Interim Methodology for Bioassessment of the Delaware River for the DRBC 2010 Integrated Assessment

DRAFT report prepared and revised by

Erik L. Silldorff and Robert L. Limbeck

24-July-2009

Delaware River Basin Commission
DELAWARE · NEW JERSEY
PENNSYLVANIA · NEW YORK
UNITED STATES OF AMERICA

Executive Summary

The Delaware River Basin Commission (DRBC) has historically assessed the "aquatic life" designated use for the mainstem Delaware River through physical and chemical water quality standards alone. Annual macroinvertebrate monitoring of the Delaware River above the head-of-tide at Trenton since 2001 provides an opportunity to directly assess the "aquatic life" designated use through measurements of the living resources in the Delaware River. Although a complete analysis of this dataset is expected to require additional time, the DRBC seeks to use an interim methodology based on results obtained to date for the 2010 Integrated Assessment of the Delaware River. This interim methodology uses a multi-metric Index of Biotic Integrity (IBI) composed of 6 ecological metrics (5 of which are currently used by one or more basin states) to set biocriteria thresholds for evaluating whether the "biological integrity" of the Delaware River has been attained. Because additional analyses will be needed to more completely evaluate the assessment methodology and the patterns in the invertebrate dataset, the DRBC proposes to limit the classification of sites based on this methodology to only Categories 1, 2, and 3 of the 5-Category Integrated Assessment. Sites not attaining the biocriteria threshold will be assigned to Category 3A (Waters of Concern), noting that "aquatic life parameters indicate a high likelihood of impairment" but where further evaluation and confirmation are needed. This report provides a simple overview of the analyses conducted to develop this methodology and an evaluation of the 6-metric IBI performance for the Delaware River.

A. Invertebrate Sampling & Processing

DRBC staff collected benthic macroinvertebrate samples annually from 2001 to 2008 at 25 fixed sites on the Delaware River from Hancock, NY to just above the head-of-tide at Trenton, NJ (see Figure 1; 2 site locations moved a short distance in 2006; no samples collected in 2004 because of high water). All samples were collected from targeted riffle habitats using a stratified sampling design to sample areas within the following limits: current velocity between 1 ft/sec and 3 ft/sec; depth between 1 ft and 2 ft; and substrate dominated by large gravel and small cobble (median between 40 and 70 mm). Sampling occurred from the last week in July to the first week in October, with the central sampling window being August and September. Samples were collected via kick samples using a 595 µm rectangular net (3 ft wide x 2 ft tall) and a 2 ft x 2 ft guide in front of the net as the sampling area. Three (3) separate 2 ft x 2 ft kicks were collected within each riffle, and the three samples were composited into a single sample for each site (note: in 2002, some samples were kept separate to evaluate within-riffle variability). Samples were preserved in the field using 95% ethanol. Additional measurements at each site included depths, current velocity, substrate particle size, and physical and chemical measurements of water quality. The full sampling methodology is provided in the DRBC Quality Assurance Project Plan (see www.state.nj.us/drbc/BioQAPP06-07.pdf)

In the lab, samples were processed by subsampling random fractions of each sample and sorting the invertebrates under 10x or greater magnification with a dissecting microscope. A target count of 500 or more organisms guided the subsampling in all years except 2001, where a 200 or more invertebrate count was used (see Metric Selection & Computation below for the methods used to compensate for this change in target counts). Identifications were made to genus for most organisms, with coarser taxonomic identifications for damaged or immature specimens and select taxonomic groups where genus-level identifications could not consistently be made (see Appendix A). The genus-level identifications included the Chironomidae midges and the aquatic mites (Hydracarina), while family-level or coarser identifications were primarily made for flatworms and oligocheate worms. Sample identifications were made by taxonomists at the USGS (Michael Bilger), EcoAnalysts, and the DRBC (Geoffrey Smith). The total number of samples collected and processed for this program to date is 163 samples.

B. Designating a “Least Disturbed” Reference Reach

To facilitate both the creation and evaluation of a multi-metric tool to assess biological integrity of the Delaware River, DRBC staff identified a zone of the river as a reference reach based on best professional judgment. This reach was selected to begin at River Mile 305 (upstream from the Callicoon Creek confluence on the Upper Delaware; see Figure 1) and extend downstream to River Mile 184 (immediately upstream from the Lehigh River confluence).

Human influence on the river exists in all reaches and includes hydrologic alteration, hypolimnetic tailwater influences, channel alteration, recreation, point source pollution, non-point source pollution, and acid mine drainage. However, two zones of the river had human influences strong enough to disqualify them from inclusion in any level of “reference” classification. First, above River Mile 305 the cold-water releases from New York City reservoirs (primarily Cannonsville Reservoir on the West Branch Delaware River) create temperature regimes that are significantly cooler than areas at and below Callicoon, resulting in part in a recreational trout fishery. Second, the relatively poor water quality of the Lehigh River and the approximately 1-to-3 ratio of discharge in the Lehigh River relative to the Delaware River at their confluence significantly increases total phosphorus and other water quality parameters in the Delaware River relative to their values above the Lehigh River confluence (see Tables 2C to 2Z in the DRBC Water Quality Regulations; www.state.nj.us/drbc/regs/WQRegs_071608.pdf). Although the cumulative effect of these water quality changes remains uncertain, the changes are strong enough to preclude this section of river in a reference designation.

The remaining contiguous reach of river from River Mile 184 upstream to River Mile 305 was therefore identified as the reference reach for the current study. In terms of the quality of this reference reach, the known human influences suggest that it is best to consider this reference reach a “least disturbed” reach of river rather than some other designation that would imply minimal human influence.

C. Data Preparation & Structure

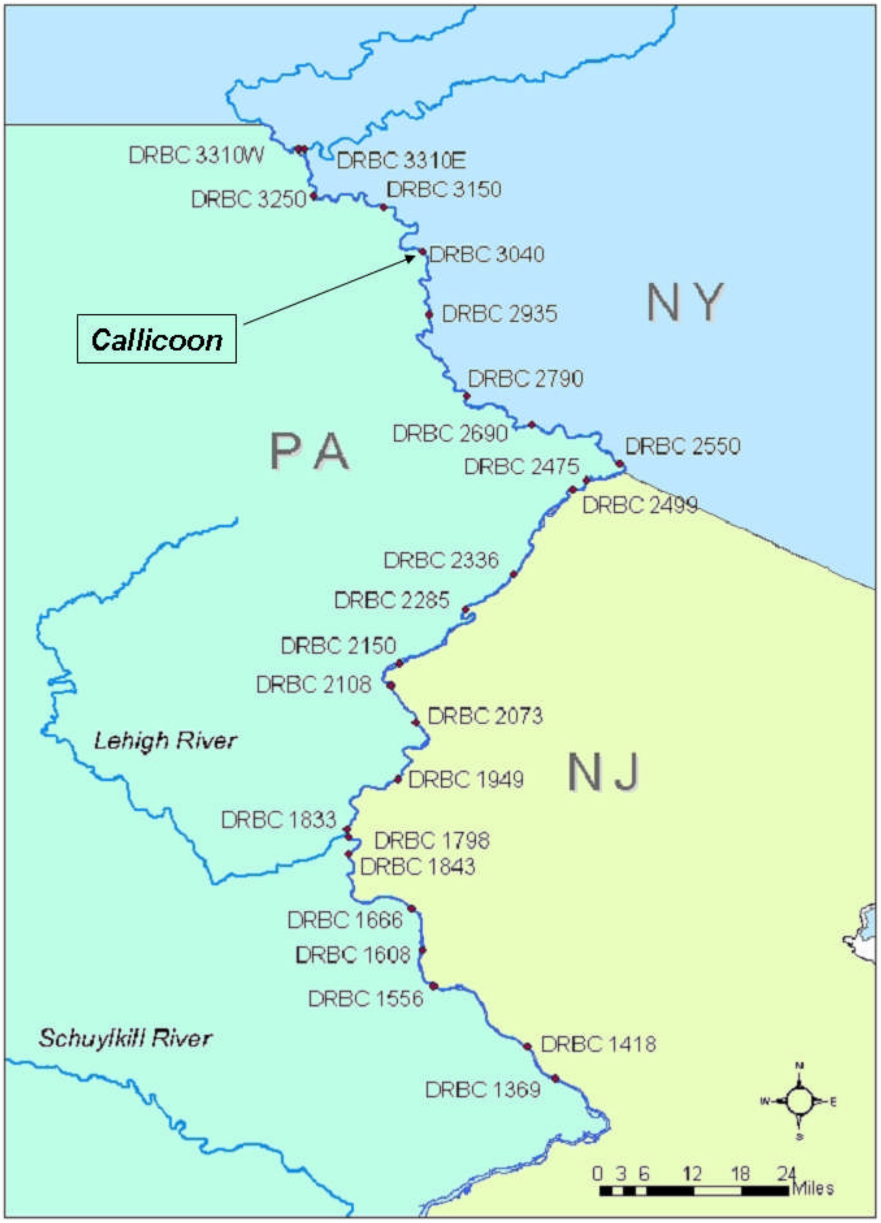
Data were entered and compiled into a Microsoft Access database. These compiled data were then exported in a standardized format to the R statistical platform for all analyses (see www.r-project.org; this software is functionally equivalent to the commercially available statistical software S-Plus). The primary effort in data preparation before any data analysis was to standardize the taxonomy across laboratories and across years to eliminate differences in what an individual invertebrate was labeled by different taxonomists. Although most identifications were consistently made using standard keys, differences did occur for select taxa. Where differences among taxonomist could not be converted in a one-to-one manner, the taxonomy for a group was backed off to a coarser taxonomic level (typically family).

Each of the 163 unique samples (distinct Station and/or Year and/or Replicate) is considered as an independent estimate of the aquatic community for the analyses in this report. Although both spatial and temporal dependencies in random errors could introduce bias in estimating variability, examination of the spatial and temporal autocorrelation validated the assumption of independence for each of these 163 samples.

D. Metric Selection & Computation

The Delaware River presents a different challenge in metric selection than is faced by most state or regional biomonitoring programs. Typically, a set of sites known to have substantial human impact can be compared to reference-quality sites in order to identify which metrics are sensitive to the human disturbances and thus show an ability to discriminate between reference and human-impacted sites. For the Delaware River, however, the human stress gradient is poorly established. There are human influences at all sites on the Delaware River, with some of the strongest human influence in the lower Delaware River below the Lehigh River confluence. Yet the entire river above the head-of-tide has been designated as anti-degradation waters by the DRBC (termed “special protection waters”) because water quality was determined to be high enough to warrant elevated protection. Thus, the Delaware River does not have any sites with unequivocally “impaired” environmental conditions that could be used to assess whether individual metrics, or composite multi-metric indices, provide the discriminatory efficiency desired in these summary statistics.

Figure 1. Original Station Locations for DRBC Bioassessment Studies of the Delaware River



Because of this inability to assess discriminatory efficiency directly, the selection of metrics for this interim DRBC methodology focused on those metrics that have proven to be sensitive to human disturbance by the individual states bordering the Delaware Basin above Trenton (i.e., Pennsylvania, New Jersey, and New York). The use of metrics already demonstrated to have strong discriminatory efficiency in streams of the basin thus provides a well-informed surrogate for directly evaluating the discrimination efficiency in the absence of strong human disturbance gradient for the mainstem Delaware River above Trenton.

Initial metric screening was completed by Dr. Joe Floetemersch (USEPA-ORD) through a Regionally Applied Research Effort grant. Metrics used by DRBC basin states and other metrics used in large, wadeable stream assessments were compiled and compared among groups. This initial list was augmented with additional metrics used by other agencies as well as specific metrics for the characteristics of the Delaware River dataset (e.g., rarefaction of richness metrics). The final set of metrics computed for the DRBC dataset are listed in Table 1, including which metrics are used by the three adjoining states in one or more of their bioassessment protocols.

As noted above, the change in target counts after the 2001 samples (from 200+ to 500+) resulted in substantial differences in the distributions of total organisms sorted and identified among samples. Because taxa richness measures typically increase with increasing effort (both in sampling area and numbers identified), ecologists have long recommended that comparisons among taxa richness be standardized for effort. One statistical approach to standardize richness measures is known as "rarefaction". In this procedure, the expected number of taxa in a sample at a lower sampling intensity (e.g., 200 individuals instead of the original 500 individuals) is calculated based on the probability theory for multinomial populations (Hurlbert 1971). Hurlbert's formula was re-coded in R and all simple richness calculations were standardized to an "expected richness" at a subsampling level of 200 individuals (Table 1). Although this method potentially weakens the data and removes information contained in the higher taxa counts, the need to include the greatest number of sampling years (i.e., including 2001) out-weighed the value obtained from the higher taxa counts.

E. Performance Evaluation of Metrics

Three primary characteristics are typically used to evaluate the performance of individual metrics: (i) the variability or noise in repeat measurements; (ii) the discrimination efficiency or ability to separate high-quality from low-quality sites; and (iii) the response along specific stressor gradients. As mentioned above, although sites with human influence exist on the Delaware River, the lack of a strong stressor gradient and the absence of known low-quality sites for the Delaware River prevented the evaluation of the latter two performance measures. As a result, this analysis focused on the variability both within-site and among-sites within defined regions of the Delaware River to determine which metrics could be considered further for inclusion in a multi-metric index of biological condition.

The primary measure of variability used to assess metrics was the among-year variability within the "least disturbed" reference reach. We used the coefficient of variation (CV = standard deviation / mean) within each site as the standardized estimate of among-year variability. The year-to-year variability in metrics scores was seen as the most comprehensive measure of variability since it would incorporate small-scale spatial variability, small-scale variability within the seasonal sampling window, and among-year variation in river populations. In addition, because the human influence on the Delaware River is relatively consistent year-to-year, metric scores with less variability among years would provide a stronger signal-to-noise ratio than similar metrics with high among-year variability.

Table 1. Metrics evaluated for DRBC assessment of the Delaware River mainstem. (Rarefaction indicates statistical computation of the expected richness at a standardized subsampling level across all samples; see Hurlbert 1971)

Metric Category	Metric	Description	Used in one or more state IBI (possibly with an alternative calculation method)			Rarefaction Used to Standardize Subsample Size
			PA	NJ	NY	
Structure	**Richness**	Number of invertebrate taxa	Y	Y	Y	x
	EPT Richness	Number of Ephemeroptera, Plecoptera, & Trichoptera taxa	Y	Y	Y	x
	Ephemeroptera Richness	Number of Ephemeroptera taxa	Y			x
	Plecoptera Richness	Number of Plecoptera taxa				x
	Trichoptera Richness	Number of Trichoptera taxa	Y			x
	Invertebrate Richness	Number of non-insect invertebrate taxa				x
Composition	**EPT Percent Abundance**	Percent of individuals belonging to Ephemeroptera, Plecoptera, or Trichoptera		Y		
	Shannon-Wiener Diversity	Diversity measure of both the richness and evenness of the invertebrate taxa	Y		Y	
	Dominance-3	Cumulative percent abundance of the three most common taxa in a sample		Y	Y	
Tolerance	**Biotic Index**	Modified Hilsenhoff Biotic Index as weigthed average of tolerance values	Y	Y	Y	
	Beck's Index	Modified Beck's Index as weighted richness for taxa with tolerance values of 0, 1, or 2	Y	Y		
	Intolerant Richness	Number of taxa with tolerance values of 0, 1, or 2				
	Intolerant Percent Richness	Percent of taxa richness from taxa with tolerane values of 0, 1, or 2				
	Intolerant Percent Abundance	Percent of individuals with tolerance values of 0, 1, or 2	Y			
	Tolerant Richness	Number of taxa with tolerance values of 8, 9, or 10				
	Tolerant Percent Richness	Percent of taxa richness from taxa with tolerance values of 8, 9, or 10				
	Tolerant Percent Abundance	Percent of individuals with tolerance values of 8, 9, or 10	Y			
Functional	**Scraper Richness**	Number of taxa belonging to the "scraper" functional feeding group		Y		
	Scraper Percent Richness	Percent of taxa richness from taxa belonging to the "scraper" functional feeding group				

Two additional measures of variability were considered along with the among-year variation. First, DRBC collected within-year spatial replicates at 9 sampling stations between 2006 and 2008 (samples collected on the same date as original). The coefficient of variation (CV) in metrics scores between the original and replicate metric scores were evaluated for these 9 samples. Lower weighting was given to this variability estimate, however, because of both the small sample size and the limited scope of this spatial variability. The final measure of variability we evaluated was the among-station variability both within and across years. Because the human influences to the Delaware River broadly affect long reaches of the river, metrics that respond more strongly to human disturbance than to natural variability were expected to show relatively low among-station variation within given river reaches. This final measure was qualitatively assessed via graphical plots of metrics scores within the "least disturbed" reference reach of the Delaware River.

A summary of these performance evaluations is presented in Table 2 for the metrics under consideration. Considerable differences exist among these metrics in terms of their performance. Four (4) of the 19 metrics performed particularly well, with consistently low among-year and among-replicate variability (within-site) as well as among-site variability: **Richness**, **EPT Richness**, **Shannon-Wiener Diversity**, and the **Biotic Index**. All four of these metrics are utilized by two or more state bioassessment programs in the areas adjoining the non-tidal Delaware River (see Table 1). Three additional metrics showed moderately strong performance in both variability categories, with two of these three metrics used by one of the three state bioassessment programs: **Dominance-3**, **Intolerant Percent Richness**, and **Scraper Richness**. Based on their individual performance for the Delaware River data (all 7) and their use in state bioassessment programs (6 of 7), these seven (7) "core metrics" were selected for possible inclusion in a multi-metric Index of Biotic Integrity (IBI).

In additional to providing strong numerical performance, these 7 core metrics also describe important aspects of the Delaware River as an ecological system. Simple taxa Richness has repeatedly been shown to decline with increasing human stress throughout the world and thus provides an overall assessment of the number of unique taxa that are found in a standardized sample. EPT Richness gives a more narrow indication of conditions by focusing on three insect orders (Ephemeroptera, Plecoptera, Trichoptera) that consistently show some of the greatest sensitivity to various forms of ecological stress to rivers and streams. Shannon-Wiener Diversity combines information on the richness and relative abundance of taxa to provide a synthetic measure of "diversity", an ecological concept that reflects not only the number of species or taxa present but also whether they exist is relatively similar abundances (higher diversity) or whether they are highly dominated by only a few taxa (lower diversity). The Biotic Index combines sensitivities to environmental stress for each taxon (termed "tolerance values"; see Appendix A) into a weighted average of the whole sample's sensitivity to stress, thus revealing (for example) whether only a few sensitive taxa exist at low abundance or whether the sample is largely composed of individuals coming from more sensitive taxa. The metric Dominance-3 indicates the degree to which the sample is dominated by the most abundant 3 taxa, with a general pattern that fewer taxa begin to dominate a system as the system undergoes greater stress. Intolerant Percent Richness provides a measure of the number of sensitive taxa but standardizes this measure among samples by presenting the intolerant richness as a percent of overall richness;

this focuses the measure on the relative composition of sensitive taxa and not just their overall number. Finally, Scraper Richness is a less-common metric but reflects the degree of overlap and richness among a particularly important group of invertebrates for a system like the Delaware River that is hard-bottomed with high water clarity, leading to benthic primary production in nearly all habitats within the river (see Appendix A for scraper designations). Across these 7 core metrics, then, many aspects of the ecological condition and health of the Delaware River are captured and quantified.

F. Re-Scaling Metric Scores

In order to combine metrics that are measured on different scales into a multi-metric IBI, the individual metrics scores for each site need to be converted to a standardized scale. Typically, this is done by re-scaling the metrics to a 0-to-10 scale by specifying metric values of the highest quality (given 10 scores) and metric values exhibiting no "biological integrity" (given 0 scores). Values between these 10-scores and 0-scores are then scaled through linear interpolation, while scores of higher quality than the 10-score benchmark are all given values of 10 and scores of poorer quality than the 0-score benchmark are all given values of 0. The manner that 10-score and 0-score benchmarks are selected is variable both among states and even within states. At times, the lowest theoretical value a metric could score is selected as the 0-score, while other programs or other metrics give 0-scores for values in a range near the lower end of the spectrum.

DRBC staff assert that "biological integrity" has typically been lost before a minimum metric score is reached (e.g., before taxa richness goes to zero). In addition, the probability that each metric achieves its theoretical minimum value under both natural and human-induced variation will vary considerably among metrics. Because of this latter inconsistency, the use of the theoretical minimum for a 0-score benchmark translates into different standardization scales among metrics. As a result of these considerations, the DRBC has chosen to identify 0-score and 10-score benchmarks based on empirical ranges that correspond to poor ecological condition and healthy ecological condition, respectively. Even within this approach, however, there are judgments that need to be made on what constitutes "poor ecological condition" and "healthy ecological condition", and how these benchmarks can be assigned consistently among the metrics considered. Frequently, "poor conditions" can be defined as a percentile in the distribution of metrics scores at ecologically stressed sites while "healthy conditions" can likewise be defined as a percentile in the distribution of metric scores for the highest quality reference sites. As noted earlier, no sites on the Delaware River are clearly in a poor condition, precluding the use of a simple distribution percentile for the 0-score benchmark. However, the DRBC has identified a reach of the Delaware River that it considers to be a "least disturbed" reference condition (see earlier discussion). Because human disturbance is acknowledged throughout this reach, higher metric scores may indeed signal healthier ecological condition. The DRBC has therefore selected the 75th percentile of scores in this reference reach as the 10-score benchmark (or the 25th percentile for reverse-scale metrics such as Biotic Index and Dominance-3 that are expected in increase with human disturbance).

Table 2. Metric Performance using Within-Site Coefficients of Variation (CV) and Qualitative Evaluation of Among-Site Variation (Tolerance Values and FFG designations given in Appendix A).

		Among-Year CV		Among-Rep CV	Among-Site Spatial Variation
Metric Category	Metric	median	75th percentile	median	(stable vs moderate vs variable)
Structure	Richness**	0.13	0.15	0.02	moderate
	EPT Richness**	0.10	0.14	0.05	stable
	Ephemeroptera Richness**	0.16	0.19	0.07	moderate
	Plecoptera Richness**	0.46	0.70	0.20	variable
	Trichoptera Richness**	0.18	0.26	0.03	variable
	Invertebrate Richness**	0.35	0.41	0.12	variable
Composition	EPT Percent Abundance	0.41	0.46	0.03	variable
	Shannon-Wiener Diversity	0.08	0.09	0.05	stable
	Dominance-3	0.19	0.22	0.10	moderate
Tolerance	Biotic Index	0.09	0.16	0.03	moderate
	Beck's Index	0.15	0.29	0.17	variable
	Intolerant Richness	0.14	0.24	0.11	variable
	Intolerant Percent Richness	0.17	0.21	0.08	moderate
	Intolerant Percent Abundance	0.32	0.42	0.16	variable
	Tolerant Richness	0.35	0.51	0.28	variable
	Tolerant Percent Richness	0.32	0.43	0.19	variable
	Tolerant Percent Abundance	0.75	0.94	0.42	variable
Functional	Scraper Richness	0.17	0.22	0.07	moderate
	Scraper Percent Richness	0.13	0.16	0.08	moderate

**** - these metrics were standardized to a 200-individual subsample via rarefaction**

For the 0-score benchmark, DRBC has selected a distributional approach based on the variation in the reference-reach scores. Specifically, the 0-score benchmark has been set as a score 5 standard deviations below the reference-reach mean. Only for the Scraper Richness metric did 5 standard deviations below the reference reach mean result in a benchmark below the theoretical minimum value of zero; for this one metric, the benchmark was therefore set at 0. A number of standard deviation factors were considered prior to selecting 5 standard deviations (ranging from 3 to 5, including fractional factors). The decision to use 5 standard deviations was based both on the theoretical probability of a reference-quality site scoring in the 0 range based on natural variation alone (roughly 1 sample in 3,500,000 at 5 standard deviations under a normal distribution) and the empirical value of these thresholds relative to extreme percentiles of both the reference and non-reference distributions (see Table 3). As seen in Table 3, these 0-score benchmarks represent extreme cases that are highly unlikely under normal conditions at the reference sites. Although this distributional approach does not establish that this 0-score benchmark is equivalent to a lack of "biological integrity", it does establish that such extreme values represent a significant departure from the healthy conditions seen at the identified reference sites. Moreover, the approach provides a consistent standardization across metrics based on the natural variation among reference sites.

Table 3. Benchmarks for Standardizing Metrics to 0-to-10 Scale. The 0-score benchmark is set at 5 standard deviations below the reference reach mean; the 10-score benchmark is set at the 75th percentile of the reference reach distribution.

		Re-Scaling Benchmarks		Reference Sites		Non-Reference Sites	
Metric Category	Metric	0-score	10-score	10th percentile	90th percentile	10th percentile	90th percentile
Structure	Richness**	7.4	32.7	24.6	35.9	20.4	38.1
	EPT Richness**	4.2	17.2	13.6	20.1	12.0	19.5
Compositior	Shannon-Wiener Diversity	1.46	2.98	2.50	3.13	2.30	3.15
	Dominance-3	89.5%	36.9%	32.8%	55.7%	33.3%	61.3%
Tolerance	Biotic Index	6.26	3.72	3.12	4.49	3.49	4.69
	Intolerant Percent Richness	0.5%	29.0%	18.9%	32.4%	16.2%	30.9%
Functional	Scraper Richness	0.0	13.5	9.0	16.0	8.0	12.8

** - these metrics were standardized to a 200-individual subsample via rarefaction

It is important to note that the benchmarks identified in Table 3 are based only on samples collected between River Miles 184 and 305 and only for samples collected in the years 2001 to 2006 (n=59). Samples from 2007 and 2008 were removed from the benchmark analysis since these data are expected to be used in the 2010 Integrated Assessment and inclusion in the benchmark calculations would create a circularity in their assessment.

G. Multi-metric Construction & Evaluation

The 7 core metrics selected for possible inclusion into a multi-metric IBI were first evaluated for the level of numerical redundancy via simple Pearson correlation coefficients among the standardized metric scores (Table 4). Moderate to strong correlations were seen among a number of the metrics, with one correlation (Dominance-3 vs Shannon-Wiener Diversity) exceeding 0.9. Because of this extremely high correlation, and because the Shannon-Wiener Diversity had among the strongest performance for all metrics, the Dominance-3 metric was dropped from further consideration.

The remaining 6 core metrics were then combined into four candidate IBIs. In all candidate IBIs, the 4 metrics with the strongest performance (Richness, EPT Richness, Biotic Index, and Shannon-Wiener Diversity) were included because of both their strong performance and because two or more of the adjacent states have incorporated these 4 metrics into their bioassessment programs, thus providing some consistency among programs. The two remaining core metrics (Intolerant Percent Richness and Scraper Richness) were then included either alone or in combination with the other 4 core metrics to create the remaining candidate IBIs (see Table 5). Construction of each IBI consisted of averaging the 0-to-10 standardized scores among component metrics and multiplying the average score by 10 to obtain an IBI ranging from 0-to-100.

Table 4. Correlation Among Standardized Scores for Seven Core Metrics (correlations corresponding to R^2 > 0.5 highlighted in bold; single extremely high correlation higlighted in bold red)

#	Metric	1	2	3	4	5	6
1	Richness **						
2	EPT Richness **	**0.72**					
3	Shannon-Wiener Diversity	**0.87**	**0.72**				
4	Biotic Index	-0.08	0.25	0.03			
5	Intolerant Percent Richness	0.08	0.50	0.10	0.41		
6	Scraper Richness	0.62	0.55	0.53	-0.03	0.20	
7	Dominance-3	0.69	0.63	**0.93**	0.10	0.12	0.41

** - these metrics were standardized to a 200-individual subsample via rarefaction

Like with the individual metrics, the evaluation of IBI performance could not include comparisons of discrimination efficiency since a strong human-stressor gradient with clearly identified impaired sites does not exist on the mainstem Delaware River. Instead, the comparison among the four candidate IBIs centered on the variability in repeat measurements within and among sites. Table 5 provides the summary measures of variability for the reference reach sites: (i) within sites and among years (coefficient of variation [CV] between years); (ii) within sites and within year (CV between replicate sample); (iii) and among sites and among years (qualitative evaluation of patterns).

Comparison among the 4 candidate IBIs reveals little difference in the variability in IBI scores within and among sites. As seen in Table 5, all 4 candidate IBIs have standard deviations that are typically 10% or less of the respective mean value, suggesting high levels of repeatability. Box-and-whisker plots of the IBI scores for each site within the reference reach are presented in Figures 2a to 2d. These figures reveal relatively stable IBI scores across the reference reach, with some increased but moderate variability in the area around Port Jervis (RM 255) as well as the in the Upper Delaware River.

Table 5. Summary of the metrics included in the candidate IBIs and their performance in three areas of among-sample variability.

							Among-Year CV			Among-Rep CV	Among-Site Spatial Variation
Multi-Metric Index	**Richness ****	**EPT Richness ****	**Shannon-Wiener Diversity**	**Biotic Index**	**Intolerant Percent Richness**	**Scraper Richness**	**25th percentile**	**median**	**75th percentile**	**median**	(stable vs moderate vs variable)
4-Metric IBI	✓	✓	✓	✓			0.08	0.10	0.13	0.04	stable
5-Metric IBI v1	✓	✓	✓	✓	✓		0.08	0.08	0.11	0.02	stable
5-Metric IBI v2	✓	✓	✓	✓		✓	0.07	0.10	0.13	0.04	stable
6-Metric IBI	✓	✓	✓	✓	✓	✓	0.07	0.08	0.09	0.03	stable

**** - these metrics were standardized to a 200-individual subsample via rarefaction**

The consistency in performance among the 4 candidate IBIs provides little basis for selecting among these IBIs for the interim multi-metric index to be used for bioassessment of the Delaware River. All 4 candidate IBIs appear suitable for such an interim index. Yet there are both statistical and conceptual advantages with using a multi-metric index based on a larger number of metrics, particularly when such an index includes metrics in additional classes or categories. The 6-metric IBI, with 2 richness measures, 1 compositional measure, 2 tolerance measures, and 1 functional measure, provides the greatest breadth and evenness across these metric classes. In addition, Table 5 suggests that the 6-metric has as low or lower variability when compared to the other 3 candidate IBIs, although the differences are minor. Finally, the use of 6 metrics rather than 4 or 5 provides some additional buffer against a single unusual metric score at any site shifting the overall IBI disproportionally and perhaps causing a change in its assessment category. ***The DRBC therefore recommends that the 2010 interim methodology for bioassessment be based on this 6-metric IBI.***

As a final evaluation of the 6-metric performance, multi-metric indices for New York DEC and Penssylvania DEP were calculated using the DRBC sample data and compared to the 6-metric DRBC index. Specifically, DRBC data at the current level of taxonomic resolution (see Appendix A) were used to calculate the Biological Assessment Profile for riffle samples (NYDEC 2002) as well as the Freestone IBI (PADEP 2007). Although the sampling methods, taxonomic resolutions, and applicable stream sizes used for the NYDEC and PADEP multi-metric indices differ from DRBC sampling on the Delaware River, the expectation would nevertheless be similar rankings and relative positions for the same samples using the different multi-metric indices. Figure 4 and Figure 5 present both the pairwise comparisons between the two state multi-metrics to the DRBC 6-metric IBI as well as the box-and-whisker plots for each state multi-metric among sites. The pairwise comparisons reveal high correspondence between the state methods and the DRBC IBI (R^2=0.69 for NYDEC vs DRBC; R^2=0.84 for PADEP vs DRBC). Likewise, the box-and-whisker plots reveal a pattern consistent with Figure 3 for the 6-metric DRBC IBI. These results reinforce the conclusion that the 6-metric IBI accurately represents the ecological conditions in the Delaware River, and that the 6-metric IBI appears to be responsive to ecological change in a manner similar to corresponding state tools.

H. Biocriteria Tresholds and Methodology for Integrated Assessment

DRBC staff are recommending the 6-metric IBI (Table 5) for bioassessment of the Delaware River in the 2010 Integrated Assessment. In order to use this multi-metric index for assessment purposes, a threshold needs to be identified for determining attainment of the aquatic life use and classification of sites needs to be mapped to the different Categories of use-attainment in the Integrated Assessment. When considering alternative thresholds, the designation of the entire Delaware River above the head-of-tide as anti-degradation waters as well as inclusion of nearly this entire section in the National Wild & Scenic system (one small segment excluded) suggests that thresholds should be chosen that are more protective of the aquatic life use than might be used on waters with less protective and less significant classifications and uses.

The DRBC therefore recommends that the 10th percentile of reference reach data for the period 2001 to 2006 be used as the threshold between “supporting” (Categories 1 and 2) and “probably not supporting” (Category 3A) for the aquatic life designated use. This 10th percentile of the 6-metric IBI equals a score of 75.6 units (see Figure 3 for all-sites, all-years plot relative to this threshold). Based on recommendations from the DRBC Biological Advisory Subcommittee, the 2010 assessment would be limited to these 3 Categories (rather than all 5) in order to provide an initial assessment of aquatic life use based on biological data but with the need and opportunity to more fully evaluate the bioassessment data and methodology as well as to gather additional data to confirm or reverse these interim assessments. Two examples are shown in Appendix B to illustrate the composition of both a typical low-scoring and a typical high-scoring sample relative to this recommended threshold (although samples vary substantially in all ranges).

The DRBC recommended methodology for aquatic life use assessment involving biological data in the 2010 Integrated Assessment consists of the following components:

- computation of the 6-metric IBI based on the taxonomic resolution in Appendix A, the metrics identified in Table 5, and standardization of metric scores with the benchmarks provided in Table 3;

- evaluation of data sufficiency with the requirement that at least 2 years of data with multiple sites per Assessment Unit be available for determining the biological condition of the Delaware River during the Assessment Period, and that methods and invertebrate identifications be compatible with those identified in the DRBC QAPP and this report;
- comparison of the 6-metric IBI scores to the assessment threshold of 75.6, with the following mapping of assessment decisions:

    i. If all 6-metric IBI scores are greater than the 75.6 threshold, the site will be classified as either Category 1 or Category 2 (“supporting”);
    ii. If all 6-metric IBI scores are less than the 75.6 threshold, the site will be classified as Category 3A (“Waters of Concern”);
    iii. If a mixture of 6-metric IBI scores are both above and below the 75.6 threshold, the classification will depend on the proportion above and below the threshold. If more than 30% of samples are below the 75.6 threshold, the sites will be classified as Category 3A (“Waters of Concern”). The 30% cut-off was selected to represent a 300% increase in the rate of samples falling below the 75.6 threshold compared to the rate reference site samples fell below the threshold in the 2001 to 2006 background period.
    iv. For all other scenarios, the site will be classified as Category 3 (uncertain status with insufficient data).

References Cited:

`

Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577-586

NYDEC 2002: Bode, R.W., M.A. Novak, L.E. Abele, D.L. Heitzman, and A.J. Smith. 2002. Quality Assurance Work Plan for Biological Stream Monitoring in New York State. New York State Department of Environmental Conservation; Albany, NY. 122 pp. (see www.dec.ny.gov/docs/water_pdf/sbuqa02.pdf)

PADEP 2007: Chalfont, B.. 2007. A Benthic Index of Biotic Integrity for Wadeable Freestone Streams in Pennsylvania. Pennsylvania Department of Environmental Protection; Harrisburg, PA. 157 pp.

Figures 2a, 2b, 2c, and 2d. Box-and-Whisker Plots of Candidate IBIs for Reference Reach.
(see following 2 pages)

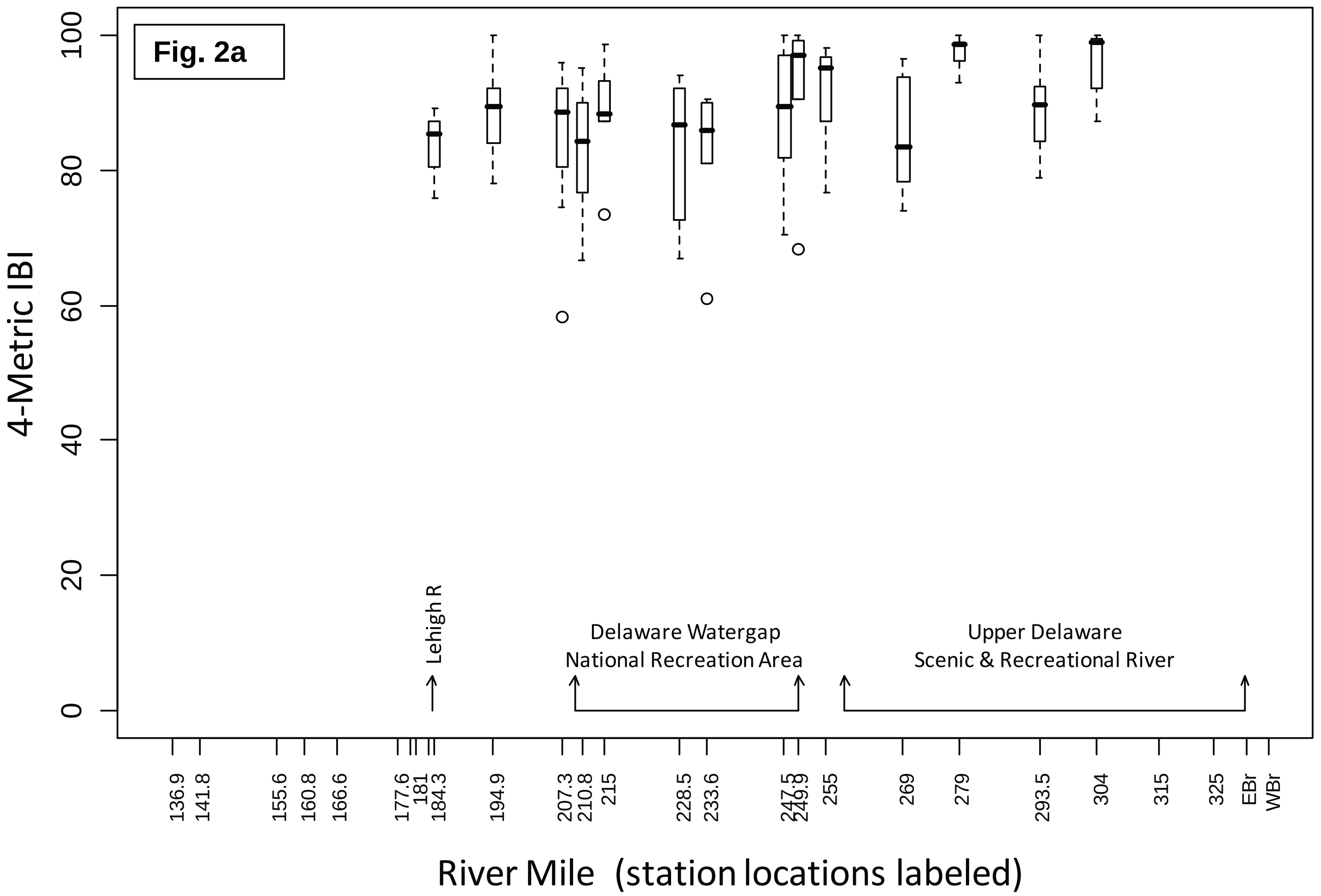


Fig. 2a
4-Metric IBI
0
20
40
60
80
100
Lehigh R
Delaware Watergap
National Recreation Area
Upper Delaware
Scenic & Recreational River
136.9
141.8
155.6
160.8
166.6
177.6
181
184.3
194.9
207.3
210.8
215
228.5
233.6
247.5
249.9
255
269
279
293.5
304
315
325
EBr
WBr
River Mile (station locations labeled)

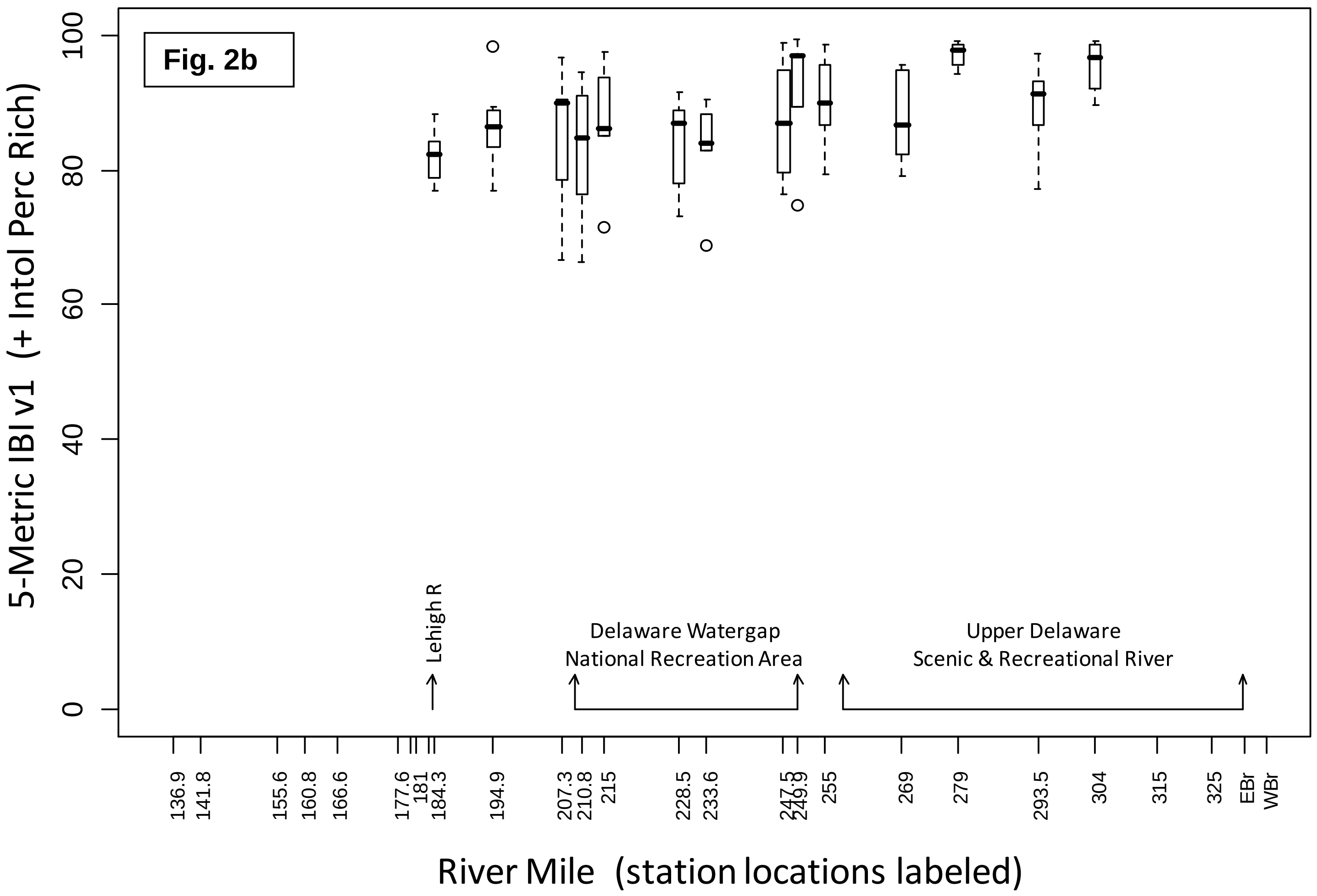


Fig. 2b
5-Metric IBI v1 (+ Intol Perc Rich)
0
20
40
60
80
100
Lehigh R
Delaware Watergap
National Recreation Area
Upper Delaware
Scenic & Recreational River
136.9
141.8
155.6
160.8
166.6
177.6
181
184.3
194.9
207.3
210.8
215
228.5
233.6
247.5
249.9
255
269
279
293.5
304
315
325
EBr
WBr
River Mile (station locations labeled)

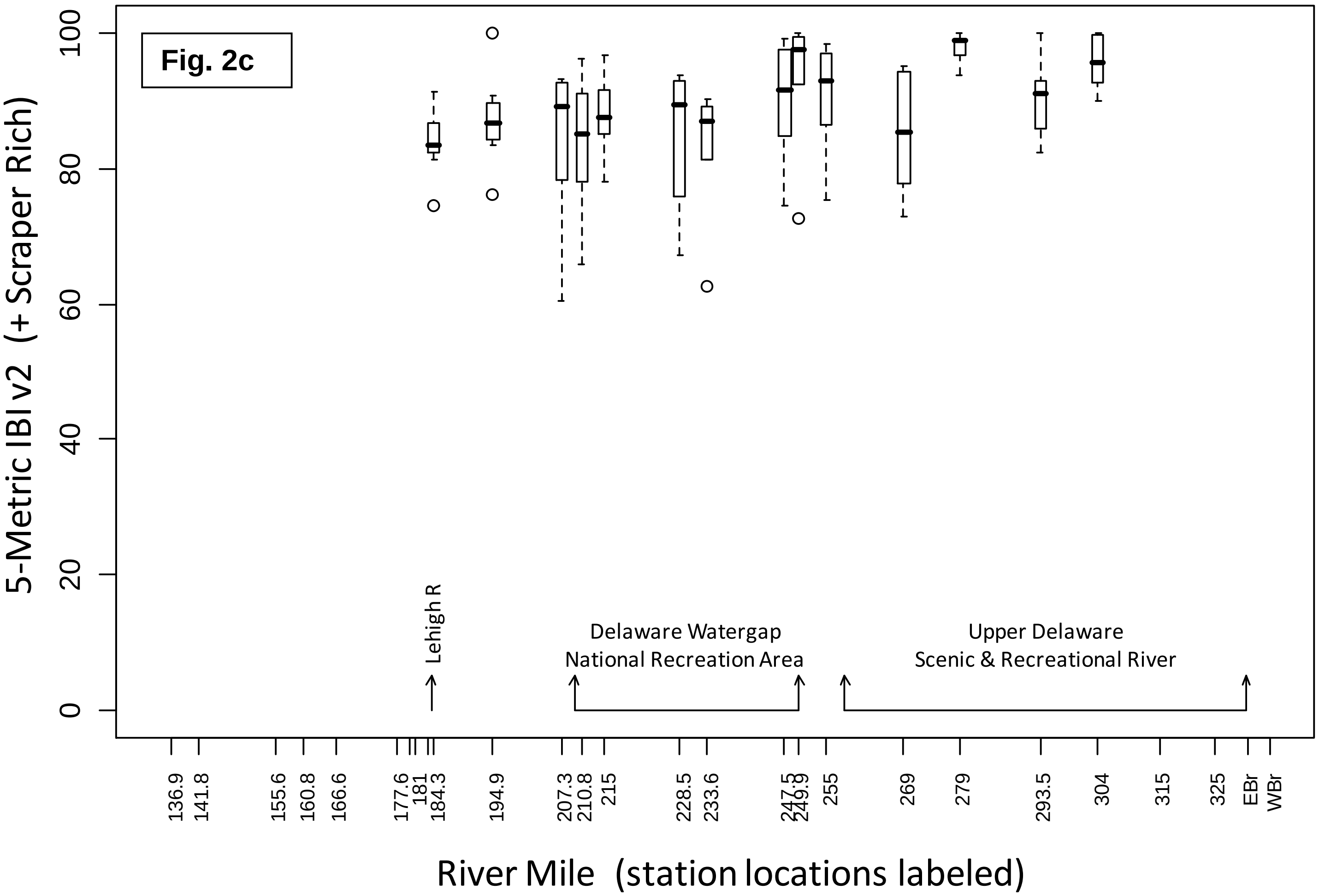


Fig. 2c
5-Metric IBI v2 (+ Scraper Rich)
0
20
40
60
80
100
Lehigh R
Delaware Watergap
National Recreation Area
Upper Delaware
Scenic & Recreational River
136.9
141.8
155.6
160.8
166.6
177.6
181
184.3
194.9
207.3
210.8
215
228.5
233.6
247.5
249.9
255
269
279
293.5
304
315
325
EBr
WBr
River Mile (station locations labeled)

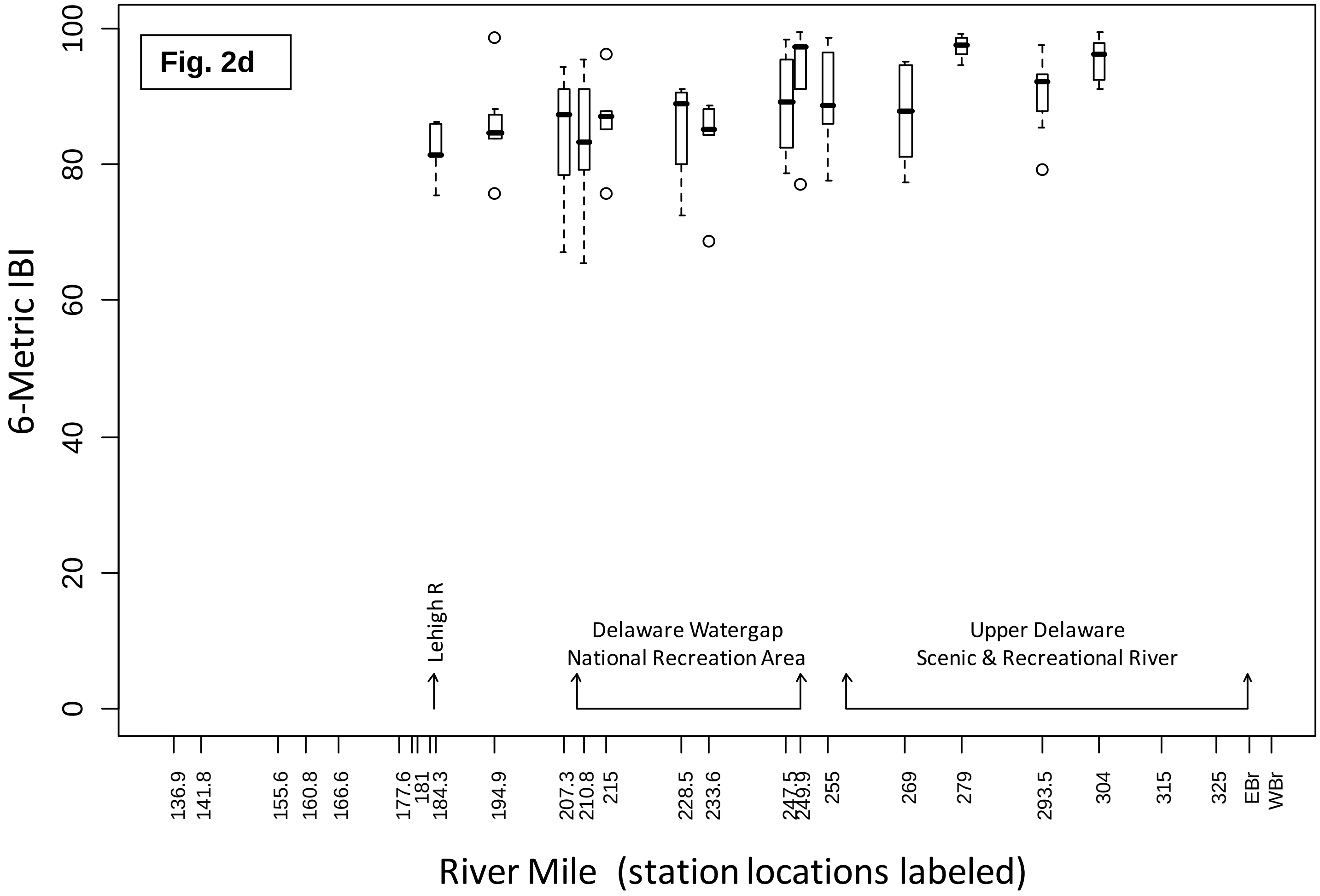


Fig. 2d
6-Metric IBI
0
20
40
60
80
100
Lehigh R
Delaware Watergap
National Recreation Area
Upper Delaware
Scenic & Recreational River
136.9
141.8
155.6
160.8
166.6
177.6
181
184.3
194.9
207.3
210.8
215
228.5
233.6
247.5
249.9
255
269
279
293.5
304
315
325
EBr
WBr
River Mile (station locations labeled)

Figure 3. Box-and-Whisker Plots of 6-Metric IBI for All Stations and All Years (dotted line indicates proposed assessment threshold)

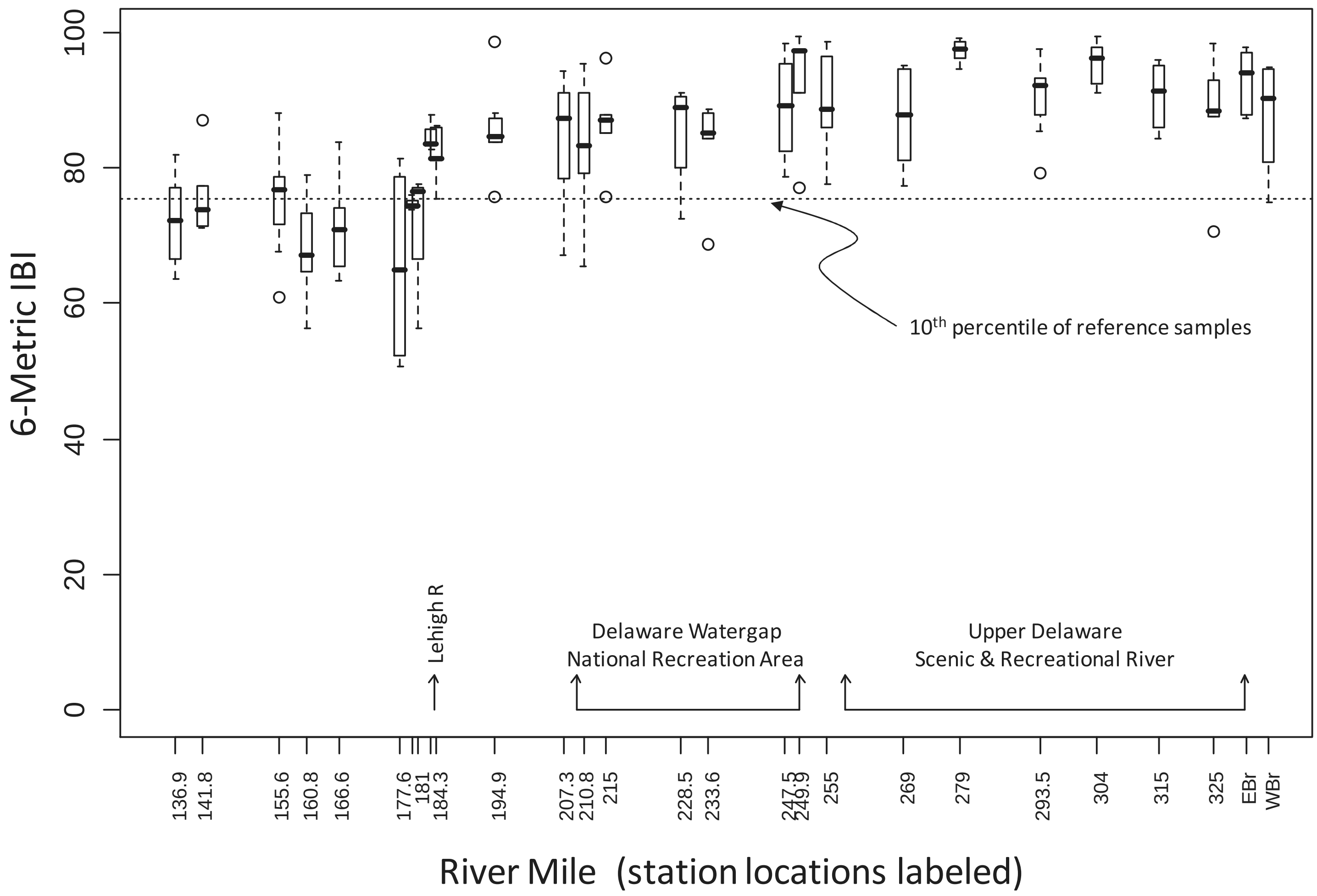


Figure 4a. NYDEC Riffle BAP score vs DRBC 6-metric IBI score using the DRBC data at the DRBC taxonomic resolution. Dotted line indicates 1:1 line.

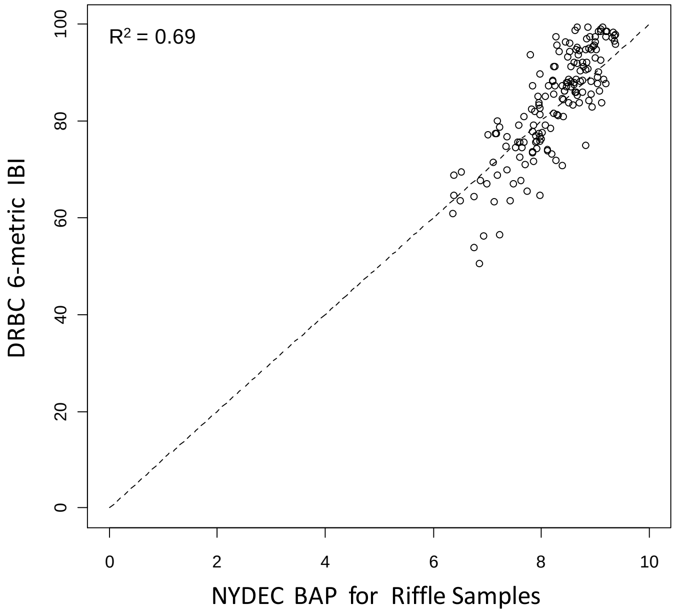


Figure 4b. NYDEC Riffle BAP score using the DRBC data at the DRBC taxonomic resolution for all samples at all stations. Dotted line indicates NYDEC threshold between "Non-Impaired" and "Slightly Impaired" condition classes.

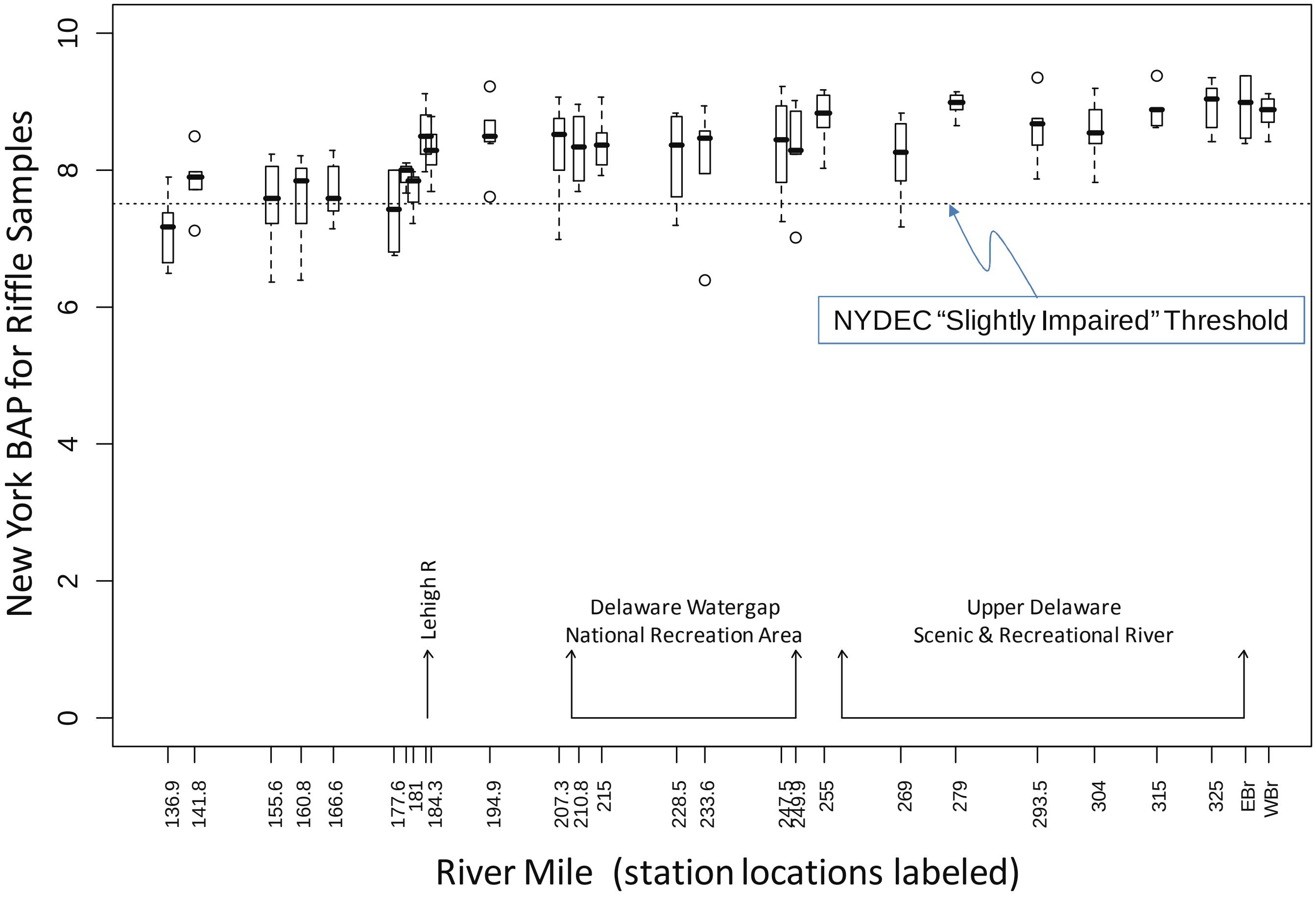


Figure 5a. PADEP Freestone IBI score vs DRBC 6-metric IBI score using the DRBC data at the DRBC taxonomic resolution. Dotted line indicates 1:1 line.

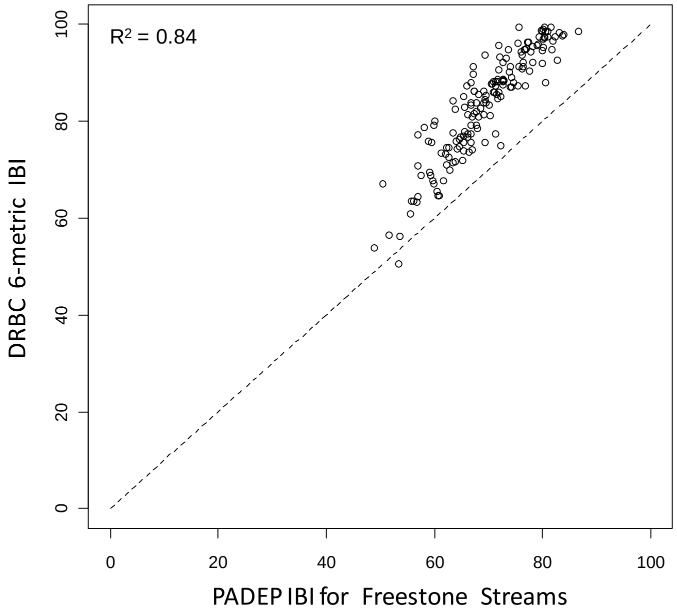
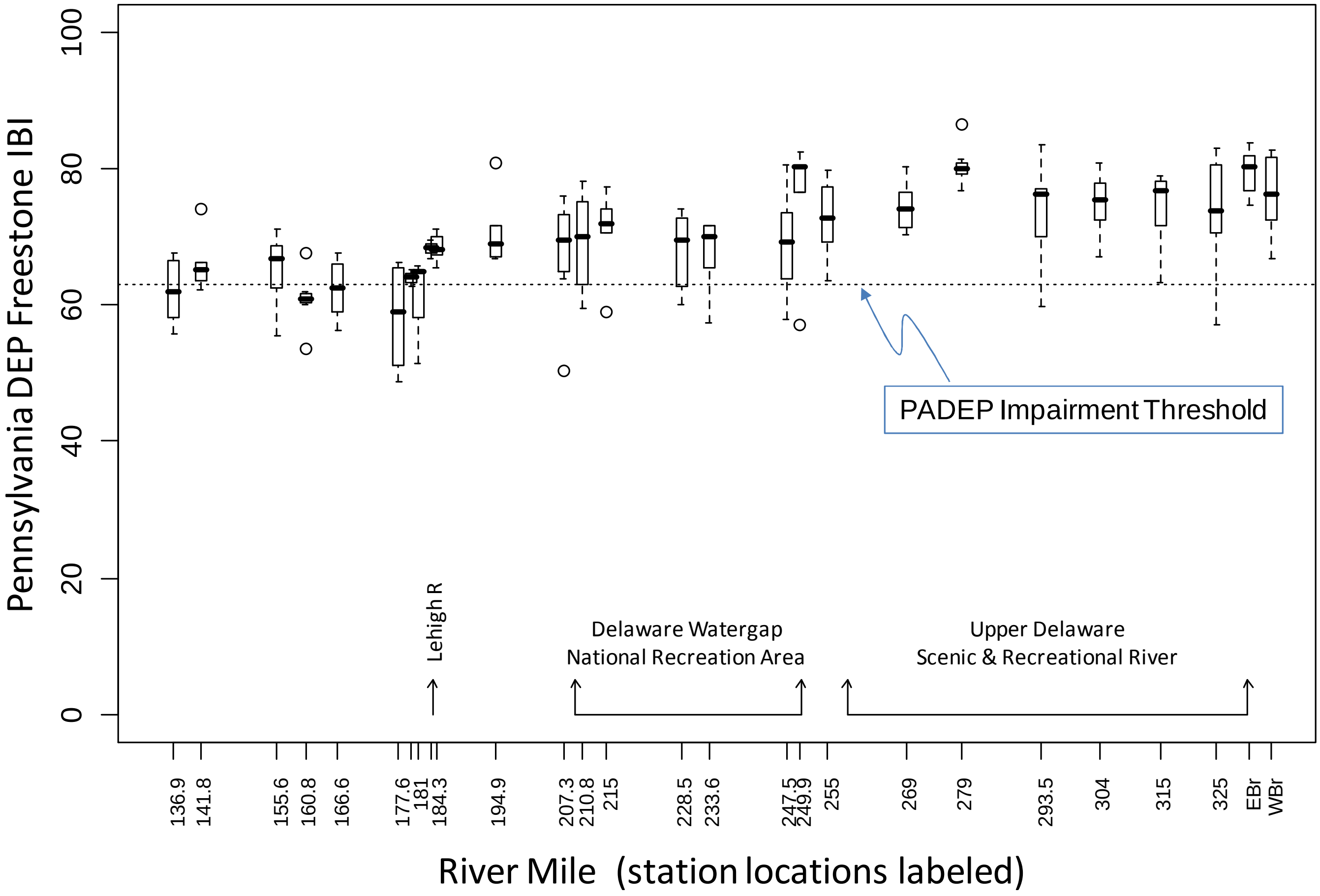


Figure 5b. PADEP Freestone IBI score using the DRBC data at the DRBC taxonomic resolution for all samples at all stations. Dotted line indicates PADEP threshold between "Supporting Use" and "Not Supporting" condition classes.



APPENDICES

Appendix A. Taxonomic resolution for final DRBC dataset, with the Tolerance Values and Functional Feeding Groups (FFG) given for each taxon. For Functional Feeding Groups, codes are: cg=collector/gatherer; f=filterer; pr=predator; sc=scraper; sh=shredder.

Taxon	Tolerance Value	FFG
Turbellaria	6	pr
Prostoma.sp.	7	pr
Manayunkia.speciosa	6	cg
Nemata	6.67	pr
Oligochaeta	8	cg
Hirudinea	8	pr
Hydrobiidae	5	sc
Pleuroceridae	6	sc
Lymnaeidae	6	sc
Ancylidae	6	sc
Ferrissia.sp.	6	sc
Laevapex.sp.	7	sc
Planorbidae	6	sc
Physa.sp.	8	sc
Corbicula.sp.	4	f
Pisidiidae	8	f
Aturus.sp.	7	pr
Sperchon.sp.	7	pr
Sperchonopsis.sp.	7	pr
Lebertia.sp.	7	pr
Hydrachna	7	pr
Rhyncholimnochares.sp.	7	pr
Protzia.sp.	7	pr
Torrenticola.sp.	7	pr
Atractides.sp.	7	pr
Hygrobates.sp.	7	pr
Mideopsis.sp.	7	pr
Oribatida	7	pr
Lirceus.sp.	8	cg
Caecidotea.sp.	8	cg
Gammarus.sp.	6	cg
Hyalella.sp.	8	cg
Crangonyx.sp.	6	cg
Heptageniidae	3	sc
Rhithrogena.sp.	0	sc
Heptagenia.sp.	4	sc
Epeorus.sp.	0	sc
Leucrocuta.sp.	1	sc
Stenacron.sp.	4	sc
Maccaffertium.sp.	3	sc
Baetidae	6	cg
Heterocloeon.sp.	2	sc
Baetis.sp.	6	cg
Acentrella.sp.	4	cg
Acerpenna.sp.	4	cg
Plauditus.sp.	4	cg
Procloeon.sp.	4	cg
Isonychia.sp.	2	f
Leptophlebiidae	4	cg
Ephemerella.sp.	1	cg

Taxon	Tolerance Value	FFG
Eurylophella.sp.	4	cg
Drunella.sp.	1	sc
Serratella.sp.	2	cg
Tricorythodes.sp.	4	cg
Caenis.sp.	7	cg
Baetisca.sp.	4	cg
Ephemera.sp.	2	cg
Ephoron.sp.	2	cg
Gomphidae	1	pr
Libellulidae	9	pr
Neurocordulia.sp.	2	pr
Somatochlora.sp.	1	pr
Macromiinae	2	pr
Argia.sp.	6	pr
Pteronarcys.sp.	0	sh
Leuctra.sp.	0	sh
Perlidae	3	pr
Acroneuria.sp.	0	pr
Paragnetina.sp.	1	pr
Agnetina.sp.	2	pr
Perlesta.sp.	4	pr
Perlodidae	2	pr
Sweltsa.sp.	0	pr
Dineutus.sp.	4	pr
Berosus.sp.	5	cg
Laccobius.sp.	5	pr
Psephenus.sp.	4	sc
Ectopria.sp.	5	sc
Stenelmis.sp.	5	sc
Dubiraphia.sp.	6	cg
Microcylloepus.sp.	3	sc
Optioservus.sp.	4	sc
Macronychus.glabratus	3	sc
Promoresia.sp.	2	sc
Oulimnius.sp.	4	sc
Sialis.sp.	4	pr
Nigronia.sp.	2	pr
Corydalus.sp.	4	pr
Climacia.sp.	5	pr
Rhyacophila.sp.	1	pr
Chimarra.sp.	4	f
Psychomyia.sp.	2	cg
Hydropsychidae	5	f
Cheumatopsyche.sp.	5	f
Hydropsyche.sp.	4	f
Macrostemum.sp.	3	f
Hydroptilidae	4	sc
Leucotrichia.sp.	5	sc
Hydroptila.sp.	6	sc
Ochrotrichia.sp.	4	sc
Oxyethira.sp.	3	sc
Mayatrichia.sp.	6	sc

Appendix A (cont)

Taxon	Tolerance Value	FFG
Apatania.sp.	3	sc
Neophylax.sp.	3	sc
Psilotreta.sp.	0	sc
Setodes.sp.	2	cg
Mystacides.sp.	4	cg
Oecetis.sp.	6	pr
Nectopsyche.sp.	3	sh
Ceraclea.sp.	3	cg
Lepidostoma.sp.	1	sh
Brachycentrus.sp.	1	f
Micrasema.sp.	2	sh
Agarodes.sp.	3	sh
Helicopsyche.sp.	3	sc
Polycentropodidae	6	f
Polycentropus.sp.	6	pr
Neureclipsis.sp.	7	f
Nyctiophylax.sp.	5	pr
Glossosomatidae	1	sc
Culoptila.sp.	1	sc
Glossosoma.sp.	0	sc
Protoptila.sp.	1	sc
Petrophila.sp.	5	sc
Tipulidae	4	sh
Tipula.sp.	4	sh
Antocha.sp.	3	cg
Hexatoma.sp.	2	pr
Dicranota.sp.	3	pr
Simulium.sp.	6	f
Tanypodinae	7	pr
Pentaneurini	6	pr
Ablabesmyia.sp.	8	cg
Conchapelopia.sp.	6	pr
Labrundinia.sp.	7	pr
Nilotanypus.sp.	6	pr
Pentaneura.sp.	6	pr
Rheopelopia.sp.	4	pr
Thienemannimyia.sp.	6	pr
Trissopelopia.sp.	6	pr
Djalmabatista.sp.	3	pr
Procladius.sp.	9	pr

Taxon	Tolerance Value	FFG
Diamesa.sp.	5	cg
Pagastia.sp.	1	cg
Potthastia.sp.	2	cg
Orthocladiinae	5	cg
Cardiocladius.sp.	5	pr
Corynoneura.sp.	4	cg
Eukiefferiella.sp.	6	cg
Lopescladius.sp.	4	cg
Nanocladius.sp.	4.5	cg
Orthocladius.Complex	6	cg
Parakiefferiella.sp.	4	cg
Parametriocnemus.sp.	5	cg
Rheocricotopus.sp.	6	cg
Stilocladius.sp.	3	cg
Synorthocladius.sp.	4	cg
Thienemanniella.sp.	6	cg
Tvetenia.sp.	5	cg
Chironominae	6	cg
Chironomini	6	cg
Cryptochironomus.sp.	8	pr
Demicryptochironomus.sp.	8	cg
Dicrotendipes.sp.	8	cg
Microtendipes.sp.	6	f
Nilothauma.sp.	4	cg
Phaenopsectra.sp.	7	sc
Polypedilum.sp.	6	sh
Robackia.sp.	4.5	cg
Stenochironomus.sp.	5	cg
Xenochironomus.sp.	2	pr
Pseudochironomus.sp.	5	cg
Tanytarsini	6	f
Cladotanytarsus.sp.	6	f
Micropsectra.sp.	7	cg
Paratanytarsus.sp.	6	cg
Rheotanytarsus.sp.	6	f
Stempellinella.sp.	4	cg
Sublettea.sp.	4	f
Tanytarsus.sp.	6	f
Atherix.sp.	2	pr
Empididae	6	pr
Dolichopodidae	4	pr

Appendix B.1. Example of a sample with low overall score. The original sample processed in the lab (551 individuals) was numerically re-sampled to illustrate what this sample would look like under both a 100-invididual and a 200-individual subsampling procedure.

SiteName	Year	River Mile
Raubs Island	2006	177.6

			Abundance in Sample		
TV	FFG	Taxon	100-bug re-sample	200-bug re-sample	original sample
5	f	Cheumatopsyche.sp.	37	82	224
4	f	Hydropsyche.sp.	16	24	59
4	cg	Acentrella.sp.	13	20	62
6	cg	Baetis.sp.	12	21	70
4	f	Corbicula.sp.	4	9	23
6	cg	Orthocladius.Complex	3	5	13
3	sc	Maccaffertium.sp.	3	12	30
4	cg	Plauditus.sp.	2	1	6
6	f	Rheotanytarsus.sp.	2	2	8
2	f	Isonychia.sp.	2	6	14
1	sh	Lepidostoma.sp.	1	1	2
5	pr	Cardiocladius.sp.	1	1	2
7	pr	Sperchon.sp.	1	2	5
6	f	Simulium.sp.	1	4	7
7	pr	Lebertia.sp.	1		3
5	sc	Petrophila.sp.	1		2
1	sc	Leucrocuta.sp.		1	2
6.7	pr	Nemata		1	1
3	pr	Perlidae		1	1
4	sc	Optioservus.sp.		1	1
3	f	Macrostemum.sp.		1	1
8	cg	Ablabesmyia.sp.		1	1
5	cg	Tvetenia.sp.		1	1
6	cg	Eukiefferiella.sp.		3	6
1	sc	Protoptila.sp.			2
2	sc	Heterocloeon.sp.			1
7	cg	Caenis.sp.			1
5	sc	Stenelmis.sp.			1
4	sc	Oulimnius.sp.			1
3	cg	Ceraclea.sp.			1
		Total # Invertebrates =	100	200	551

Final Score
6-Metric IBI
53.8

Metric Scores					
Shannon Diversity	Biotic Index	Richness (200-bug rarefaction)	EPT Rich (200-bug rarefaction)	Intolerant % Rich	Scraper Rich
2.10	4.71	20.4	10.5	16.7%	8

Appendix B.2. Example of a sample with high overall score under the same format.

SiteName	Year	River Mile
Ascalona	2008	279

			Abundance in Sample		
TV	FFG	Taxon	100-bug re-sample	200-bug re-sample	original sample
5	sc	Hydrobiidae	11	21	68
3	sc	Maccaffertium.sp.	8	22	58
4	cg	Plauditus.sp.	7	13	42
6	cg	Baetis.sp.	7	11	33
1	sc	Culoptila.sp.	6	16	50
4	f	Hydropsyche.sp.	6	12	39
1	sc	Leucrocuta.sp.	6	11	35
2	f	Isonychia.sp.	6	6	25
2	sc	Heterocloeon.sp.	5	7	18
2	cg	Serratella.sp.	4	11	19
6	f	Rheotanytarsus.sp.	4	10	24
6	sc	Hydroptila.sp.	4	4	15
4	f	Chimarra.sp.	3	3	11
6	sh	Polypedilum.sp.	3	3	11
5	sc	Stenelmis.sp.	3	2	8
5	f	Cheumatopsyche.sp.	2	10	30
8	f	Pisidiidae	2	8	13
1	sc	Protoptila.sp.	2	2	8
4	cg	Acerpenna.sp.	2	1	12
6	cg	Orthocladius.Complex	2	1	11
2	pr	Agnetina.sp.	1	2	8
7	f	Neureclipsis.sp.	1	2	6
8	cg	Oligochaeta	1	1	3
5	sc	Petrophila.sp.	1	1	2
1	sh	Lepidostoma.sp.	1		3
3	f	Macrostemum.sp.	1		1
6	pr	Oecetis.sp.	1		1
6	sc	Planorbidae		6	11
4	cg	Tricorythodes.sp.		2	7
3	sc	Apatania.sp.		2	6
2	sc	Promoresia.sp.		2	3
3	cg	Ceraclea.sp.		1	4
1	f	Brachycentrus.sp.		1	4
8	sc	Physa.sp.		1	2
0	sc	Epeorus.sp.		1	2
1	pr	Gomphidae		1	1
4	pr	Corydalus.sp.		1	1
2	cg	Psychomyia.sp.		1	1
2	sh	Micrasema.sp.		1	1
5	cg	Tvetenia.sp.			5
4	sc	Heptagenia.sp.			3
0	pr	Acroneuria.sp.			3
7	pr	Atractides.sp.			2
6.7	pr	Nemata			1
9	pr	Libellulidae			1
6	sc	Mayatrichia.sp.			1
6	pr	Empididae			1
		Total # Invertebrates =	100	200	614

Final Score
6-Metric IBI
99.3

Metric Scores

Shannon Diversity	Biotic Index	Richness (200-bug rarefaction)	EPT Rich (200-bug rarefaction)	Intolerant % Rich	Scraper Rich
3.23	3.82	35.9	23.5	31.9%	16