable measure for a site, but stream width at a site and between sites of the sample drainage area can vary with stream flow state, channel characteristics, and location and number of samples used to estimate the mean width. Although fish species distributions are known to vary by season related to spawning migrations and winter movements, sampling for the purposes of IBI calculation is typically limited to summer and early fall periods when such widespread movements are more limited. Our question here is whether changes in stream width within the typical summer-fall sampling period are a better variable for gaging species richness expectations than drainage area.

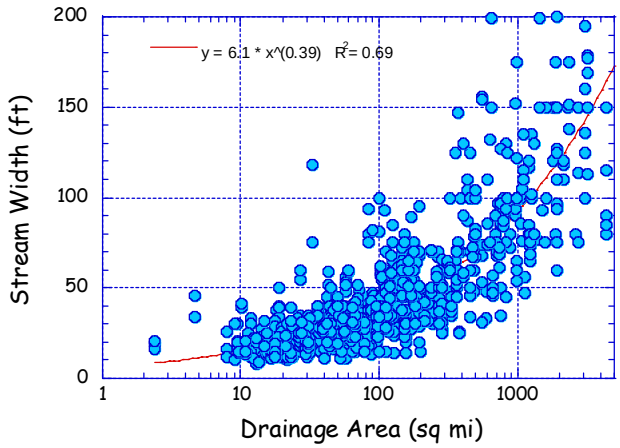


Figure 1. Plot of drainage area (sq mi) vs. stream width (ft) for sites in Illinois with two or more sampling events.

Mechanism of Species Richness Change with Stream Size

The River Continuum Concept provides a framework for attempting to investigate the mechanisms why biological components and processes of rivers show changes along a longitudinal downstream continuum. There are a number of seminal papers that looked at changes in species richness in streams as size increases with explanations including changes in habitat diversity and habitat volume (Shelford 1911; Sheldon 1968).

When using stream width as a predictor of species richness it is assumed that a wider stream will provide more habitat for species than a narrower stream at a given site. While this is consistent with drainage area predictions of species richness in natural streams it can be confounded by artificial widening caused by channelization or urbanization which often results in less diverse habitat and fewer species. For example, Galster et al. (2008) identified a widening of stream channels below an urban area in Pennsylvania. Simon and Rinaldi (2006) reported variation in stream width changes that can vary as much as 0.5–1.0 m/y (TN) to 10–20 m/y (WA) in examples associated with the varying resistance of bed and bank materials to widening. In Illinois where channelization is widespread, widening is artificial, but channels can remain wide (with poor habitat) compared to natural streams of similar drainage area. Because of low stream power substantial fine sediment deposition can occur, maintaining poor habitat conditions until "the inset channel formed by this deposition has enough power to transport the supplied sediment" (Simon and Rinaldi 2006). Conversely, during other stages of channel evolution in response to disturbance, channels may narrow due to downcutting, also known as incision (Simon and Rinaldi 2006).

Habitat, Stream Size and Niche Theory

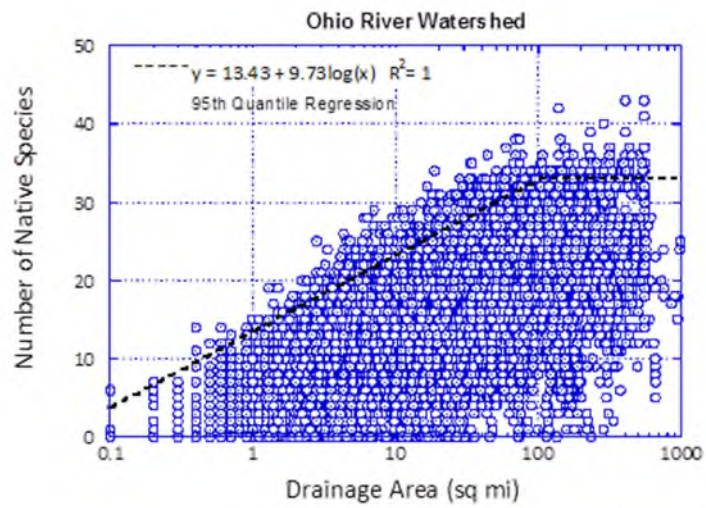


Figure 2. Scatter plot of native species richness vs. drainage area (sq mi) (HW, WD) with 95th quantile regression in streams of the Ohio River basin of Ohio (from Rankin 2010).

The concept that species richness increases with increasing habitat diversity and/or volume is hypothesized as a key mechanism explaining increasing species richness with stream size. Thresholds between maximum fish species richness and stream size as measured by drainage area are particularly distinct in Ohio streams. Figure 2 illustrates a plot of fish species richness vs. drainage area for headwater and wadeable streams in the Ohio River basin in Ohio (Rankin 2010). In a study to examine associations between low flow and species richness in Ohio, we used USGS models to determine mean September flows for ungaged stream stations. The hypothesis underpinning these analyses is that habitat conditions limit maximum species richness expected at a site and low-flow conditions represent the environmental "bottleneck" that determines the habitat available for species at a site, at least during the summer-fall low flow period. September flows represent the lowest average stream flow period across Ohio under typical conditions. The generally sedentary nature of warmwater fish species during the summer/fall period reinforces the view that most species do not respond quickly to short-term changes in stream width that might occur during the summer/fall period.

To test the relationship of species richness with base flow, Qualitative Habitat Evaluation Index (QHEI) data attributes were used to calculate the number of different habitat "niches" that existed at a given site (Rankin 2010).

To estimate the number of habitat niches for streams of a given size we used attributes of the QHEI to "count" habitat types related to pool habitats, stream current types, riffle habitats and physical

Table 1. Summary of general process to classify number of niches at sites with natural channels in Ohio.

Total Niches	Pool Niches	Riffle Niches	Current Niches	Structure Niches
Score 0-63	Increase with depth and with increase depth of riffle and run connecting features (Score 0-15)	Increase with riffle depth and increasing coarseness of substrates (Score 0-15)	Increase with diversity and velocity of flows (0-15)	Increase with variety and amount of cover types (Score 0-17)

structure (i.e., cover) in a stream (Table 1) and calculated the average number of niches by intervals of drainage area (Figure 3). The analysis included streams with largely natural channels. The QHEI contains metrics that measure habitat types (e.g., structure) and other metrics that measure condition and for this analysis relied on habitat "type" attributes to illustrate how available niches generally increase with stream size or stream flow.

As illustrated in Figure 3, the mean number of habitat niches is strongly related to stream size (left) and mean September flow (right) strongly suggesting that habitat features contribute to the increasing species richness in streams as stream size and base flow, as measured by mean September flows, increase.

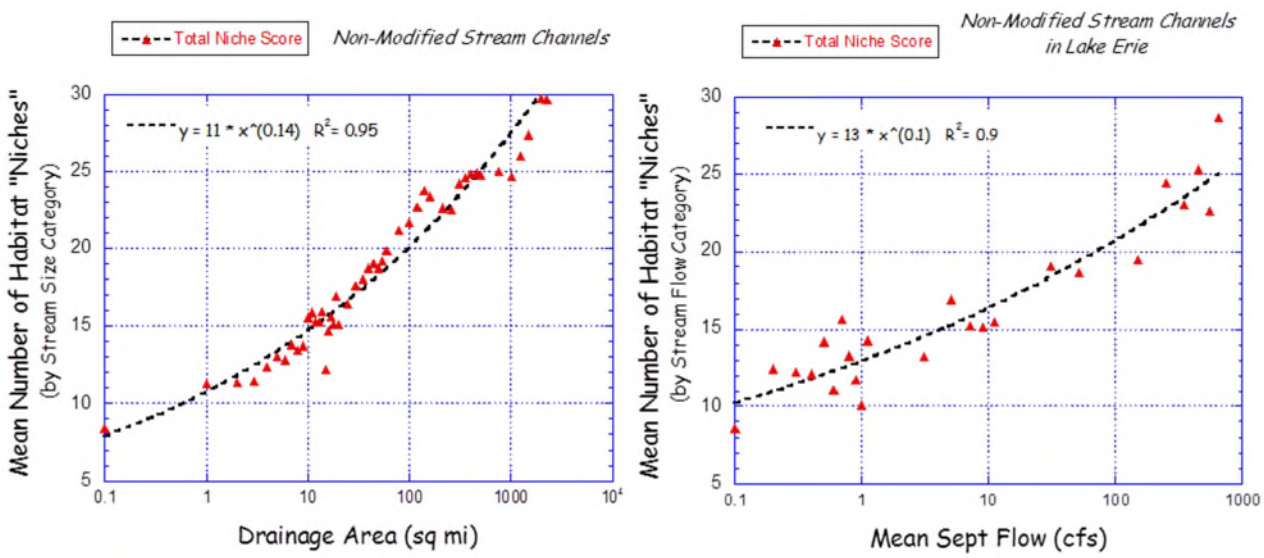


Figure 3. Plots of number of habitat niches (total) by drainage area category statewide (left) or by mean September stream flow category in the Lake Erie drainage (right). Data was restricted to streams with primarily natural channels.

A plot of the 90th percentile value of sensitive fish species by ranges of total average niche scores (Figure 4) shows a strong threshold for number of these species, most of which are habitat specialist and fluvial specialist species. Thus is appears that stream size, through its influence on flow and depth, limits habitat features which in turn places a ceiling on the number of specialist species that may inhabit a site.

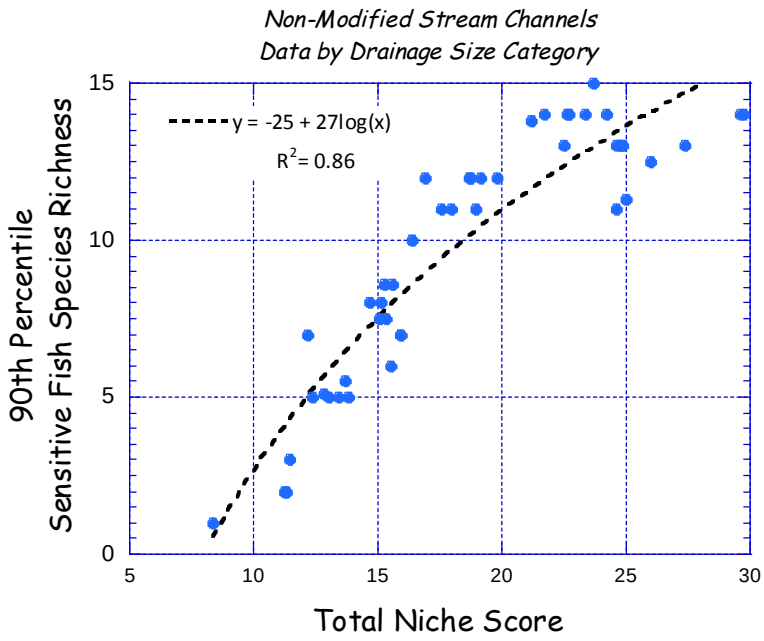


Figure 4. Plot of total habitat niche score vs. the 90th percentile of sensitive Ohio fish species along a gradient of drainage categories. Data from Ohio streams with non-modified channels.

Stream Width in Channelized Streams

The analyses conducted above focuses on streams with natural channels to avoid confounding by stream channelization where wetted channel width can often be artificially controlled by characteristics of the channel work, age and maintenance conducted in the channel, and bed and bank characteristics. Channelization typically results in an increase in stream gradient leading to channel evolution that can result in channel narrowing due to down-cutting or channel widening either during construction or during the "rejuvenation" cycle of channel evolution (Simon 1989; Hupp 1992). Channelization in warmwater streams leads to declines in species richness of habitat specialist species (Gorman and Karr 1978) and declines in indices such as the IBI (Rankin 1995, Sullivan et al. 2003, Lau et al. 2006). This may cause a problem in the use of stream width as a primary measure of stream size because increased width, which may be a result of channelization, makes the stream appear "larger" in relation to an IBI calibration and will potentially over rate the level of degradation. Similarly, artificial downcutting or incision could result in a narrower channel, resulting in stream width potentially under-rating the level of disturbance. The River Continuum Concept predicts that increasing stream size results in the addition of greater habitat volume, i.e., more habitat types or niches (as shown above). In contrast, artificial widening associated with channelization typically results in shallow channels with poor habitat features.

Alternately, watershed alteration can also result in a narrowing of stream channels under certain conditions or during differing periods in response to channel evolution back to a more stable, natural state. Where stream bottoms are highly erosive, down-cutting may occur in response to increased flow from development, resulting in a narrower stream channel. Other studies have shown increasing stream width from urbanization (Galster et al. 2008). In New Zealand, Davies-Colley (1997) found that stream channels narrowed as land use changed from forest to pasture. In each of these cases stream width is varying due to anthropogenic changes in the watershed, potentially confounding stream width as a method to assign metric scores for size dependent IBI metrics which are usually species richness measures.

Trend Analyses

Another concern with the use of stream width is confounding of biological trends assessment at sites. Because stream drainage area is invariable over time periods, it does not contribute to variation in the IBI as can changes in stream width either because stream width varied due to changes from habitat modification or urbanization effects (e.g., bank erosion and channel widening from increased peak flows). As environmental flow data become more readily available (e.g., modeled daily flows) the influence of flow during a sampling event as well as a variety of other indicators (e.g., peak flows, flashiness, etc.,) can result in more refined classification of stressor measures.

Frequency of Direction in Species Richness Measure Related to Stream Width

If fish species respond strongly to the stream width measure taken during sampling then it could be important to account for these changes in stream width at a site from one sample to another if the pattern and direction of species richness is strong and consistent. At a specific site, stream width changes are primarily caused by changes in stream flow. The calibration of metrics along a stream width gradient assumes an increase of species richness with increasing stream width. As a “between-site” measure of habitat volume this hypothesis is consistent with the predictions associated with drainage area or flow yield. At a single site; however, should there be a strong presumption of species richness increase with stream width which is most often caused by flow increases? There is only a sparse literature relating sampling efficiency of electrofishing along a gradient of flow conditions, but the literature that does exists generally documents a lower success rate of electrofishing at high flow conditions. In the Neuse River of North Carolina minor increases in flow result in decreases in sampling effectiveness for the abundance of a target species, largemouth bass and CPUE was greatest when flow was lowest (Homan and Barwick 2010). Pierce et al. (1985) show similar results for multiple species on the Mississippi River.

Although very high flow may decrease sampling efficiency and reduce species richness in samples, increases in stream width at a site with more moderate flow changes could result in species richness increases if fish are sufficiently mobile. The data to support the mobility of

species in warmwater streams is somewhat variable, but tends to support a more sedentary nature than highly mobile one for most species (e.g., Gerking 1959; Hill and Grossman 1987, Gatz and Adams 1994, but see Funk 1957; Gowan et al. 1994). Environmental characteristics, such as length of riffles, can act to restrict movements of fishes (Lonzarich et al. 2000) during low flows, but storm events in spring may allow for greater movement or mixing of stream fish (Funk 1957).

One way to explore the strength of stream width as a predictor of species richness is to examine the direction of change of species richness at sites with repeat samples with individual width estimates for each example. We took Illinois data with at least two samples at a station and compared the species richness at the sample with the smallest stream width to replicate samples with the same or greater width (Figure 5). We examined this for sites with any increase in width (Figure 5, top) and also where the width increase had be to at least 10% greater (Figure 5, middle) or at least 25% greater (Figure 5, bottom) than the minimum width. The expectation would be that higher species richness counts should occur more frequently as stream width increases. As can be seen in Figure 5, there is only a minor increase in the frequency of sites with more native species where stream width was greater than the minimum width at a station. It would take more detailed data to examine whether factors such as channel modification confound these analyses.

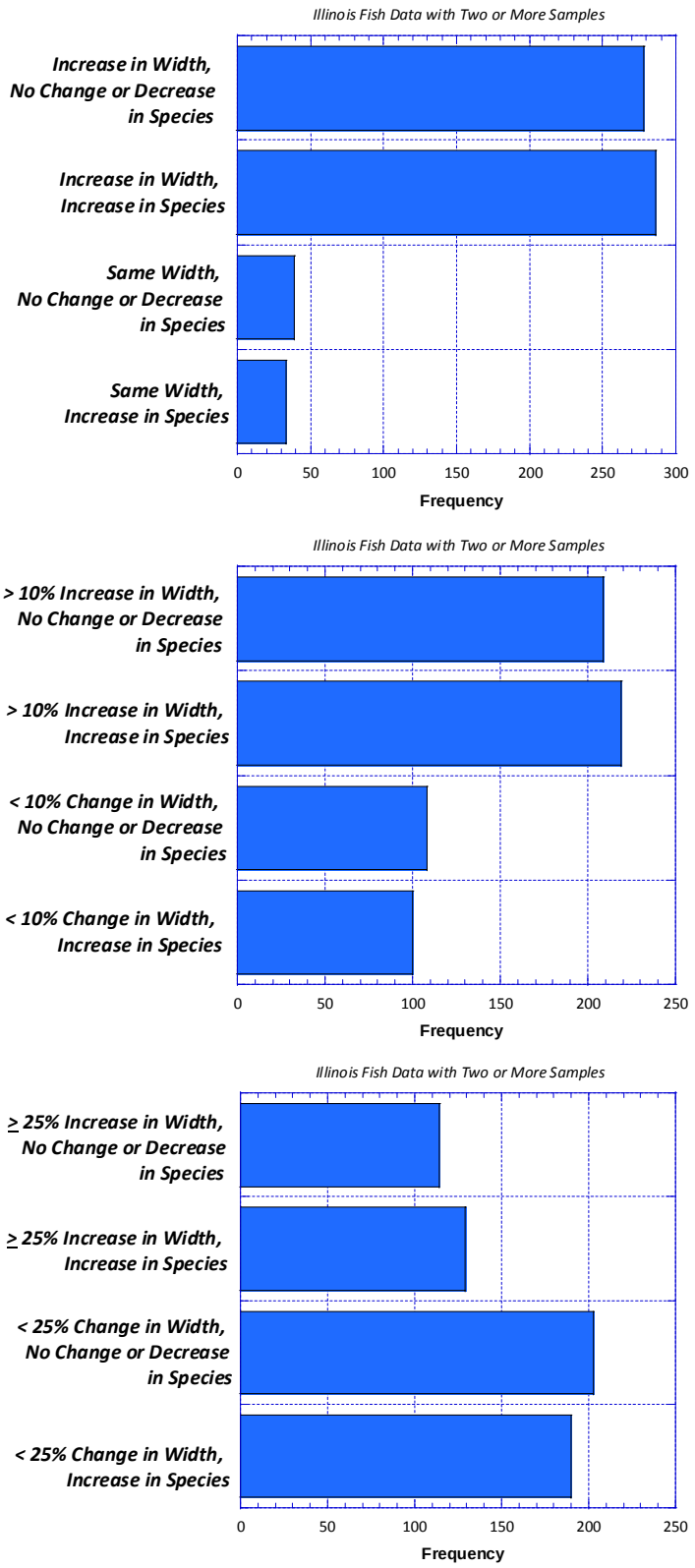


Figure 5. Column charts illustrating the frequency of native species richness increases vs. no change or decreases in species richness with samples showing increases in stream width compared to sample with narrowest width. Top chart represents any increase in width, middle chart at least 10% increase in width, and bottom chart at least 25% increase in width.

Error in Estimates of Mean Wetted Width

Although measurement of wetted stream width can be made with some precision, an estimate of stream width does have some error depending on methodology, number of measurements, and variability of channel width at a site. The IBI calculator on Illinois DNR web site (http://dnr.illinois.gov/IBICalculation/NewSampleForm.aspx) indicates that three bank-to-bank measurements of wetted width are needed when imputing data into the online calculator for the Fish IBI. One issue with the error that might occur during different samples is some of the variation in stream width may be measurement or user error, rather than true flow differences in the estimate of stream size. Several investigators have quantified precision in stream width measurements (Kauffman et al. 1999; Pert and Dauwalter 2001) and although the measure is considered precise, these analyses have been conducted with more than three width measures.

Best Professional Judgement and Alternate Stream Size Measures.

The establishment of standard method of stream size such as drainage area should be amenable to site-specific situations where a specific measure may fall short in a unique setting. The addition of flow from a gravel mining operation or the withdrawal of flow for a water supply or for cooling could result in conditions where drainage area may not reflect the best measure of habitat volume for a site. Thus a mechanism for selecting an alternate stream size measure may be appropriate as a substitute. The association between stream width and drainage area could be used to select an alternate stream size measure for IBI metric calibration where such arguments can be supported.

Conclusions

Although stream width is correlated with drainage area, there is enough variation in this relationship to have some concern with confounding metric scores due to width differences from channel alteration which is widespread in Illinois. There is also the possibility that width measurements themselves (e.g., variability along a channel) could influence metric calculation. Although drainage area is invariable as measured, it may not be a perfect measure compared to the conceptual “habitat volume” of the River Continuum Concept. If fish are responding to a wetter year at a site with increasing species richness and higher IBI, with the use of drainage area rather than stream width as the calibration variable, then stream flow differences can be used as an explanatory variable. Although not readily available at this time, the ability to derive daily flow estimates from modeling could allow an even better measure of environmental stream size that could incorporate a flow yield measure for calibration or to calculate a measure of flow that could act as a covariable to explain IBI variation.

References

Allan, J.D. and M.M. Castillo. 2007. Stream Ecology: The Structure and Function of Running Waters. Springer Press.

Angermeier, P.L. and I. J. Schlosser. 1989. Species-area relationship for stream fishes. Ecology 70:1450–1462.

Compton, M.C., G J. Pond, and J. F. Brumley. 2003. Development and application of the Kentucky Index of Biotic Integrity (KIBI). Kentucky Department for Environmental Protection, Division of Water, Water Quality Branch, 14 Reilly Rd., Frankfort, KY 40601

Davies-Colley, R.J. 1997. Stream channels are narrower in pasture than in forest, New Zealand Journal of Marine and Freshwater Research, 31:5, 599-608

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

Faustini , J.M. , P. R. Kaufmann and A. T. Herlihy. 2009. Downstream variation in bankfull width of wadeable streams across the conterminous United States. Geomorphology 108: 292–311.

Funk. J.L. 1957. Movement of Stream Fishes in Missouri. Transactions of the American Fisheries Society. 85(1): 39-57.

Galster, J.C., F. J. Pazzaglia, and D. Germanoski. 2008. Measuring the impact of urbanization on channel widths using historic aerial photographs and modern surveys. Journal of the American Water Resources Association 44(4): 948-960.

Gatz, A.J. Jr. and S. M. Adams. 1994. Patterns of movement of centrarchids in two warmwater streams in eastern Tennessee. Ecology of Freshwater Fish. 3(1): 35–48.

Gerking, S.D. 1959. The restricted movement of fish populations. Biol. Rev. Camb. Philos. Sot. 34: 221-242.

Gorman, O. T. & J. R. Karr. 1978. Habitat structure and stream fish communities. Ecology 59: 507–515.

Gowan, C., Young, M.K., Fausch, K.D., and Riley, S.C. 1994. Restricted movement in resident stream salmonids: a paradigm lost? Can. J. Fish. Aquat. Sci. 51: 2626-2637.

Hill, J.H. and Grossman, G.D. 1987. Home range estimates of North American stream fishes. Copeia, 1987: 19-25.

Homan, J.M. and R.D. Barwick. 2010. Relationship between Discharge and Electrofishing Catch-per-unit-effort of Largemouth Bass in the Neuse River, North Carolina. Proc. Annu. Conf. Southeast. Assoc. Fish and Wildl. Agencies 64:184–189.

Hughes, R. M. and J. M. Omernik. 1983. An alternative for characterizing stream size. Pages 87–102. in Fontaine, T. D. and S. M. Bartell. (editors). Dynamics of lotic ecosystems. Ann Arbor Press. Ann Arbor, Michigan.

Hupp, C.R. 1992. Riparian vegetation recovery patterns following stream channelization: A geomorphic perspective. Ecology 73(4): 1209-1226.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. Assessing biological integrity in running waters: A method and its rationale. Illinois Natural History Survey Special Publication 5, Urbana, IL.

Kaufmann, P.R., P. Levine, E.G. Robison, C. Seeliger, and D. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. Washington, DC: U.S. Environmental Protection Agency. http://nepis.epa.gov/Adobe/PDF/300042RU.pdf (PDF) (149 pp, 910K).

Lamouroux, N., N. Leroy Poff, and P. L. Angermeier. 2002. Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients. Ecology, 83(7): 1792–1807.

Lau, J.K., T. E. Lauer and M. L. Weinman, 2006. Impacts of channelization on stream habitats and associated fish assemblages in east central Indiana. The American Midland Naturalist 156(2): 319-330

Lonzarich, D.G., M.R. Lonzarich, and M.L. Warren, Jr. 2000. Effects of riffle length on the short-term movement of fishes among stream pools. Can. J. Fish. Aquat. Sci. 57: 1508-1514.

McCormick, F.H., R. M. Hughes , P. R. Kaufmann , D. V. Peck , J. L. Stoddard & A. T.Herlihy. 2001. Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands Region, Transactions of the American Fisheries Society 130(5): 857-877.

McGarvey, D.J. 2014. Moving beyond species—discharge relationships to a flow-mediated, macroecological theory of fish species richness. Freshwater Science 33(1):18-31.

MPCA. 2014. Development of a fish-based Index of Biological Integrity for assessment of Minnesota's rivers and streams. Document number wq-bsm2-03. Minnesota Pollution Control Agency, Environmental Analysis and Outcomes Division, St. Paul, MN

Ohio EPA (Environmental Protection Agency). 1989. Biological criteria for the protection of aquatic life, volume II. User's manual for biological field assessment of Ohio surface waters. Ohio EPA, Columbus. Ohio, USA.

Pert, E.J. and D. C. Dauwalter. 2001. Development of an Index of Biotic Integrity for Fish Assemblages for Wadeable Streams in Arkansas' Ozark Highlands Ecoregion. Prepared for Tetra Tech, Inc., or the subcontract for EPA Contract #68-C98-111 WA#1-10, University of Arkansas at Pine Bluff, Department of Aquaculture/Fisheries.

Pierce, R. B., D. W. Coble, and S. D. Corley. 1985. Influence of river stage on shoreline electrofishing catches in the upper Mississippi River. Transactions of the American Fisheries Society 114:857–860.

Poff, N. L. & J. D. Allan, 1995. Functional organization of stream fish assemblages in relation to hydrological variability. Ecology 76: 606–627.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The natural flow regime. BioScience 47:769–784.

Pont, D., R.M. Hughes, T.R. Whittier and S. Schmutz. 2009. Predictive index of biotic integrity model for aquatic-vertebrate assemblages of Western U.S. streams. Transactions of the American Fisheries Society 138: 292–305.

Rankin, E. T. 1995. The use of habitat assessments in water resource management programs, pages 181-208. in W. Davis and T. Simon (eds.). Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, FL.

Rankin, E.T. 2010. Calibration of the Ohio IBI and ICI using continuous scoring methods. Prepared for Ohio EPA by ILGARD, Voinovich School, Building 21, The Ridges, Ohio University, Athens, Ohio 45701

Schlosser, I.J. 1982. Fish community structure and function along two habitat gradients in a headwater stream. Ecological Monographs 52:395–414.

Sheldon, A.L. 1968. Species Diversity and Longitudinal Succession in Stream Fishes. Ecology 49(2): 193-198.

Shelford, V.E. 1911. Ecological Succession: Stream fishes and the method of physiographic analysis. Biological Bulletin 21(1): 9-35.

Simon, A., 1989a. A model of channel response in disturbed alluvial channels. Earth Surface Processes and Landforms 14, 11–26

Simon, A. and M. Rinaldi. 2006. Disturbance, stream incision, and channel evolution: The roles of excess transport capacity and boundary materials in controlling channel response. Geomorphology 79 (2006): 361–383.

Sullivan, B.E., Rigsby, L.S., Berndt, A., Jones-Wuellner, M., Simon, T.P., Lauer, T., and Pyron, M., 2003, Habitat influence on fish community assemblage in an agricultural landscape in four east central Indiana streams: Journal of Freshwater Ecology, 19(1): 141-148.

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

Walters, D.M. 2006. Development of a Fish Index of Biotic Integrity to Assess the Condition of West Virginia Streams: Technical Support Document. EPA/600/R-06/010. U.S. Environmental Protection Agency, Cincinnati, Ohio.

Whittier, T. R., R. M. Hughes, J. L. Stoddard, G. A. Lomnicky, D. V. Peck, And A. T. Herlihy. 2007. A structured approach for developing indices of biotic integrity: three examples from streams and rivers in the western USA. Transactions of the American Fisheries Society. American Fisheries Society, Bethesda, MD, 136: 718-735.