An Index of Biological Integrity for "True" Limestone Streams

April 2009

The Spring at Big Spring Creek

<u>**An Index of Biological Integrity (IBI) for "True" Limestone Streams**</u>
April 2009

Prepared by:

William Botts
PA Department of Environmental Protection
Bureau of Water Standards and Facility Regulation
Rachel Carson State Office Building
Harrisburg, PA 17105

Acknowledgments

Ron Hughey and John Ryder of the PA DEP Northcentral Regional Office collected and processed many of the samples used to develop the IBI and provided extensive help in the development process. Michael Barbour of Tetra Tech and Maggie Passmore of EPA Region 3 provided information that was critical in developing an IBI that conformed to current methodologies. Rod Kime and Brian Chalfant PA DEP Central Office, Robert Schott, Marvin Hammond, and Mark Embeck of PA DEP Southcentral Regional Office also provided guidance and support through the years of developing and refining the IBI.

Executive Summary

In 2006 Pennsylvania Department of Environmental Protection (PA DEP) published a protocol to assess limestone streams. The data used to develop that assessment protocol was collected from 1998 to 2003 with approximately 90 % of data collected from 2000 to 2003. PA DEP has collected additional data from 2004 through 2008 and used that data to analyze and update the original 2006 document "An Index of Biological Integrity (IBI) for "True" Limestone Streams." Additional sample data collected from 2004 to 2008 were primarily from streams that were sampled for the original protocol. This was due to the lack of limestone streams in the state and because those streams are still being monitored for anthropogenic impacts. From 2004 through 2008 there were 94 new samples collected from previously sampled sites (n=36), new sites on previously sampled streams (n=47) and sites on streams that were not sampled before (n=12). There are now 188 samples in the limestone data set. The Limestone IBI has proven to be an effective assessment tool. It has been used extensively to assess with NPDES discharges to limestone streams and to assess the overall ecological health of some limestone streams. The current Limestone Protocol has good precision and accuracy, but advances in biological assessment science and doubling the sample data indicated some minor adjustments should be implemented. This paper will present the original concepts concerning "True" limestone streams with adjustments instituted due to analysis of additional data using the most current biological assessment science.

In the second limestone stream analysis process a thorough attempt to connect abiotic factors to stream changes was tried again. All the water in a limestone stream comes from deep limestone aquifers. Surface water of substantial quantities rarely enters these streams and on a yearly or even monthly time period the stream flow must consist of mainly deep ground water. There should be no network of even intermittent streams entering a limestone stream unless they are also deep groundwater supported streams. There may be ephemeral stream beds connected to a limestone stream, but even ephemeral stream beds should be few. Exceptions may only occur if the limestone flow completely dominates the surface flow. Limestone stream classification mandates impacts, natural or anthropogenic, can only occur when a surface flow or piped flow is connected to the stream. Exceptions are sink holes or underground cave streams that direct surface flows to the stream or if the deep aquifer is substantially polluted. These facts mean land use cannot be used as an abiotic indicator of reference conditions or impairment. Water is purified as it percolates to the deep groundwater level and pollutants are diluted by the quantity of groundwater. Unfortunately, this only applies to impacts to macroinvertebrates. Many aquifers contain high nitrates and pesticides, but it would be very rare for an aquifer to contain enough of a pollutant to impact the biota

Habitat scores are also of reduced value in evaluating a limestone stream. The loss of trees or shrubs in the riparian zone is not always a reason for concern. Shading can actually be detrimental due to lost productivity. Historically some limestone streams were probably shaded and some were not. The streams are low gradient so they have poor riffle development and moderate to high sedimentation problems. In limestone streams the sensitive or good quality taxa are adapted to living with sedimentation. This is analogous to stoneflies in acidic mountain streams.

Habitat scores can be of some use. Reference sites generally score above 150, but so can many impaired streams. There is too much overlap in habitat scores between reference sites and impaired sites to make it a statistically useful parameter. However, when a site has a very low score as 120 it is probably impaired. Riparian land use is detrimental if it allows surface flows (sheet or concentrated) to reach the stream or there is a substantial disturbance that extends to the edge of the stream. An example would be crop farming to the edge of the stream or pastures and barnyards on the stream.

The main abiotic source of pollution to a limestone stream occurs when the surrounding watershed is directly connected to the stream. Even though limestone watersheds are generally extensively farmed the connection is usually a point source type discharge. Examples of point source discharges are: municipal or industrial permitted discharges, urban and industrial storm water discharges, areas of concentrated farm animals, and sink holes. Water withdrawals are another threat to limestone streams. Abiotic parameter to measure possible impacts to limestone streams should include the magnitude of the pollutant, including concentration and volume and the distance of the discharge to the sample site. It is important to consider that the assimilation of nutrient and organic wastes is different in the productive ecological system of limestone streams.

An Index of Biological Integrity (IBI) for "True" Limestone Streams

April 2009

"True" limestone streams, limestone spring streams, or simply limestone streams are very unique. These streams are formed by large alkaline springs or they are streams maintained by many large alkaline springs. Pennsylvania has approximately 83,000 miles of streams and there are probably less than 800 miles of limestone streams. However, this small subset of streams is of great ecological and economical importance. Limestone streams like the Letort Spring Run and Spring Creek are world famous trout fishing streams attracting anglers from around the country and from many nations. The ecological integrity of limestone streams must be assessed correctly if they are going to be properly protected. These streams have fairly low gradient, constant temperatures, high alkalinity and are highly productive. Their unique physical and chemical characteristics produce a unique macroinvertebrate community. The lack of diversity in habitat, temperature and water chemistry produces a macroinvertebrate community with low diversity. The highly productive water chemistry produces a high density of macroinvertebrates. The end result is a community with a low number of taxa that is generally dominated by a few taxa. In fact five taxa, *Lirceus*, *Gammarus*, *Ephemerella, Optioservus* and Chironomidae, accounted for about 79.2 % of the total organisms collected in the 188 sample data set. The unique macroinvertebrate communities created by these unique aquatic environments make it essential that a separate Index for Biological Integrity (IBI) be developed for limestone streams. If limestone streams are assessed with an IBI for freestone streams even the very best sites would look impaired. On the other hand, if a freestone stream is assessed using the limestone IBI an impaired stream could easily pass as unimpaired. This makes it very important for streams to be correctly classified as limestone streams. A mistake in stream classification will make it impossible to properly assess the stream's ecological condition. The EPA publications detailing the development of Rapid Bioassessment Protocols (Plafkin et al. 1989; Barbour et al. 1999) were a major source for the development of the limestone stream IBI.

Stream Classification and Reference Criteria

Limestone streams are streams formed by large limestone springs or are very strongly influenced by limestone springs. However, a stream located in limestone geology that appears to originate from spring sources does not guarantee it should be classified as a limestone stream. Limestone streams are always in limestone geology, but all streams in limestone geology are not limestone streams. The two most important characteristics in the classification of a limestone stream are temperature and alkalinity. The sampling of Pennsylvania limestone streams indicates the alkalinity should be maintained above 140 mg/l throughout the year. Many streams may yield high alkalinity results for much of the year, but if there are any periods where the alkalinity fluctuates below 140 mg/l the stream should be examined very carefully. Groundwater temperatures are approximately 50 to 55 degrees Fahrenheit (F). Streams strongly influenced by groundwater will maintain temperatures near 50 degrees F. Many macroinvertebrates need fluctuating temperatures to complete their life cycles so if temperatures fluctuate too much the diversity of the macroinvertebrate community increases and it no longer is a distinct limestone community. These two criteria may require the investigator to have year round data on the stream to correctly classify it as limestone. Table 1 lists the criteria for limestone streams and reference limestone streams. Note for a stream to qualify as reference it must qualify for at least High Quality under the Chapter 93.4b antidegradation requirements.

Table 1.

Limestone Streams Criteria

Parameter	Criterion	Explanation
Alkalinity	Minimum 140 mg/l	Stream must maintain high alkalinity throughout the year
Temperature	40 to 65 deg. F 4 to 18 deg. C	Constant temperatures are very important, check to see if the stream is ice free in the winter
Stream originates from limestone springs or very strongly influenced by limestone springs		
Drainage Area	Maximum 20 sq. miles Surface drainage area	There maybe exception to this parameter as long as all other criteria are met
Designated Water Use	Cold Water Fishery (CWF)	Must be designated a CWF in Chapter 93

Reference Criteria

Designated Water Use	Cold Water Fishery (CWF) HQ or EV	Chapter 93 HQ or EV. Existing Use would also be acceptable
Dissolved Oxygen	Minimum 6.0 mg/l	Taken from Chapter 93.7 Specific water quality criteria (CWF)
Trout Population	Must be Class A	Some streams may not have been sampled by the PFBC and would require an assessment
pH	From 6.0 to 9.0	Taken from Chapter 93.7 Specific water quality criteria
No point source discharges upstream of site		
No obvious non-point-source pollution (NPS) impacts		
Evaluate anthropogenic activities in the drainage area	Limestone streams are isolated from their drainage area because they maintain flow almost exclusively from ground water. They develop problems when surrounding land uses get directly connected to the stream. Examples: quarries, activities near sink holes, storm water pipes etc.	
Total Habitat Score	Minimum 150	There maybe exception to this parameter

Field Sampling and Sample Processing Methods

Macroinvertebrates were collected using the PA Modified RBP 3 method. Using this method samples must be collected January through May. Each macroinvertebrate sample was collected with a D-Frame net, disturbing a minimum area of 1 x 1 meters at two-selected representative riffle/run areas. Each sample consisted of two kicking efforts, one collected from a fast riffle/run habitat and one from a slow riffle/run habitat. In low gradient limestone streams it is often difficult to locate a riffle so it is appropriate to sample runs or the best available rock substrate. The two kicking efforts were composited, fixed in a solution of 95 % ethanol alcohol and returned to the laboratory for processing. The entire sample was placed in an 18” x 12” x 3.5” pan containing 28-2 inch square grids. A four square inch area circular cookie-cutter was used to randomly select and remove 4 grids. These 4 grids were placed in a second grided pan. Grids were randomly selected and picked until a 300 organism (+/- 20%) subsample was obtained. Organisms in the subsample were identified and enumerated. The midges were identified to the family level. Flatworms and aquatic earthworms were identified to the class level. Proboscis worms and roundworms were identified to the phylum level and all other macroinvertebrates were identified to the genus level. The data from the subsamples were used in the computation of the metrics. An explanation of the sampling and processing protocols is provided in Appendix A.

Selection of Metrics

The multimetric approach was used to develop the IBI for limestone streams. Each metric should measure the impact of pollution on different attributes of the macroinvertebrate community. The measured impact from each metric is added together to generate an IBI score that characterizes the ecological integrity of the site. Multimetrics work best (Gerritsen et al. 2000) when:

- Metrics have good discrimination efficiency (DE) between reference and impaired sites.
- Metrics are not redundant.
- Metrics are designed to measure different aspects of the macroinvertebrate community (richness, tolerance, composition).

In the literature there are long lists of metrics (Barbour et al. 1996; Barbour et el. 1999; Gibson et al. 1996). However, a large number of the metrics did not appear to be appropriate. The composition of limestone stream macroinvertebrate communities definitely influences the selection of metrics. The communities have low diversity and a few taxa with high density. This is true to some degree in both reference and impaired streams. Table 2 lists the five most common taxa collected from limestone streams and shows how the composition of the macroinvertebrate communities changes from reference sites to impaired sites. This table clearly shows how different limestone macroinvertebrate communities are and how these differences could affect metric selection.

Table 2.

Average Percent of Organisms of a Taxa Collected Per Sample

Taxa	Common Name	TV	Reference Sites	Attaining Sites	Impaired Sites
Lirceus	Sowbugs	8	9.4 %	29.5 %	52.9 %
Gammarus	Scuds	6	25.0 %	10.7 %	12.7 %
Ephemerella	Mayfly	1	12.0 %	12.4 %	1.2 %
Optioservus	Riffle Beetle	4	11.6 %	5.8 %	1.8 %
Chironomidae	Midges	6	15.7 %	14.8 %	15.5 %
Total %	**Organisms**		**73.6 %**	**73.2 %**	**84.0 %**

These 5 taxa account for 79.2 % or 45,967 out of 58,010 organisms collected from 188 samples. Tolerance Value (TV)

The six metrics currently in use are: Total Taxa, EPT Taxa, % Intolerant (TV $\leq$ 3), Shannon Diversity, % Tolerant (TV $\geq$ 7) and HBI. These work well, but a new analysis would demonstrate if that group of metrics was still the best. The same community characteristics as before were considered. Low diversity influences metrics that involve counts. Pollution-sensitive taxa and pollution-tolerant taxa found in limestone streams tend to be in the same functional feeding groups so functional feeding group metrics do not work well. Based on the unique macroinvertebrate communities found in limestone streams and best professional judgment, 15 possible candidate metrics were selected. These 15 metrics were tested to determine the optimal set of metrics to assess the ecological condition of limestone stream sample sites.

Metric Discrimination Efficiencies (DE)

A metric must have the ability to differentiate between a reference site and an impaired site. The ability of a metric to determine impairment is measured by calculating the discrimination efficiency (DE) (Barbour et al. 1999; Gerritsen et al. 2000). The distribution of metric scores from reference sites were used to set thresholds for metrics that increase with pollution at the 75th percentile and metrics that decrease with pollution at the 25th percentile. Using the reference thresholds and known stressed sites (303d listed sites) the DE formula below was used to find a metric's DE. An example using Total Taxa is shown in the box and whisker plot in Figure 1. Twelve of the 96 impaired sites exceeded the reference distribution's 25th percentile. In other words, the Total Taxa metric correctly assessed the impaired sites 87.5 % of the time. Table 3 contains the DE for all the candidate metrics. Very strict stream classification, careful reference site selection and a rigorous metric pre-selection process generated many good DE. This made it possible to select metrics with high DE values. The analysis showed 11 of the 14 metrics had very good DE. The box and whisker plots for the selected metrics are found in Appendix B.

The DE formula: DE = 100 x a/b

For metrics that decrease with pollution.

a = the number of impaired sites scoring below the 25th percentile of the reference distribution.

b = the total number of impaired samples.

For metrics that increase with pollution.

a = the number of impaired sites scoring above the 75th percentile of the reference distribution.

b = the total number of impaired samples.

Graph 1. Total Taxa DE = 87.5

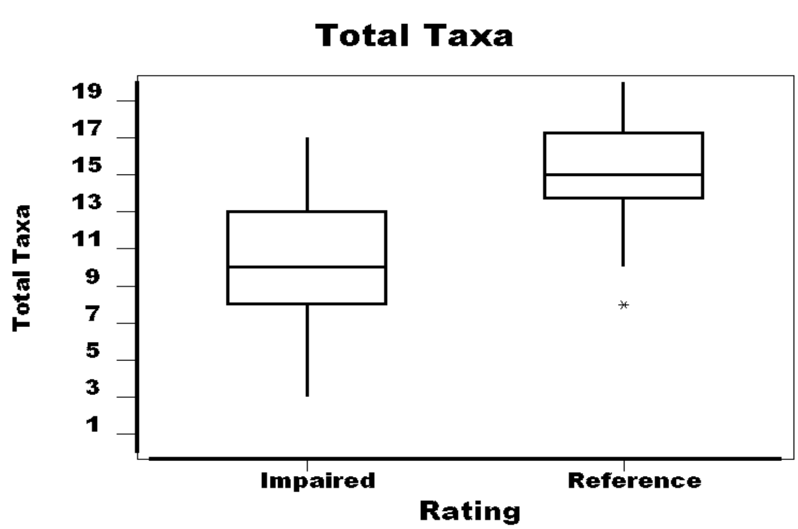


Table 3. A list of candidate metrics, discrimination efficiency (DE) and reason for selection.

Metric by Category	Response to Pollution	(DE)	Used in Original IBI	Selected For New IBI
<u>Taxonomic Richness</u>				
Total Taxa	Decreases	87.5 %	**Yes**	**Yes**
EPT Taxa	Decreases	85.4 %	**Yes**	**Yes**
EPT Taxa (TV ≤ 4)	Decreases	90.6 %		
<u>Taxonomic Composition</u>				
Shannon Diversity	Decreases	84.4 %	**Yes**	**Yes**
% Mayflies	Decreases	86.5 %		
% Mayflies (TV ≤ 4)	Decreases	92.7 %		
% EPT	Decreases	83.3 %		
% EPT (TV ≤ 4)	Decreases	88.5 %		
% Limestone Macros	Decreases	86.5 %		
<u>Tolerance/Intolerance</u>				
Beck's Index, version 3	Decreases	94.8 %		
Beck's Index, version 4	Decreases	86.5 %		**Yes**
HBI	Increases	94.8 %	**Yes**	**Yes**
% Intolerant (TV ≤ 4)	Decreases	87.5 %		
% Tolerant (TV ≥ 7)	Increases	90.6 %	**Yes**	**Yes**
% Intolerant (TV ≤ 3)	Decreases	88.5 %	**Yes**	

<u>Redundant Metrics</u>

All fifteen candidate metrics had DE values greater than 80 %, demonstrating the ability to assess impairment, but some of the metrics measure the same community attribute. Metrics measuring the same attribute are redundant and may be eliminated. Pearson Correlation Coefficients were calculated for the 15 metrics (Table 4) to find which metrics were highly correlated and may be redundant. Metrics may be fairly well correlated and still be retained if they measure different community characteristics and enhance the overall DE of the IBI score. A useful tool in evaluating a metric are scatter plots of highly correlated metrics and a good understanding of what the metrics are measuring in the context of limestone streams.

Table 4. Pearson Correlation Coefficients for 15 candidate metrics. All the samples (n=188) were used for the correlation.

Pearson Correlation r

	Total Taxa	**EPT Taxa**	**Beck,4**	**HBI**	**%Tol (≥ 7)**	**Shan Divers**	EPT Taxa≤4	%Lime Macros
EPT Taxa	**0.85**							
Beck's Index 4	0.84	**0.89**						
HBI	-0.54	-0.56	-0.63					
% Tolerant ≥ 7	-0.52	-0.49	-0.58	**0.89**				
Shannon Diversity	**0.85**	0.73	0.75	-0.61	-0.61			
EPT Taxa ≤ 4	0.76	**0.90**	**0.91**	-0.57	-0.49	0.64		
% Limes Macros	0.23	0.25	0.32	-0.70	-0.69	0.26	0.28	
Beck's Index 3	0.70	0.79	**0.91**	-0.58	-0.51	0.58	**0.89**	0.37
% Intolerant ≤ 3	0.43	0.44	0.50	-0.84	-0.54	0.49	0.48	0.55
% Intolerant ≤ 4	0.34	0.36	0.44	-0.78	-0.73	0.35	0.39	**0.96**
% Mayflies ≤ 4	0.34	0.33	0.36	-0.78	-0.50	0.42	0.36	0.58
% Mayflies	0.36	0.33	0.37	-0.70	-0.50	0.50	0.34	0.43
% EPT	0.51	0.52	0.53	-0.75	-0.59	0.65	0.50	0.40
% EPT ≤ 4	0.41	0.41	0.43	-0.80	-0.51	0.46	0.43	0.58

	Beck,3	%Intol (≤ 3)	%Intol (≤ 4)	%May (≤ 4)	%May	%EPT
% Intolerant ≤ 3	0.50					
% Intolerant ≤ 4	0.47	0.66				
% Mayflies ≤ 4	0.35	**0.89**	0.59			
% Mayflies	0.32	0.77	0.43	**0.87**		
% EPT	0.46	0.74	0.43	0.81	**0.95**	
% EPT ≤ 4	0.42	**0.92**	0.61	**0.98**	**0.85**	0.82

The 15 metrics contain groups of metrics that are very similar and measure approximately the same community reaction to stress. Similar metrics are EPT Taxa ≤ 4, % Intolerant ≤ 3, % Intolerant ≤ 4, % Mayflies ≤ 4, % Mayflies, % EPT, and EPT Taxa; however, each metric may measure stressors a little differently. Several metrics are correlated or highly correlated. Metrics with a Pearson Correlation r of +/- 0.85 are in bold font and are correlated. Because this is an update of an already successful IBI, as long as the new data does not substantially change the DE of the metric or the DE and precision of the total IBI score, the original metrics are favored to be retained. The one exception was % Intolerant (TV ≤ 3) which had a fairly high coefficient of variation and high enough number of quartile outlier values that an outlier value, not the 95th percentile, was used for the Best Value. Also, in this analysis of metrics Beck's Index, version 3 and Beck's Index, version 4 were added to the candidate group. Redundancy of selected metrics will be explained after performance characteristics are analyzed.

Performance Characteristics

In the original Limestone IBI Protocol paper performance characteristics of metrics were included with performance analysis of the IBI after the protocol metrics were selected. However, the performance characteristics of each metric and how various groups of metrics perform is very important in metric selection and will be analyzed as part of the metric selection process. Table 5 contains a lot of information that can assist in the selection of a metric or the final group of metrics. The first 6 metrics in the table were the metrics that were ultimately selected. One reason they were selected is because they produced a very good precision estimate (PE) for intra-site spatial variability of the total IBI score. Other combinations of metrics were analyzed, but with all factors considered this combination showed the best ecological assessment potential. The calculations for Table 5 were based 63 replicate samples from a combination of reference, attaining and impaired sites.

Table 5. Performance Characteristics

Metric	Average Variance or RMSE	Standard Diviation	P.E. 90% CI +/-	Approx. Mean	Approx. Coefficient of Variation (%)	DE	Category
Total Taxa	**2.00**	**1.414**	**1.81**	**12.59**	**11.24**	**87.5**	Richness
EPT Taxa	**1.56**	**1.247**	**1.60**	**4.87**	**25.59**	**85.4**	Richness
Beck's Index 4	**2.25**	**1.501**	**1.92**	**6.73**	**22.31**	**86.5**	Tolerance
HBI	**0.103**	**0.321**	**0.544**	**6.12**	**5.24**	**94.8**	Tolerance
% Tolerant ≥7	**58.75**	**7.665**	**9.81**	**42.89**	**17.87**	**90.6**	Tolerance
Shannon Diversity	**0.037**	**0.192**	**0.246**	**1.418**	**13.56**	**84.4**	Composition
Total Score	**32.25**	**5.68**	**7.27**	**58.24**	**9.75**	**97.9**	
EPT Taxa ≤4	**0.92**	**0.959**	**1.23**	**2.60**	**36.86**	**90.6**	Richness
% Lime Macro	**53.12**	**7.288**	**9.33**	**32.92**	**22.14**	**86.5**	Composition
Beck's Index 3	**2.67**	**1.633**	**2.09**	**4.24**	**38.53**	**94.8**	Tolerance
% Intolerant ≤3	**22.31**	**4.724**	**6.05**	**12.54**	**37.68**	**88.5**	Tolerance
% Intol ≤4	**32.71**	**5.719**	**7.32**	**37.98**	**15.06**	**87.5**	Tolerance
% May ≤4	**33.83**	**5.816**	**7.44**	**8.06**	**72.20**	**92.7**	Composition
% May	**38.67**	**6.218**	**7.96**	**10.79**	**57.65**	**86.5**	Composition
% EPT	**39.60**	**6.293**	**8.05**	**15.96**	**39.44**	**83.3**	Composition
% EPT ≤4	**27.70**	**5.263**	**6.74**	**9.41**	**55.94**	**88.5**	Composition

Boxes around CV percents indicate high values
Standard Diviation (s.d.) = square root of RMSE or Avg. Variance
90% Confidence Interval = s.d. x 1.28, Coefficient of variation = s.d./mean

All the metrics show good or adequate discrimination efficiency so that was not a reason to eliminate any metrics. All of these metrics are ecologically relevant to limestone stream biological assemblages (Flotemersch et al. 2006). It appears there are a lot of metrics that could be used, but as stated before, several metrics are very similar and generally only one metric from a similar group should be selected. The tendency is to select the metrics with the highest DE values, which is a good place to start. However, gaining a few percent in the final IBI score DE and increasing the PE for intra-site spatial variability should be avoided. A good measure of how an individual metric will impact the IBI's precision is the metric's Coefficient of Variation (CV). The CV indicates the "relative variability" of a metric. Therefore, metrics with high CV percents will increase variability in the total IBI scores of replicate samples and, more importantly, lower the protocol's ability to repeat samples with precision. Table 5 shows that most of the metrics that were rejected had high CV percents. The % Intolerant (TV $\leq$ 4) had a good CV and good DE, but it is in the metric category Tolerance Measure and there are already three Tolerance Measures in the selected group of metrics. Among the four Tolerance Measures, the % Intolerant (TV $\leq$ 4) metric is the least ecologically relevant to limestone streams and was rejected.

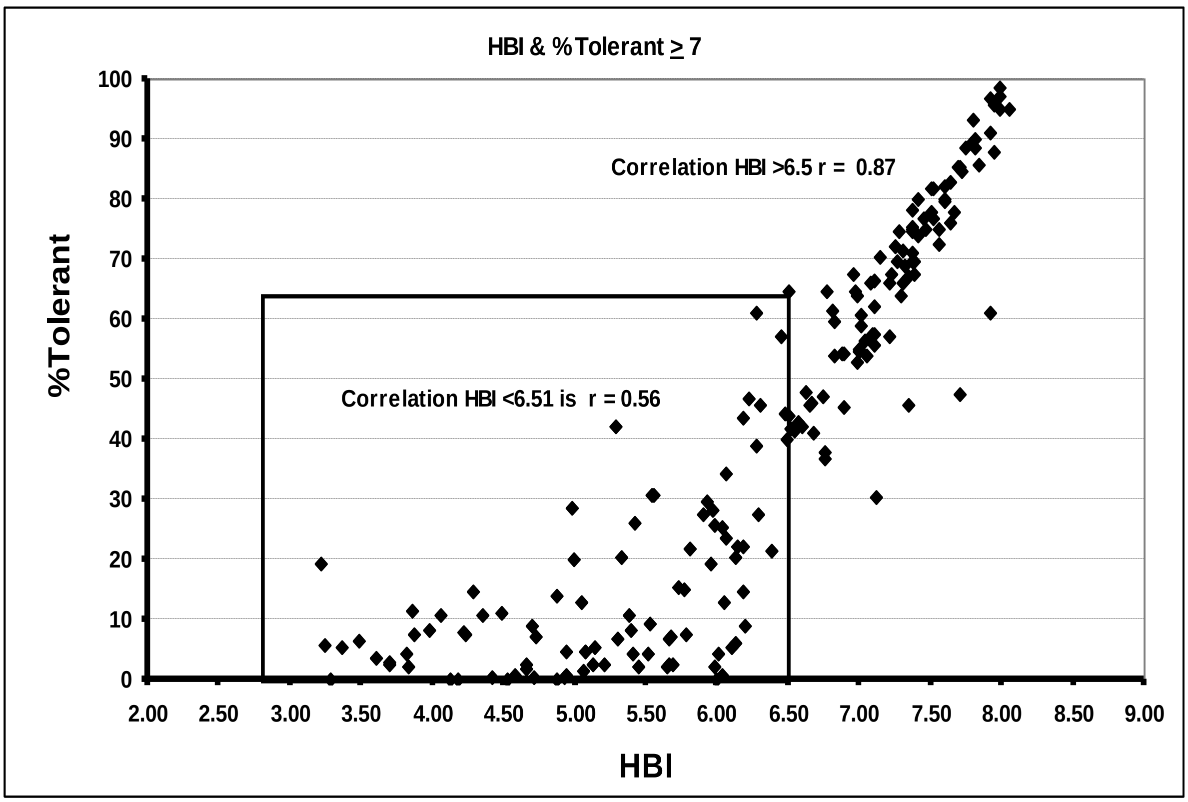
The % Limestone Macros (macroinvertebrates) was a metric developed to exploit the simplistic dynamics of community composition from reference sites to impaired sites. This fact was always apparent, but a useable, consistently predictable metric could not be developed. This was because a taxon that was known to be strongly associated with limestone streams was ignored and the relationship of that taxon between reference sites and impaired sites was not noted. Table 2 shows the obvious utility of *Optioservus*. A relationship between *Lirceus*, *Gammarus*, and *Ephemerella* was noticed almost from the start of limestone stream IBI development, but you cannot make a good metric when components of the metric move in different directions in response to a stressor. However, *Optioservus, Gammarus,* and *Ephemerella* all decrease with anthropogenic impacts and *Optioservus* shows approximately the same response as *Ephemerella* from reference sites to impaired sites. The three taxa are 27.5 % of all the organisms collected and 48.6 % of the organisms collected at reference sites. The % Limestone Macros metric is simply the number of *Optioservus*, *Gammarus*, and *Ephemerella* individuals divided by the total number of organisms in the sub-sample. The metric has great potential, but at this time the biological significance of *Optioservus* is not known and more time is needed to evaluate this. Currently the metric may be useful to help analyze a sample site. The metric seems to respond fairly early in a stream recovery scenario.

The six metrics that were selected have good performance characteristics and DEs. However, there are four pairs of metrics that need to be analyzed for redundancy. Although there is no absolute threshold, r > 0.9 is generally not expectable (Paul and

Gerritsen 2005). The metric pairs with r > 0.85 were examined with scatter plots to determine if their relationships are nonlinear. If the scatter plot shows a curvilinear relationship, then both metrics may be retained because each one contributes information in a different part of the range (Paul and Gerritsen 2005).

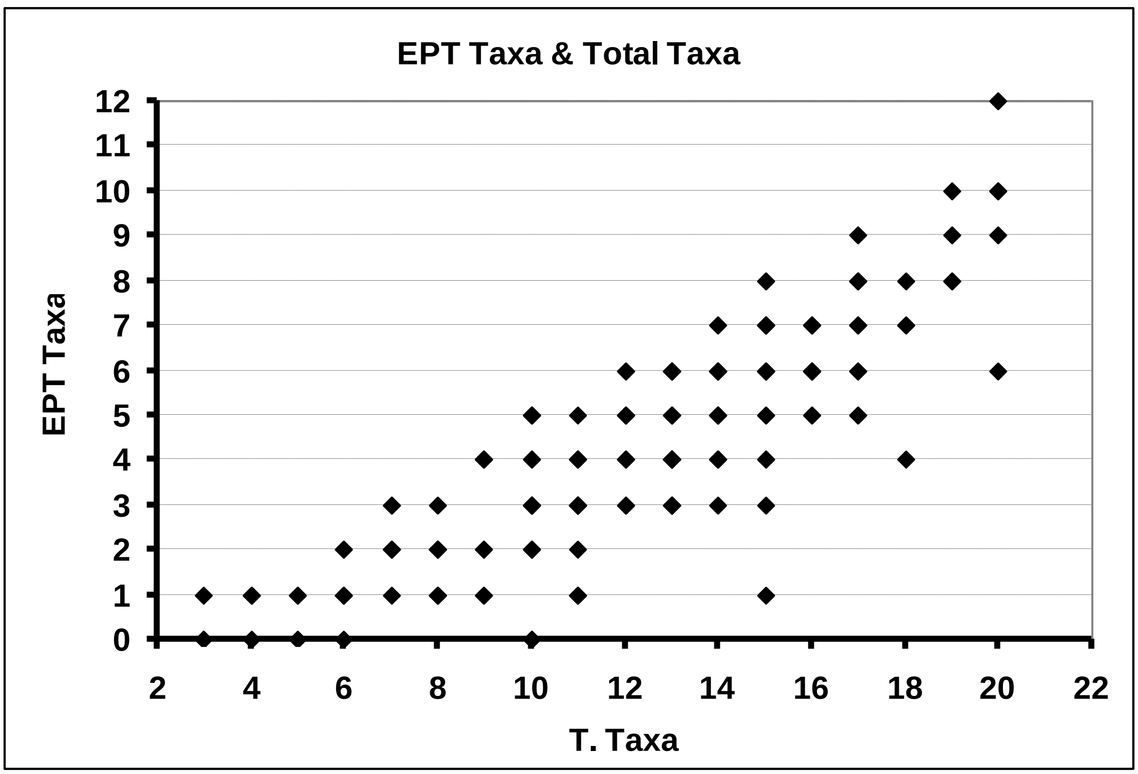
The metrics HBI and % Tolerant (TV $\geq$ 7) had a correlation of r = 0.89. Graph 1 shows that the relationship was strongly linear in the very impaired sites with pollution tolerant organisms greater than 60 % of the subsample, but nonlinear for sites that would qualify as attaining or reference. Each metric demonstrates its ability to contribute different information through the range of anthropogenic stress where it is most needed.

Graph 1. Scatter Plot of HBI and % Tolerant $\geq$ 7



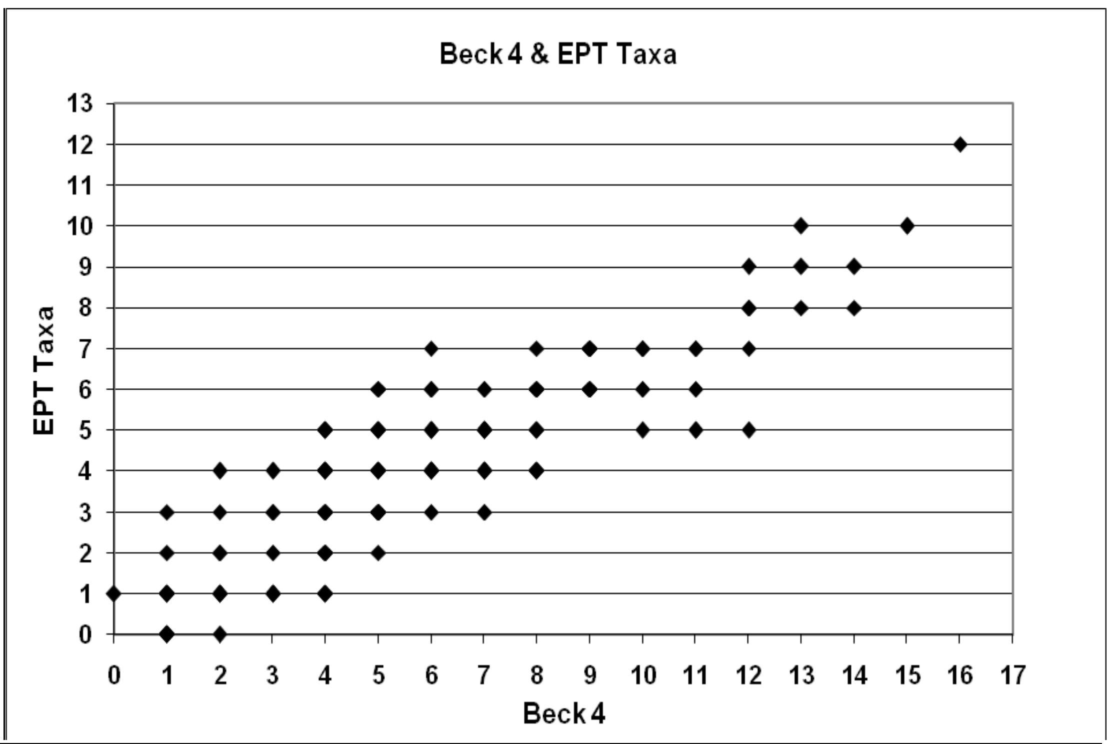
Total Taxa and EPT Taxa are both used almost universally in metric analysis. They are good metrics at assessing richness. As expected the metrics are correlated to each other with an r = 0.85. However, the scatter plot shows they complement each other. In some cases a sample with a fairly high number of taxa may have a low number of EPT Taxa and, more rarely, may have a high number of EPT Taxa and a lower number of Total Taxa. Total Taxa and EPT Taxa both move in the same direction with an increase or decrease in pollution, but they do not always move at same rate. This increases the discriminator power of the IBI in the midrange conditions. Total Taxa and EPT Taxa have good DE's and contribute independently to the IBI.

Graph 2. Scatter Plot of EPT Taxa and Total Taxa



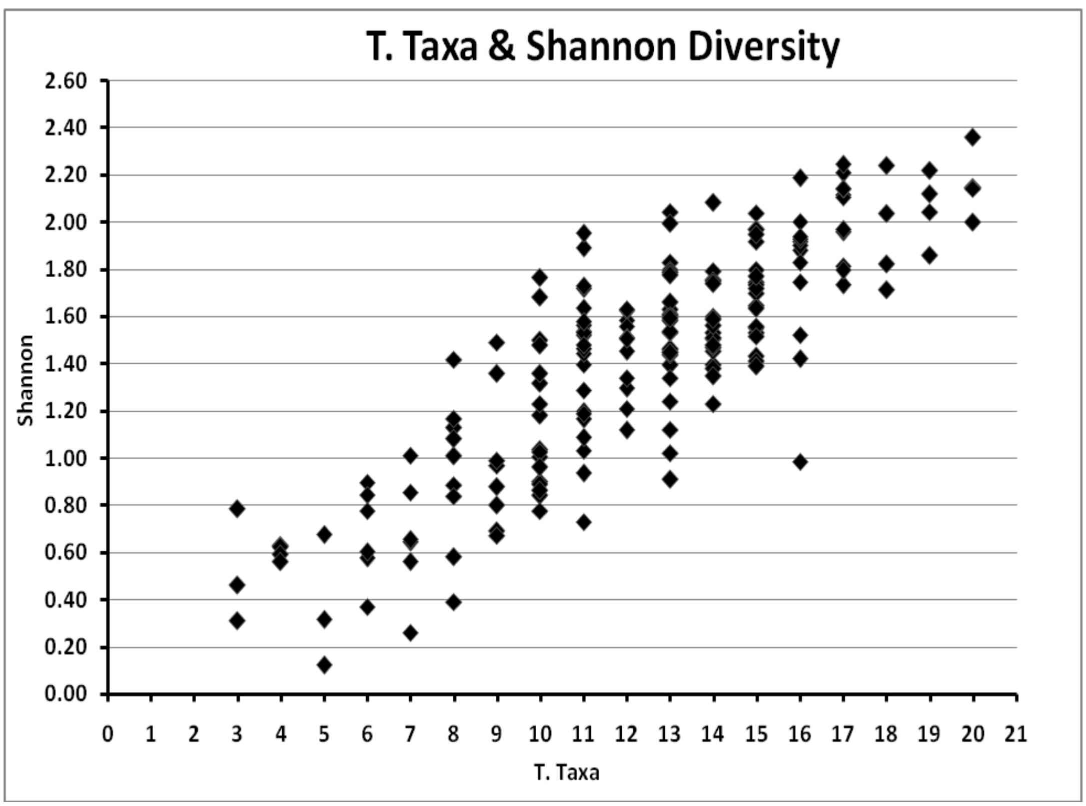
Beck's Index 4 is a new metric to the Limestone IBI that replaces % Intolerant (TV ≤ 3). The Beck's 4 metric is a pollution weighted taxa richness measure based on taxa with low tolerance values (TV). It is not limited to EPT taxa, but many taxa used in the Beck's 4 metric calculation are EPT taxa. It would be expected that the two metrics would be correlated (r = 0.89), but there are a few taxa that are collected in at least 50 % of the samples that have an important impact on each metric. There are two EPT taxa that have a tolerance values higher than 4, *Hydropsyche* and *Baetis,* and three non-EPT taxa that have a tolerance values less than or equal to 4, *Optioservus*, *Promoresia* and *Antocha*. There are also many taxa that are collected less often that will only fit one of the metrics. Graph 3 shows a good scatter of points indicating both metrics add unique sensitivity properties to the IBI at many sample sites.

Graph 3. Scatter Plot of Beck's 4 and EPT Taxa



Shannon Diversity and Total Taxa had correlation r = 0.85. Shannon Diversity uses total taxa as one of the major components in the diversity calculation so it is often correlated to Total Taxa. Graph 4 shows the points are scattered and each one contributes information sensitive to different community attributes. The metrics are also in different category measures; Shannon Diversity is a composition measure and Total Taxa is a richness measure.

Graph 4. Scatter Plot Total Taxa and Shannon Diversity



The "new" set of 6 metrics are listed in Table 6 with a brief definition. A more complete definition of each is found in Appendix C.

Table 6. The New 6 Metrics

Category	**Metric**	**DE**	**Definition**	**Response to Pollution**
Richness Measure	Total Taxa	87.5 %	Number of taxa in the subsample	Decreases
	EPT Taxa	85.4 %	Number of taxa in the orders Ephemeroptera, Plecoptera and Trichoptera	Decreases
Tolerance/ Intolerance Measures	Beck's Index 4	86.5 %	Modified Beck's Index giving taxa with a TV of 0 or 1 two points and TV of 2, 3, or 4 one point.	Decreases
	% Tolerant	90.6 %	Percent of organisms considered to be tolerant of pollution, TV $\geq$ 7	Increases
	HBI	94.8 %	The biotic index and abundance of each taxa are used to find a biotic index for the sample	Increases
Composition Measures	Shannon Diversity	84.4 %	Uses both taxa richness and abundance to measure general diversity and composition	Decreases

Best Values/Standardizing Scores

In a multimetric approach an aggregate index score is generated for each site. It is meaningless to add a Total Taxa score of 12 to a HBI score of 4.55. To facilitate calculating an index score that gives each metric approximately equal weight the raw metric scores must be standardized. To standardize raw metric scores a continuous scoring method is used. Each raw metric score is converted to a scale of 0 to 100 with 0 representing the worst ecological condition and 100 representing the best ecological condition. Metrics react to perturbation by either increasing in value or decreasing in value, so to standardize scores two different formulas were needed.

Metrics such as Total Taxa, EPT Taxa, Beck's Index 4 and Shannon Diversity decrease with greater impairment. The higher the score for these metrics the better the ecological condition, so the best raw metric score should equal 100. However, to reduce the impact of outliers on the best value the 95th percentile of all sample data was used. Metric scores between the minimum (usually 0) and the 95th percentile value (standard or best value) were scored proportionally from 0 to 100 based on the formula below (Gerritsen et al. 2000).

$$\text{Score} = (X/X_{95} - X_{min}) \times 100$$

Where:
- X = metric value
- X_{95} = 95th percentile value
- X_{min} = minimum possible value, usually 0

Metrics	Standard (best value) X_{95}	X_{min}	Standardization Formula
Total Taxa	18.0	0	Score = (X/18.0) x 100
EPT Taxa	8.0	0	Score = (X/8.0) x 100
Beck's 4	12.0	0	Score = (X/12.0) x 100
Shannon Diversity	2.13	0	Score = (X/2.13) x 100

Metrics such as % Tolerant and HBI increase with greater impairment. The lower the score for these metrics the better the ecological condition so the lowest metric score should equal 100. However, to reduce the impact of outliers on the best value the 5th percentile of all sample data was used. Metric scores between the maximum (100 %, percentage metrics, 10 TV) and the 5th percentile value (standard or best value) were scored proportionally from 0 to 100 based on the formula below (Gerritsen et al. 2000).

$$\text{Score} = (X_{max} - X/X_{max} - X_5) \times 100$$

Where:
- X = metric value
- X_5 = 5th percentile value
- X_{max} = maximum possible value, 100% for percentage metrics, 10 for TV

Metrics	Standard(best value) X_5	X_{max}	Standardization Formula
% Tolerant $\geq$ 7	1.5	100	Score = (100 - X/100 - 1.5) x 100
HBI	3.84	10	Score = (10 - X/10 - 3.84) x 100

The Best Values should be calculated from a continuum of scores from the worst ecological sites to the best ecological sites (Barbour et al. 1999) so sites from impaired, attaining and reference were used. However, there was concern with using replicate sample data. Using three replicate samples from the same date and location would overly weight those sites. For sampling events where replicate samples were collected only replicate one was used to calculate Best Values. The Best Values were calculated from 146 samples out of 188 total samples. Table 7 below is an example of standardizing a sample's raw metric scores and generating an IBI score. The final IBI score is the average of the six metric scores. Also note that the maximum score for a metric is 100.

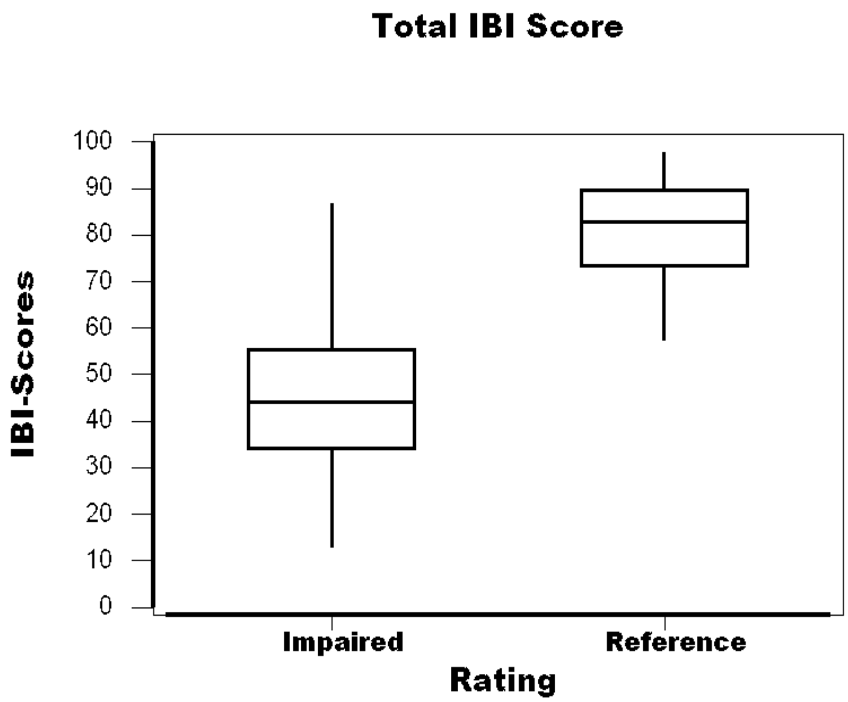
Table 7. Sample from Letort Spring Run

Metrics	Raw Metric Score	Standardization Formula	Standardized Score
% Tolerant	10.0	Score = (100 - **10.0**/100 - 1.5) x 100	91
HBI	4.33	Score = (10 - **4.33**/10 - 3.84) x 100	92
Total Taxa	13	Score = (**13.0**/18.0) x 100	72
EPT Taxa	6	Score = (**6.0**/8.0) x 100	75
Beck's 4	15	Score = (**15.0**/12.0) x 100	100
Shannon Diversity	2.02	Score = (**2.02**/2.13) x 100	95
IBI Score			87.5

The Limestone IBI Protocol Performance Characteristics

Performance characteristic data were presented previously to analyze individual metrics. Now the performance of the IBI protocol will be analyzed. A standard DE analysis was calculated for total IBI scores. Using the 25th percentile (73.1) of reference of sites as a threshold, 94 impaired sites out of a total of 96 sites were correctly identified for a DE of 97.5.

Graph 5. Box Plot of the Total IBI Scores Reference and Impaired DE 97.5



Precision Estimate of Intra-Site Variability

The precision of the method should be evaluated for a continuum of ecological conditions so samples were collected from reference, non-reference and impaired sites. The precision estimate (PE) for intra-site variability was calculated from replicate samples (63 sample pairs). The precision estimate was found by first calculating the variance for each set of replicate samples. Then the average variance of all the replicates was calculated and the square root of the average variance was calculated to find the standard deviation. The standard deviation was multiplied by the 1-tailed Z value of 1.28 to find the precision estimate, or the 90 % confidence interval. Table 8 shows the PE for the total IBI scores was very good at 7.27 and the CV or "relative variability" was only 9.75 %. The 2006 Limestone IBI had a PE of 8.30. The low PE indicates the Limestone IBI Protocol can produce repeatable sample results.

Table 8. Precision Estimates and Coefficient of Variation for Metrics and IBI Scores

Metric	Average Variance or RMSE	Standard Deviation	P.E. 90% CI +/-	Approx. Mean	Approx. Coefficient of Variation (%)
Total Taxa	2.00	1.414	1.81	12.59	11.24
EPT Taxa	1.56	1.247	1.60	4.87	25.59
Beck's Index 4	2.25	1.501	1.92	6.73	22.31
HBI	0.103	0.321	0.544	6.12	5.24
% Tolerant $\geq$7	58.75	7.665	9.81	42.89	17.87
Shannon Diversity	0.037	0.192	0.246	1.418	13.56
Total IBI Score	**32.25**	**5.68**	**7.27**	**58.24**	**9.75**

Temporal Precision Estimate

There were 187 temporally paired samples with some sites sampled over a 10 year period. Some examples of temporal data are 3 sites sampled 7 times over 10 years, 6 sites sampled 5 or 6 times over 7 years and 9 sites sampled 4 times over 4 years. This is temporal data that should show temporal precision close to the extreme extent of natural variability. Table 9 shows a precision estimate of 11.43. This is a little high, but is fairly good considering the number of anthropogenic changes and natural catastrophic events that could occur over a 7 to 10 year period. Samples from reference, attaining and impaired sites were included in this analysis. The estimate of IBI precision incorporates natural intra-site spatial variability and methodological variability, however, temporal intra-site spatial variability includes possible site selection from one riffle to another riffle in the area and changes in field and laboratory staff in 7 to 10 years. The protocol must be very resilient to maintain a PE close to the expected natural variability.

Table 9. Temporal Precision

Metric	Average Variance or RMSE	Standard Deviation	P.E. 90% CI +/-	Approx. Mean	Approx. Coefficient of Variation (%)
Total Score	79.67	8.93	11.43	61.91	14.42

Threshold for Impairment

Limestone streams represent a small subset of Pennsylvania streams, so to qualify as a limestone stream means a stream is already quite unique. From this small group of streams reference sites were carefully selected. All the reference streams are designated as high quality (HQ-CWF) or exceptional value (EV-CWF) in PA Chapter 93 and there was extensive knowledge of any possible current and historic anthropogenic impacts at the sites. The intensive reference qualification process produced a high quality reference condition. Sample sites that generate an IBI score greater than the 1st quartile (73.1) of reference site index scores should be reference sites and qualify for Special Protection under PA Chapter 93. The threshold for limestone streams was developed based on the examples presented in the EPA publication, "Best Practices for Identifying Reference Condition in Mid-Atlantic Streams." (U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC 20460, August 2006.)" Other Region 3, Mid-Atlantic States such as Virginia, Maryland, and West Virginia used similar methods. The West Virginia Condition Index (West Virginia; Gerritsen et al. 2000; West Virginia, DEP web site) was an important model for the development of the Limestone Streams Protocol.

The method used to determine a threshold for impairment must take the quality of the reference condition into account (Barbour et al. 1995). In the Limestone Stream IBI development the expectation was that all the reference samples were high quality and unimpaired. When reference samples of high quality are used the 5th percentile is an appropriate starting point. The impairment threshold was determined by calculating the 5th percentile of reference samples (59.5).

Since the introduction of the Limestone Protocol, EPA has published, "Best Practices for Identifying Reference Condition in Mid-Atlantic Streams" (EPA 2006). MAIA established two benchmarks based on the distribution of reference sites. The 25th percentile value set the lower limit for "good" condition. The 1st percentile was used as the threshold below which values were deemed "poor." Values between the 1st and 25th percentiles were designated as "marginal." The Limestone IBI method to select an impairment threshold is in the range of EPA's methods to determine "poor, marginal and good" and the EPA methods did not require their reference sites to be Special Protection. Pennsylvania does not use the terms "poor, marginal and good"; Pennsylvania uses impaired, attaining and special protection.

Table 10.

IBI Scoring Thresholds

		Attaining	Impaired CWF	
Classification	**Reference**	**CWF**	**Moderately**	**Severely**
IBI Score	**>73**	**73 - 60**	**<60 - 30**	**<30**

Less Than <60 is impaired

The ability of the protocol to correctly identify impaired sites was measured by calculating the discrimination efficiency (DE) using the 5th percentile of reference as the threshold. Sample sites included in the impaired group were sites previously listed on the 303d list due to criteria independent of this protocol. There were 96 samples in the impaired group and 82 scored below the impairment threshold of < 60, for a DE of 85.4 %. The impairment threshold agreed with an independent impairment criterion a very high percentage of the time. The DE is lower because there have been remediation at some impaired site due to use of the IBI and improved NPDES discharges.

The DE formula: DE = 100 x a/b

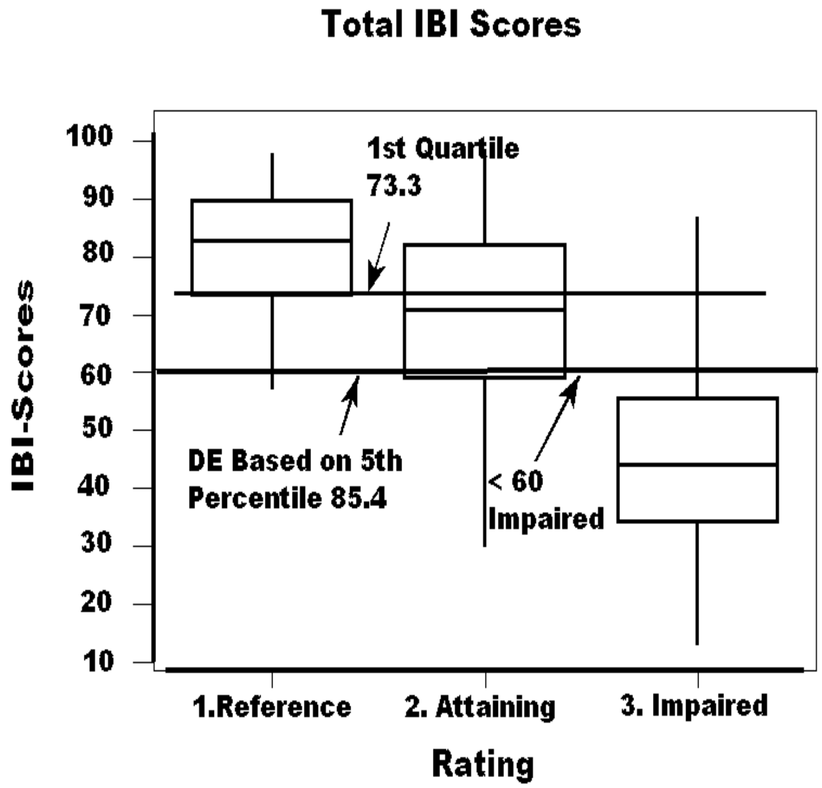
Where:

a = the number of impaired sites scoring above the impairment threshold of < 60 (n=82)

b = the total number of impaired samples (n=96)

DE = 85.4

Graph 6. PA DEP Threshold for Impairment



Conclusion

Several years ago the Department (DEP) needed to assess a limestone stream. It was recognized that limestone streams had different macroinvertebrate communities than other stream types in Pennsylvania, but there was no good understanding of what a high quality limestone stream macroinvertebrate community looked like. Over the next eight years potential limestone stream sites were investigated. Most of the areas with limestone geology in Pennsylvania were investigated for possible limestone stream sites. Criteria capable of properly classifying limestone streams were developed. Many streams were eliminated because they did not meet the criteria for limestone streams. Sample collection was expanded to include a continuum of ecological conditions. Replicate samples were needed to test the accuracy and repeatability of the method. Years of sampling and continuing development have generated an excellent assessment tool. The classification criteria will identify "true" limestone streams. The selected metrics are tailor-made to assess the unique macroinvertebrate communities of limestone streams. The continuous scoring method provides a good way to generate an aggregate score that is easy to understand and compare to other sites. The performance of the method was tested and showed accuracy and repeatability in accrual assessments to determine stream quality at NPDES discharges. The well-established regional reference condition data provides documentation to set a threshold for impairment and in the future can be used to establish a benchmark for Special Protection or an antidegradation threshold.

LITERATURE CITED

Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M.L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *J. N. Am. Benthol. Soc*. 15(2):185-211.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. Second Edition. EPA/841-B-99-002. U.S. EPA, Office of Water, Washington, D.C.

Barbour, M.T., J.B. Stribling, and J.R. Karr. 1995. The multimetric approach for establishing biocriteria and measuring biological condition. Pp. 63-76. In W.S. Davis and T.P. Simon, editors. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Ann Arbor, Michigan.

Flotemersch, J. E., J. B. Stribling, and M. J. Paul. 2006**.** Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers**.** EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.

Gerritsen, J. J. Burton and M. T. Barbour. 2000. A stream condition index for West Virginia wadeable streams. Prepared for USEPA Office of Water and USEPA Region 3. EPA-822-B00-001. U.S. EPA, Office of Water, Washington, D.C.

Gibson, G.A., M.T. Barbour, J.B. Stribling, J. Gerritsen, and J.R. Karr. 1996. Biological criteria: Technical guidance for streams and rivers. EPA/822-B-94-001. U.S. Environmental Protection Agency (US EPA), Office of Science and Technology, Washington, D.C.

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

Paul, M. J., J. Gerritsen. 2005. Unpublished. Draft Statistical Guidance for Developing Indicators for Rivers and Streams A Guide for Constructing Multimetric and Multivariate Predictive Bioassessment Models. Prepared for US EPA, Office of Research and Development and Office of Water.

USEPA (US Environmental Protection Agency).2006. Best Practices for Identifying Reference Condition in Mid-Atlantic Streams. EPA-260-F-06-002. Office of Environmental Information, Washington, DC 20460

APPENDIX A

Field Sampling and Laboratory Sample Processing

Field Sampling and Laboratory Sampler Processing

Net Mesh Considerations

All limestone stream samples collected for the development of this document used net mesh in the 800-900µ range. In recent years, many state water quality programs, federal agencies (e.g. EPA, USGS), and other water quality monitoring organizations began using net sampling devices with 500µ mesh nets. Field sampling comparisons have shown that the 500µ mesh size blocked quickly preventing macroinvertebrates and vegetation from entering the net resulting in a poor sample. In order to insure an accurate assessment 800-900µ net mesh must be used to collect samples.

D-Frame Net

The handheld D-frame sampler consists of a bag net attached to a half-circle ("D" shaped) frame that is 1 ft. wide. The net is employed by one person facing downstream and holding the net firmly on the stream bottom. One "**d-frame effort**" is defined as such: the investigator vigorously kicks an approximate area of $1m^2$ (1 x 1 m) immediately upstream of the net to a depth of 10cm (or approximately 4", as the embeddedness of the substrate will allow) for approximately one minute. All benthic dislodgement and substrate scrubbing should be done by kicks only. Substrate handling should be limited to only moving large rocks or debris (as needed) with no hand washing. Since the width of the kick area is wider than the net opening, net placement is critical in order to assure all kicked material flows toward the net. Avoiding areas with crosscurrents, the substrate material from within the 1 m^2 area should be kicked toward the center of the square meter area.

Semi-Quantitative Method (PaDEP-RBP):

In Plafkin (1989), EPA presented field-sampling methods designed to assess impacts normally associated with pollution impacts, cause/effect issues, and other water quality degradation problems in a relatively rapid manner. These are referred to as Rapid Bioassessment Protocols (RBPs). The PaDEP-RBP method is a bioassessment technique involving systematic field collection and subsequent lab analysis to allow detection of benthic community differences between reference (or control) waters and waters under evaluation. The PaDEP-RBP is a modification of the EPA RBP III (Plafkin, et al; 1989); designed to be compatible with Pennsylvania's historical database. Modifications include: 1) the use of a D-frame net for the collection of the riffle/run samples, 2) different laboratory sorting procedures, 3) elimination of the CPOM (coarse particulate organic matter) sampling, and 4) metrics substitutions. Unlike the EPA's RBP III methodology, no field sorting is done. Only larger rocks, detritus, and other debris are rinsed and removed while in the field before the sample is preserved. While EPA's RBP III method was designed to compare impacted waters to reference conditions (cause/effect approach), the PaDEP-RBP modifications were designed for un-impacted waters, as well as impacted waters.

Sample Collection

The purpose of the standardized PaDEP-RBP collection procedure is to obtain representative macroinvertebrate fauna samples from comparable stations. The PaDEP-RBP assumes the riffle/run habitat to be the most productive habitat. Riffle/run habitats are sampled using the D-frame net method described above. For limestone stream surveys, two paired D-frame efforts are collected from each station - one from an area of fast current velocity and one from an area of slower current velocity within the same riffle. Limestone streams have low gradient often making it difficult to locate well developed riffles. If there are no riffle in the sample area use a run or the best rock substrate available. The resulting "D-frame efforts" (two) are composited into one sample jar (or more as necessary). Care must be taken to minimize "wear and tear" on the collected organisms when compositing the materials. It is recommended that the benthic material be placed in a bucket and filled with water to facilitate gentle stirring and mixing. The sample is preserved in ethanol (95%) and returned to the lab for processing.

Sample Collection Period

Samples must be collected from January through May. All samples used to develop this IBI were collected in this time period. Limestone streams have a low number of sensitive taxa and only a few of these taxa are generally found larger numbers. One very important sensitive taxon is *Ephemerella*. A good population of *Ephemerella* generally indicates better water quality. The three species of *Ephemerella*: *invaria, rotunda* and *dorothea* found in limestone streams emerge in May and June and are normal difficult or impossible to collect from June through December. Collecting samples from January through May ensures this very important ecological indicator taxa will not be missed.

Sample Processing

Samples collected with a D-frame net are generally considered to be qualitative. However, the preserved samples can be processed in a manner which yields data that is "semi-quantitative" - data that was collected by qualitative methods but gives information that is almost statistically as strong as that collected by quantitative methods.

The following procedure is adapted from EPA 1999 RBP methodology and used to process qualitative D-frame samples so that the resulting data can be analyzed using benthic macroinvertebrate biometric indices (or "metrics"). Equipment needed for the benthic sample processing are:

- 2 large laboratory pans gridded into 28 squares (more gridded pans may be necessary depending on the size of the sample). White polyethylene pans 18"L x 12"W x 3.5"D were used, but any similarly sized pan with 28 equal grids may be used.
- Illuminated magnifying viewer. (optional)
- Slips of paper (numbered from 1 to 28) for drawing random numbers, and
- Forceps (or any tools that can be used to pick floating benthic organisms),
- Grid cutters made from tubular material that approximates an inside area of 4 in^2.

The targeted sub-sample size is 300 for Limestone surveys (± 20%), (240 to 360 organisms). Samples must be properly prepared for sub-sampling. Macroinvertebrates tend to clump so the sample should be mixed in the sample container or the sub-sample pan to make it as homogenous as possible. If necessary the sample maybe mixed in a bucket prior to being placed in the pan. In order to further reduce the effect of clumping a two-tiered sub-sampling technique is employed. A minimum of 4 grids must be selected from the first pan.

Tier 1 – Rinse the sample in a standard USGS No. 35 sieve to remove fine materials and residual preservative. During the rinse larger rocks, sticks, and leaves maybe removed making sure to retain all the macroinvertebrates. Place the sample in a 28-square gridded pan (Pan1) and add enough water to distribute the sample evenly. Randomly select 4 grids using the 28 random number set and, using the grid cutters, remove the debris and organisms entirely from within the grid cutter and place in a second gridded pan (Pan2). Selecting a minimum of 4 grids reduces the effect of clumping. Do a visual scan of Pan2 to ensure that there are enough identifiable (this excludes pupae, extremely small instar larvae, and empty shells or cases) organisms to reach the targeted sub-sample size (300 +/- 20%). If there do not appear to be enough organisms randomly select additional grids until there appears there are a minimum of 300 +/- 20% organisms.
Note: In limestone streams we have never needed more than 4 grids.

Tier 2 –Randomly select grids from pan2 removing all the organisms from each grid until there is a sub-sample of 300 +/- 20%. If it appears that the number of benthic organisms from the last grid will cause the sub-sample to exceed it's target size by more than 20% (>360 organisms), count them and place in a clean gridded pan (Pan3) with enough water to facilitate gentle stirring and even distribution. Randomly select grids from Pan3 and remove individuals until the count of organisms remaining in Pan3 falls within the +20% upper limit.

Comments:

1. If the sample is too large to fit in pan 1 evenly divide sample into 2 or more pans. Randomly select a minimum 4 grids from each pan and place them in a pan.
2. The benthic material remaining after the target sub-sample has been picked can be returned to its original sample jar and preserved. They shall be retained in accordance with QA retention times as specified for this respective survey type.
3. Any grid chosen must be picked in its entirety.

Identification, Taxonomic Level

The level of identification for most aquatic macroinvertebrates will be to genus. Some individuals collected will be immature and not exhibit the characteristics necessary for confident identification. If an individual cannot be confidently identified to the proper level it should be discarded. All pupae are discarded. Certain groups are identified to a higher taxonomic level as follows:

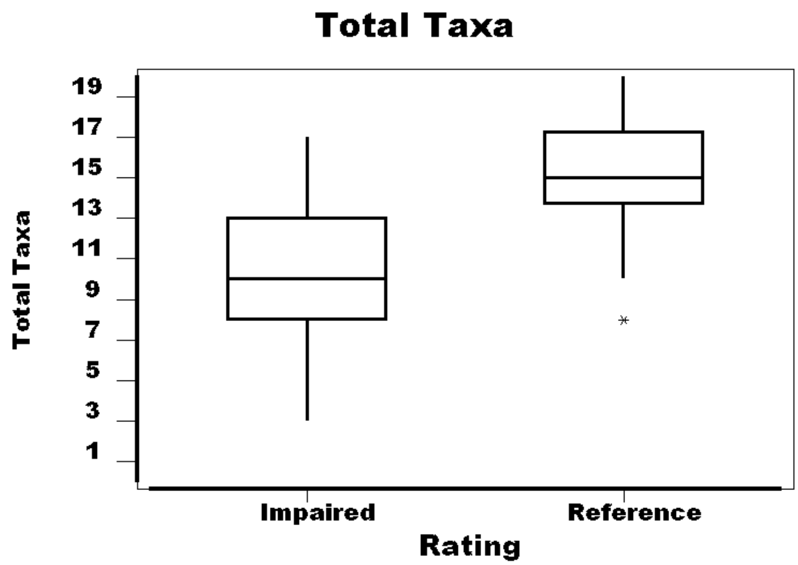
Flatworms (Turbellaria) - Phylum Turbellaria
Segmented worms (Annelida) aquatic earthworms & tubificids - Class Oligochaeta
Proboscis worms – Phylum Nemertea
Roundworms - Phylum Nematoda
Water mites - "Hydracarina" (an artificial taxonomic grouping of several mite superfamilies)
Midges – Family Chironimadae

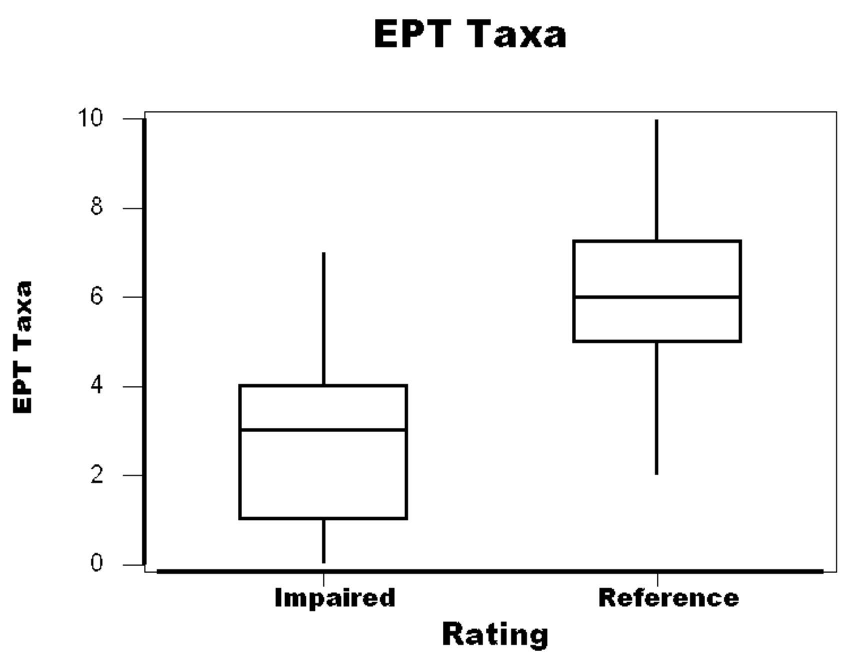
APPENDIX B

Supporting Graphs

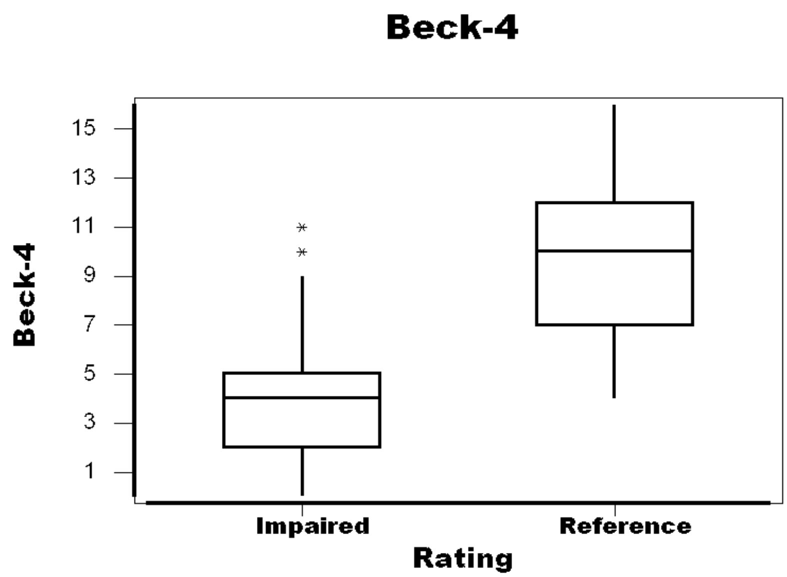
Box & Whisker Plots of DEs

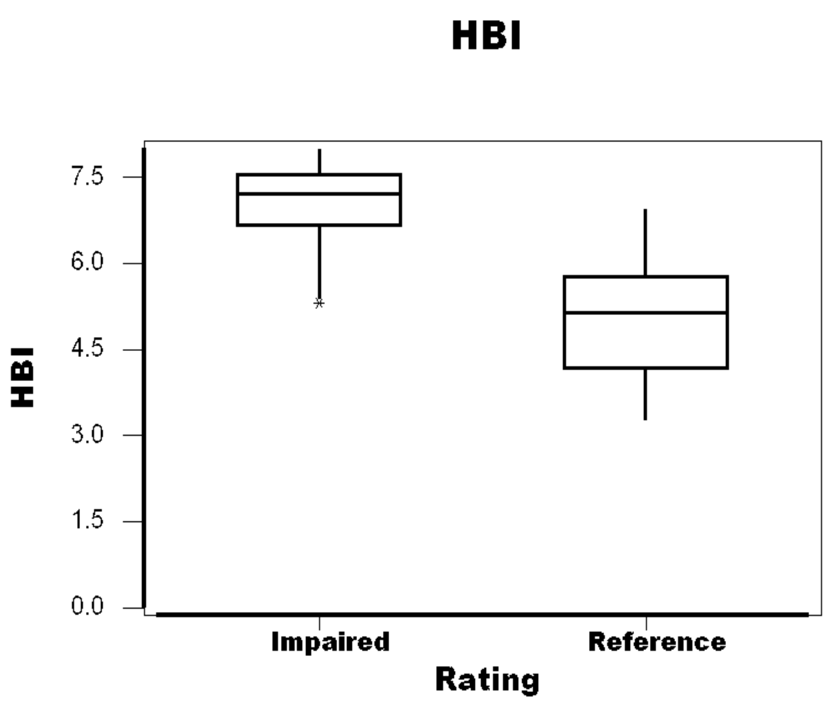
Box & Whisker Plots Total Taxa DE = 87.5 and EPT taxa DE = 85.4



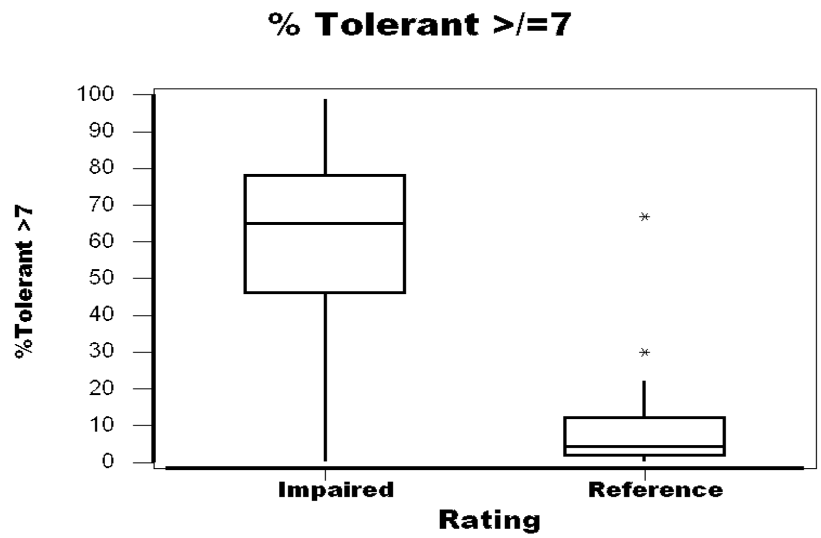


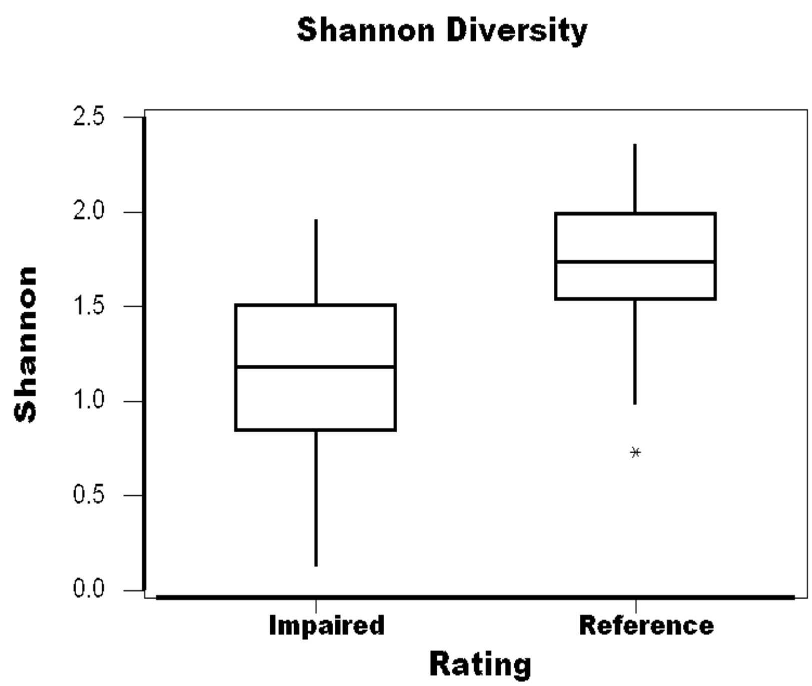
Box & Whisker Plots Beck's Index 4 DE = 86.5 and HBI DE = 94.8





Box & Whisker Plots % Tolerant ≥ 7 DE = 90.6 and Shannon Diversity DE = 84.4





APPENDIX C

Definition of Metrics

The Six Metrics Selected for the Limestone IBI

Total Taxa
Discrimination Efficiency = 87. 5 %

This richness measure is a count of the total number of taxa in a sub-sample. This metric is expected to decrease with increasing anthropogenic stress to a stream ecosystem, reflecting loss of taxa and increasing dominance of a few pollution-tolerant taxa. It is commonly used in biological monitoring and assessment programs.

EPT Taxa Richness
Discrimination Efficiency = 85.4 %

This richness measure is a count of the number of taxa belonging to the orders Ephemeroptera, Plecoptera, and Trichoptera (EPT) in a sub-sample. Common names for these orders are mayflies, stoneflies, and caddisflies, respectively. The aquatic life stages of these three insect orders are generally considered sensitive to pollution. This metric is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting the loss of taxa from these largely pollution-sensitive orders. In addition to discriminating between reference and stressed sites well, EPT has a history of use in biological monitoring and assessment programs.

Beck's Index version 4
Discrimination Efficiency = 86.5 %

Beck's Index 4 is an intolerant/tolerant measure which "weights" the count of pollution sensitive taxa. The number of taxa with a tolerance value of 0 or 1 multiplied by two, taxa with a tolerance value of 2, 3 or 4 are multiplied by one and the values are totaled. This metric is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting the loss sensitive taxa with a higher proportion loss placed on the most sensitive taxa.

HBI - Hilsenhoff Biotic Index
Discrimination Efficiency = 94.8%

This intolerant/tolerant measure is calculated as an average pollution tolerance value weighted by the number of individuals of each taxa in the subsample. The Hilsenhoff Biotic Index generally increases with increasing ecosystem stress. Although the HBI generally only measures the impacts of organic and nutrient pollution it is almost always a good metric and it is a major component of many of other useful metrics.

$$= \sum [(i * n_{indvTVi})] / N$$

where $n_{indvPTVi}$ = the number of individuals in a sub-sample with pollution tolerance value (TV) of i and N = the total number of individuals in a subsample

<u>% Tolerant Individuals </u>(tv <u>></u> 7)
Discrimination Efficiency = 88.7%

This intolerant/tolerant measure is the percentage of individuals with pollution tolerance values of 7 or greater in a subsample and is expected to increase in value with increasing anthropogenic stress to a stream ecosystem. Percent Tolerant Individuals tend to be *Lirceus* which are 36.2 % of the total organisms collected in the 188 samples.

<u>Shannon Diversity Index</u>
Discrimination Efficiency = 84.4 %

This composition measures calculates a value based on taxonomic richness and evenness of individuals in a subsample. This metric is expected to decrease in values with increasing anthropogenic stress to a stream ecosystem, reflecting loss of pollution-sensitive taxa and increasing dominance of a few pollution-tolerant taxa.

$$= -\sum (n_i / N) \ln (n_i / N)$$

where n_i = the number of individuals in each taxa (relative abundance); N = the total number of individuals in a subsample; and Rich = the total number of taxa in a sub-sample (total taxa richness)

This is a new formula for Shannon Diversity. The formula was changed to match the Shannon Diversity formula used in DEP's other IBIs. It will not impact previous assessments because all values change equally. The old formula is below:

Shannon-Wiener Diversity

$$H'=C/N[NLog_{10}N-Sum(n_iLog_{10}n_i)]$$

H'=Diversity C=3.321928

n_i=Total number of individuals in the i^{th} taxa
N= Total number of individuals