Proposed Methods for Selecting and Scoring Metrics for Chessie BIBI Index

DRAFT 1/13/2016

A subset of the Chessie BIBI database (n > 24,000 records) will be created for the purpose of refining the index. The subset consists of samples that occur in streams with a Strahler order ≤ 4, are collected with a kick-net, and include all of the in-stream and water quality data necessary to categorize the sample as representing a Reference, Mixed or Degraded site. These data are classified into six bioregions (i.e., Coastal Plain, Northern Appalachian Plateau and Uplands, North Central Appalachians, Piedmont, Ridges, and Valleys), which are based on EPA Level IV Ecoregions. Reference, Mixed, and Degraded categories are then assigned to each sampling event using site classification criteria, some of which are bioregion-specific.

Discrimination efficiency

We will apply the following approach on this data subset to test the ability of more than 50 macroinvertebrate metrics to differentiate (discriminate) between Reference and Degraded conditions. The biological metrics are classified into five types: diversity, composition, tolerance, functional feeding groups, and habit. The objective is to include at least one highly discriminatory metric from each metric type in the index. Including several types of metrics in an index is assumed to make the index sensitive to different forms of degradation. Tolerance metrics, for example, typically indicate nutrient pollution and an index that contains only tolerance metrics could potentially overlook degradation unrelated to nutrient levels, such as habitat alteration or elevated specific conductivity.

Discrimination efficiency (DE) is used to characterize how well a metric can discriminate between Reference and Degraded conditions. These two conditions represent the ends of our defined range of best and poor conditions. Percentiles of a metric's distributions in these conditions are used to establish a standard scale for scoring the metric. The Mixed condition is intermediate to Reference and Degraded conditions. Its metric values broadly overlap Reference and Degraded distributions, and the Mixed condition provides no clear thresholds delineating the best and poorest conditions. Thus, the Mixed category is not used in establishing the standard scoring scales.

The DE statistic is the frequency of correctly classified Reference and Degraded sites for a given metric value, or threshold (Equation 1).

Equation 1. $$DE = \frac{\%Ref + \%Deg}{2}$$

where %Ref is the percentage of Reference sites correctly identified at the given threshold, and %Deg is the percentage of Degraded sites correctly identified at the given threshold. High DE's indicate good separation.

We employ a simple technique for testing and finding the threshold associated with each metric's highest DE. For each metric, a table is created representing the Reference population's percentile values. If a metric decreases with disturbance, the 0-50 percentiles of the metric's

Reference values (whole numbers) are tested as thresholds. If a metric increases with disturbance, the 50-100 percentiles of the Reference values are tested as thresholds. The %Ref and %Deg are calculated for each percentile tested and the metric value corresponding to the highest DE is identified as the optimal discrimination threshold.

It is possible for multiple thresholds to return the same high DE value. When two or more thresholds produce identical, high DE's, Equation 2 is used to select the final discrimination threshold.

Equation 2.

$$Diff = |\%Ref - \%Deg|$$

The threshold with the smallest Diff value is considered more discriminatory and is selected as the optimal discrimination threshold. Equivalent DE values indicate the same total percentage of samples are correctly identified in all instances but the distribution of correctly identified samples in each data pool differs. Therefore, the more balanced distribution is favored over a distribution skewed towards a particular data pool (i.e., Reference or Degraded).

Metric redundancy

Pearson correlation is used to assess metric redundancy within each metric type. Including two significantly correlated metrics ($r \geq 0.70$ or ≤ -0.70) in the final index would be comparable to doubling the value of a single metric within the index. Pairwise comparisons of the DE's of redundant metrics are done to select the metric with the better discriminatory power (higher DE). The metrics were then divided into one of the five metric types and arranged in descending order based on the optimal DE value. An effort was made to select a metric from each metric type to include in the index. However, if all metrics from a metric type have low (e.g., <75%) DE values no metric is selected. The metric with the largest optimal DE value from each metric type is included in the final index. Other non-redundant metrics with optimal DE value ≥ 75% are also considered for inclusion in the final index.

Metric scoring

Buchanan *et al.* (2011) tested four procedures for scoring metrics. The authors concluded that the following binning approach provided the most accurate metric scores. The optimal DE value is used as the threshold to separate the poor condition from the intermediate condition (Bin 1 and 2, respectively, in Figure 1). If the metric decreased with disturbance, values above the reference median are categorized as the best condition (Bin 3 in Figure 1). If the metric increased with disturbance, values below the reference median are categorized in Bin 3. If a station's raw metric score falls within the range of Bin 1 or Bin 3, the station receives a score of 0% or 100% respectively. Stations with raw metric scores in Bin 2 are scored using Equation 3, if the metric decreases with disturbance, or Equation 4, if the metric increases with disturbance. Bin 2 scores range between 1-99% ($0 < x < 100$).

Equation 3. $$Score = \frac{X - X_T}{X_M - X_T} * 100$$

X = the raw metric score

X_T = the optimal discrimination threshold

X_M = the reference median value

Equation 4. $$Score = \frac{X_T - X}{X_T - X_M} * 100$$

X = the raw metric score

X_T = the optimal discrimination threshold

X_M = the reference median value

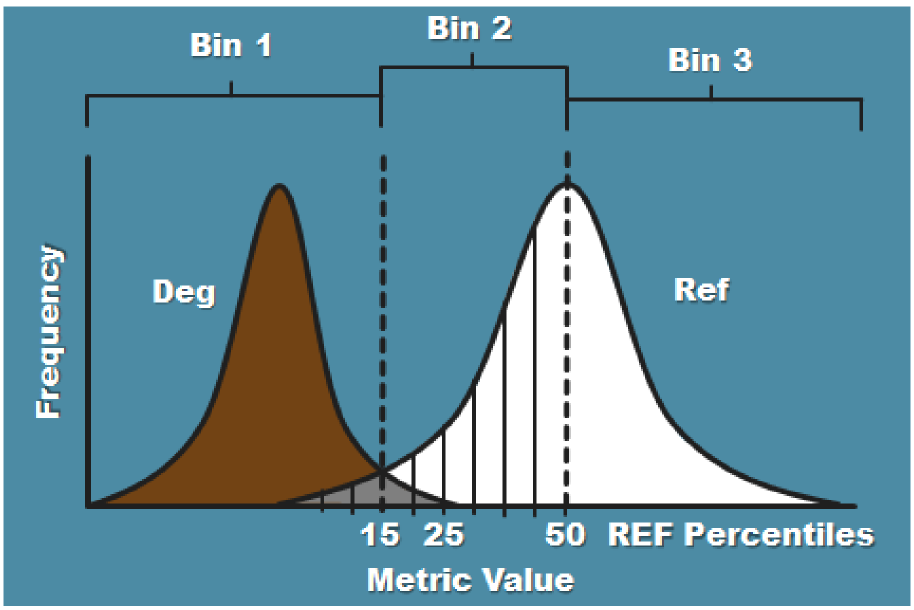

Figure 1. Illustration of bins established for scoring a metric that decreases in value with degradation. Dashed line separating bins 1 and 2 indicates the metric value (threshold) associated with optimal discrimination efficiency (DE) between Reference (Ref) and Degraded (Deg) conditions. Dashed line separating bins 2 and 3 indicates the Ref 50th%ile.