A Multimetric Biological Condition Index for Rhode Island Streams

Prepared for:

The Rhode Island Department of Environmental Management
Office of Water Resources

and

The U.S. Environmental Protection Agency
Region 1 (New England)

Prepared by:

Tetra Tech, Inc. – The Center for Ecological Science

Final Report

March, 2012

State of Rhode Island Department of Environmental Management
Office of Water Resources

United States Environmental Protection Agency

TETRA TECH, INC.

Executive Summary

The Rhode Island Department of Environmental Management (RIDEM) Office of Water Resources and U.S. EPA Region 1 are developing macroinvertebrate sampling, criteria and assessment methods to incorporate bioassessment information into RIDEM Office of Water Resources programs for managing aquatic resources in streams. The Biological Condition Gradient provides a consistent scale for assessing biological conditions throughout the United States. With this study, regional BCG models and interpretations were used to refine assessment tools specific to Rhode Island.

The preferred biological assessment tool for streams is the Multi-metric Index (MMI), which uses a combination of measures from benthic macroinvertebrate samples, calibrated to natural and stressor gradients, to indicate biological conditions relative to the highest quality samples observable throughout the State. The MMI was developed to replace current bioassessment methods, which were based on a single reference station for each assessment. The index development and calibration process involves identification of the least and most disturbed sites in the monitoring dataset (reference site designation), investigation of variability among minimally disturbed sites (site classification), metric testing for sensitivity to stress, and index compilation.

The reference site designations of the sites were determined using criteria related to land use and water quality characteristics in the sites and their catchments. This was an objective scale of site condition to which metrics were tested for sensitivity to stressors. The least disturbed sites were essential for examining possible site classes because the stressor effects in these sites were less influential on sample composition and natural differences could be discerned.

In the site classification exercise, we found that samples appeared to be similar among undisturbed sites in the higher gradient Coastal Plains and Hills area of the state. Southern coastal and eastern sites were apparently different either because the natural habitat differed between the regions or because sites had different disturbance levels from more intense land uses. Therefore, the index was developed only for the higher gradient sites in the Coastal Plains and Hills where there was a sufficient number of samples available.

Sixty eight (68) metrics were tested for responsiveness along the stressor gradient. Of these, 17 strongly responded to stressors and were considered for inclusion in the index. Along the same stressor gradient, metric combinations were tested that included non-redundant metrics with diverse types of information. The best performing and most ecologically meaningful combination was selected as the Macroinvertebrate Biological Condition Index (MBCI). It included six metrics from all five metric categories, had a discrimination efficiency of 100%, and

a *Z*-score of 3.0. The metrics in the MBCI are Total Taxa, % Non-insect, Beck's Index, Clinger Taxa, % Predators, and % Filterers.

The Connecticut (CT) Biological Condition Gradient (BCG) model was calculated for the same set of Rhode Island calibration samples and the indicators were compared to each other and to the stressor gradient. If the BCG model for CT is valid in Rhode Island streams, then the multimetric index can be used to estimate biological condition on the standard BCG scale. Applicability of the CT BCG model should be further investigated, but the high correlation among MBCI and BCG indicators (Figure ES-1) suggests that establishing condition thresholds for bioassessment based on MBCI scores is defensible.

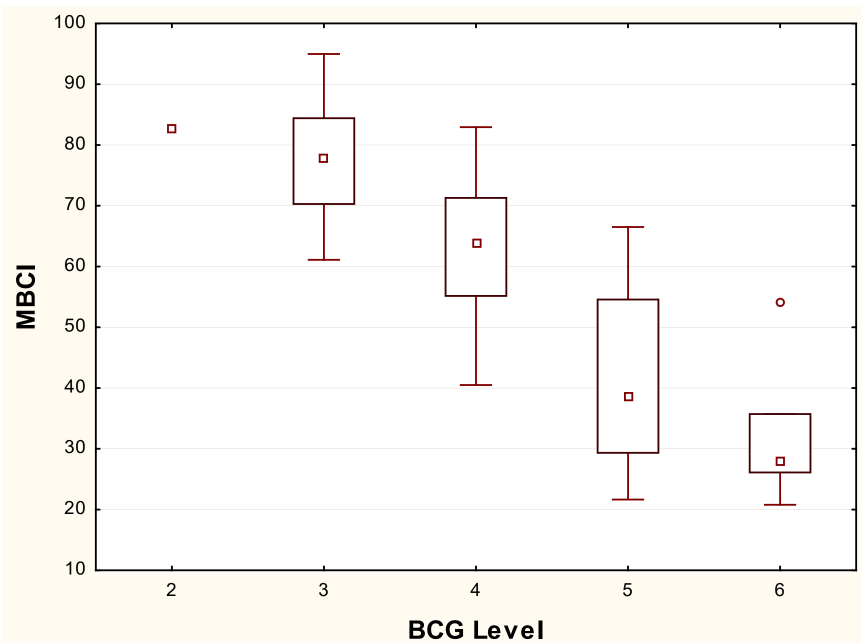

Figure ES-1. Rhode Island MBCI value distributions by Connecticut BCG levels estimated from the decision analysis model

Acknowledgements

This project was supported through a contract with the U.S. Environmental Protection Agency, Health and Ecological Criteria Division (Use of Rhode Island (RI) Bioassessment Information and Tools to Support Water Quality Management Programs, Contract #EP-C-09-001, Work Assignment #2-38). The contract was administered by the Great Lakes Environmental Center, Inc. (GLEC), with Tetra Tech, Inc. as a subcontractor.

The work group involved with technical project execution and oversight included Ben Jessup, Jen Stamp, and Jeroen Gerritsen of Tetra Tech; Connie Carey, Katie DeGoosh, and Sue Kiernan for the Rhode Island Department of Environmental Management; and Dave MacDonald of EPA Region 1. Dennis McIntyre of GLEC assisted with contract administration.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

The Clean Water Act (CWA) directs EPA and states to restore and maintain the biological integrity of the Nation's waters. Under the CWA, the EPA and RI have established Water Quality Standards (WQS) Programs to help achieve this objective and develop water quality criteria to assess the condition of the state's waters. The Rhode Island Department of Environmental Management (RIDEM) Office of Water Resources and U.S. EPA Region 1 are developing and testing macroinvertebrate sampling, criteria and assessment methods to support incorporation of bioassessment information into RIDEM OWR programs for managing aquatic resources in streams. The Biological Condition Gradient (BCG; Davies and Jackson 2006) provides a consistent scale for assessing biological conditions throughout the United States. Considerable effort has been spent on specifically calibrating and interpreting the BCG for macroinvertebrates and fish in New England higher gradient streams (Gerritsen and Jessup 2007, Snook et al. 2007). With this study, we used regional BCG models and interpretations to refine assessment tools specific to Rhode Island.

A multimetric biological index is a combination of responsive measurements that indicates degrees of biological stress in an assemblage (Barbour et al. 1999). In 2002, Tetra Tech developed and populated an EDAS-based database with the Rhode Island biological data (called BioQual) to manage data, automatically calculate metrics and run a preliminary multimetric index. Tetra Tech analyzed RIDEM biomonitoring data and developed a preliminary biotic index called the RI Wadeable Stream Condition Index (WSCI). The WSCI multi-metric index was developed using a set of data collected from 42 stations in 2001 with macroinvertebrates identified to genus. Using best professional judgment (because GIS and water quality data were unavailable at that time), a calibration set of stations was chosen and stations were designated either reference, average, or stressed. The preliminary WSCI index ranked several reference stations with low (poor) index scores and stressed stations with high index scores indicating a potential error in reference station designation, sampling error, or natural variability.

Although the RI WSCI index could not be confidently used as the only guidance to make assessments, Tetra Tech made recommendations to refine and improve the WSCI biotic index. The suggestions included identifying macroinvertebrate taxa to genus levels or better; collecting more data (three sampling seasons; new stations); selecting objective criteria to designate calibration stations as reference or stressed to eliminate bias or error; and re-developing the biological index after implementing these recommendations. Starting in 2002, RIDEM has worked to re-sample some stations each year (to collect more multi-year data at those stations), and collect data from new stations throughout the state. This has resulted in a sample size increase from 42 samples used to develop the original WSCI to 382 samples available to refine the index. These new macroinvertebrate samples were identified with increased resolution, to genus or species levels. Technological advances and increased availability of Geographic Information System (GIS) data (at point and catchment scales) as well as the addition of water quality and habitat data collected at each site has facilitated the ability to objectively set quantitative stressed or reference station criteria to designate the index calibration stations.

In 2009, Tetra Tech updated the BioQual database to a more recent platform, importing all the newly collected macroinvertebrate and habitat data as well as creating Access macros to automate the data migration process in the future. Benthic taxa lists were updated to reflect 149 taxa in the original lists that did not match any existing taxa in the database, however taxa tolerance values, FFG, TFI and Beck values were not able to be updated at that time. New queries were also added to facilitate display of macroinvertebrate presence/absence or relative abundance data matrices.

Goals of this current project are to continue building upon this previous work to re-develop a biological index using recently collected data, incorporate the new biotic index into the database queries and reports and apply the new RI index to the BCG. The resulting Multimetric Biological Condition Index (MBCI) will be an assessment tool to aid RIDEM to accurately and precisely evaluate biological conditions of waterbodies.

2.0 Methods

We developed the MBCI by following a series of steps. These steps are outlined below and further described in subsequent sections.

1. **Data Organization**: Compile and organize data pertaining to benthic macroinvertebrates, water quality, habitat, and site characteristics
2. **Stressor Gradient**: Identify the stressor gradient of the sampled sites, categorizing them as high quality reference or stressed
3. **Site Classification**: Examine site characteristics in relation to biological community types to identify major sources of natural variability
4. **Biological Metrics**: Calculate counts, percentages, and indices based on taxonomic attributes
5. **Sensitivity Analysis**: Identify those metrics that respond to the stressor gradient as expected and would be meaningful in an index
6. **Index Compilation**: Calculate and evaluate multiple index alternatives using the responsive metrics
7. **Apply the Index**: Assess index performance and limitations as they relate to objectives of the biomonitoring program

2.1 Data Organization

Data used for index development were based on 278 macroinvertebrate samples collected throughout the state in 128 wadeable streams between 2002 and 2009 from mid-July to early-mid September (Appendix A). RIDEM staff were familiar with the sites and recommended excluding from analysis 72 sites that were atypical, such as non-wadable, flat, and ephemeral streams (Appendix A). Data that were consistently collected during field visits included measures of benthic macroinvertebrates, water quality, flow and the physical habitat (ESS Group 2007). Data were organized in BioQual, a relational database with metric calculation functions customized for RI. Water quality variables related to physical characteristics, flow, ions, nutrients, metals, and microbiology were recorded with the samples, though data were incomplete or sparse for some variables (**Table 1**).

Metrics were calculated based on the identifications as entered in the database, which were mostly to the genus or species level. Taxa identified at a course level were counted in richness metrics only if no other taxon of the same group was identified at a more refined level in the same sample.

Table 1. Water quality variables associated with the biological samples.

Water Quality	Nutrients
Temperature (˚C) [1]	Nitrate/Nitrite (mg/L as N) [2]
Dissolved oxygen (mg/L) [1]	Dissolved Ammonia (mg/L as N) [3]
Dissolved oxygen (%) [1]	Total Ammonia (mg/L as N) [3]
Turbidity (NTU) [1]	TKN (mg/L as N) [3]
pH (su) [1]	Total Nitrogen (mg/L as N) [2]
Hardness (mg/L) [2]	Orthophosphate (mg/L as PO4) [2]
Total Suspended Solids (mg/L) [2]	Total Phosphorus (mg/L) [2]
Volatile Suspended Solids (mg/L) [3]	Metals
Water Quantity	Cadmium (ug/L) [2]
Flow (cfs) [1]	Copper (ug/L) [2]
Ions	Iron (ug/L) [3]
Conductivity (umhos/cm) [2]	lead (ug/L) [2]
Specific Conductance (umhos/cm) [2]	Zinc (ug/L) [3]
Salinity (ppt) [2]	Micro-biology
Chloride (mg/L) [2]	Biological Oxygen Demand (mg/L) [3]
Sodium (mg/L) [2]	Enterococcus (MDN100) [2]
	Fecal coliform (MDN100) [3]

1: Complete – data were available for all samples
2: Incomplete – data were missing for some samples
3: Sparse – data were missing for most samples

Taxonomic attributes for taxa recently added into the database and for BCG attributes were incomplete. Therefore, attributes were completed when information was available from literature sources, other regional databases, or by association with similar taxa (Appendix B). The literature included the Rapid Bioassessment Protocols (Barbour et al. 1999), which compiled information from multiple regions of the U.S. The regional databases with similar information included those from the DEP of both CT and ME. BCG attributes were entirely from the CT database. When the literature or databases were used to update the attributes, they were listed as sources. The third source was listed "by association", meaning that attributes associated with a genus were assumed to apply to a species of the genus or that concurring attributes of several genera within a family could be applied at the family level.

Only one randomly selected sample per site was used in site classification and index calibration analyses to avoid any potential bias that might occur if multiple samples were used in some sites and not others. Replicate samples within sites were used only in analyses of metric and index precision.

Each sampling site was geo-referenced, making it possible to analyze site characteristics that were available as geographic information layers. RIDEM analyzed site characteristics at the site or catchment scale (**Table 2**) and the information was used for either classification of site types or evaluation of site conditions (reference or stressed). Rhode Island lies entirely within the Northeast Coastal Zone Level III ecoregion and is subdivided by parts of three level-IV ecoregions. While the majority of the state is characterized as Southern New England Coastal Plains and Hills (CPH), the smaller coastal lowland areas are separated between sections of the Narragansett/Bristol Lowland, and the Long Island Sound Coastal Lowland Level IV ecoregions (Griffith et al. 2009).

Table 2. Variables derived from GIS analyses and used for either classification of site types or evaluation of site conditions. Variables preceded with a ‘c’ were determined for the upstream catchment of the site. Those preceded with a ‘p’ were determined at the sampling point.

Classification variables	Condition variables
c Total Area (sq. mi) of upstream catchment	c % natural cover
p Elevation (ft.)	c % forest
p Strahler Stream Order	c % crop agriculture
c Density of Streams in watershed (mi/mi2)	c % developed land
p Bedrock Geology	c Impervious surface %
p Glacial Deposit Type (outwash or till)	p Distance to dam (300, 500, 1000, 2000m)
c % wetlands	c Density of sewer overflows (#/sq.mi)
p Slope in a 500m reach (mean, min, and max)	c Density of CERCLIS (#/sq. mi.) (Superfund)
p Slope in a 1000m reach (mean, min, and max)	c Density of permitted discharges (#/sq. mi.)
p Hydrologic Unit (HUC8)	c Density of Outfalls (#/sq. mi.)
p Level 4 Ecoregion	

2.2 Stressor Gradient

The reference condition concept is one in which the acceptable indicator conditions are defined by the conditions observed in sites with minimal disturbance (Barbour et al. 1999, Stoddard et al. 2006). Index values that are not similar to those observed in reference sites indicate effects of stressors at the site. For multimetric indices, defining the most disturbed (or stressed) sites as well as the least disturbed is necessary to establish clear signals of metric and index responsiveness. Reference sites in Rhode Island are not *un*disturbed. Rather, they are the best available – meaning that they have the least amount of disturbance relative to other sites sampled in the dataset.

Reference and stressed sites were defined using water quality (conductivity and dissolved oxygen, from water quality monitoring samples, averaged per site) and GIS derived measures of the intensity of human activity in the watersheds. Water quality records were incomplete for some variables, such as nutrients, metals, and microbiology, which were therefore not used.

Criteria for reference (best available) and stressed (degraded) site conditions were established using distributions of variables related to intensities of land uses and direct stressors that were assumed to arise from human activities. By plotting distributions of these variables, we were able to set thresholds of disturbance for defining relative stress levels for each variable. We expect that the indicators resulting from these analyses will be most responsive to the stressors used in defining reference and stressed sites.

In defining reference sites for streams, we intended to recognize overarching patterns of land use intensity as we set reference criteria. In this way, we could assure that reference sites would be distributed spatially throughout the state. At the same time, we concede that reference conditions are not identical across the state. The intention was to have representative sites for all natural stream types, and to recognize where the reference sites are less than natural, to a degree appropriate to their locations. We only accept less-than-natural conditions as reference where truly natural sites could not be found or are too remote to be used in valid comparisons. In other words, we did not want to compare streams in the agriculturally dominated areas to forested mountain streams far away, so we sought the best streams that were in the same geographical setting as the agriculturally dominated areas. In this way the reference sites have climate, geology, and other controlling natural conditions that are similar to the test sites that are compared to them.

Land uses in Rhode Island appear to be somewhat aligned with the natural settings, such that the lowlands in the east and south of the State have been more conducive to development than the hillier northwest. The close relationship between land uses and natural landscape characteristics requires that we recognize some aspects of site classification at the same time we are identifying reference and stressed sites. Natural ecoregions (Griffith et al. 2009) are therefore a reasonable framework for recognizing human development patterns as they relate to variations in expectations for degrees of disturbance across the State. In areas of the state with ubiquitous development or disturbance, we considered relaxing reference criteria so that each potentially unique stream type could be recognized during index development.

2.3 Site Classification

Site classification is the process by which natural gradients among sites are examined to identify appropriate classes or “bins” of sites with similar indicator characteristics. The purpose of classification is to minimize within-class natural variability of biological indicators so that the signal of human disturbance can be recognized with less background noise. Potential site classification variables and biological structure of samples were analyzed simultaneously to identify patterns of covariance. In most analyses, only reference sites were used for site classification so that the patterns in the natural settings of Rhode Island could be discerned with less influence from human disturbances. Environmental factors useful for classification include measures that are not likely to change with human activities, such as latitude, longitude, ecoregion, elevation, slope, catchment size, and catchment basin. Resulting site classes were the framework within which multimetric indices were calibrated.

We hypothesized that of the variables considered for classification (**Table 2**), level-4 ecoregions (Griffith et al. 2009), hydrological units (USGS), catchment size, and elevation would be important determinants of natural biological conditions. Ecoregions were integral to the reference site selection process and were therefore somewhat entwined in the classification determination, in which reference site data are key. We used several techniques to help discern which of the environmental factors might account for biological variability in reference sites. The community structure of benthic macroinvertebrate samples was explored using Non-metric Multidimensional Scaling (NMS) ordination and cluster analysis. Environmental factors were related to the major ordination axes or site groupings. Because some areas of Rhode Island had few reference sites, we also used Principal Components Analysis (PCA) of environmental variables in all sites. By using only variables that are unresponsive to human activities, the natural settings of the sites could be distinguished regardless of human activities.

2.3.1 Operational Taxonomic Units (OTU)

In preparation for ordination and clustering of taxa, in which sites are grouped by taxonomic similarities, taxonomic identifications were examined to reduce uncertainties and increase distinctiveness. Taxa were aggregated into Operational Taxonomic Units (OTU, Cuffney et al. 2007) or eliminated from the analysis. Aggregation and elimination of ambiguous macroinvertebrate taxa were necessary in the specific analyses for two reasons. First, rare taxa can influence ordination results to a degree greater than their actual significance in the ecological settings, and should therefore be removed from analysis. Second, the ordination and clustering routines assumes that each taxonomic identification is unique. However, that assumption may be incorrect because of our inability to reliably identify all individuals at the species level. For instance, a family level identification is interpreted to be different from a genus within that taxon, though the family level identification is ambiguous and may well represent a member of the same genus. This taxonomic uncertainty can lead to meaningless ordination configuration, and therefore *must* be resolved. Aggregation and elimination of taxa was performed so that the least amount of taxonomic information would be lost in the analysis. The OTUs were only used for site classification analyses and were not used in index calculations.

Taxa were considered rare and removed or aggregated when they occurred in 5% or less of the reference sites. When there were several rare taxa in a taxonomic group and no common taxa in that group, all taxa in the group were re-assigned to the next higher taxonomic level (e.g. species lumped into the parent genus). If the common taxa outnumbered the rare taxa in a group, the rare taxa were eliminated. When there were higher level identifications for a group, they were eliminated from analysis if the lower level identifications in the same group were common and numerous.

2.3.2 NMS Ordination

Similarity among reference macroinvertebrate samples was determined using the Bray-Curtis (BC) similarity measure in a non-metric multidimensional scaling (NMS) ordination. Sites were arranged in ordination space based on a site-by-site matrix of BC similarity. Sites with similar taxonomic composition were plotted in close proximity and those with less similarity were plotted at a distance. Multiple dimensions were compressed into two or three dimensions that we can perceive. The stress associated with this compression indicates how closely the Bray-Curtis distance is reflected in the plot. Interpretation of the ordination diagram with respect to taxa within the samples and characteristics of the sites takes place through visual inspection of variable overlays and correlation along the ordination axes.

The NMS ordination was performed using PC-Ord software (McCune and Mefford 2006). A site-by-taxon matrix was compiled with abundance of each OTU in one sample of each site. A preliminary Detrended Correspondence Analysis (DCA) was performed to establish stable starting coordinates for the NMS ordination. Ordination was performed using taxa presence and relative abundance.

2.3.3 Cluster Analysis

Cluster analysis allows comparisons of OTU in samples and links similar samples in a tree structure. This allows a clear distinction of biological site types which can then be related to environmental factors through Discriminant Function Analysis (DFA) and display onto the NMS diagram. Cluster analysis used taxa presence, the Bray-Curtis distance measure, and a flexible beta agglomeration method with beta set at -0.50.

2.3.4 Principal Components Analysis (PCA)

PCA identifies similarities among many measures and reduces them to a few factors with variance maximized on each successive axis. It was used to identify the major factors of two data sets: natural variability among sites throughout the state and metric variability among reference samples. With these factors it was possible to associate reference metric distributions with influential environmental factors and to identify which axes were most important for classifying sites. The steps of the analysis were to 1) conduct a PCA of environmental factors in all sites, 2) conduct PCA of metrics in reference sites, 3) relate metric components to environmental components, and 4) use the environmental components that are related to reference metrics to refine site classes. The natural variables used in the PCA included latitude, longitude, catchment area, elevation, stream order, stream length, slope, and pH. Variables were transformed using logarithms to approximate normal distributions when necessary.

2.3.5 Correlation Analysis and Remaining Variability

Correlation analysis was used to relate reference metrics to environmental variables as single factors. Non-parametric Spearman rank correlations were used. Based on correlations, we examined the potential for adjusting individual metrics to continuous gradients of conditions, as is typically done for adjusting fish richness metrics to site drainage area (McCormick et al. 2001, Fausch et al. 1984). After sites were classified into distinct bins, we examined the remaining variability of each metric with a set of natural variables, including elevation, catchment size, and percent water and wetlands. This analysis also considered the pragmatism of adjusting metrics to natural gradients if the natural gradients were short or were not ecologically meaningful.

When correlations were significant ($p<0.05$), the strongest relationships were examined in bi-plots to determine which of multiple possible relationships could be reasonably estimated. Relationships that were consistent along the environmental gradient (not driven by outliers), were linear (or could be estimated with non-linear relationships), and that could result in relatively precise reference distributions after adjustment (not wedge-shaped), were considered for adjustment.

2.4 Biological Metrics

Benthic macroinvertebrate metrics were organized into five categories: taxonomic richness, assemblage composition, feeding group, habit (methods of attachment or locomotion), and pollution tolerance. Each category addresses aspects of the sample that are expected to change with general or specific stressors. Richness is high when habitats are complex and water quality does not limit sensitive taxa. Homogeneous habitats within a sampling reach or polluted water can limit taxonomic diversity overall or in specific groups of taxa. Composition of taxa, numbers of individuals in various groups, can vary with stressor intensity depending on the tolerances or opportunistic abilities of each group. Feeding group and habit metrics exhibit patterns when niche space in stressed sites is limited due to food resource quality or habitat types. Tolerance metrics are based on standardized scales of pollution tolerance to which each taxon is measured. Typically, general types of pollution are incorporated into the scale, including nutrients, sediments, and organic pollutants.

Metrics were calculated using BioQual based on sample taxonomic lists and taxa attributes. All richness metrics (e.g., total taxa or EPT taxa) were calculated such that only unique taxa were counted. Taxa that were identified at higher taxonomic levels because of damage or under-developed features were not counted as unique taxa if other individuals in the sample were identified to a lower taxonomic level within the same sample. The level of identification reported in the database was used in metric calculations, which assumes that taxonomic resolution has been consistent over the study period and will continue to be so.

2.5 Sensitivity Analysis

2.5.1 Discrimination Efficiency

The ability of each metric to distinguish between reference and stressed sites within a site class was measured as discrimination efficiency (DE) (Flotemersch et al. 2006). For metrics with a pattern of decreasing value with increasing environmental stress, DE is the percentage of stressed values below the 25^{th} percentile of reference site values. For metrics that increase with increasing stress, DE is the percentage of stressed sites that have values higher than the 75th percentile of reference values. DE can be visualized on box plots of reference and stressed metric or index values with the inter-quartile range plotted as the box (**Figure 1**). Higher DE denotes more frequent correct association of metric values with site conditions. DE values ≤25% show no discriminatory ability in one direction. DE values ≥50% are generally adequate for consideration in an index. However, because we placed emphasis on finding metrics in each metric category, metric adequacy was usually dependent on relative DE values within a metric category. Metric scores are directly related to metric values except that extreme values are scored at the scoring range limits (0-100 index points, see Section 2.6). For the non-parametric DE value derived from the 25^{th} percentile of values (which ignores the differences at the extremes), evaluation of sensitivity is identical using either metric values or metric scores. The sensitivity analysis was conducted with subsets of reference and stressed sites that were randomly selected for index calibration.

2.5.2 *Z*- scores

A second measure of metric discrimination was the *Z*-score, which was calculated as the difference between reference and stressed metric or index values divided by the standard deviation of reference values. There is no absolute *Z*-score value that indicates adequate metric performance, but among metrics or indices, higher *Z*-scores suggest better separation of reference and stressed values. We evaluate metric values instead of metric scores because the values reflect the true variability in metrics; whereas scores artificially reduce variability by re-valuing the extremes (see Section 2.6).

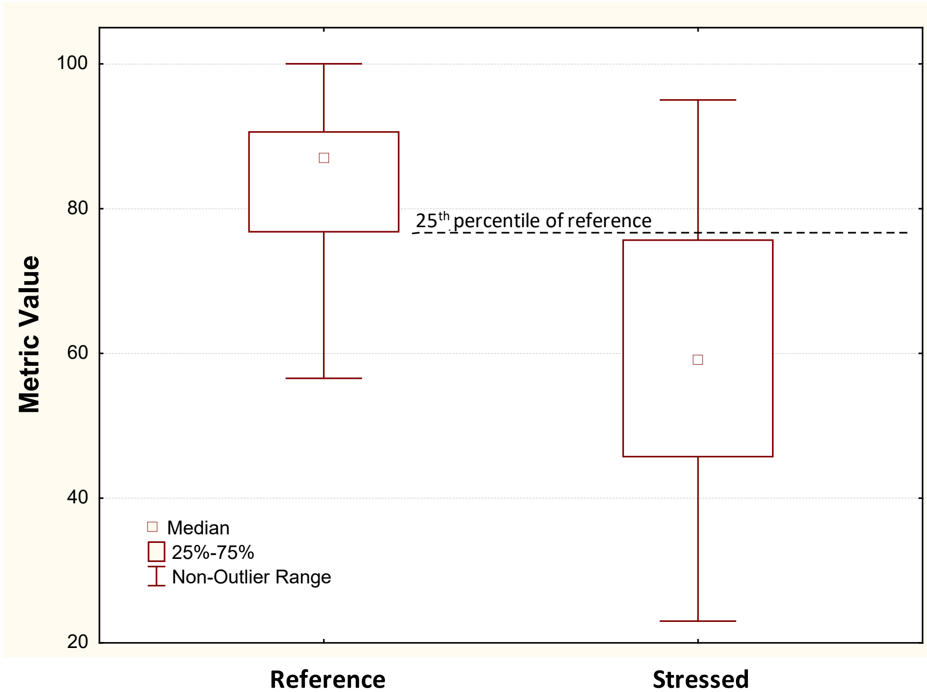

Figure 1. Box and whisker plot illustrating a metric that decreases with increasing stress and that has a DE slightly greater than 75%.

2.5.3 Metric Precision

Metric precision describes the reliability of a metric for indicating the same conditions regardless of the micro-site sampling differences, sampling error, or temporal sampling conditions (Stribling et al. 2008). When selected for an index, more precise metrics limit some of the variability inherent in sampling protocols and lead to more precise index results. Metrics were evaluated for precision in "same day" replicates and in replicates collected over time within sites. Analysis of Variance (ANOVA) using sites as grouping variables and metrics as independent variables yielded the Mean Squared Error (MSE). MSE is an estimate of variance within replicated samples and is the basis for other precision statistics: the Root MSE (RMSE), the Coefficient of Variability (CV), and the confidence interval (CI). The RMSE is the square root of the MSE and is an estimate of standard deviation within groups (Zar 1999). The CV is the quotient of RMSE by the mean of all replicated samples, expressed as a percentage (CV = 100*RMSE/mean). The CI is based on the RMSE and a level of confidence, such as 90%. The positive and negative 90% CI around an observation is the range within which the true mean is expected to fall in 90% of cases. It is calculated as ±1.645*RMSE (the constant is related to the area under a normal curve for the selected confidence). Metrics with small RMSE, CV, and CI90 are relatively precise.

2.6 Index Composition

Metrics were scored on a continuous scale from 0 to 100 prior to combination in an index. The optimal score is determined by the distribution of metric values (Blocksom 2003). For metrics that decrease with increasing stress, the 95th percentile of all data within the site class was considered optimal (to lessen the influence of outliers [Barbour et al. 1999]), and scored as 100 points using the equation:

$$MetricScore = \frac{100 * MetricValue}{95^{th}\ Percentile}$$

Metrics that increase with increasing stress (reverse metrics) were scored using the 5th percentile of data as the optimal, receiving a score of 100. Decreasing scores were calculated as metric values increased to the 95th percentile using the equation:

$$MetricScore = \frac{100 * (95^{th}\ Percentile - MetricValue)}{95^{th}\ Percentile - 5^{th}\ Percentile}$$

In some cases, percentiles other than the 95th were used in the equation above to reduce the effects of a skewed distribution. The metric scoring range was from 0 to 100. Scores outside of this range were re-set to the nearest extreme before the index was calculated.

Conceptually, the best index alternatives were those that included the metrics with the highest DEs within each metric category. Using an iterative process with R software (R Development Core Team 2010), more than 85,000 index alternatives with different metric combinations were calculated. The index alternatives were evaluated using the following considerations.

- Discrimination of reference from stressed sites using the DE and Z-score;
- Inclusion of individual metrics that discriminated reference and stressed sites;
- Inclusion of metrics that were ecologically meaningful;
- Inclusion of metrics that represented diverse metric categories;
- Exclusion of metrics that were redundant with other metrics in the index; and
- Index precision for replicated samples.

The RIDEM workgroup reviewed indices with similar performance characteristics to select a final index that included metrics that were meaningful to their programs. As many metric categories as practical were represented in the index alternatives so that signals of various stressor-response relationships would be integrated into the index. While several metrics should be included to represent biological integrity, redundant metrics can bias an index to show responses specific to certain stressors or taxonomic responses. Redundancy was evaluated using a Spearman rank order correlation analysis. In this index development effort, we excluded metrics that were redundant at $|r| > 0.85$.

Index performance was validated with a set of samples that were not used in index calibration. Validation data were expected to perform as well as calibration data or to have reference and stressed values similar to the calibration values. Performance was judged by comparing index values relative to the site reference and stressed designations. An evaluation threshold was established for these comparisons so that index error could be quantified in the calibration and validation datasets.

2.7 Applying the Index

Application of the index follows an established set of steps, which are outlined here and detailed in the results section.

1. Determine whether a sample falls within index application limitations
2. If the index is applicable, calculate the metrics
3. Score the metrics
4. Combine the metric scores into an index
5. Compare the index value to condition thresholds
6. Communicate assessment results and associated performance characteristics

Performance of the recommended index was quantified in terms of accuracy and precision of repeated measures. With these performance characteristics, the reliability and uncertainty of assessments based on the index can be communicated (Stribling et al. 2008). Accuracy was quantified with Type 1 and Type 2 errors in relation to potential impairment thresholds. The DE is an expression of accuracy based on the assessment results in stressed sites.

Precision of repeated measures within sites was quantified by calculation of confidence intervals based on index variability observed in sites over time. As described above for metric values, index values from repeated samples within sites were entered into an ANOVA, from which the variability of indicator values was evaluated. While index precision should not be used to qualify site assessment results, it can be used to justify repeated sampling for specific sites in future sampling efforts.

RIDEM had specific questions relevant to index application that were also addressed, as follow.

a. What are appropriate applications of the index and what are the limitations of applicability?

b. Does the index appear to be biased by catchment size?
c. Does the index appear to be biased by stream slope?
d. How does the multimetric index compare to the Connecticut BCG model?
e. What is the variability of the index for samples collected within one Waterbody Identification unit (WBID)?

3.0 Results

3.1 Stressor Gradient Results

Preliminary reference and stressed thresholds were established for each stressor variable based on the distribution of values in all sites. Thresholds were set so that several sites passed the reference threshold and a few site failed the stressed threshold. When combining criteria, a site must pass all reference criteria to be categorized as reference. Failure of any of the stressed criteria results in a site being categorized as stressed. The resulting reference and stressed sites were reviewed for confirmation by the RIDEM staff, reasonable spatial distribution, and sufficient sample size for index calibration. Adjustments were then made to the criteria as needed.

Criteria for reference and stressed site designations differed between the CPH and the two Lowlands ecoregions (**Table 3**). This was done because no lowland sites were identified as reference using the initial criteria and relaxing the criteria for all regions would have resulted in lower quality reference sites in the CPH. The stressed site criteria were equally applied across the state. Although a site could be designated stressed for failing any single stressed site criterion, sites with more than 85% natural land cover were not designated as stressed regardless of the other criteria. Application of the reference and stressed criteria resulted in 32 reference sites and 9 degraded sites in the CPH and 9 lower quality reference sites and 25 degraded sites in the two Lowland ecoregions. The CPH and Lowland distinctions were useful for reference site identification but were not assumed to be site classifications.

Table 3. Criteria for reference and stressed sites. A site must meet all reference criteria to be reference. If any stressed criteria are met, the site is deemed stressed (unless % natural is >85).

Reference	CPH	Lowlands	Stressed
% natural	>90	>70	<50 and not >85%
% impervious surface	<5	<12	>15
Conductivity[a]	<150	<200	>300
Density of permitted discharges	<0.1	<0.1	>0.3
Density of sewer overflows	0	0	>0.1
Density of CERCLIS	0	<0.5	>0.5
Dissolved oxygen	>5	>5	<5

[a] This is the geometric mean of conductivity. Where conductivity and specific conductance were both collected, they were averaged. Otherwise, criteria were applied to the recorded measure (either conductivity or specific conductance).

The "density of outfalls" variable was not used to evaluate site conditions as the metadata for the GIS coverage did not clearly indicate equal georeferencing across the state and some RI municipalities appeared underrepresented. The "distance to a dam" variable was also not used to define reference sites due to the prevalence of dams throughout the state. However, this variable may be significant in relation to stream flow during dry years when low lake levels prevent flow over a dam and result in desiccated stream beds.

The sites selected as reference quality showed some biases in environmental characteristics when compared to the characteristics in all sites (**Table 4**). For example, CPH reference sites were higher in elevation and smaller in catchment size, stream length, and flow when compared using mean values. As expected, differences in stressor characteristics were also apparent (e.g. % imperviousness, % natural or developed land cover, and conductivity).

Table 4. Distributions of selected variables in all sites and in reference and stressed sites. Samples sizes are as follows: All – 128, Reference – 32, and Stressed – 26.

		min	25th	mean	75th	max	sd
Latitude	All	41.35	41.52	41.69	41.85	42.05	0.19
	Reference	41.46	41.54	41.66	41.71	42.05	0.16
	Stressed	41.46	41.68	41.76	41.85	41.96	0.12
Longitude	All	-71.82	-71.72	-71.59	-71.48	-71.13	0.14
	Reference	-71.77	-71.73	-71.67	-71.62	-71.37	0.08
	Stressed	-71.69	-71.49	-71.49	-71.44	-71.40	0.08
Elevation (ft)	All	0.0	64.5	169.5	262.0	462.0	128.7
	Reference	66.0	158.0	263.3	328.0	462.0	110.8
	Stressed	0.0	37.0	96.9	109.0	349.0	90.0
Catchment Area (mi2)	All	0.3	2.7	18.1	9.9	433.8	56.1
	Reference	0.5	2.0	5.2	6.2	23.9	5.0
	Stressed	0.3	4.3	9.7	11.0	48.5	10.1
Stream Length (mi)	All	0.3	4.5	39.0	22.4	1047.8	134.6
	Reference	0.6	4.1	9.6	11.7	60.0	11.0
	Stressed	0.3	7.3	19.1	19.8	106.9	22.0
Stream Slope (m/100m)	All	1.4	4.1	5.5	6.7	10.0	1.8
	Reference	3.0	5.3	6.1	7.2	9.0	1.5
	Stressed	1.4	4.5	5.4	6.6	9.4	1.8
% Impervious surface	All	0.1	1.2	7.7	11.7	47.6	9.7
	Reference	0.1	0.5	1.1	1.5	6.8	1.2
	Stressed	1.6	13.8	18.5	22.8	36.1	9.1
% Natural Land Cover	All	9.2	63.4	76.6	91.6	99.1	20.1
	Reference	90.4	91.7	93.3	95.8	99.1	4.9
	Stressed	29.6	47.8	55.9	60.7	83.6	14.8

		min	25th	mean	75th	max	sd
% Developed Land Use	All	0.3	3.8	16.7	25.1	82.7	18.7
	Reference	0.3	1.1	3.0	4.3	17.9	3.2
	Stressed	5.3	31.4	37.4	46.1	68.4	15.1
Water Temperature (˚C)	All	13.7	17.8	19.0	20.2	24.4	1.9
	Reference	14.3	17.0	18.4	20.1	22.8	2.1
	Stressed	16.4	18.1	18.9	19.9	21.7	1.4
Dissolved oxygen (mg/L)	All	5.1	7.5	8.1	8.6	17.1	1.2
	Reference	6.8	7.6	8.2	8.7	9.6	0.7
	Stressed	5.6	7.5	8.3	8.7	17.1	2.0
pH (su)	All	5.9	6.4	6.8	7.1	7.8	0.5
	Reference	5.9	6.2	6.5	6.6	7.8	0.4
	Stressed	6.2	7.0	7.1	7.5	7.8	0.5
Conductivity (umhos/cm)	All	15.0	74.3	169.5	245.4	789.6	128.3
	Reference	15.0	42.8	69.6	90.0	175.2	36.3
	Stressed	129.5	282.5	349.3	387.1	789.6	126.2
Spec. Cond. (umhos/cm)	All	35.5	86.5	199.8	273.1	946.7	147.3
	Reference	35.5	53.0	80.8	99.6	193.4	37.1
	Stressed	218.6	328.7	395.2	435.1	946.7	142.1
Outfall Density (#/mi^2)	All	0.0	0.4	3.1	4.4	15.2	3.8
	Reference	0.0	0.0	2.5	4.5	12.7	3.3
	Stressed	0.0	1.2	4.7	7.5	15.2	4.8
RIPDES Density (#/mi^2)	All	0.0	0.0	0.1	0.0	0.7	0.1
	Reference	0.0	0.0	0.0	0.0	0.0	0.0
	Stressed	0.0	0.0	0.1	0.2	0.4	0.1
CERCLIS Dens. (#/mi^2)	All	0.0	0.0	0.2	0.2	3.1	0.4
	Reference	0.0	0.0	0.0	0.0	0.0	0.0
	Stressed	0.0	0.0	0.7	0.9	3.1	0.8
Flow (cfs)	All	0.1	2.4	17.0	14.2	326.3	45.5
	Reference	0.1	1.7	5.7	6.2	34.3	7.0
	Stressed	0.5	4.4	10.6	17.8	26.5	8.3

3.2 Site Classification Results

We examined site classification in the CPH and not in the Lowland ecoregions because all of the high quality reference sites were in the CPH. With 32 reference sites, there were enough to split the data into at most two site classes and still calibrate an index in each. At the outset, the Lowlands were considered a separate site class based on two factors: 1) the reference site criteria were relaxed in the Lowlands and 2) Lowland ecoregions have essentially different biological and physical stream types compared to non-Lowland regions (Jessup et al. 2000). We assumed that the biological metrics in the reference Lowland sites would be different from those in the CPH because the reference sites in the lowlands were less strictly defined. In addition, only 9 lower quality reference sites were identified in the Lowland ecoregions, which is insufficient to confidently calibrate an index.

3.2.1 NMS

The NMS ordination of taxa presence/absence resulted in a diagram where most of the variation in sample taxonomic composition was on the first axis. This was also the axis to which most of the metrics were correlated (**Figure 2**). The environmental variables associated with the first axis were land cover (% water, % emergent wetland, and scrubland) and water temperature. Although these were highly correlated with the axis, they were not convincing classification variables. For example, only a few sites had high percent water and those were not consistently in one part of the diagram (**Figure 3**). Temperature is not a good site classification variable because it is dependent on date and time of sampling. Slope and elevation were not strong determinants of the biological similarities. Stressor variables (and longitude) were associated with the third axis, but none of the metrics were. This is probably because among reference sites, the stressor gradient is not extreme. Hydrologic units (12-digit HUC) were identified as a potential classifying variable, but evidence of strong classification value was not apparent in the diagram (**Figure 4**).

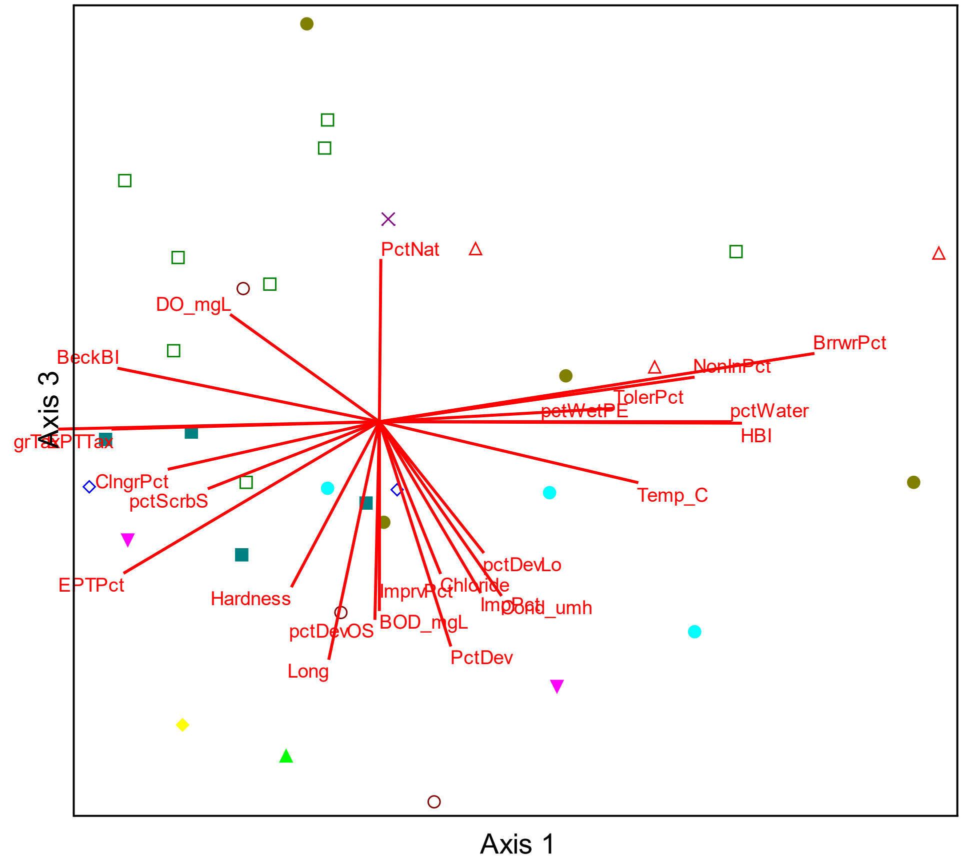

Figure 2. NMS diagram using presence of taxa in CPH reference samples, showing variable vectors in relation to the first and third axes.

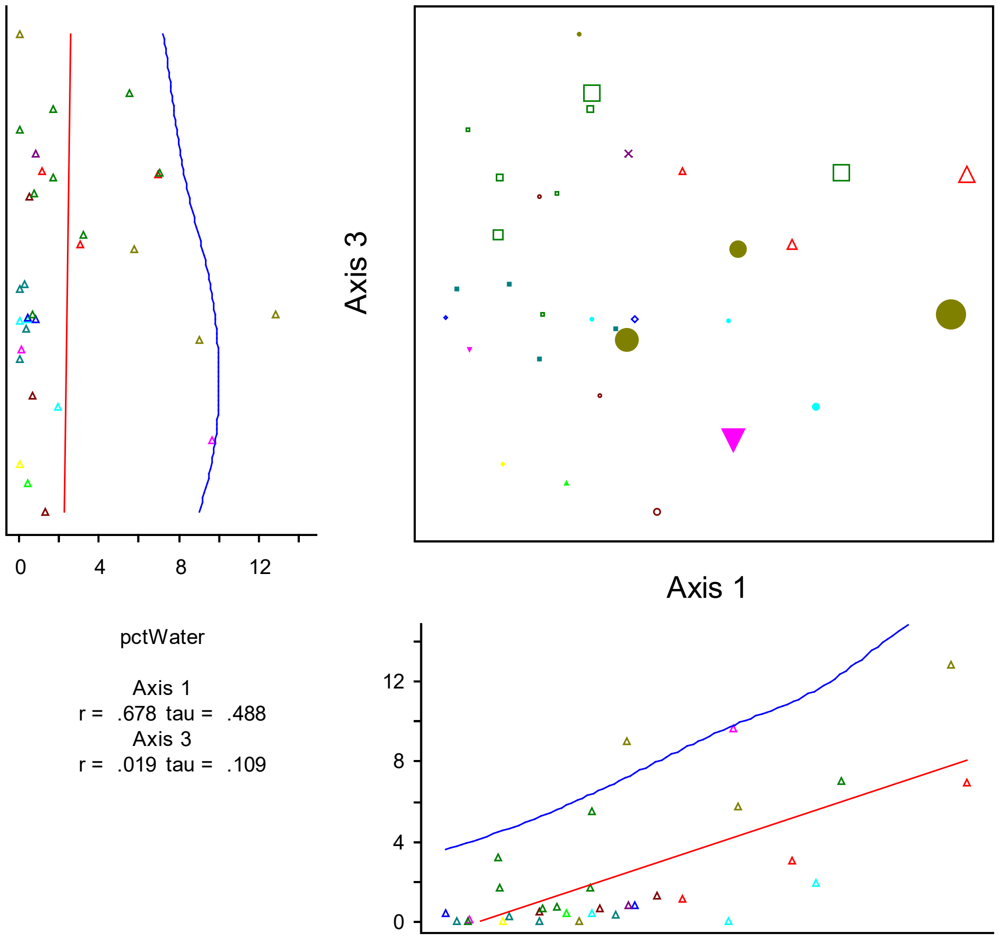

Figure 3. NMS diagram using presence of taxa in CPH reference samples, showing relationship of samples in taxa space and the percentage of water in the catchments of the sites.

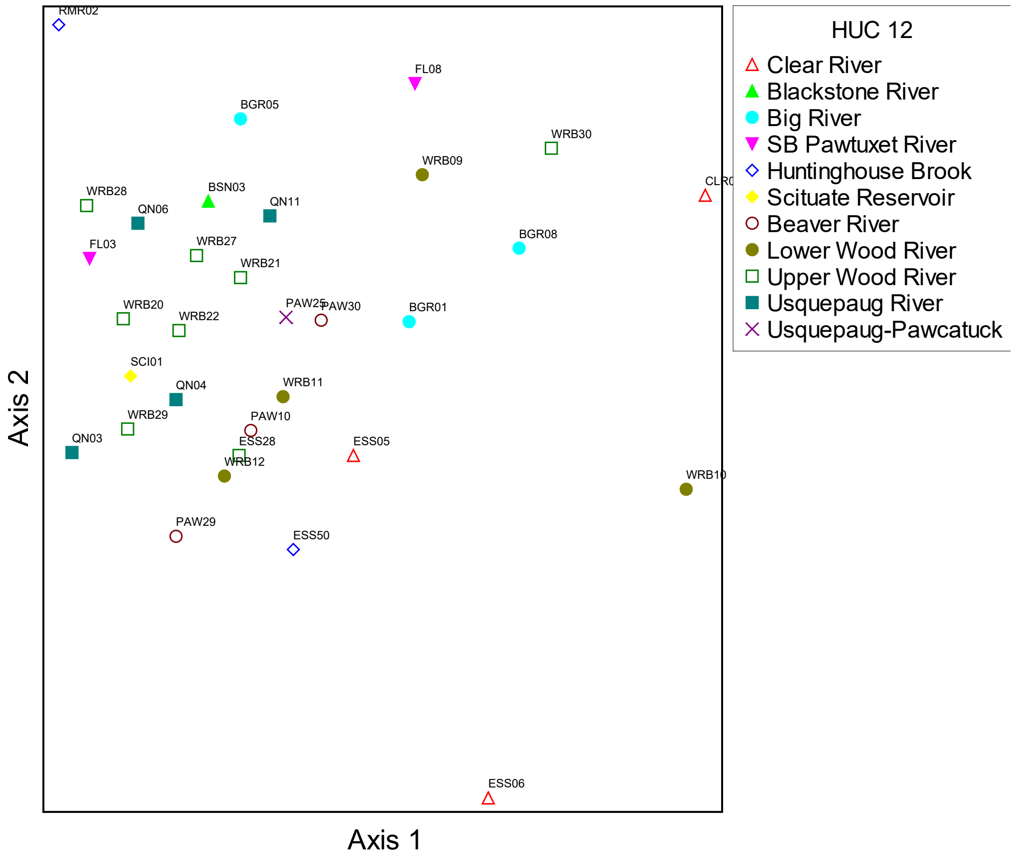

Figure 4. NMS diagram using presence of taxa in CPH reference samples, identified by site name and hydrologic unit (12-digit HUC).

The NMS ordination using relative abundance of taxa showed most of the variation in sample taxonomic composition on the second axis, though this was not the axis to which most of the metrics were correlated (**Figure 5**). The second axis was related to longitude, % forest type, and slope. The association with slope is not consistent enough to suggest distinct site classes (**Figure 6**). The associations with forest type variables appeared to be driven by a few outlier points. The environmental variables associated with the first axis were longitude and % water. Neither of these appeared to adequately distinguish the group of sites that are distinct in the diagram.

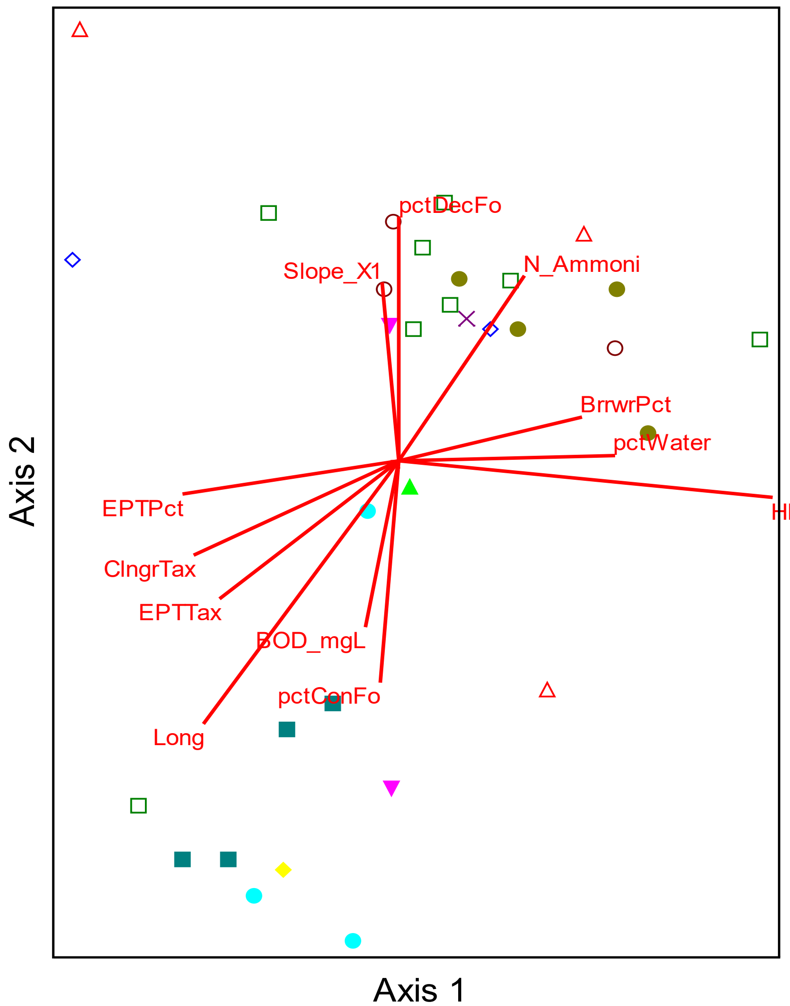

Figure 5. NMS diagram using relative abundance of taxa in CPH reference samples, showing variable vectors in relation to the first and second axes.

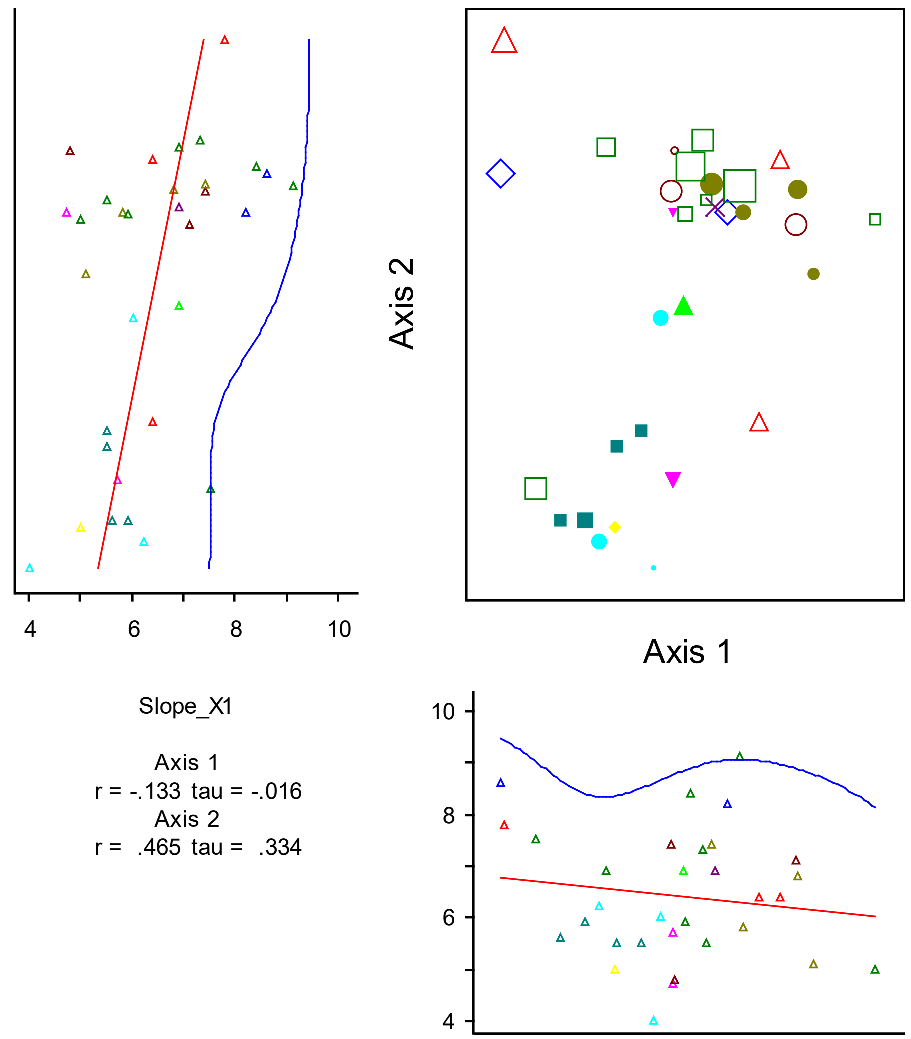

Figure 6. NMS diagram using relative abundance of taxa in CPH reference samples, showing relationship of samples in taxa space and the stream slope.

3.2.2 Cluster Analysis

A cluster analysis with taxa presence/absence suggested that three major site groupings could be distinguished. When superimposed on the NMS diagram, one group has sites with low EPT, high non-insects, and high HBI in sites with higher % water in the catchment (**Figure 7**). The other two groups are split into eastern and western sites with the westerns sites having more crane flies (Tipulidae and *Hexatoma*) and riffle beetles (*Optioservus*).

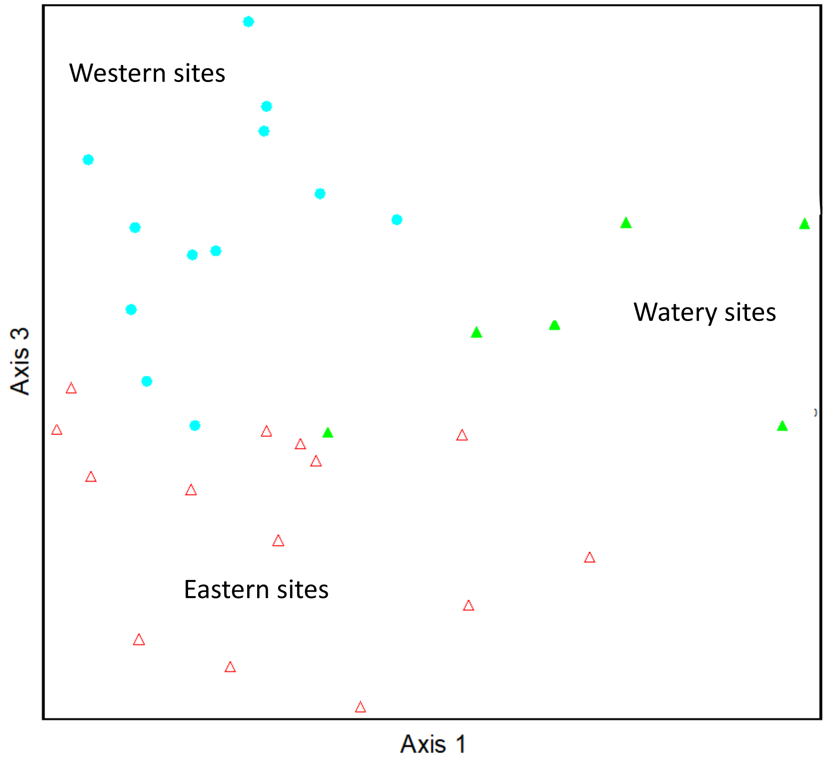

Figure 7. Clusters shown on the NMS diagram for taxa presence data.

In the cluster analysis using relative abundance of taxa, two groups were clearly distinguished. The smaller group, with nine sites, had relatively high EPT, low non-insects, and low HBI. In the NMS diagram (**Figure 8**), these sites were related to eastern streams with lower slopes.

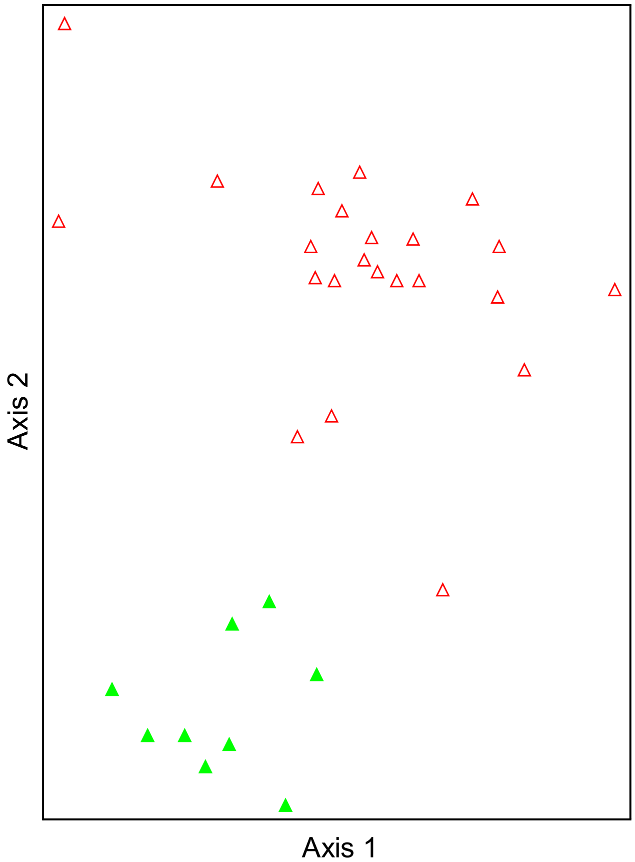

Figure 8. Clusters shown on the NMS diagram for relative abundance data.

3.2.3 PCA Analysis

In a PCA of natural environmental variables in all sites, the first three factors explained 81% of the variance (**Table 5**). The first factor was related to stream size, the second was related to slope and elevation, and the third was related to longitude. In a separate PCA of biological metrics in reference sites, the first three factors explained 44% of the variance (**Table 6**). The first factor was related to richness of sensitive taxonomic groups, the second factor was related to composition of Diptera and collectors, and the third factor was related to richness of collector taxa and composition of Crustacea and Mollusca.

Table 5. Factor scores for natural variables in a PCA of all sites.

Natural Variable	Factor 1 35%	Factor 2 29%	Factor 3 17%
Latitude	0.65	-0.23	0.19
Longitude	0.52	0.38	**0.55**
Elevation (ft)	-0.21	**-0.68**	-0.22
Catchment Area (mi2)	**0.75**	0.29	-0.51
Stream Order	0.62	0.21	-0.49
Stream Length (mi)	**0.77**	0.31	-0.48
Stream Slope (m/100m)	0.26	**-0.81**	-0.31
pH (su)	0.52	0.37	0.29
% wetland and water	-0.58	0.43	-0.55
% wetland	-0.60	0.40	-0.50
% water	0.02	0.11	-0.17

Table 6. Factor scores for biological metrics in a PCA of CPH reference sites.

Factor 1 (22%)		Factor 2 (12%)		Factor 3 (10%)	
EPT Taxa	-0.94	% Diptera	-0.87	Collector Taxa	0.69
Clinger Taxa	-0.89	% Chironomidae	-0.82	% Bivalvia	0.67
Beck's Index	-0.87	% Collector	-0.81	% Crustacea & Mollusca	0.66
Intolerant Taxa	-0.84	% Filterer	0.74	Crustacea & Mollusca Taxa	0.64
Scraper Taxa	-0.83	% Clinger	0.73	% Non-Insect	0.59
Trichoptera Taxa	-0.83	% Tanytarsini	-0.71	% Plecoptera	-0.59
Total Taxa	-0.80	% Trichoptera	0.69	% Intolerant	-0.57
Ephemeroptera Taxa	-0.71	% EPT	0.60		
% Burrower	0.70				
Hilsenhoff's Index	0.60				

Of the three principal metric factors, the second factor was most highly correlated with the first factor of the environmental factors (**Table 7, Figure 9**), indicating that percent EPT was positively related to catchment size and stream length. This relationship is not ecologically meaningful and suggests that it is spurious, perhaps due to a relatively small range of values in the environmental factors among reference sites.

Table 7. Correlations among principal factors for natural environmental variables and biological metrics in reference sites of the CPH.

	Ref. Metrics PCA Factor 1	Ref. Metrics PCA Factor 2	Ref. Metrics PCA Factor 3
Natural PCAFactor 1	-0.01	**0.48**	0.02
Natural PCAFactor 2	-0.10	-0.17	0.02
Natural PCAFactor 3	-0.15	0.03	0.03

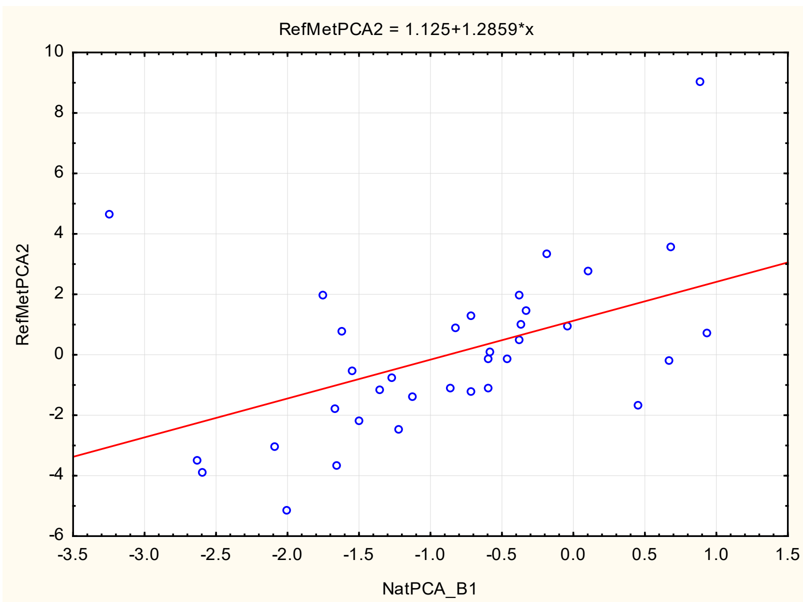

Figure 9. Positive relationship between environmental and biological PCA factors in reference sites.

3.2.4 Correlations

Significant and strong (|r|>0.45) correlations between metrics and potential classification variables were fairly common (**Table 8**). However, we did not adjust metrics to these variables for several reasons. First, some of the environmental variables have short gradients in the reference data set that are not meaningful and should not be extrapolated to the larger range of values found in all sites (see ranges in Table 4). For example, the biological effects of elevation over a range of 100 – 400 feet are presumed to be negligible, despite some significant correlations with metrics. In addition, most of the correlations have absolute correlation coefficients of 0.50 or less, which is not very strong. Some of the relationships are due to a few outliers and the rest of the points do not show strong patterns. After considering adjusting metrics to environmental variables, the idea was abandoned.

3.2.5 Classification Conclusions

The NMS and cluster analyses were inconclusive regarding potential site classes. For the environmental variables related to taxa differences in reference sites, threshold values for distinguishing one site class from another were uncertain. For some variables, the gradient of the values was not long enough to be ecologically meaningful, such as longitude, which separates far western sites from central sites in a single level 4 ecoregion, a distance of at most 20 miles.

- There are no overwhelming or obvious patterns to suggest distinct site classes among the 32 reference sites in these analyses
- While some patterns are evident with longitude, % water cover in the catchment, elevation, and slope, none are strong or consistent throughout the analyses
- Adjustment of individual metrics with environmental variables was not reasonable due to short environmental gradients that were not consistent among reference and non-reference data sets.

Table 8. Spearman Rank order correlation coefficients relating biological metrics with classification variables in reference sites.

Biological Metric	Classification Variable	Correlation Coefficient
Insect Taxa	Elevation	-0.45
Ephemeroptera Taxa	Elevation	-0.48
EPT Taxa	% water	-0.45
Plecoptera Taxa	% water	-0.49
% Non-Insect	Elevation	0.52
% Non-Insect	Latitude	0.48
% Non-Insect	% wetlands	-0.49
% Oligochaeta	Stream Order	-0.45
% Collector	Latitude	-0.45
% Burrower	Longitude	-0.58
% Burrower	% water & wetlands	0.46
% Burrower	% water	0.68
% Sprawler	pH	-0.50
Sprawler Taxa	pH	-0.55
Swimmer Taxa	Elevation	-0.49
Beck's Index	Elevation	-0.50
Intolerant Taxa	Elevation	-0.51
Intolerant Taxa	Latitude	-0.46
Intolerant Taxa	% water	-0.46
Tolerant Taxa	Total Area	-0.51
Tolerant Taxa	Stream Order	-0.45

The decision to include all CPH sites in a single site class was reasonable to RIDEM staff. Without additional reference sites outside of the CPH, the decision on site classification for Lowland sites was based on professional knowledge of the site types in the state. Those sites that were in the Lowland ecoregions but with substantial portions of their watersheds in the CPH were grouped with CPH sites. Sites with more than 50% of their watershed in the lowland ecoregions were thought to be a different site type. When applying this logic, the numbers of reference and stressed sites in the CPH-associated region became 32 and 26 respectively. Sites in the core Lowland regions included four lower-quality reference sites and nine stressed sites.

3.3 Results of Sensitivity Analysis

In each metric category, at least one metric in each class had a DE greater than 75% (**Table 9**). Six metrics had DE >90%, including EPT taxa; Ephemeroptera taxa; % Ephemeroptera and Plecoptera, excluding Baetidae; clinger taxa; Beck's index; % intolerant; and intolerant taxa. In the richness metric category, Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa and Ephemeroptera taxa were the most responsive to stress with DE of 94%. Trichoptera taxa were less responsive to stress, which was surprising because they are usually considered a sensitive group. Coleoptera taxa decreased with stress, but the range of values in this metric was small and therefore the metric scores would be highly variable with small changes in taxa richness.

In the composition category, the % Ephemeroptera and Plecoptera, excluding Baetidae metric had the highest DE. Percent Ephemeroptera and Plecoptera, % Plecoptera, and % Coleoptera were nearly as responsive. The Shannon-Wiener index, % non-insects, % Oligochaeta, and % Ephemeroptera, were also strongly responsive based on both DE and *Z*-scores. Most of the metrics showed responses that were consistent with expectations. However, % Diptera and % Chironomidae had generally higher values in reference sites in comparison to values in stressed sites. Because Diptera and Chironomidae are commonly tolerant of pollution, this response was not expected. Though the response signal was fairly strong (DE = 82%), we were reluctant to use these metrics in index trials because of this unexpected response and the mechanism by which these groups responded to stress was unknown. The range of values for % Plecoptera was too small to be considered in index trials (**Appendix C**).

In the feeding group category, % filterers, % predators, predator taxa, and scraper taxa were most responsive. In the habit category, only clinger taxa responded strongly and as expected. Percent clingers generally decrease with increasing stress, but they were weak increasers in this dataset (DE: 53%). At this DE, they were not considered for inclusion in the index. In the tolerance category, sensitive taxa and individuals were more responsive than tolerant ones. Hydropsychidae of EPT increased with stress, suggesting that these relatively tolerant clinging net-spinners may have affected some other metrics strongly, such as % EPT and % clingers.

Table 9. Metric descriptions and sensitivity in calibration dataset. Metric codes are in bold-type if the metric was included in index trials. DE and Z-scores are bold-typed if they are relatively high (DE>75, |Z|>1.5). Trend indicates whether the metric was decreasing (dec), increasing (inc), or not responsive (NR) with increasing stress.

		CPH		
Metric Name	Metric Code	DE	z	trend
Richness				
Total Taxa	**TotalTax**	**76**	**2.0**	dec
Insect Taxa	**InsecTax**	**88**	**2.2**	dec
EPT Taxa	**EPTTax**	**94**	**1.5**	dec
EP Taxa	EPTax	**94**	**1.7**	dec
Ephemeroptera Taxa	**EphemTax**	**94**	1.4	dec
Plecoptera Taxa	PlecoTax	**88**	1.2	dec
Trichoptera Taxa	TrichTax	53	1.1	dec
Non-insect % of Taxa	**NonInsPT**	**76**	**-1.8**	inc
Chironomidae Taxa	ChiroTax	59	1.4	dec
Coleoptera Taxa	ColeoTax	**88**	1.4	dec
Crustacea & Mollusca Taxa	CrMolTax	35	-1.0	inc
Diptera Taxa	DipTax	**76**	1.4	dec
Oligochaeta Taxa	OligoTax	47	-0.6	inc
Orthocladiinae Taxa	OrthoTax		-0.7	NR
Tanytarsini Taxa	TanytTax			NR
Composition				
Shannon-Weiner Index (base 2)	**Shan_2**	**76**	**1.6**	dec
% Dominant Taxon	1DomPct	53	-0.9	inc
% EPT	EPTPct	47	-0.7	inc
% Ephemeroptera	**EphemPct**	**82**	1.3	dec
% Plecoptera	PlecoPct	**88**	0.9	dec
% Trichoptera	TrichPct	53	**-1.6**	inc
% EPT, no Hydropsychidae & Baetidae	pEPTnoHB	**76**	1.1	dec
% EP, excluding Baetidae	**pEPnoB**	**94**	1.3	dec
% EP	pEP	**88**	1.3	dec
% Diptera	DipPct	**82**	1.4	dec
% Chironomidae	ChiroPct	**82**	**1.6**	dec
% Cricotopus & Chironomus of Chiron.	CrCh2ChiPct			NR
% Orthocladiinae of Chironomidae	Orth2ChiPct	53	-0.5	inc
% Tanytarsini	TanytPct	71	0.8	dec
% Tanytarsini of Chironomidae	Tnyt2ChiPct	47	0.3	dec
% Coleoptera	ColeoPct	**88**	1.0	dec
% Odonata	OdonPct	**76**	0.6	dec

Metric Name	Metric Code	CPH DE	z	trend
% Non-Insect	**NonInPct**	**88**	**-2.2**	inc
% Amphipoda	AmphPct	47	**-7.1**	inc
% Bivalvia	BivalPct	29	-0.3	inc
% Crustacea & Mollusca	CrMolPct	41	**-1.9**	inc
% Gastropoda	GastrPct	41	**-3.1**	inc
% Isopods	IsoPct		-1.3	NR
% Oligochaeta	OligoPct	**76**	**-1.9**	inc
<u>Feeding</u>				
% Collector	CllctPct	35	0.4	dec
% Filterer	**FiltrPct**	71	**-1.8**	inc
% Predator	**PredPct**	**88**	**2.1**	dec
% Scraper	ScrapPct	**76**	1.0	dec
% Shredder	ShredPct	53	0.4	dec
Collector Taxa	CllctTax	29	0.6	dec
Filterer Taxa	FiltrTax	29	0.4	dec
Predator Taxa	**PredTax**	**88**	**2.8**	dec
Scraper Taxa	**ScrapTax**	**82**	**1.5**	dec
Shredder Taxa	ShredTax	29	0.8	dec
<u>Habit</u>				
% Burrower	BrrwrPct	53	-0.8	inc
% Climber	ClmbrPct	53	-0.2	dec
% Clinger	ClngrPct	53	-0.6	inc
% Sprawler	SprwlPct	35	**-1.7**	inc
% Swimmer	SwmmrPct	71	0.0	dec
Burrower Taxa	BrrwrTax	29	0.5	dec
Climber Taxa	ClmbrTax	35	0.3	dec
Clinger Taxa	**ClngrTax**	**94**	**1.5**	dec
Sprawler Taxa	SprwlTax		0.3	NR
Swimmer Taxa	SwmmrTax	71	1.0	dec
<u>Tolerance</u>				
Beck's Index	**BeckBI**	**94**	**2.3**	dec
Hilsenhoff's Index	HBI	35	-0.2	inc
% Baetidae of Ephemeroptera	Baet2EphPct		-1.0	inc
% Hydropsychidae of EPT	Hyd2EPTPct	**82**	**-1.5**	inc
% Hydropsychidae of Trichoptera	Hyd2TriPct	47	-0.5	inc
% Intolerant	**IntolPct**	**94**	**1.5**	dec
% Tolerant	TolerPct	53	-0.9	inc
Intolerant Taxa	**IntolTax**	**94**	**2.3**	dec
Tolerant Taxa	TolerTax	29	-0.1	dec

3.4 Index Composition Results

Of the 69 metrics calculated, we selected 17 that were candidate for inclusion in the index because they were responsive to stress and represented various components of the assemblage (**Table 9**). We evaluated more than 25,000 index alternatives with 3-9 metrics that excluded redundant metric pairs. In the calibration data, 536 index alternatives had excellent performance characteristics, with DE = 100, Z-score > 2.9, all five metric categories included, and no redundant or conceptually redundant metrics such as EPT taxa and Ephemeroptera taxa. The final index was selected based on the best performance statistics, considerations of the metrics included, and consensus among analysts.

The selected index included metrics from all five metric categories (**Table 10**), had a calibration DE of 100%, and a *Z*-score of 3.0. Discrimination of reference from stressed sites was adequate in calibration data (**Figure 10**). As illustrated in the figure, independent verification data confirmed that the index was robust in distinguishing reference from stressed sites. All of the stressed verification sites are below the 25th percentile of calibration reference sites and all of the verification reference sites are within the non-outlier range of the reference sites. Assessment error for the MBCI is evaluated in Section 3.5.3. The index metrics were not redundant, with the highest correlations existing between total taxa, Beck's index, and clinger taxa (**Table 11**).

Table 10. Metrics of the Macroinvertebrate Biological Condition Index (MBCI), including discrimination efficiency (DE), trend with increasing stress (increasing or decreasing), and scoring formula.

Metric	Metric Category	DE	Trend	Scoring Formula[a]
Total Taxa	Richness	76	Decreasing	100*metric value/32.8
% Non-insect	Composition	88	Increasing	100*(46.3-metric value)/(46.3)
Beck's Index	Tolerance	94	Decreasing	100*metric value/24.8
Clinger Taxa	Habit	94	Decreasing	100*metric value/18
% Predators	Feeding Group	88	Decreasing	100*metric value/22.7
% Filterers	Feeding Group	71	Increasing	100*(83.1-metric value)/(80.8)

a: If the calculated score was outside of the valid scoring range of 0-100, the score was re-set to the nearest extreme before averaging all scores to arrive at the index score.

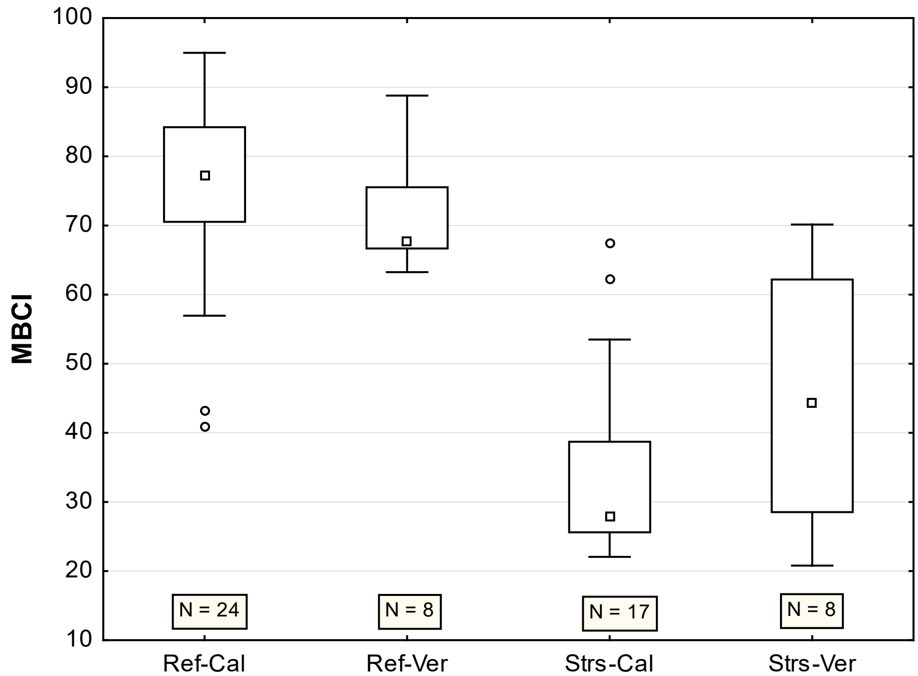

Figure 10. Distributions of MBCI values in categories of reference (ref), stressed (strs), calibration (cal), and verification (ver).

Table 11. Correlations among index metrics (Spearman rank r).

	TotalTax	NonInPct	BeckBI	ClngrTax	PredPct	FiltrPct
Total Taxa	1					
% Non-insect	-0.03	1				
Beck's Index	0.84	-0.36	1			
Clinger Taxa	0.75	-0.42	0.81	1		
% Predators	0.49	0.20	0.50	0.27	1	
% Filterers	-0.39	-0.17	-0.36	0.00	-0.56	1

3.5 Results for applying the Index

The index recommended in this report is appropriate for assessing benthic macroinvertebrate conditions in wadeable streams of RI in which the dominant ecoregion in the watershed is the coastal plains and hills (CPH) ecoregion. Samples must be collected and processed in accordance with RIDEM QAPP for Taxonomic Identification of Benthic Macroinvertebrates, RI (ESS 2007; http://www.dem.ri.gov/pubs/qapp/taxbenth.pdf) and taken during adequate flow conditions (aquatic base flow [ABF] must be ≥ .32 cfsm). Metric calculations must use conventions and taxonomic attributes as defined by RIDEM. Conventions include discounting ambiguous taxa for richness metrics and considering appropriate assessment for small samples (**Appendix D**). Impairment or condition designations are recommended in this report (Section 3.5.3), though these recommendations must be approved by RIDEM before they can be used to justify management actions. The index should not be used for impairment designations in streams that are predominantly in the lowland ecoregions (with <50% of their watersheds in the CPH). However, because we expect that the CPH index will give consistently lower scores in the Lowlands, Lowland samples with index scores above impairment thresholds used in the CPH could be assumed to be meeting at least those acceptable conditions. The danger is in identifying impairment in the Lowlands, where more error is expected at this time.

At the outset of the analysis, RIDEM had questions about the applicability of the index in streams with small catchments and below dams. Because reference sites included streams with small catchments and below dams and the site classification exercise did not identify catchment size as a determinant of conditions, the index can be applied in small streams and below dams. The range of stream catchment sizes in the sites used in calibration was from 0.3 to 48.5 square miles. The effect of dams was not addressed in the classification exercise. Dams are so common in Rhode Island that excluding reference sites based on their presence would limit the analysis excessively. To control for drought conditions that may keep lake levels low and restrict flow over dams, application of the index is limited to streams below dams that are meeting ABF.

In addition, the reference sites were smaller than non-reference sites on average and reference streams with large catchments were not represented in the calibration data (**Table 4**). As catchment size increases in RI, it is more likely that development has intensified around the stream sites, especially in catchment sizes greater than 20 square miles. This does not necessarily prevent assessment of large catchments using the index. Instead, it suggests that conditions in streams with large catchments are usually less than reference quality in RI. Other variables that are under-represented in the reference sites compared to non-reference sites are sites in low elevations (<60 feet), although these elevations tend to fall within the coastal lowland ecoregions. Future sampling efforts should target these under-represented types to confirm that streams with large catchments or low elevations and minimal disturbance are not of different site classes than the streams with smaller catchments and higher elevations.

3.5.1 MBCI Application Process

1. Collect and process benthic samples with ancillary site observations and GIS analyses
2. Calculate the MBCI
    a. Enter/import data into BioQual and calculate the MBCI using the analytical functions, or

b. Calculate the MBCI manually
    i. List taxa, numbers in the sample, and attributes provided by RIDEM.
    ii. If there are <80 or >120 individuals in the sample, see **Appendix D**.
    iii. Calculate metrics, excluding redundant taxa (see Section 2.4).
    iv. Score metrics according to **Table 10**, on a scale of 0-100.
    v. Average scores to arrive at the MBCI.

3. Consider application caveats for sites in the core-Lowland areas, large catchments (>20 square miles), low areas (<60 feet), or flow below ABF.
4. Rate the site biological conditions by comparing the MBCI score to the BCG model (**Table 15**) or other impairment thresholds approved by RIDEM.
5. Report the MBCI score and rating with performance characteristics (DE, confidence intervals, and Type 1 and 2 error rates associated with impairment thresholds)

3.5.2 Precision Analysis

Repeated samples that spanned all geographic regions of Rhode Island and all reference conditions were used to test the reliability of metrics and the index. Replicates were collected at the same site on the same day to estimate sampling variability (same-day replicates, 18 samples from 9 sites) or at the same site in different years to estimate temporal variability (annual replicates, 214 samples from 91 sites). In addition, variability of index scores was described among waterbody IDs, using data from 74 sites in 30 waterbody IDs.

For the MBCI, the CV and CI90 were 15 percent and ±12.1 index units, respectively, for repeated samples on the same day. For annual replicates, the CV and CI90 were 15 percent and ±13.4 index units, respectively. Therefore, we can expect that the true mean MBCI value from a single observation will be within 13 index points of the observation in 90% of the cases. Variability was higher for almost all component metrics compared to the MBCI (**Table 12**). Based on CV, the most precise metrics for same-day replicates were total taxa, the Shannon-Wiener index, % EPT, % Trichoptera, filterer taxa, % clingers, and clinger taxa. For annual replicates all metrics had higher CV than the index. Metrics with the greatest variability relative to the mean were those counting percentages of uncommon taxonomic groups or percentages in relative to certain groups.

Table 12. Metric and index precision for samples replicated over multiple years (annual) or on the same day. The Root Mean Square Error (RMSE) is an estimate of the standard deviation within replicate sets. The 90% confidence interval (CI90) is the range around a single observation in which we expect to find the true mean value in 90% of cases. The mean is for those samples included in the analysis (Annual $N = 91$ sets, Same Day $N = 9$ pairs). The coefficient of variation is a standardized measure of variability (CV = RMSE*100/mean). Low CV are preferable for index development and are highlighted in blue. High CV are in red.

	Annual				Same Day			
	RMSE	CI90	Mean	CV	RMSE	CI90	Mean	CV
TotalTax	4.3	7.0	20.0	21.3	1.9	3.1	18.5	10.1
EPTTax	2.2	3.6	7.3	30.4	1.7	2.7	6.7	25.0
EphemTax	0.8	1.4	1.9	45.7	0.9	1.6	2.0	47.1
PlecoTax	0.7	1.2	0.8	87.4	0.6	0.9	0.6	103.9
TrichTax	1.6	2.6	4.6	33.6	0.9	1.5	4.1	21.5
InsecTax	3.9	6.4	17.1	22.9	2.6	4.3	16.2	16.2
NonInsPT	8.6	14.2	16.0	54.0	6.8	11.2	14.0	48.4
ChiroTax	1.6	2.6	3.8	41.8	0.9	1.5	3.8	23.3
ColeoTax	1.2	2.0	2.8	43.1	0.9	1.6	2.7	35.4
CrMolTax	1.0	1.7	1.4	73.6	0.7	1.2	0.8	95.8
DipTax	2.0	3.3	5.6	35.6	0.9	1.5	5.6	16.3
OligoTax	0.7	1.1	0.8	85.2	0.6	0.9	0.9	65.0
OrthoTax	0.9	1.5	1.3	69.3	0.8	1.3	1.5	52.1
TanytTax	0.6	1.0	1.0	61.4	0.5	0.8	1.1	42.4
Shan_2	0.5	0.8	3.5	13.7	0.3	0.5	3.4	9.3
AmphPct	5.0	8.2	2.1	240.1	0.4	0.6	0.1	357.9
BivalPct	3.9	6.5	1.7	229.9	1.8	3.0	1.0	185.9
ChiroPct	14.1	23.2	23.8	59.4	7.9	13.0	14.2	55.5
ColeoPct	8.1	13.4	12.3	66.0	5.6	9.3	9.7	58.0
CrCh2ChiPct	6.0	9.8	1.0	612.4	11.9	19.6	3.2	372.0
CrMolPct	11.2	18.5	7.5	149.3	2.1	3.4	1.4	153.1
DipPct	15.5	25.4	29.3	52.7	9.9	16.3	20.2	49.1
EphemPct	4.8	7.9	7.0	68.7	4.2	6.9	5.7	73.1
EPTPct	15.9	26.2	42.6	37.4	7.9	13.1	59.0	13.5
pEP	5.3	8.8	8.8	60.4	4.1	6.7	6.7	61.0
pEPnoB	4.4	7.3	7.0	63.3	1.9	3.1	5.1	36.4
pEPTnoHB	9.1	15.0	18.7	48.7	6.7	11.0	18.9	35.6
GastrPct	6.7	11.0	1.7	393.6	0.4	0.7	0.2	182.4
IsoPct	6.3	10.4	1.9	337.3	0.2	0.3	0.0	424.3
NonInPct	13.9	22.9	13.2	105.7	10.8	17.7	9.1	118.4
OdonPct	1.6	2.7	1.2	138.9	0.6	1.0	0.5	126.2
OligoPct	7.5	12.4	4.6	162.3	9.8	16.1	6.3	154.7
Orth2ChiPct	24.9	41.0	40.3	61.9	22.1	36.3	46.7	47.3

	Annual				Same Day			
	RMSE	CI90	Mean	CV	RMSE	CI90	Mean	CV
PlecoPct	2.4	4.0	1.8	132.5	1.0	1.7	1.0	108.0
TanytPct	7.9	13.0	7.2	109.8	3.8	6.2	5.2	72.0
Tnyt2ChiPct	22.7	37.3	26.9	84.4	15.1	24.8	34.1	44.2
TrichPct	14.9	24.5	33.8	44.0	6.4	10.5	52.3	12.2
CllctPct	16.8	27.6	32.2	52.1	9.6	15.8	17.7	54.1
FiltrPct	17.1	28.1	39.4	43.3	9.8	16.2	61.2	16.1
PredPct	4.2	6.9	7.6	55.2	4.1	6.8	5.4	76.3
ScrapPct	11.5	18.9	17.4	66.1	4.9	8.1	11.5	42.8
ShredPct	4.7	7.8	4.2	113.6	4.1	6.8	4.0	102.3
CllctTax	1.9	3.1	5.5	34.5	1.3	2.1	4.5	28.2
FiltrTax	1.6	2.6	4.8	33.5	0.7	1.2	5.6	12.6
PredTax	1.7	2.8	4.1	41.7	1.2	1.9	3.5	33.7
ScrapTax	1.5	2.5	3.9	38.7	0.6	1.0	3.1	20.4
ShredTax	1.1	1.8	1.6	66.7	1.0	1.7	1.7	59.7
BrrwrPct	9.4	15.4	8.9	105.5	10.2	16.7	8.7	117.0
ClmbrPct	3.8	6.3	2.5	154.0	3.9	6.5	2.0	197.9
ClngrPct	17.3	28.4	58.6	29.4	10.0	16.4	76.7	13.0
SprwlPct	6.8	11.1	5.5	124.3	2.2	3.7	4.2	52.9
SwmmrPct	3.5	5.8	2.7	129.1	3.3	5.4	2.2	149.5
BrrwrTax	1.4	2.3	2.4	56.6	1.0	1.7	2.2	47.4
ClmbrTax	0.8	1.3	0.9	85.7	0.8	1.3	0.9	91.9
ClngrTax	3.0	4.9	10.4	28.6	1.6	2.7	10.9	15.0
SprwlTax	1.5	2.5	2.0	78.4	0.7	1.2	2.3	31.9
SwmmrTax	0.6	1.0	0.9	72.2	0.7	1.2	0.8	95.8
BeckBI	3.6	5.9	11.3	32.0	3.5	5.8	10.1	34.9
HBI	0.7	1.2	4.3	17.1	0.8	1.3	3.8	21.0
1DomPct	11.4	18.8	28.0	40.7	6.2	10.2	27.1	22.8
Baet2EphPct	26.0	42.8	22.5	115.6	26.8	44.0	14.4	186.1
Hyd2EPTPct	22.8	37.5	48.6	47.0	11.9	19.5	60.4	19.6
Hyd2TriPct	25.3	41.5	61.9	40.8	14.8	24.3	70.7	20.9
IntolPct	9.3	15.3	20.7	44.9	5.6	9.2	12.4	45.0
TolerPct	12.5	20.6	13.8	90.9	9.3	15.4	8.8	106.2
IntolTax	2.4	3.9	6.5	36.2	1.9	3.1	5.6	33.9
TolerTax	1.4	2.2	2.9	46.3	1.1	1.8	2.4	45.2
MBCI	8.1	13.4	54.3	15.0	7.3	12.1	48.2	15.2

The RMSE of the MBCI for waterbody units was 10.3, which yields a 90% confidence interval of ±16.9 index units. Power analysis was used to estimate the number of samples necessary to detect a 20 percent difference in mean MBCI values among waterbody units or in a single unit over time. With 90% confidence and 80% power, 10 sites per waterbody unit should be sampled.

Because the waterbody units are small and such intensive sampling is not feasible, alternatives should be considered. The assessment unit for statistical comparisons could be expanded to cover a larger land area and more sites. If a river basin or hydrologic unit was used as an assessment unit, then it is more likely that RIDEM has already collected enough samples to make comparisons with the appropriate power, confidence, and detection limits. We used the first seven characters of the WBID code to denote groups of related sites and calculated an RMSE of 16 index points. This yields a 90% confidence interval of ±26.3 index units for a single observation. To detect a 20 point difference among such site groups with 90% confidence and 80% power, 24 sites per site group should be sampled.

If RIDEM reports assessment results in terms of the percent of a waterbody unit that is impaired, then the number of samples collected per assessment unit will affect the certainty of the reported percentage. The variability associated with proportional statistics is dependent on the sample size and the reported proportion. The standard deviation of a proportion is calculated as:

$$s = \left(\frac{p\,(1-p)}{n}\right)^{0.5}$$

where p is the reported proportion and n is the sample size that went into calculating the proportion (Snedecor and Cochran 1967). From this equation, it is possible to calculate confidence intervals around estimates of the percentage of impaired water resources in a set of samples. **Table 13** displays estimates of the confidence intervals around percentages of impaired sites for three sampling scenarios. Because the confidence interval is so broad for small sample sizes, RIDEM would have more confidence in proportions derived for more samples, larger areas of the state, or multiple years of sampling. For example, in a typical sampling year including 30 sites, if 30% of the resource is impaired, then the 90% confidence interval is from 16.2 – 43.8% (30 ±13.8%). These estimates assume that there is no bias in the samples assessed and would be best applied in a probability-based monitoring design.

3.5.3 BCG and MMI Concordance

The BCG model developed for high gradient streams in CT is applicable in RI CPH streams. However, the CT model was calibrated on a 200 organism subsample and the RI samples target only 100 organisms. Therefore, the model rules that refer to counts of taxa could over-estimate the BCG level assignment (indicating worse conditions) for RI samples. Without a recalibration of the BCG to a smaller sample size, these rules cannot be adjusted to account for the difference in sample sizes. Instead, we applied the CT BCG rules (**Table 14**) without adjustment and interpret results in light of the possible error. For each BCG level, a rule in the model concerns the total number of taxa in a sample. An additional rule for BCG levels 3 and 4 addresses the count of sensitive taxa with Attributes II and III. Because the shape of taxa saturation curves tend to rise steeply and gradually level off, we expect that BCG level assignments may be at most one level higher in a 100 organism sample compared to a 200 organism sample. The observed relative differences between 100-organism sample BCG levels are expected to be comparable to differences that would be observed with 200-organism samples.

Table 13. Variability (s = standard deviation, CI90 = 90% confidence interval) associated with percent impairment in different assessment scenarios.

	For 10 sites		For 30 sites		For 150 sites	
% Impaired	s	CI90	s	CI90	s	CI90
10	9.5	± 15.6%	5.5	± 9.0%	2.4	± 4.0%
20	12.6	± 20.8%	7.3	± 12.0%	3.3	± 5.4%
30	14.5	± 23.8%	8.4	± 13.8%	3.7	± 6.2%
40	15.5	± 25.5%	8.9	± 14.7%	4.0	± 6.6%
50	15.8	± 26.0%	9.1	± 15.0%	4.1	± 6.7%
60	15.5	± 25.5%	8.9	± 14.7%	4.0	± 6.6%
70	14.5	± 23.8%	8.4	± 13.8%	3.7	± 6.2%
80	12.6	± 20.8%	7.3	± 12.0%	3.3	± 5.4%
90	9.5	± 15.6%	5.5	± 9.0%	2.4	± 4.0%

Table 14. Decision rules for Connecticut High Gradient Streams (Gerritsen and Jessup 2007). Ranges in parentheses denote fuzzy membership function.

Attributes	BCG Level					
	1	2	3	4	5	6
0 General		2.1 Total taxa > (25–30) 2.2 count > (50–60%) of target	3.1 Total taxa > (19–23) 3.2 count > (50–60%) of target	4.1 Total taxa > (17–21) 4.2 count > (50– 60%) of target	5.1 Total taxa > (8–12) 5.2 count > (50– 60%) of target	Total taxa < (8–12) count < (45–55%) of target
I *Endemics*		(no rule)	(no rule)	(no rule)	(no rule)	
II *Highly sensitive taxa*		2.3 Taxa II > (3–5)				
III *Sensitive taxa*		2.4 % Taxa (II+III) > (45–55%) 2.5 % Indiv (II + III) > (30–40%)	3.3 Taxa (II+III) > (8–10) 3.4 % Indiv (II+III) > (30–40%)	4.3 Taxa (II+III) > (3–5) 4.4 % Indiv (II+III) > (10–20%)		
IV *Intermediate tolerant taxa*		(no rule)	(no rule)	(no rule)	(no rule)	
V *Tolerant taxa* (all)		2.6 % Indiv V < (10–15)%	3.5 % Indiv V < (40–50%)	4.5 % Indiv V < (65–75%)		
Indicator Taxa		[E taxa > 2]		[E taxa > 0]		
Combining Rule		2.1, 2.2, 2.3, 2.4 and (2.5 or 2.6)	Fails any of rules 2.2-2.6, and 3.1, 3.2, 3.3, and (3.4 or 3.5)	Fails any level 2 rules 2.2–2.6 and fails level 3 rules 3.3–3.5 and 4.1, 4.2, 4.3, and (4.4 or 4.5)	Fails level 2 rules 2.2–2.6, and level 3 rules 3.2–3.5 and level 4 rules 4.2–4.5, and 5.1 and 5.2	Fails all higher levels

Sites that were defined as reference, other, or stressed based on environmental criteria had median BCG scores of 3, 4, and 5, respectively (**Figure 11**). At level 2, there are minimal changes in the structure of the biotic community and minimal changes in ecosystem function. Only a few samples in the CPH of RI were assessed at level 2 using the CT BCG model. Several of the reference samples were assessed as level 3, in which changes in the structure of the biotic community are evident and changes in ecosystem function are minimal. Even after accounting for the potential error associated with the smaller subsample target in RI, many of the reference sites are probably in BCG level 3. On the opposite end of the scale, most stressed sites were assessed at BCG level 5, which indicates major changes in structure of the biotic community and moderate changes in ecosystem function. If there were assessment errors due to subsample target size, then some of these might be assessed as level 4, though it is likely that the median would remain at level 5.

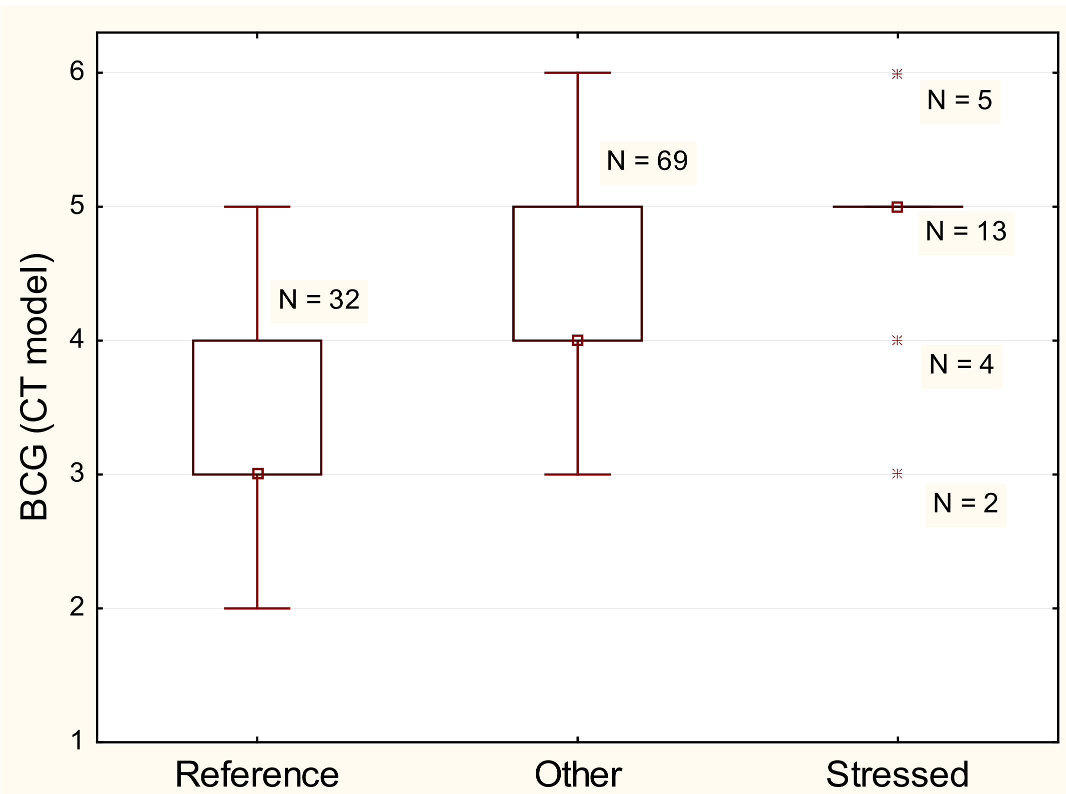

Figure 11. BCG levels calculated for Rhode Island samples using the Connecticut BCG model.

The MMI uses metrics that are similar in objective to the BCG rules, but calculated somewhat differently (e.g., EPT taxa in the MMI includes taxa considered to be Attributes II, III, IV; and Attribute II includes taxa from the EPT orders, as well as a few dipteran and beetle taxa). The total MMI score is based on the average of all metrics, while BCG decisions are based on decision-specific critical attribute groups; e.g., Attributes II and III for the higher tiers and Attribute V for lower tiers. Concordance of the two assessment endpoints is strong, with higher MMI scores associated with lower BCG levels (**Figure 12**). In application of the MBCI, MBCI scores could be used to estimate BCG levels, using ranges of MBCI scores to indicate the levels (**Table 15**). The estimated BCG levels and MMI scores could then be interpreted as degrees of biological impairment to inform watershed assessments and management actions. The scoring ranges were derived from the inter-quartile ranges of MBCI values in each BCG level seen in **Figure 12** with rounding adjustments.

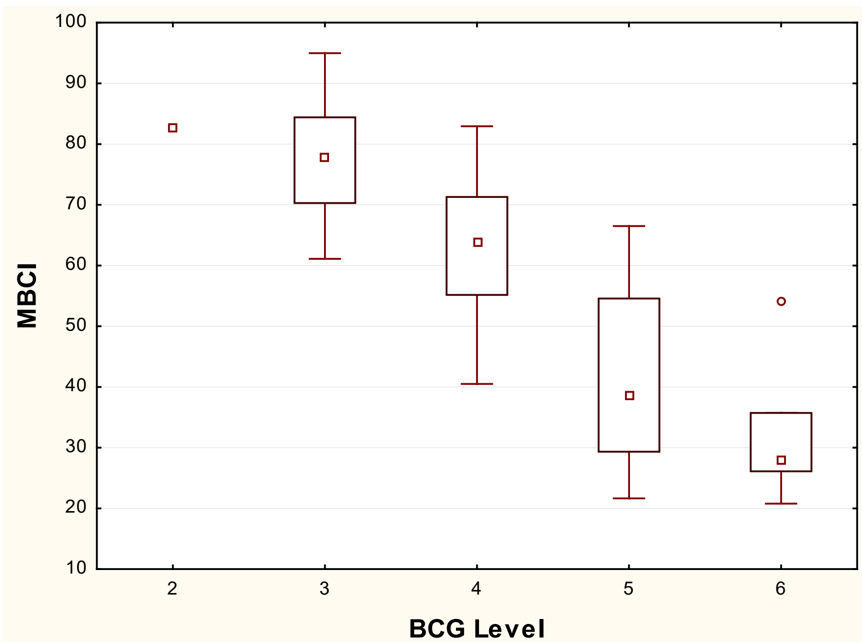

Figure 12. Rhode Island MBCI value distributions by Connecticut BCG levels estimated from the decision analysis model (Gerritsen and Jessup 2007).

Table 15. Ranges of MBCI values corresponding to BCG levels.

BCG Level	MBCI Scoring Range	Narrative Interpretation of Biological Conditions
1, 2	MBCI > 85	Natural conditions
3	85 ≥ MBCI > 70	Slightly impacted but functional
4	70 ≥ MBCI > 55	Moderately impacted but functional
5	55 ≥ MBCI > 35	Impacted with loss of functions
6	MBCI ≤ 35	Severely impacted and nonfunctional

Using the thresholds suggested in Table 15, we can quantify Type I and Type II error for each threshold and for calibration and verification data (**Table 16**). This serves two purposes. It helps verify that the model performs adequately and it gives a measure of uncertainty to communicate with assessment results. For model verification, the Type I and Type II errors for the threshold between levels 3 and 4 are 62.5 and 12.5, respectively. Between levels 4 and 5, the Type I and Type II errors are 0% and 37.5%, respectively. The verification data for reference sites are clumped just below the 2-3 threshold (see **Figure 10**). This gives a high Type I error rate if the level 2-3 threshold is used to indicate impairment, though the errors are minor in magnitude. In application, index values that are close to a threshold may be given less credibility until they are confirmed with a second or third sample). If the BCG framework is used for assessment, then the differences in management actions triggered at each level are not as stark as those that might be associated with a one threshold system of attainment and impairment. In other BCG models, errors of one level difference are most common (e.g., a level 2 sample predicted by a model to be level 3) (Gerritsen and Jessup 2007, Snook et al. 2007). We suspect this to be the case with the Rhode Island index as well, where many of the reference verification index values are just below the 2-3 threshold (which is also the 25th percentile of the calibration reference values).

Table 16. Sample enumeration by reference status, calibration status, and the BCG level indicated by index results.

	BCG	Reference		Stressed	
	Level	Calibration	Verification	Calibration	Verification
MBCI > 85	2	6	1	0	0
85 ≥ MBCI > 70	3	12	2	0	1
70 ≥ MBCI > 55	4	4	5	1	2
55 ≥ MBCI > 35	5	2	0	4	2
MBCI ≤ 35	6	0	0	11	3

3.5.4 Lowland Stream Assessments

Streams in the east and south of the state were represented by few reference sites and were assumed to be of a different site type. The streams that primarily lay within the Narragansett-Bristol Lowlands and the Long Island Sound Coastal Lowland ecoregions are low gradient, slow velocity, sandy or silty, and generally surrounded by intensive land uses. While there were four second-tier reference sites and nine stressed sites identified in these regions, these numbers are only adequate to give anecdotal evidence that an index may be effective in future assessments.

The evidence is that the index developed for the CPH is also responsive to stress in the Lowland regions (**Figure 13**), with a DE of 67%. This response should be tested with additional data before any index could be used confidently in assessments of Lowland streams. We suspect that Lowland streams may score lower than CPH streams with comparable stressor intensities because of natural site differences. This suspicion is difficult to test without adequate reference site representation in the Lowland areas. The CT BCG model does not help in characterizing conditions in the Lowland regions because it was calibrated for high gradient streams only.

Figure 13. MBCI value distributions in reference and stressed Lowland sites.

The Massachusetts DEP performed a pilot study in which an index was calibrated for the Narragansett Bristol Lowlands in MA (Jessup et al. 2000). That analysis was also based on a small number of samples. The metrics used in the MADEP index were only partly responsive in the small RI Lowland dataset (**Table 16**). However, when applied in the Rhode Island Lowland sites, the MADEP Lowland index had a respectable DE of 78% (**Figure 14**).

Table 16. Metrics proposed in the MADEP pilot Lowland index (Jessup et al. 2000), with discrimination efficiency (DE) and *Z*-score in the RI Lowland sites.

Metric	DE	*Z*-score
Number of insect taxa	77.8	2.13
Number of non-insect taxa	55.6	-0.42
% Plecoptera individuals	0.0	0.55
Hilsenhoff's Biotic Index	77.8	-1.62
% filterers	33.3	-0.21
% predators	33.3	0.45

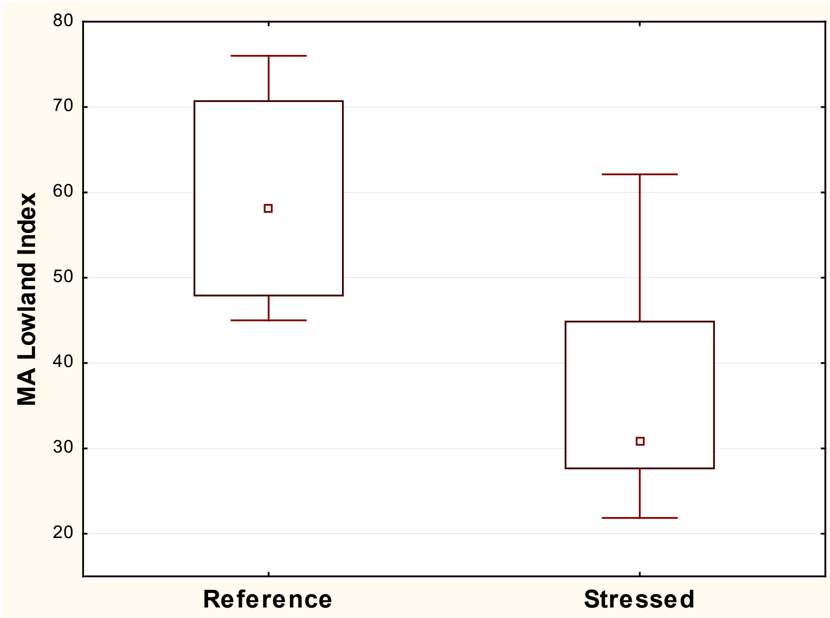

Figure 14. MADEP Lowland Index value distributions in reference and stressed Lowland sites.

Development of benthic macroinvertebrate indices for low gradient streams in the northeast has been slow relative to development of indices for higher gradient streams. The CT and New England BCG models (Gerritsen and Jessup 2007, Snook et al. 2007) were calibrated with high gradient streams only. The MADEP Lowland index (Jessup 2000) was calibrated with too few sites to confirm its reliability. New Hampshire has a coastal plains site class, but the index was not calibrated specifically to that region (David Neils, NHDES, personal communication). Maine DEP assessment methods do not recognize stream gradient as a classification issue, either because it not an issue in Maine or because their methods transcend stream gradients. Vermont recognizes “slow winders” as a stream type, but does not have enough reference site data among those streams.

Because of the lack of emphasis on low gradient streams in New England, additional effort is needed in Rhode Island and the region. Identification of low gradient reference streams is a current obstacle for index development. If RIDEM targeted Lowland streams with minimal disturbance for sampling, index development may be possible in the future. Based on general principles, a minimum of 10 reference sites would be required. An optimal number for calibration and validation would be closer to 30 for reference sites and an additional 30 for stressed sites. Cooperative efforts with CT and MA may be warranted to identify appropriate reference sites in the Lowland regions, where sufficient numbers of undisturbed sites may be difficult to find within any single state. Existing RIDEM protocols may be most appropriate for riffle dominated systems, which are not consistently encountered in low gradient streams.

Therefore, regional cooperation will also be useful in establishing appropriate sampling protocols for low gradient streams in New England, as it has in the Mid-Atlantic States (Maxted et al. 2000, U.S. EPA 1997).

4.0 Discussion and Conclusions

The MBCI is a robust indicator that RIDEM could use to assess streams for attainment of aquatic life uses. The MBCI can be compared to the BCG scale so that multiple degrees of use attainment or impairment could be discerned. While these analyses show that the MBCI is correlated with stressor gradients, the index could still be improved through future monitoring, analysis with stressors, and index recalibration.

We did not use field habitat data to define reference sites. Because the reference-stressed gradient was defined without consideration of habitat conditions, the habitat scores for individual variables and the total habitat score could be evaluated for responsiveness along this scale. The total habitat score could be refined to indicate reference or stressed conditions. In addition, the MBCI was compared to the existing total habitat score (**Figure 15**). The MBCI is apparently limited by habitat conditions, as shown by the slope of the upper values in the bi-plot. For example, if the total habitat score is below 120, we do not expect MBCI scores above 70 index units. Values in the "heel of the wedge" (lower right side in the figure) are probably stressed by something other than or in addition to habitat.

Metals and nutrient data were collected at some sites but were not used in reference site designations. When those data were compared to the MBCI, it was clear that some variables could be stressors to the benthic assemblage (**Appendix E**). If and when the MBCI is recalibrated, consideration of water chemistry variables might allow for a more robust definition of the reference and stressed sites. The associations illustrated in the figure may also serve to suggest which of the water chemistry variables are associated with biological stress and at which concentrations the stresses become persistent.

As with habitat and water chemistry, base flow conditions were associated with the MBCI and could be used in the future to refine the stressor gradient for index recalibration. The RI Aquatic Base Flow of 0.32 cfs/mi^2 is a benchmark of minimal flow. When plotted against the MBCI, ABF was not strongly associated with biological conditions and several reference samples had high MBCI scores even when ABF was low (**Figure 16**). In assessing sites, RIDEM requires the site to meet ABF before making an assessment, so if it is a drought year, the assessment is considered to be tentative.

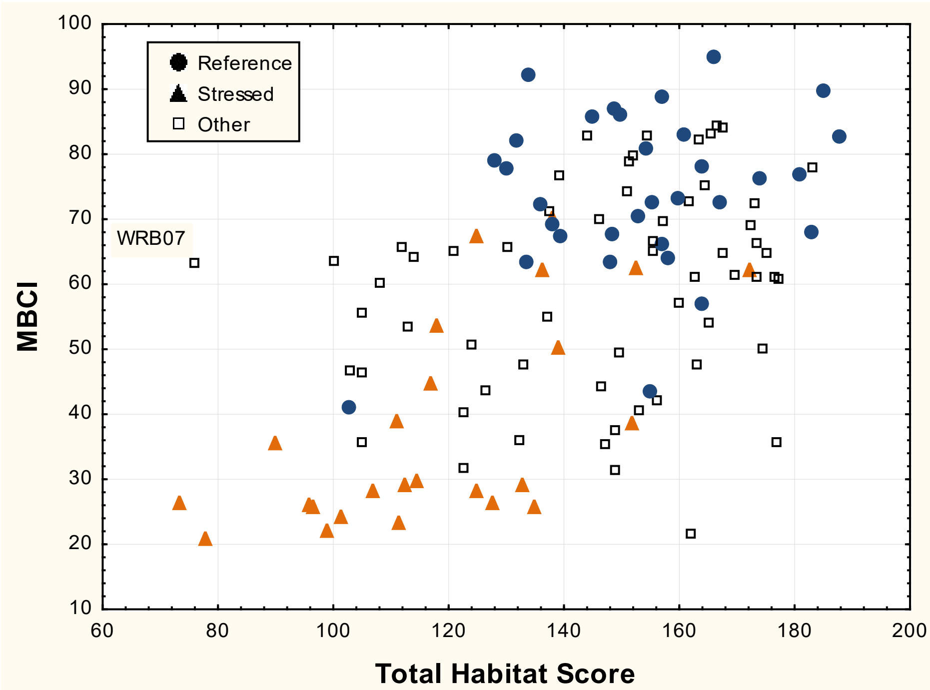

Figure 15. MBCI values in relation to total habitat scores. The outlier is a site in the Lower Wood River basin (WRB07).

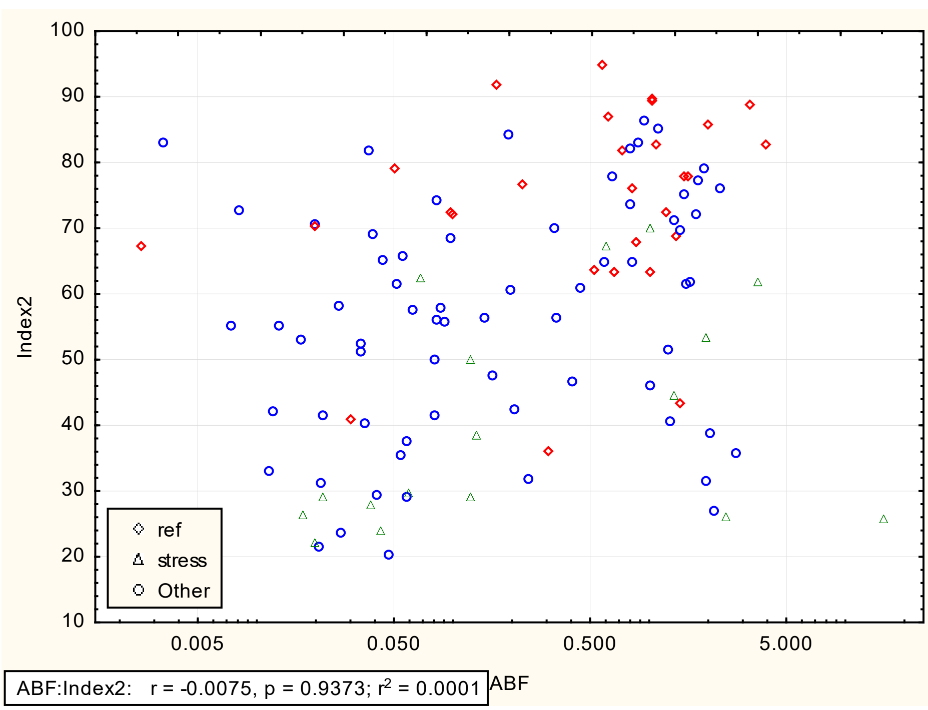

Figure 16. MBCI values in relation Aquatic Base Flow (ABF).

We can set expectations for all sites based on the observed condition in the best available sites, but the absolute conditions in the sites are best interpreted in the context of the BCG. The BCG model developed in Connecticut for similar samples (albeit with a larger subsample size) can be used to interpret biological condition and ecosystem function in sites that are less than pristine or even disturbed to greater degrees. The Connecticut BCG model applied in the CPH-associated RI sites indicated that the biological conditions were generally in accordance with the reference and stressed designations (**Figure 11**). The RI and CT sites may differ in other ways that could make the comparison questionable. For example, the CT samples may be from generally higher gradient streams than those sampled in RI. If that was found to be true, then it may be that the BCG levels predicted through association with the MBCI would be inaccurate and could bias biological indications towards higher BCG levels (worse conditions). Such possible bias could be investigated in the future through interstate cooperation on monitoring and analysis.

In the future, sites should be targeted to facilitate comparisons of environmental effects with minimal human disturbance. In particular, the site types that were noticeably missing and which could not be analyzed for site classification included relatively undisturbed CPH-associated sites from large catchments (>20 square miles) and low elevations (<60 feet). In addition, core-Lowland sites with minimal disturbance have not been identified in sufficient numbers for index development. As monitoring continues in Rhode Island, additional samples from these types of sites will facilitate refined site classification or metric adjustment and development of an index for core-Lowland sites.

5.0 References Cited

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

Blocksom, K. (2003). A performance comparison of metric scoring methods for a multimetric index for mid-Atlantic highlands streams. Environmental Management, 31(5):670 –682.

Cuffney, T.F., Bilger M.D., & Haigler, A.M. (2007). Ambiguous taxa: effects on the characterization and interpretation of invertebrate assemblages. *Journal of the North American Benthological Society* 26(2):286-307.

Davies, S. and S. Jackson 2006. The Biological Condition Gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251–1266.

ESS Group. 2007. Quality Assurance Project Plan for taxonomic identification of macroinvertebrates, Rhode Island. Prepared for RIDEM, Providence, RI. http://www.dem.ri.gov/pubs/qapp/taxbenth.pdf

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

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. U. S. EPA, Office of Research and Development, Cincinnati, OH.

Gerritsen, J. and B.K. Jessup. 2007. Calibration of the Biological Condition Gradient for high-gradient streams of Connecticut. Prepared for U.S. EPA Office of Science and Technology, and Connecticut DEP.

Griffith, G.E., Omernik, J.M., Bryce, S.A., Royte, J., Hoar, W.D., Homer, J.W., Keirstead, D., Metzler, K.J., and Hellyer, G., 2009, Ecoregions of New England (color poster with map, descriptive text, summary tables, and photographs): Reston, Virginia, U.S. Geological Survey (map scale 1:1,325,000).

Jessup, B.K., J.S. White, J. Gerritsen and M.T. Barbour. 2000. Analysis and interpretation of pilot study data for the biocriteria for streams and small rivers in Massachusetts. Prepared for Massachusetts Department of Environmental Protection.

Maxted, J. R., M. T. Barbour, J. Gerritsen, V. Poretti, N. Primrose, A. Silvia, D. Penrose, and R. Renfrow. 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. Journal of the North American Benthological Society 19(1):128-144.

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

McCune, B. and M. J. Mefford. 2006. PC-ORD. Multivariate Analysis of Ecological Data. Version 5.18. MjM Software, Gleneden Beach, Oregon, U.S.A.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Snedecor, G.W. and W.G. Cochran.1967. Statistical Methods (Sixth Edition). Ames, IA, Iowa State University Press.

Snook, H., S.P. Davies, J. Gerritsen, B.K. Jessup, R. Langdon, D. Neils, E. Pizutto. 2007. The New England Wadeable Stream Survey (NEWS): Development of Common Assessments in the Framework of the Biological Condition Gradient. Prepared for USEPA Office of Science and Technology and USEPA Office of Watersheds Oceans and Wetlands, Washington, DC.

Stoddard, J.L, D.P. Larsen, C.P. Hawkins, R.K. Johnson, R.H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications: 16(4):1267-1276.

Stribling, J.B., B.K. Jessup, and D.L. Feldman. 2008. Precision of benthic macroinvertebrate indicators of stream condition in Montana. Journal of the North American Benthological Society 27(1):58–67.

U.S. EPA (US Environmental Protection Agency). 1997. Field and laboratory methods for macroinvertebrate and habitat assessment of low gradient, nontidal streams. Mid-Atlantic Coastal Streams (MACS) Workgroup Environmental Services Division, Region 3, USEPA, Wheeling, West Virginia, and Mid-Atlantic Integrated Assessment Program, USEPA, Fort Meade, Maryland. (Available from: Environmental Services Division, Region 3, USEPA, 303 Methodist Building, 11th and Chapline Streets, Wheeling, West Virginia 26003 USA, and Mid-Atlantic Integrated Assessment Program, USEPA, 701 Maples Road, Fort Meade, Maryland 20755 USA.).

Zar, J.H. 1999. Biostatistical Analysis. 4th edition. Prentice Hall. Upper Saddle River, New Jersey. 663pp. + appendices and literature cited.