Development of a Benthic Macroinvertebrate Index of Biological Integrity (IBI) for Stream Assessments in the Eastern Sierra Nevada of California

David B. Herbst
Sierra Nevada Aquatic Research Laboratory
University of California
Route 1, Box 198
Mammoth Lakes, CA 93546
herbst@lifesci.ucsb.edu

and

Erik L. Silldorff
Delaware River Basin Commission
P.O. Box 7360, 25 State Police Drive
West Trenton, NJ 08628

Final Report

December 2009

TABLE OF CONTENTS

ABSTRACT

The assessment of biological integrity of streams is mandated through the Clean Water Act as a component of water quality regulation and protection. While various types of aquatic organisms have been used as indicators of biological integrity, benthic (i.e., bottom-dwelling) macroinvertebrates have been used most extensively and have been shown to provide a reliable measure of stream health.

Using collections of macroinvertebrates from streams of the eastern Sierra Nevada, this report details how data from these samples were used to develop a quantitative Index of Biological Integrity (IBI). The IBI is an index composed of multiple metrics (much like composite economic indicators) that can be used to accurately and cost-effectively assess stream health.

Component metrics were selected for inclusion in the IBI based on performance indicators such as sensitivity in response to disturbance stressors, high signal-to-noise ratio (strong response to stress with low variation), and little redundancy with other metrics. Ten metrics were selected through this process and were compared to different combinations and numbers of metrics. Classification structure from the 10-Metric IBI was also compared to the performance of a multivariate (RIVPACS-type) predictive model, and to a 9-Metric IBI based on lower taxonomic resolution. We documented a high degree of conformity in the assessment results produced by the different approaches.

Thresholds for assessment of biological impairment were based on reference streams of the region, defined as those least influenced by land use disturbances. To identify reference streams, we used criteria such as low levels of exposure both to the density of upstream road crossings in the watershed, and local reach-scale bank erosion. Streams not conforming to the reference site selection criteria were designated as test sites. The IBI scores of test sites were evaluated relative to the distribution of IBI scores for reference sites to determine whether biological integrity was impaired (according to 5 condition classes). We found that sediment-related stressors were among the most important sources of disturbance impacting streams in the region.

The IBI and alternative analytical methods presented here may be used to assess stream condition within the eastern Sierra represented by these surveys – from the Upper Owens River drainage in the south to the Truckee River drainage in the north. All streams surveyed are listed according to reference or test grouping, within-site and between-year variability, and cross-comparisons of scores and impairment assessments from the alternative methods.

Future refinements of these recommended biocriteria may include contrasts with independent validation data sets (i.e., stream surveys not used to develop the metrics), comparisons to periphyton (i.e., algae) indicators, use in conjunction with other data to develop biocriteria for the entire Sierra Nevada, combination with water chemistry and physical/habitat measures to permit integrated assessments of water quality, and development of additional options to apply these data for regulatory decision-making.

1. INTRODUCTION

Stream water quality conditions are often evaluated using chemical criteria, and stream habitat conditions are often evaluated using measures of the physical form and stability of channels and the quantity and distribution of riparian vegetation. These features provide useful information about the environmental setting of streams but fail to evaluate the biological health or integrity of stream ecosystems. A direct measure of the ecological suitability of aquatic habitats can be obtained by sampling the varied forms of life found on the stream bottom. Aquatic insects and other invertebrates are the most common organisms used for such biological assessments. Some of these organisms can live and even thrive under polluted conditions, but many others require a clean water environment to survive. The various types (i.e., the "assemblage") of organisms present in a stream can be used to indicate the health of the habitat. Assessing stream water and habitat quality based on the kinds of organisms living there is called "bioassessment."[1]

In recent years, bioassessment has been used throughout the United States, Canada, and many other nations to determine whether chemical water quality standards are sufficiently protective of actual instream biological conditions. Often it has been found that bioassessment provides a more integrated view for detecting impaired water quality or demonstrating improvements in environmental quality. Most states now use stream invertebrates as a regular part of their monitoring programs. Several states (such as Ohio, Maine, North Carolina, Florida, and others) have established regional "reference conditions," based on biological condition at relatively undisturbed stream sites, which are then used as biological standards for determining compliance with the Clean Water Act's mandate to protect the biological integrity of the nation's waters. Such programs are resulting in better means for detecting pollution, guiding abatement projects, and determining compliance points. Volunteer monitoring groups are also becoming active in using bioassessment through community programs such as "adopt-a-stream," local school education projects, and stream restoration work.

The objectives of the bioassessment program described here are to provide the foundation for developing biological criteria for water quality in the Lahontan Region (Figure 1, map of region and sampling locations). This effort has emphasized development of a database of reference streams that can be used to set biological expectations, or "biocriteria," for the Region's wadeable streams.[2] In addition, bioassessment data can also be used to aid in listing and de-listing decisions pursuant to the Clean Water Act Section 303(d), for reporting the condition of wadeable streams pursuant to the Clean Water Act Section 305(b), for monitoring the progress of restoration projects, for evaluating the effectiveness of management measures and permit conditions, and for setting biological targets to guide the management and improvement of water quality by all interested stakeholders.

---

[1] For more background information and references on bioassessment, see the USEPA's Bioassessment and Biocriteria Homepage at: http://www.epa.gov/waterscience/biocriteria/

[2] "Wadeable" streams are those that are small enough to be sampled without a boat (1st to 5th order here).

The Lahontan Region encompasses all watersheds on the eastern slope of the Sierra Nevada and includes many relatively pristine areas but also a variety of point- and nonpoint-source pollution problems. Among the land uses and/or features that may contribute to water pollution are roads and accelerated slope erosion, livestock grazing, construction activities, urban runoff, drainage from active or abandoned mines, stream flow diversions and channelization, and various other land and forest management activities. Erosion and sedimentation are widespread problems but are difficult to detect and evaluate with only chemical or physical assessments. Bioassessment can be used to detect changes in streams related to sedimentation because scouring and burial of the stream-bed habitat can affect the assemblage of instream organisms. Biological signals include changes in the number of different types of invertebrates (diversity), their relative tolerance of environmental stress and pollution (sensitivity), and their functional organization (role in the food web).

Monitoring stream invertebrates in comparison to "reference sites" (i.e., areas having little or no impact but similar physical setting), and/or over time at targeted sites, provides an estimate of impact or recovery in response to changing land use. Bioassessment may be used together with traditional stream chemistry and riparian monitoring to provide a more robust tool that measures the response of stream life to habitat changes.

When pollution does not originate from a single point (i.e., "non-point"), it can be difficult to measure using chemical methods because this type of pollution usually does not occur continuously and could be missed in a single water sample. Further, water quality problems that exist upstream of a location may not be obvious in the channel form or riparian conditions at that site. The advantage of using stream invertebrates is that they live in the stream and experience everything that flows over and around them and so incorporate and embody changes in water quality that occur in both local and upstream areas of the watershed.

Though the Lahontan Region covers a primarily arid landscape, it is comprised of a wide variety of watersheds, including both the mountainous Sierra Nevada and the deserts of the Great Basin and Mojave. Many streams flow from mesic, forested mountain slopes to the xeric conditions of high-elevation desert, crossing from the Sierra Nevada into the Great Basin ecoregions. The data summarized in this report on recommended stream biological standards (Index of Biological Integrity, or IBI) comes from stream surveys conducted in the central Lahontan Region, from the Truckee River watershed in the north, to the Upper Owens River watershed in the south (see map at Figure 1). Data collected from these sites include physical, chemical and biological information, intended for implementing the following objectives:

- Develop appropriate regional biological criteria or reference stream conditions (using samples from a network of least-impaired streams) for streams within the defined geographic area. This database may be used as guidance in determining the status of streams that may have degraded ecological integrity relative to the defined standards, or biocriteria. The biological standards may then be used to assess the extent of degradation (or absence of impact) and a target for gauging the progress/success of ecological recovery following restoration or implementation of management measures.

- Provide site-specific baseline data for evaluating local restoration projects or management programs directed at alleviating specific pollution source problems. Examples include: improved livestock grazing management (e.g., fencing, rest-rotation, varied stocking levels) at the West Walker River and Upper Owens River; control of acid mine drainage (AMD) in the Leviathan Mine watershed (e.g., chemical and biological treatments of AMD), channel restoration along the Upper Truckee River (e.g., erosion control and geomorphology), and TMDL target development for sediment problems in selected watersheds (e.g., Squaw Creek, Heavenly Valley Creek, Middle Truckee River).

The approach taken here recognizes the ecological diversity and wide variety of distinctive environments in California. The geographic restriction of the data set within the central Lahontan Region is expected to produce a high-resolution coverage of stream ecosystems with shared environmental conditions and biological composition, allowing more precise and more relevant assessments of water quality as it relates to the biological integrity of streams. A longer-term goal of this program is to integrate these regional data, to the extent feasible, with bioassessment surveys from elsewhere in the Sierra Nevada and throughout other regions of California.

2. METHODS

2.1 Stream Reach Selection, Reference Classification, and Repeat Sampling

Locating streams that may be used as “reference sites” for defining the unimpaired state of aquatic life in flowing waters is the first step to developing biological standards for water quality. Control sites are often chosen based on location above sources of impact, prior to an impact, or following subjective evaluations of what is believed to be an “undisturbed” condition. It is more desirable and defensible to use an objective selection procedure that may be applied over a range of geographic conditions to assess stream potential. Outlines for a systematic approach to reference site selection have been proposed (Hughes et al. 1986, Hughes 1995, Bailey et al. 2004) and have been used elsewhere in California to establish an Index of Biological Integrity (Ode et al. 2005). The approach used here was to define reference conditions according to “least disturbed” conditions (Stoddard et al. 2006) using disturbance measures at both the upstream watershed-scale, and for local-scale reach conditions at each site.

Most potential sources of human-related disturbance within any landscape require road access, so a first approximation used for identifying reference sites was derived from road coverage found on USGS maps. The extent of upstream watershed development and disturbance due to human sources was defined for the purposes of this analysis using the density of road-stream crossings upstream of each survey site (i.e., road crossings per km of stream above each site).[3] As an additional criterion, local reach-scale disturbance was gauged by the extent of bank erosion measured during site surveys (often attributable to livestock grazing). The criteria for categorizing reference and test sites are:

- Criteria for defining least disturbed conditions for designation of reference sites: ≤0.2 upstream road crossings /km and <25% local bank erosion, unless another known local disturbance or pollution source exists
- For test sites: ≥0.2 crossings /km and >25% bank erosion (if bank erosion less than 25% and no other disturbance known, this exception then designated reference), OR if site has ≤0.2 upstream crossings /km but bank erosion exceeds 25% and known pollution source exists (if no known local pollution source then designated as reference)

More than 60 percent of the reference designations (26 of 42 reference sites) were established based on sites meeting both the minimal road crossing and bank erosion criteria; the remainder were based on the exceptions (meeting one criterion). The reference sites defined by these criteria (42 sites total) were used to develop the IBI biological criteria. In order to establish the stability of assessment scores among reference streams, a subset of this group (15) was sampled in multiple years to obtain a measure of temporal variability. Test reaches were selected in part to examine a range of stream conditions where documented and known disturbance existed, and in other cases where in-stream impacts were uncertain despite potential sources of degradation.

2.2 Physical Habitat and Water Chemistry Sampling

Each site was defined as a 150-meter length study reach, located by GPS-UTM coordinates and elevation (near lower end of each site). The longitudinal distribution and length of riffle and pool habitats were first delineated, and flagged for marking transect locations. The slope of each reach was measured with an autolevel and stadia rod, and sinuosity estimated as the ratio of the 150-meter thalweg length to the linear distance between the upper and lower ends of the reach. Bank and channel habitat were measured over the length of each reach along 15 transect cross-sections spaced at 10-meter intervals. Water depth, substrate type and current velocity were measured at five equidistant points on each transect along with stream width, bank structure (cover/substrate type and stability rating), riparian canopy cover, and bank angle. Bank structure between water level and bankfull channel level was rated as open, vegetated, or armored (rock or log), and as stable or eroded (evidence of bank erosion, collapse or scour scars). Bank angles were scored as shallow, moderate, or undercut (<30°, 30-90°, and >90°, respectively), and riparian cover was measured from vegetation reflected on a grid in a concave mirror densiometer (sum of grid points for measurements taken at each stream edge and at mid-stream, facing up- and downstream). The type and amount of

---

[3] The density of road crossings was used instead of cumulative road crossings because our analysis showed that cumulative road crossings tended to eliminate the lowest portions of drainages from consideration as reference condition.

riparian vegetation along the reach was also estimated by qualitative visual evaluation. The embeddedness of cobble size substrate was estimated as the volume of the rock buried by silt or fine sand for 25 cobbles (encountered during transect surveys or supplemented with random selected cobbles). Discharge was estimated for each transect as the sum of one-fifth the width times depth and current velocity at each of the five transect points, and averaged. A suite of basic water chemistry and related parameters were also measured at each site: dissolved oxygen, conductivity, alkalinity, pH, temperature, and turbidity. Photographs were also taken to document habitat conditions, from mid-channel looking upstream at 0, 50, and 100 meters, and downstream at 150 meters.

2.3 Stream Macroinvertebrate Sampling

Benthic macroinvertebrate sampling consisted of five separate replicate samples taken in riffle zones using a 30-cm wide D-frame kick-net, having a 50-cm length bag with 250-µm mesh. Each replicate was comprised of a composite of three 30.5 x 30.5-cm sampled areas (0.093 m^2 each, 0.279 m^2 total), taken across the riffle transect (or in upstream series for small streams) over zones of varied depth, substrate and current (so can be considered a targeted-riffle sample type). Sample transects were selected using a random number table for locations corresponding to a delineated riffle segment. Each "kick sample" was most often taken manually to dislodge, turn over, and rub substrates thoroughly (usually for 30 seconds to one minute), so that both mobile and attached invertebrates were washed off and into the downstream net being held against the bottom. Actual kicking of the sample area was used sometimes in deeper water, standing upstream of the D-net. The composite samples, consisting of differing microhabitats, was intended to combine varied riffle conditions and so homogenize the variation that can exist among stream bed patches.

Samples were processed in the field by washing and removing large organic and rock debris in sample buckets followed by repeated elutriation of the sample to remove invertebrates from remnant sand and gravel debris. The remaining rock and gravel debris was inspected in a shallow white pan to remove any remaining organisms including caddisflies with stone cases and shelled snails or other molluscs. Elutriated and inspected sample fractions were then preserved in ethanol, and a small volume of rose bengal stain was added to aid in lab processing. Invertebrate field samples were subsampled in the laboratory using a rotating drum splitter (i.e., Folsom plankton splitter), sorted under a stereo microscope at 10X magnification, and identified to the lowest practical taxonomic level (usually genus; species or species groups when possible based on the availability of taxonomic keys, including midges and mites. Only oligochaetes and ostracods were not identified to further sub-divisions). The taxonomic identification level followed the California Aquatic Macroinvertebrate Laboratory Network (CAMLnet) standard effort protocols (Ode 2003) where possible, except that *Chironomidae* midges were identified to the genus/species level.[4] A minimum count of 250 organisms was removed from each replicate for identification, although typically more than 250 organisms were examined for each sample (median of 400 organisms). All sample sorting was conducted to achieve <5% error in removal, and quality control verifications of every taxon identified in every sample were performed by the lead author. Unprocessed sample remnants were also searched (using a 3X magnification visor) for rare and large taxa not encountered in the processed sample, and single counts of those found were added to the total.

Note: In February 2007, the State of California's Surface Water Ambient Monitoring Program (SWAMP) adopted standardized field sampling protocols for the collection of benthic macroinvertebrates in freshwater wadeable streams throughout California (State of California 2007). Those new protocols, known as reach-wide benthos (RW), rely on an approach that differs from the methods described above (TR) in that collections from multiple habitat types are combined according to their occurrence within a study reach, rather than being stratified by riffle habitat type. While detailed comparisons of differing targeted-riffle methods indicate that they are comparable (see, for example, Herbst and Silldorff 2006), and some studies suggest TR-RW data are interchangeable (Rehn et al. 2007), conversions between the two methods may need to be developed for montane Sierra stream habitats and particular reach geomorphic types.

2.4 Data Preparation

Based on comparative studies among sampling and analytical methods for California stream bioassessment protocols (including components of these Lahontan data), a consensus has been reached on the most cost-effective yet statistically powerful means to standardize California sampling methods. (See, for background, Ode et al. 2005, Herbst and Silldorff 2006, State of California 2007.) Two primary standardizations were the focus of this recent decision: (1) uniformity in technique: using both a single composite sample of (a) targeted riffle habitat, and (b) reach-wide or multiple habitats, taken at regular intervals for each assessed stream reach; and (2) uniformity in count: each reach-level composite sample will be subsampled as a 500 organism fixed-count in the laboratory (in practice a 550 minimum, randomly re-sampled to 500 for statistics).

Because the bioassessment methods used in the Lahontan Region have followed a different standard operating procedure since 1998 (when that Region's biomonitoring efforts were initiated), a conversion method for the existing data was required in order to ensure that the biological thresholds (i.e., biocriteria) developed through this research are applicable to all future sampling efforts. We examined a number of conversion algorithms for their ability to replicate a single composite sample collected from a reach followed by a 500-organism fixed count in the lab. A number of trade-offs existed among the algorithms, and the procedure which provided equal representation for each original replicate sample in the final composite was selected primarily because it gave equal

---

[4] CAMLnet has recently been replaced by the newly-incorporated nonprofit Southwestern Association of Freshwater Invertebrate Taxonomists (SAFIT). For more information about SAFIT and/or its standard taxonomic effort (STE) documents, see SAFIT's website: http://www.safit.org/ste.html.

probabilities for inclusion of organisms from different sections of the sampled stream reach. Specifically, 100 organisms were sampled (without replacement) from each of the five replicates collected for the Lahontan Region streams. The 100 statistically re-sampled organisms were then pooled to provide a single 500-organism composite sample for each site on each sampling date. Although this method did not provide an exact duplication of the expected sampling protocols that have now been adopted as the standard, the algorithm yielded a 500-organism sample for each reach that represented the diversity of habitats, environmental conditions, and the resulting invertebrate communities present across the entire study reach. Re-sampled data also showed community similarity between methods nearly equal to the original within-method replicate similarity.

2.5 Selection and Evaluation of Candidate Summary Metrics

Our calculation of a multimetric index (referred to herein as an Index of Biological Integrity, or IBI) closely follows the recommendations and procedures outlined in the U.S. Environmental Protection Agency's Rapid Bioassessment Protocols guidance document (Barbour et al. 1999).

A broad array of summary statistics for benthic macroinvertebrates (commonly called metrics) has been developed by researchers around the world. These metrics attempt to summarize important aspects of the biological community that are responsive to human disturbance while eliminating some aspects of the background variability and reducing the complexity of multidimensional community data. Although some metrics extract a clear signal with low background variability from most environmental settings and regions, the actual performance of a given metric cannot be determined *a priori* for any specific data set. As a result, the exact metrics that produce clear signals for human disturbance will often vary from region to region, and a broad suite of metrics needs to be evaluated in order to determine the most efficient means of extracting the usable information from benthic macroinvertebrate data in that region.

For the current study, a total of 71 metrics were evaluated for their ability to provide a clear signal with relatively little background noise between minimally-disturbed reference sites and a subset of the "test" sites where clear evidence of human-caused disturbance was present. The initial screening of the full suite of metrics utilized a subset of the current data: all collected in year 2000 from 24 minimally disturbed reference sites and 16 sites with some known source of human-caused impacts based on watershed conditions and/or physical/chemical stressors present in the sampling reach (referred to as "test" sites).

The initial screening of metrics focused on graphical and quantitative measures of the overlap between the distribution of metric scores at the reference sites and the test sites. Metrics were eliminated from further consideration if they clearly provided little or no separation between reference sites and test sites, or if the metric represented a slightly modified form of the standard metric calculation and the modified version yielded similar or poorer discrimination among the reference and test streams. From this broader set of metrics, a subset of 30 metrics was selected for a more rigorous examination. The full suite of candidate metrics evaluated in this study is provided in Appendix I (with the 30 metrics evaluated in greater detail noted by an asterisk). In addition, the formulas for calculating the core 12 metrics used in the alternative IBI formulations are given in Appendix II.

The subset of 30 metrics was then evaluated using data from all sites and dates (a total of 134 site-date combinations) in order to determine the best complement of metrics for biological monitoring standards in the Lahontan Region. This included 42 minimally disturbed reference sites, with 15 of these sites sampled in multiple years. In addition, 39 stream reaches with varying degrees of local- and watershed-scale disturbance were included in the analysis, with 21 of these 39 stream reaches sampled in multiple years. These latter 39 stream reaches will be referred to as the full set of "test" reaches evaluated in this study.[5]

Six criteria were used to quantitatively and qualitatively evaluate the ability of these 30 metrics to provide clear and unique discrimination of human-induced degradation of the benthic invertebrate communities:

1. The **background variability** or **noise** in the reference distribution was measured using the coefficient of variation (i.e., standard deviation divided by the mean for all reference sites). Metrics with a coefficient of variation less than 0.20 were rated the highest for this factor.
2. The **signal** for human disturbance was measured as the ratio between the mean score for the reference sites and the mean score for the test sites. Metrics with a ratio of means greater than 1.5 (or less than 0.67 for reverse-scale metrics) were rated highest for this factor.
3. The **signal-to-noise ratio** was measured as the difference between the means of the reference and test distributions divided by the standard deviation in the reference site scores for that metric. Metrics with a signal-to-noise ratio of greater than 1.5 were rated highest for this factor.
4. The **discrimination efficiency** was measured as the percentage of test sites having metric scores greater than (or less than for reverse-scale metrics) standard percentiles of the reference site distribution for that metric. The percentiles evaluated were the 10th, 25th, and 50th (or median), with highest ratings given to metrics whose percent overlap was less than 50%, 35%, and 25% for the respective percentiles.
5. Descriptive **properties of both the reference and test distributions** were visually examined using box-and-whisker plots. The symmetry or normality in the distributions, the lack of repeated extreme values (e.g., multiple observations of a 0 score), the extent of outliers in the distributions, and the degree of overlap

---

[5] A total of 80 stream reaches were included in this study, with 1 stream reach being used in the reference distribution for 1999 and among the test sites in 2002 because the location was burned by wildfire just before the 2002 samples were collected.

between the distributions were subjectively evaluated and ranked across the 30 candidate metrics.

6. The **correlation and pattern in the relationship between metric scores and three gradients of potential human disturbance** were then qualitatively and quantitatively evaluated. The three environmental gradients were: (1) the percent of fines, sands, and gravels in the substrate; (2) the specific conductivity of stream water at the site; and (3) the extent of eroded stream banks within the study reach. Metrics with correlations greater than 0.5 were rated the highest, with greater ranking given to metrics with strong relationships across more than one environmental gradient. In addition to the quantitative evaluation, the scatter plot between each environmental variable and each metric was examined to determine whether correlations were driven by outliers and/or points with high leverage, or whether correlations were weak because of non-linear relationships. These qualitative evaluations were then used to modify the ranking based on the quantitative correlations alone.

The combined rankings across all six factors were then evaluated for each metric. Metrics that did not rank high in any of the six categories were eliminated from further consideration in the final IBI multimetric score. Eight (8) of the 30 metrics were eliminated during this step, and the remaining 22 metrics were included in further analyses to determine the combination of metrics that, when added together on standardized scales, led to the best composite multimetric IBI score.

These remaining 22 metrics were then re-scaled following the recommended procedures contained in the USEPA’s Rapid Bioassessment Protocols (Barbour et al. 1999). Specifically, each metric was placed on a continuous scale from 0 to 10, with a zero score representing extreme impairment and 10 representing reference quality conditions. Because the set of reference streams used in this study represented sites with least human influence, the median score for each metric among the reference sites was selected as the benchmark for a 10 on the new scale. Conceptually, this scoring identifies the mid-point value among the reference sites’ scores as being sufficient to receive the best rating, with scores higher than the median not receiving higher scaled metric scores because they represent natural variability among reference sites with low levels of human disturbance rather than sites with increasingly more pristine conditions. Where the reference distribution is composed of sites that have substantial human disturbance, higher metric scores may in fact represent higher ecological condition. In these situations, selection of the 75th or higher percentile of the reference distribution for receiving the maximum scaled score (e.g., a value of 10) might be warranted (see, for example, Ode et al. 2005). But the reference sites in our study that had metric scores above the median were not sites with less disturbance than others, so scoring these sites higher than the median reference sites was deemed inappropriate.[6]

---

[6] There are alternative metric scoring systems, such as scaling metrics relative to the 5th to 95th percentile range over <u>all</u> sites, and adjusting metric scaling by significant environmental co-variates such as watershed area (Barbour et al. 1999). However, the method chosen here was deemed appropriate for the Lahontan Region because it is more consistent with other IBIs developed in California and we found no evidence that

On the lower end of the scoring range, the 10th percentile of the test distribution (or the 90th percentile for reverse-scale metrics such as the Biotic Index) was selected as the benchmark for obtaining a score of 0 on the scaled metric scores. The 10th percentile of the test distribution has been used elsewhere in California to set the benchmark for a minimal score (*see* Ode et al. 2005). Although scores less than the 10th percentile could indeed represent poorer ecological conditions, and thus could receive lower scores, the statistical estimation for percentiles less than the 10th becomes increasingly imprecise. For instance, the minimum observed value across all test streams is a conceptually attractive value to choose for identifying the zero-point for the scaled metric scores. However, the actual value of the minimum score is highly variable, posing the potential for both random and systematic errors to cause significant shifts in the overall scoring of all sites (e.g., outliers may skew the scores). The 10th percentile was therefore chosen as a compromise between setting the zero-point at the worst observed condition and identifying a benchmark that is relatively insensitive to being skewed by the empirical data collected during this study.

Once the benchmarks for a score of 10 and a score of 0 were identified for each of the final 22 metrics, the actual metric scores for all streams were transformed to this 0-10 scale. Any site with a metric score equal to or greater than the benchmark for receiving a 10 would also receive a score of 10 for that metric. Likewise, any site with a metric score equal to or less than the benchmark for receiving a 0 would also receive a score of 0 for that metric. For reverse-scale metrics, the procedure was reversed. Metric scores between the 0 and 10 benchmarks were linearly interpolated along a continuous scale. (The benchmarks for the final 12 metrics selected for the candidate IBIs are given in Section 3.3 along with formulas for converting metric scores to this 0-10 scale.)

The following calculation shows the conceptual and numeric procedures for scaling one of these metrics scores. Total taxa richness among all reference sites had a median value of 50 taxa, and the 10th percentile among all test sites was 30 taxa. The standardized scoring for total taxa richness was therefore:

$$\text{Scaled Richness Score} = \begin{cases} 0 & for\ rich \leq 30\ taxa \\ 10 & for\ rich \geq 50\ taxa \\ 10 \cdot \dfrac{Richness - 30}{50 - 30} & otherwise \end{cases}$$

The final procedure for selecting the metrics that would be added together to yield the final candidate multimetric IBI scores involved a combination of quantitative and subjective decisions. First, quantitative correlations among the final 22 metrics were computed to determine the degree of numerical redundancy among different metrics. Second, subjective evaluations were made about the conceptual overlap among metrics,

---

metrics were strongly related to natural environmental gradients (such as elevation, watershed area, or channel slope).

the need to incorporate metrics representing the four primary aspects of the invertebrate community (richness, composition, tolerance, and functional), and the empirical performance of different multimetric IBIs with similar qualities with respect to these other considerations. A total of 20 different combinations of metrics into a single IBI score were evaluated, with the number of metrics included ranging from 5 up to 20 metrics. From these 20 different IBI formulations, three (3) candidate IBIs were selected for presentation in this report. These three IBIs represent slightly different philosophies for a multimetric index, and the performance of each IBI based on standard measures is presented in the results. Based on the performance of each IBI and additional considerations for consistency with other multimetric IBIs, a desire to provide a degree of overlap and redundancy among the IBI metrics, and the fulfillment of a broad-based measure of the invertebrate community structure and function, one of these three IBIs is recommended as the best candidate for assessing the ecological condition of wadeable streams and rivers in the Lahontan Region.

2.6 Quality Assurance and Quality Control (QA/QC)

A detailed Quality Assurance Project Plan (QAPP) was prepared for this project (*see* Herbst 2002). This project followed all applicable QA/QC procedures specified in the QAPP, as well as all applicable QA/QC procedures then mandated by the State of California's Surface Water Ambient Monitoring Program (SWAMP). The SWAMP's QA/QC procedures are detailed in a separate Quality Assurance Management Plan (*see* State of California 2002).

3. RESULTS & DISCUSSION

3.1 Metric Evaluation

Three (3) alternative multimetric Indices of Biological Integrity (IBIs) were selected for further consideration, as described above. The three IBIs were derived using different numbers of metrics (12 metrics, 10 metrics, and 8 metrics), with a total of 13 individual metrics included in one or more of the IBIs. Prior to presenting final evaluations of the three IBIs, we next examine the quantitative and graphic evaluations for these 13 metrics as a means of describing the rationale for their inclusion in the different IBIs, and to provide a means of juxtaposing the performance of the different metrics included in the final three multimetric IBIs.

The 13 metrics (and their contracted names for presentation purposes) that were included in one or more of the final three IBIs are:

1. Total Taxa Richness (***rich***)
2. Ephemeroptera (E) Richness (***ephem rich***)
3. Plecoptera (P) Richness (***plecop rich***)

4. Trichoptera (T) Richness (***trichop rich***)
5. Acari Richness (***acari rich***)
6. Chironomidae Richness as a percent of Total Richness (***chiro perc rich***)
7. Intolerant Taxa Richness as a percent of Total Richness (***intol perc rich***)
8. Tolerant Taxa Richness as a percent of Total Richness (***tol perc rich***)
9. Number of Predator Taxa, or Predator Richness (***pred rich***)
10. Abundance of EPT as a percent of Total Abundance (***ept abund***)
11. Abundance of Shredder Taxa as a percent of Total Abundance (***shredder***)
12. Dominance as measured by the Abundance of the 3 most common taxa as a percent of Total Abundance (***dominance 3***)
13. Biotic Index – modified Hilsenhoff type (***bi***)

Note: The metrics based on tolerance values used the values for each taxon given in the California Aquatic Macroinvertebrate Laboratory Network standard taxonomic effort list (Ode 2003, now updated by SAFIT).[7] In addition, the Intolerant Taxa Richness was based on those taxa with tolerance values less than or equal to 2, while the Tolerant Taxa Richness was based on taxa with tolerance values greater than or equal to 7. Box-and-whisker plots for each metric are presented in Figure 2.

The first criterion by which these metrics were evaluated was the background level of variation among the reference streams, which was measured by the coefficient of variation (standard deviation divided by the mean) among scores for the reference streams alone. Table 1 presents the coefficient of variation ranked from smallest to largest. These data show that the majority of these metrics had low levels of background variability in their reference distributions, with a number of metrics showing a CV less than 0.20. The most variable metric in terms of this background level of noise was the percent of shredders at each site. For this metric, the standard deviation exceeded the mean value for reference streams, as indicated by a CV greater than 1.0.

The second criterion by which the metrics were evaluated was the strength of the signal between reference streams and test streams exposed to varying levels of human-caused disturbance. This "signal" was measured as the ratio of means (the reference site mean divided by the test site mean), and the results for each metric are presented in Table 2 in ranked order. Like the CV, the ratio of means shows variable results among the metrics. Of particular note is that some metrics with a large ratio (and thus a strong "signal" for impairment) had poor rankings for the CV (high levels of "noise"), and vice versa.

The third criterion used to evaluate the different candidate metrics was the signal-to-noise ratio, measured as a standardized difference (difference between the test and reference means divided by the standard deviation of the reference streams). As with the other two measures, some metrics had large standardized differences and thus suggested a clear separation between reference streams and test streams when taking into account

---

[7] CAMLnet has recently been replaced by the nonprofit Southwestern Association of Freshwater Invertebrate Taxonomists (SAFIT). For more information about SAFIT and/or its standard taxonomic effort (STE) documents, see SAFIT's website: http://www.safit.org/ste.html.

the background variability of reference stream scores. However, other metrics (e.g., percent abundance of shredders) had weak signal-to-noise ratios, indicating that the signal was somewhat compromised by the large amount of noise.

The fourth criterion to evaluate each metric was the discrimination efficiency (*see* Blocksom 2003). For this empirical evaluation, the overlap of the test and reference distributions were measured as the proportion of test streams having scores greater than (or less than for reverse-scale metrics) a given percentile of the reference site distribution. Three percentiles were used to evaluate this overlap: the 10th, the 25th, and the 50th (or median) percentiles (or the complementary percentiles for reverse-scale metrics). Like the signal-to-noise ratio, overlap between the test and reference distributions was expected because many of the test streams had uncertain, if any, biological impairment. As a result, the expectation was not for a 100% separation between these distributions, but rather a clear pattern separating a portion of the test streams from the streams known to be unimpaired (i.e., the reference streams). Table 4 presents the results of these overlap calculations for each of the 13 candidate metrics, with the ordering based on the overlap for the 10th percentile. These data continue to show that different criteria for evaluating candidate metrics results in different rankings of the 13 metrics.

The fifth criterion for metric evaluation was a more subjective evaluation of the distributions of the metric scores between reference sites and test sites. This was evaluated with box-and-whisker plots for each metric, and these graphs are presented in Figure 2. For all metrics, distinct differences can be seen between the center of the distribution as defined by the median score (represented by the white line within the box). In addition, the central distribution range box (25th to 75th percentiles) differed between the test streams and the reference streams for most metrics. Finally, the distributions for most metrics exhibited symmetrical or normal shapes, with few if any outliers.

The sixth criterion used to evaluate the suitability and performance among the different metrics was the relationship of the metric scores to three gradients in environmental conditions that are strongly related to human disturbance. These three gradients were: (1) the percent of the substrate composed of fines, sand, and gravels, combined; (2) conductivity; and (3) the percent riparian cover (see Methods for a description of how these were measured). Figure 3 presents each of the three plots of environmental factor vs. metric score for the 13 candidate metrics. Like the measures of performance for the metric scores alone, these relationships between metrics and environmental gradients suggest some metrics have stronger relationship than others, highlighting their relevance for inclusion in the final multimetric index. Yet, just as each of the performance measures describes only one aspect of the overall usefulness of these metrics, these relationships show only partial community responses to environmental stress gradients.

3.2 Multimetric Creation

The process of multimetric index creation entails the selection of component metrics to be added together to yield a single composite score for each site. Here, the selection process utilized the above evaluations for each metric as the starting point for identifying suitable metrics. Two additional factors were then incorporated into the final metric selection process. First, the pairwise correlations among metrics were evaluated in order to reduce unintentional redundancy among the metrics included in final multimetric indices. Second, the type of information each metric extracted from the invertebrate data was considered in order to include measures that focused on different aspects of both the structure and function of the invertebrate community. These two additional evaluation processes are described below.

First, the correlations among metrics were examined to identify suites of metrics that minimized unintentional redundancy within the multimetric IBIs. No clear and unambiguous criterion exists for determining how strong a correlation introduces excessive redundancy within a multimetric IBI. Correlations greater than ±0.70 to ±0.80 have often been used to identify variables with relationships that are particularly strong, and which therefore should be considered for exclusion. These values for the correlation coefficient correspond to R^2 values between 0.49 and 0.64, indicating that 49% to 64% of the variation in one variable can be explained by the other. In this analysis, we considered any correlation coefficient between -0.707 and 0.707 ($R^2 \leq 0.50$) sufficiently low to present little or no problem with introducing unwanted redundancy. Correlations between ±0.707 to ±0.80 were carefully scrutinized, and attempts were made to eliminate and minimize the number of such correlations among component metrics. Correlations between metrics greater than 0.80 or less than -0.80 suggested strong relationships that would typically be unsuitable for use together in a multimetric index. For metrics exhibiting such strong relationship, the inclusion of both metrics in a multimetric index was considered at times, but only when a clear conceptual difference (metric type) existed between the metrics.

Table 5 presents the pairwise correlation coefficients (r) among the 13 candidate metrics. The correlations in **bold** among these final 13 metrics show those pairs that have high levels of redundancy with one another (the greatest being that between total taxa richness and predator richness). The final three multimetric IBIs (Table 6) included different sets of these metrics; one included no metrics with pair-wise correlations greater than ±0.707 (the 8-Metric IBI), one included one metric pair (the 10-Metric IBI), and the third included 4 metric pairs with moderate to strong correlations (12-Metric IBI).

The second additional consideration for selecting metrics for the final IBI was the type of information extracted by each metric. There are four general classes of biological metrics: (1) richness or diversity measures; (2) composition measures; (3) tolerance measures; (4) functional measures (i.e., trophic or habit). Table 6 presents the 13 core metrics and the categories to which they belong. Six of the metrics are direct measures of richness-diversity, five metrics show some aspect of community composition, four metrics evaluate the tolerance of the invertebrate community to pollution and human disturbance, and two metrics measure some functional aspect of the community (both are measures of feeding groups). Note that four of the metrics fall into two categories.

The last step in creating multimetric IBIs was to select metrics to combine into different versions of a multimetric Index of Biological Integrity (IBI). Over 20 different IBIs, each with a different complement of candidate metrics, were computed and evaluated. From these different formulations, three (3) multimetric IBIs provided suitable performance, minimal redundancy (inter-correlation) among variables, and representation from the four metric types. These three IBIs used 12 metrics, 10 metrics, and 8 metrics, respectively, in their computation (see Table 6).

The 12-metric IBI includes 6 richness measures, 4 compositional measures, 3 tolerance measures, and 2 functional measures (3 metrics overlap categories). The median magnitude of correlations among metric scores included in this IBI is 0.41, with a maximum correlation of 0.90 for the strong relationship between predator richness and total taxa richness. In addition, three correlations among metrics were moderately strong (*pred rich* vs. *plecop rich*, *ept abund* vs. *bi*, *pred rich* vs. *acari rich*). We decided to include the predator richness metric despite its high correlation with total richness because the variety of predators present in a community is indicative of trophic complexity as well as taxonomic diversity.

The 10-metric IBI includes 5 richness measures, 3 compositional measures, 3 tolerance measures, and 1 functional measure. The median magnitude of correlations among metric scores is 0.39, with the correlation between *ephem rich* and *rich* being the only correlation greater than 0.707 in magnitude.

The metrics in the 8-metric IBI were specifically chosen to eliminate any correlations among metrics greater than 0.707 in magnitude, while also maintaining a diversity of metric types and a conceptually satisfactory group of metrics. For the 8-metric IBI, there are 3 richness metrics, 3 composition metrics, 3 tolerance metrics, and 1 functional metric. The median magnitude of correlation is 0.33, with the maximum correlation between metrics being 0.7069 for the correlation between the biological index (*bi*) and the percentage of taxa within the most intolerant group (*intol perc rich*).

3.3 Selection of Biocriteria Thresholds and Evaluation of Final 3 IBIs

3.3(a) IBI Performance

A number of analyses were used to evaluate the performance of the three final IBI candidates (12-, 10-, and 8-Metric IBIs). First, a suite of IBI performance measures were calculated and evaluated directly, with the performance measures compared among the three IBI versions. Second, the sites were classified into condition categories based on the thresholds identified in Section 3.3b below (16th percentile reference for full-support/partial boundary, and 5th percentile reference for the partial/non-support boundary). The agreement and distribution among classification categories based on the three alternative IBIs were then compared.

First, the box-and-whisker and cumulative distribution plots for each of the final three IBIs are presented in Figures 4, 5, and 6, with reference and test stream scores displayed separately. These figures display both the pattern of the distributions for reference and test streams as well as the overlap in distributions for these two stream classes. The results indicate that all three IBIs meet the expectations for an index that discriminates differences between reference and test scores, and that little distinction exists among the three IBI formulations. The reference and test distributions in each case are uniformly distributed across their expected ranges, without clumping of sites around any scores or uneven representation toward one end of the range. The one distinction among the results is the relatively clear break in low scoring reference sites for the 12-metric and 10-metric IBI, with the 8-metric IBI results showing a more continuous gradation on the lower tail of the reference site distribution.

Second, quantitative summaries of the results are presented in Table 7. Overall, the results indicate that there are only minor differences in the performance among the three IBI formulations. For instance, the coefficient of variation for each reference stream distribution (the measure of "noise") is nearly indistinguishable among the three, with all three IBIs showing low variability relative to the mean value. The only notable difference is in the theoretical overlap in the distributions, as measured by the standardized difference between the means (the measure of signal-to-noise ratio). For this one performance characteristic, the 12- and 10 Metric IBIs had a slightly larger standardized difference compared to the 8-Metric IBI. However, the discrimination efficiency measure of overlap (quantified as the percent of test streams scoring higher than the 25th percentile of the reference stream scores) suggests a contrary pattern, with the 12-Metric IBI actually having more overlap and less of a difference compared to either the 8-Metric or 10-Metric IBIs. This underscores the general pattern in Table 7 that the performance of these three IBI formulations is essentially indistinguishable. All three multimetric IBIs provide a clear separation between the test and reference streams, with both a strong signal and relatively low background variation.

It is important to note that there remains substantial overlap in the distributions of reference and test streams for all three IBI formulations. Rather than providing an indication of poor performance, this overlap captures a valid and important pattern in the data. Specifically, a consistent fraction of the test streams did not score differently from the biological condition score of the reference streams. The current analysis is, in part, intended to identify the extent to which disturbance-exposed test streams are resistant to biological degradation. These data suggest that approximately 20% of the test streams have invertebrate communities that are largely indistinguishable from 75% of the reference stream communities. This overlap is discussed in greater detail below.

The third method for evaluating and comparing the IBIs was to use the final results from the IBI, specifically the categorization of sites into condition classes, to determine the consistency of the IBIs in identifying the degree of degradation across both the reference and test sites. Table 8 presents a summary of this cross-classification. The table reveals that the 12-metric and 10-mertric IBIs classified reference sites into the same categories for all 62 sites, and that only 1 site was classified lower by the 8-Metric

IBI compared to the 10-Metric and 12-Metric IBIs. The results from the test sites also showed high concordance among the IBI classifications with all but one site graded the same by the 10- and 12-metric IBIs, and the 8-metric differing in 6 cases (of 72) between the 12- and 10-metric IBIs. The total non-correspondence rate (defined simply as the percentage of sites classified by two methods into different categories) therefore ranges between 1.4 to 9.7%. An additional pattern in Table 8 is the increasing proportion of sites classified into the middle category as fewer component metrics are used (8 for the 12-Metric, 9 for the 10-Metric, 14 for the 8-Metric). This pattern suggests that the distribution of test sites perhaps becomes more concentrated toward the intermediate and more indeterminate condition values with the use of fewer metrics. One interpretation of such a pattern is that the IBIs with fewer metrics yield greater ambiguity about the degree of degradation by simply classifying a greater proportion of test sites into an intermediate category rather than identifying them clearly as within the typical range of reference site scores or clearly outside the range of reference site scores. Adding metrics to the IBI permits clearer assessments of stream condition as the signal-to-noise ratio is strengthened. Beyond marginal improvements in index performance, however, the further addition of metrics becomes redundant.

The final consideration for selecting one of these three IBIs as the standard means of evaluating biological condition in the eastern Sierra portion of the Lahontan Region was an evaluation of the complement of metrics used in each IBI (Table 6). The 12-Metric IBI has 6 richness metrics, 4 composition metrics, 3 tolerance metrics, and 2 functional metrics (with 3 metrics falling within 2 categories). The 10-Metric IBI eliminates one richness and one functional metric, and the 8-Metric IBI further reduces the number of richness metrics by 2, and adds a different composition / tolerance measure (intolerant richness percent).

Like other aspects of the IBI development process, the final selection among alternative IBIs involves trade-offs. Some researchers have suggested that a minimum set of metrics with the least amount of overlap is the best approach for determining an appropriate multimetric IBI. However, there are compelling reasons for including a moderate amount of redundancy among the component metrics for an IBI, and there is no clear threshold for determining what is sufficient redundancy compared to excessive redundancy. Our detailed evaluation of the three multimetric IBIs presented in this report, and our evaluation of the conceptual overlap among the component metrics, leads us to conclude that the 10-Metric IBI is the most useful among these three candidate IBIs. First, the 10-Metric IBI performed as well or better than the other two IBIs in all performance characteristics evaluated. Second, the classification of sites among three condition categories suggested that the 8-Metric IBI tended to distribute sites with moderate degrees of degradation across a broader range of scores, both placing them in the supporting category and the non-supporting category compared to the classifications from the 10-Metric and the 12-Metric IBIs. Finally, representation from each metric category (mostly richness metrics but including functional, composition, and tolerance measures) with the 10-metric IBI, while eliminating metrics with higher intercorrelation (predator richness and EPT abundance), yields the potential for a more robust bioassessment tool that may accurately measure both degraded sites and sites sustaining biological integrity even beyond the range of conditions evaluated in the existing data set. The risk exists that, with a smaller number of metrics representing fewer aspects of the community composition and function, an unusual site may score either higher or lower than its degree of human-caused disturbance would justify simply because one or a small number of the metrics scores exceedingly well or exceedingly poorly. Including more metrics than needed to produce an optimal calibration that separates reference from test sites is also undesirable because such an index would contain an excess of redundant information. Thus, we conclude that the 10-Metric IBI provides an optimum set of metrics because it has both a low risk of misclassifying sites and a low risk of excessive redundancy. In any case, the process of metric selection for IBIs involved a balancing of decisions that usually resulted in multiple similar outputs.

The 10-Metric IBI with the following scoring rules is suggested as the preferred procedure for assessing biological integrity of stream macroinvertebrate communities in wadeable streams and rivers in the eastern Sierra portion of the Lahontan Region (included for purposes of comparisons are two additional metrics that were used in our analysis: predator richness and EPT abundance):

EASTERN SIERRA MACROINVERTEBRATE IBI – SCALED CALCULATIONS
Sum of the 10 scaled metrics below = Eastern Sierra IBI for benthic invertebrates (Level II taxonomy)

Metric	Score & Scaled-Metric Calculation	Metric Value Range
Total number of taxa found in sample		
1. Total Taxa Richness	0 if	≤30
	10 if	≥50
	10*(Rich - 30)/(50 - 30)	If between 30 - 50
Total number of Ephemeroptera – mayfly – taxa found in sample		
2. Ephemeroptera Richness	0 if	≤3
	10 if	≥9
	10*(E.rich - 3)/(9 - 3)	If between 3 - 9
Total number of Plecoptera – stonefly – taxa found in sample		
3. Plecoptera Richness	0 if	≤1
	10 if	≥6
	10*(P.rich - 1)/(6 - 1)	If between 1 - 6
Total number of Trichoptera – caddisfly – taxa found in sample		
4. Trichoptera Richness	0 if	≤2
	10 if	≥8
	10*(T.rich - 2)/(8 - 2)	If between 2 - 8
Total number of Acari – water mite – taxa found in sample		
5. Acari Richness	0 if	≤1
	10 if	≥6
	10*(A.rich - 1)/(6 - 1)	If between 1 - 6
Percent of taxa in sample that are Chironomidae – midges (%midge taxa of total taxa)		
6. Chiro Perc Richness	0 if	≥43.4%
	10 if	≤26.4%
	10*(43.4 - C.rich.%)/(43.4 - 26.4)	If between 26.4 - 43.4%
Percent of taxa in sample that are tolerant (TVs=7-10) out of total taxa		
7. Tol. Taxa Rich. Perc.	0 if	≥34.1%
(TV=tolerance value)	10 if	≤18.7%
	10*(34.1 - Tol.rich.%)/(34.1 - 18.7)	If between 18.7 - 34.1%
Percent of total abundance (500) in sample comprised by individuals that are shredders		
8. Shredder Abund. Perc.	0 if	0
	10 if	≥2.7%
	10*(Shred.abund.% - 0)/(2.7 - 0)	If between 0 - 2.7%
Percent of total number in sample (500) that come from the 3 most common taxa		
9. Dominance Perc 3	0 if	≥65.9
	10 if	≤42.9
	10*(65.9 – Dom.3.%)/(65.9 – 42.9)	If between 42.9 - 65.9%
Composite community tolerance (sum product TV x relative abundance over all taxa)		
10. Biotic Index	0 if	≥5.79
(modified HBI)	10 if	≤4.05
	10*(5.79 – BI)/(5.79 – 4.05)	If between 4.05 - 5.79
Predator Richness		
	0 if	≤ 7
	10 if	≥16
	10*(pred rich - 7)/16 - 7	If between 7 - 16
EPT abundance as percent of total abundance		
	0 if	≤ 17.5%
	10 if	≥ 59.1%
	10*(EPT abund – 17.5)/(59.1 - 17.5)	If between 17.5 - 59.1%

Because there are 10 metrics used in the final IBI, a simple summation of the scaled metric scores will result in an overall IBI score that ranges between 0 and 100. Thus, beyond the summation of these 10 scores, no additional computations are necessary for calculating the 10-Metric eastern Sierra IBI score.

A final important note about the above scoring criteria is that: 1) the calculation of scores for each of these metrics requires that invertebrates be identified to the same taxonomic level, and 2) the same tolerance values need to be used as in our study in order to be consistent with the formulas presented above. For almost all groups, we have used the Level II taxonomic identification as specified by the California Aquatic Macroinvertebrate Laboratory Network (CAMLnet) standard effort protocols (Ode 2003),[8] and further specify that Chironomid midges were identified to the genus and sometimes species-group levels for these data. (The taxonomic levels used here are listed in Appendix IV.) An alternative to this system is to use lower resolution (Level I) taxonomy. Comparative results using simplified (i.e., Level 1) taxa identifications are presented in Section 3.6. For Tolerance Values, Appendix IV presents the values used in the calculations for this report. These values were likewise obtained from the CAMLnet accepted protocols.

Interannual variability in the condition of both test and reference sites highlights the importance of periodic re-examination of the IBI model and its constituent data set. While this IBI tool has been thoroughly evaluated and is now ready for use within the eastern Sierra portion of the Lahontan Region, continued stream surveys could add data permitting more complete description of both the spatial and temporal variability of the reference condition. The biological thresholds (i.e., biocriteria) proposed here can be updated and refined over time to produce improved precision and geographic coverage of biocriteria, and ongoing evaluations of the progress of water quality improvement at test sites where restoration measures or permit conditions have been implemented. The collection of additional samples over a broad range of stream types, habitat conditions, and locations could further ensure that the models produced are based on reference streams that are representative of the variety of flowing waters present throughout the region.

3.3(b) Thresholds

Having selected a multimetric IBI, regulatory application requires setting thresholds along the IBI scale that discriminate different classes of ecological condition and different levels of water quality attainment. Our threshold identification employed both theoretical considerations and empirical properties of the data set as a means of selecting values along the continuous IBI scale where a justifiable break could be identified for separating more homogeneous sets of sites with similar levels (or lack) of ecological integrity.

---

[8] CAMLnet has recently been replaced by the nonprofit Southwestern Association of Freshwater Invertebrate Taxonomists (SAFIT). For more information about SAFIT and/or its standard taxonomic effort (STE) documents, see SAFIT's website: http://www.safit.org/ste.html.

From a more theoretical perspective, the water quality and ecological integrity thresholds can be selected along the IBI continuum in a way that balances resource protection against the risk of incorrectly identifying streams as impaired. On the one extreme, a subset of the highest scoring sites can be used to characterize the range of acceptable conditions, with the threshold for classifying a site as not-supporting biological integrity set at a high standard. Such an approach provides a high level of protection for the stream resources because even slight reductions below these high scores are flagged as exhibiting signs of degradation. However, such an approach also increases the frequency of incorrectly classifying sites as degraded, with the restrictive threshold forcing sizable fractions of reference sites (where minimal human alteration exists) to be classified as "not-supporting" (i.e., Type I statistical error). On the other extreme, all sites with little or no human influence (i.e., the reference sites) can be used to set an impairment threshold at the limits of variation across the entire group. Such an approach leads to reduced rates of misclassifying acceptable reference-quality sites as degraded, and in this way also protects the regulated community from incorrect classifications of sites as degraded when in fact they are not. But this approach also leads to weaker protection of the stream resources because an individual site might experience substantial degradation before it scores below the minimum range of reference conditions (i.e., Type II statistical error). As a result, impaired condition may not be detected, and when it is identified, it is less feasible to address and may sometimes be irreversible.

Unfortunately, there are direct trade-offs in these two directions, and high levels of resource protection cannot be attained without increasing the degree of falsely classifying sites as impaired. In statistical terms, the Type I and Type II errors are inversely related to each other, and decreases in the Type I error rate (falsely classifying a site as impaired) cannot be attained, all else equal, without increasing the Type II error rate (identifying a site as unimpaired when in fact it has a degraded condition). Thus, the selection of numerical threshold values is in part a technical choice between which types of errors are more acceptable or less costly, but decisions ultimately lie with how regulatory agencies represent and balance the environmental, economic, and aesthetic values and concerns of society.

We evaluated a number of approaches that varied in their Type I and Type II error rates, and selected an approach that was intermediate among those considered. Specifically, we first established two primary categories (unimpaired or supporting, and impaired or partially-supporting and non-supporting) using properties of the reference site distribution. We then created a 5-class system by partitioning the lower reference range into an intermediate class, then equally dividing both the remaining "unimpaired" and "impaired" into upper and lower classes.

The threshold between unimpaired and impaired, the most critical regulatory limit, was set at the 5th percentile of the references site scores (= 63.23 for the 10-metric IBI). By definition, such a threshold will always incorporate nearly all reference site scores, but will exclude aberrant outliers. Thus, this threshold leads to very few classifications of the *a priori* reference sites into the impaired "partially-supporting" class (i.e., 5% of sites by definition).

Within the reference range, a threshold between acceptable and uncertain classes was set at the 16th percentile of the reference site scores (= 75.0 for the 10-metric IBI). This value was selected based on two criteria. First, a reasonable balance between Type I and Type II errors was achieved for thresholds between the 10th and the 20th percentiles of the reference site distribution (Herbst and Silldorff 2006). Specifically, Herbst and Silldorf (2006) showed that Type II errors (or "false positives") could be minimized with the Type I error rate set at the 16th percentile. This means that most errors in not detecting degraded condition when it actually occurs can be nearly eliminated when the lowest 16% of reference sites are regarded as uncertain, or requiring further evaluation. Second, the data we collected showed a sharp discontinuity in the reference-site IBI scores at approximately the 16th percentile. As a result, the 16th percentile met the objectives of achieving a balance between keeping a low Type I error rate and minimizing Type II error rates while recognizing important empirical properties of the collected data. In practice, this low-end reference range represents a zone of uncertainty with regard to judging impairment. Though conforming to reference selection criteria, either natural sources of disturbance or low-level stress from human-related causes may compromise the integrity of these sites or place them in these transitional states of biological condition – "supporting" but uncertain.

Finally, the two broad classes of "supporting and acceptable" and "not supporting and unacceptable" were subdivided into an upper and a lower range to designate higher/lower quality within these broad classes. The threshold between the upper and lower range for the "supporting and acceptable" class was set at the median value of reference site scores within this class (i.e., the median of just those sites scoring above the 16th percentile; equal to 87.45 for the 10-metric IBI). Similarly, the threshold between the upper and lower range for the "not supporting" class was set at the median value of test site scores within this class (i.e., the median of just those sites scoring below the 5th percentile reference; equal to 42.24 for the 10-metric IBI). The upper group within the broader "not supporting" class was then designated as "partially supporting" (consistent with the Clean Water Act Section 305(b) reporting language used by the USEPA), and the lowest group as not-supporting. Both designations are considered unacceptable conditions, but may be used to set different priorities for regulatory actions and/or restoration efforts. The median value of the observed data was chosen in each case because it represents a robust statistic that is relatively insensitive to both extreme values and the amount of data collected.

The condition classes defined above can be equated to aquatic life use "tiers" that correspond to changes over a biological condition gradient (*see* Davies and Jackson 2006) created by responses to a human disturbance gradient (Figure 7). For the purposes of reporting, it may be useful to make the names of these tiers interchangeable with condition classes that have been created for other applications. IBIs have often used terms such as "good," "fair," and "poor" as descriptors of condition class (e.g., Barbour et al. 1999, Ode et al. 2005). And, using the analogy of "report card" grades for expressing condition quality, the 5 tiers could also be called A-B-C-D-F. [**Note**: although the thresholds described above are reliable, we decided to use the more common levels of 5th and 25th reference percentile to define impaired and intermediate supporting levels; Appendix V.]

3.4 Spatial and Temporal Variability of the 10-Metric IBI

The results and discussion presented above regarding the recommended 10-Metric IBI clearly show that this multimetric index for assessing the ecological integrity of stream macroinvertebrate communities synthesizes the complex and often noisy invertebrate data into a single number with three important characteristics: low background variability; an ability to correctly assign minimally disturbed sites to a Fully Supporting category; and a discriminatory power that separates sites with varying levels of human disturbance into other categories of ecological condition. Attaining these characteristics is the specific goal of the multimetric creation process. However, once the multimetric IBI is selected there are a number of additional means of assessing the performance of that IBI. In this section, we explore the performance of the recommended 10-Metric IBI in terms of three additional performance measures: (1) small-scale within-site sampling variability; (2) among-year temporal variability; and (3) use of variability measures to derive a minimum detectable difference (MDD) to characterize the statistical reliability of condition class designations based on the repeatability of site IBI scores.

As additional steps in the cross-validation of assessment outcomes, classification structure from the 10-Metric IBI was compared to the performance of a multivariate ("RIVPACS"-type) predictive model (Section 3.5), and to an IBI based on lower taxonomic resolution and fewer metrics (Section 3.6).

3.4(a) Within-Site Sampling Variability

As described in the Methods section, the standard protocol for all sampling sites evaluated in this report was to collect and process 5 separate composite samples during each sampling event at every site. In order to facilitate the use of this IBI tool with the new protocols adopted by the State of California's Surface Water Ambient Monitoring Program (SWAMP),[9] a random sample of 500 invertebrates from these 5 replicate samples was used as the basis for all calculations presented thus far in the report. However, because each of the 5 replicates were processed and tabulated separately, an evaluation of replicate sampling variation at each station can be performed.

Each of the 5 replicates can be used to compute an IBI score for that site-date combination, and the within-site variability can then be quantified and evaluated. Such an analysis is a measure of riffle-scale variability within the reach rather than the full reach-scale sample variability. In the 5 riffle-scale samples here, the threshold for sample processing was to sort and enumerate a minimum of 300 invertebrates from each replicate, so these samples are frequently based on samples of less than the 500 individuals used to develop the 10-Metric IBI. Across the 670 individual samples used for this analysis (5 replicates in 134 site-surveys), the median number of invertebrates

---

[9] *See* State of California (2007)

processed was 393 individuals, with 221 (33%) of the samples based on 500 or more invertebrates. The replicate samples were also derived from pooling of 3 separate sampling locations within the same riffle area rather than the 8 composite samples collected from over the entire reach now used in standard California SWAMP bioassessment protocols (targeted riffle composite method). The effect of these discrepancies compared to a fully replicated reach-scale re-sampling is to increase the expected variability among replicate samples in our data set. Thus, the estimates for the within-site sampling variability presented below should be viewed as upper limits to the expected reach-scale sampling variability. Repeated statistical re-sampling of these 5 replicates to achieve a 500 fixed-count can be used to derive a lower limit estimate of sampling variability.

The among-replicate standard deviation of 10-Metric IBI scores was used as a measure of variability for within-site sampling variability. The standard deviation is an unbiased estimator of the within-site variability and has minimal variability among estimators of the variance. The following is a summary of the 134 independent estimates of within-site variability (standard deviation) across the 134 site-date combinations of both reference and test sites:

Statistic	Value
Minimum	1.8
25th Percentile	5.6
Median	7.4
Mean	8.0
75th Percentile	9.9
Maximum	22.3

These summary statistics indicate that, across 134 sampling surveys, the individual replicates typically score within 8 points of the group mean. However, there are sites where the variability among replicates was substantially greater than or less than this typical variability. The typical variation in IBI scores among replicate samples compares well with the among-year variability (see discussion below) and is low on an absolute scale, as well. In addition, analyses for reference and test sites separately reveal that there is no difference in the amount of variability seen for these two classes of streams.

3.4(b) Among-Year Temporal Variability

Among-year temporal variability is perhaps the most complete and robust assessment of a bioassessment method because it includes the spatial variation in macroinvertebrate communities that would be captured using within-site replication, and because it captures the natural temporal variability in the structure of the macroinvertebrate community that results from deterministic and stochastic factors affecting population sizes and distributions. Thus, if a site has no change in human disturbance and scores within a narrow range on a given bioassessment scale when sampled at multiple points in time, that bioassessment method has attained the criterion of “repeatability.”

The existing data set for the Lahontan Region, where 36 of the 70 sites were sampled in 2, 3, or 4 separate years, thus provides an extraordinary opportunity to assess the degree of repeatability for the recommended bioassessment tool, the 10-Metric IBI presented above. As with the within-site variability, we focused this analysis on the standard deviation among repeated sampling events at a site as the measure of temporal variability in IBI scores. The distribution of the resulting 36 standard deviation estimates is characterized below (scores for each year at each site are presented in Appendix III):

Statistic	Value
Minimum	0.3
25th Percentile	4.0
Median	7.7
Mean	9.1
75th Percentile	13.2
Maximum	34.5

These measures of temporal variability in IBI scores are very similar to those presented earlier for within-site variability. Specifically, the mean and median standard deviation for both measures of variability range from 7.4 to 9.1 points on the 100-point IBI scale. However, the range of standard deviations among sites sampled in multiple years is slightly larger, with a lower minimum, higher maximum, and greater spread in the interquartile range (i.e., the range between the 25th percentile and the 75th percentile). Nevertheless, IBI scores for a site sampled in multiple years typically varied from the mean of scores by approximately 10 points or less, suggesting high consistency for most sites in their scoring for the recommended bioassessment tool (the 10-Metric IBI).

3.4(c) Minimum Detectable Difference (MDD) Estimates

Using the between-riffle replicate sampling variability (see Section 3.4a above) as an upper limit of variation associated with repeat sampling, and repeated statistical re-sampling of these replicates to obtain a 500-count fixed sample (Section 2.4) to represent a lower limit of sampling variability, estimates of minimum detectable difference were obtained for a 2-sample comparison (*see* Zar 1999, Fore 2004). This MDD calculated here specifies the difference that must exist between two samples for there to be an 80% probability ($\beta = 0.20$) of detecting a significant difference between the samples (at $\alpha = 0.05$ level). The value found here, based on the MSE of the ANOVA, is 18.08 (MSE = 77.84, based on n=5 sample replicates, and $t_{\alpha\ 0.05(2),8} = 2.306$, $t_{\beta\ 0.2(1),8} = 0.936$). This would permit detection of 5 condition classes over a 100-point IBI scale. For repeated re-sampling to 500 fixed-count composite samples, the MSE = 17.63 and the MDD = 8.61. This estimate of MDD would permit detection of 10 condition classes on a 100-point scale. The mid-range between these maximum and minimum estimates of detectable difference is about 13, which is similar to the interval between the supporting condition classes (roughly 90-100, 80-90, 63-80; Figure 7).

3.5 Classification Structure Compared to RIVPACS-type Model

To further examine whether assessment depends on the method of data analysis, a RIVPACS-type multivariate predictive model (e.g., Hawkins et al. 2000) was developed for the 62 separate site-date combinations of minimally-disturbed reference sites using the protocols outlined in the methods comparison study for the Lahontan basin (*see* Herbst and Silldorff 2006). An entirely new model was developed for the Level II taxonomic data set, and this model was applied to both the reference and test sites to estimate the observed-to-expected (O/E) ratio of taxa richness for common invertebrates.

A probability of 0.50 was selected as the threshold for “common” invertebrates, with any taxon predicted at a site with a probability of less that 0.50 eliminated from the calculation of that site’s score. The cluster analysis of reference sites used an amalgam of clustering groups derived from simple Unweighted Pair Groups Mean Averaging (UPGMA) as well as flexible-beta Weighted Pair Groups Mean Averaging (WPGMA) with beta values ranging from -0.1 to 0.0 (note: clusterings based on Ward’s algorithm and flexible-beta clustering, with beta ranging from -0.5 to -0.2, did not yield cluster groups that were as well-supported by the invertebrate data or the pairwise similarities within the presence/absence similarity matrix). The final clustering utilized 5 groups of site-dates, with between 5 and 20 site-dates within each cluster grouping. The discriminant analysis models that we evaluated used multiple predictor variables of environmental and location data, with the final model selected having the following three predictor variables to assign sites to cluster groups:

1. Latitude
2. Stream Width
3. Estimated Yearly Precipitation

The Estimated Yearly Precipitation was derived from the CLIMOD predictions for California, and these data were log-transformed in order to better approximate a normal distribution. The Stream Width was also transformed to better approximate a normal distribution, with a square-root transformation used for this variable. This model was selected based on a parsimonious description of the variability in site identifications with low values of apparent and cross-validation error rates (<20% for the final model).

The final RIVPACS-type model had a median score for reference sites of 1.03, with a mean value of 1.02, and maximum and minimum scores of 1.22 and 0.72, respectively. The coefficient of variation of O/E scores for reference sites may serve as the standard measure of model performance for RIVPACS-type models (Van Sickle et al. 2005). For the model developed during this study, this CV was 0.114 for the reference sites, which is on the low end of the range reported for RIVPACS models (*see* Herbst and Silldorff 2006).

Although this RIVPACS-type model can be used separately, the current analysis focuses on the comparison between the O/E and IBI classifications of sites into three categories:

1. Supporting: ≥ the 16th percentile (or 17th for RIVPACS) of Reference Sites
2. Supporting-Intermediate: ≤ the 16th percentile and ≥ 5th percentile Reference
3. Not-Supporting: < 5th percentile of Reference Sites

Table 11 summarizes the classifications based on the recommended 10-Metric IBI and the final RIVPACS-type model. Among reference sites, the IBI and RIVPACS methods produced overall similar rates of classification into the Supporting and Intermediate categories, but there were differences in the specific sites placed within the Intermediate category. This led to an overall difference in classification of 26% for these reference sites. Among the Test sites, the overall difference in classification was slightly lower (21% of sites classified into different categories) but still relatively high compared to the classification matches among the 3 IBIs considered. Thus, the RIVPACS assessments provided relatively good correspondence with the IBI assessment (74-79%) but the misclassification rate was substantially greater than seen when different versions of the IBIs were compared.

It should be noted that RIVPACS-type models are known to perform best when data are available from a large number of sites, though 25 may be considered a minimum (Bailey et al 2004). This model was constructed with what is considered to be about the minimum number of sites for a model of this type. Therefore, while the model by all available measures demonstrated good performance characteristics, the model's outputs cannot reasonably be expected to exactly match the IBI's categorical classifications. The key point here is that the RIVPACS-type model and the recommended 10-Metric IBI are in close agreement regarding the fundamental designations of impaired versus unimpaired condition.

While we believe that the recommended 10-Metric IBI possesses performance characteristics that allows its use as a stand-alone tool, the use of these analytical tools in conjunction with one another could permit the application of separate lines of evidence in making evaluations of biological integrity, and provide more certainty in making decisions regarding regulatory actions. However, operating the RIVPACS-type model would require specialized expertise, and would therefore be more complex, time-consuming, and expensive to use in standard applications than calculating the 10-Metric IBI score.

In sum, our analysis using the RIVPACS-type model confirms our belief that the 10-Metric IBI has adequate performance characteristics to be used as a stand-alone assessment and regulatory tool. But due to the resources required to operate and maintain a RIVPACS-type model over time, other options should be considered if multiple lines of evidence are desired. For example, it may be more cost-effective to add a second ecological indicator (such as algae assemblages) to stream assessments rather than a second method of analyzing benthic macroinvertebrate communities.

3.6 An IBI Based on Simplified Taxonomic Resolution

The 10-metric IBI was developed based on taxonomic resolution that identified specimens to what is known as Level II Standard Taxonomic Effort (STE) as defined by SAFIT (Southwestern Association of Freshwater Invertebrate Taxonomists - the successor to CAMLnet; listed in Appendix IV). At this level, most taxa are identified at least to genus and many to species where reliable keys are available. The final data set analyzed was based on this STE Level II (refer to the Master Taxa List for Lahontan Region, Appendix IV; and to documentation at http://www.safit.org/ste.html), but as some taxa were not always of certain status ("non-distinct"), some Level II designations were held at Level I so that (a) uncertain identifications were not included in the data set, and (b) no "phantom" taxa were created – those resulting from distinct specimens being recorded separately from non-distinct and being counted as different taxa. In the Level II data set, this was reflected most commonly by dropping species level designations for *Baetis*, *Serratella*, *Zapada*, and many gastropod snails being left at genus (that is, where species identifications were made, they were "backed-off" to genus when entering these data for analysis).

An important consideration in the practical use and application of IBIs is the uncertainty that identifications are consistent when produced by different laboratories, and that data compilation and computation of metrics and scores by different analysts has no ambiguity. One way to minimize potential discrepancies is to simplify the taxonomy to lower levels of resolution. To achieve this simplification, and compare differences in assessment of biological integrity and impairment, we converted the Level II data set to SAFIT Level I resolution. (*See* Appendix V.) The main differences in resolution of identifications used are summarized in Table 9, and show that collapse of the family *Chironomidae* (midges) from 100 genus/species designations to the single family, and *Rhyacophila* from 17 species groups to one genus, accounts for most of the decrease in resolution in going from Level II to Level I.

In making a direct conversion from Level II to Level I we also used the same suite of metrics but eliminated one metric no longer relevant at the Level I taxonomic resolution (chironomid percent of richness). This resulted in a 9-metric IBI as the recommended index for biological assessments when using the Level I taxonomic resolution (*see* Appendix V for more details). As an additional step in making assessments more consistent, thresholds for defining condition classes were aligned with other IBIs and assessment tools developed in California. Specifically, the limit between the intermediate class of "supporting intermediate" to "fully supporting" was changed to 25th percentile of the reference distribution (from the 16th used in the Level II IBI). The impairment threshold remained at the 5th percentile of reference, and the upper and lower condition classes (supporting and impaired) were split at the median values of scores within these defined ranges (i.e., median of reference sites > 25th percentile, median of test sites < 5th percentile reference) (Table 10).

The Level I IBI showed similar performance to the Level II IBI and to the RIVPACS O/E scores in terms of signal:noise ratio (Table 10), and with final biological condition assessments (Appendix V, Tables A.8a and A.8b). Assessments for reference sites were in broad agreement among methods, particularly when comparing the Level I and Level II IBIs. Assessments of test sites also showed close matching among the differing approaches (Appendix V, Table A.8b). The 47 Test sites judged impaired (red color) by the Level II IBI agreed in all but three cases with the Level I IBI, and these primarily reflected borderline sites. All sites classified as impaired by the Level I IBI were also classified as impaired by the Level II IBI. The relatively rare discrepancies of judged impairment show that the results obtained between the Level I and Level II IBIs are usually in accordance, with most differences resulting from borderline assessment decisions. Test sites judged unimpaired were also typically in agreement between the Level I and Level II IBIs, but the RIVPACS O/E model had a more mixed correspondence to both IBIs (see also Section 3.5, above) reflecting the distinct nature of this predictive-model assessment relative to the two multi-metric assessment.

To further test the performance of a component metric (taxa richness) and the 9-Metric (Level I) IBI in relation to a stress gradient, we examined responses to erosion and sedimentation. Using either the percent of smallest particle sizes (fines, sand, gravel), or the D-50 median particle size as measures of sedimentation exposure, the Level I richness and IBI declined as sediment content increased (Figure 8). Effect thresholds of biological response were apparent at %FSG > 50-60%, and at a D-50 value < 50 mm, and are consistent with those observed for the Level II IBI.

We conclude that the 9-Metric (Level I) IBI possesses adequate performance characteristics to be used as a stand-alone tool. It has the added benefits of being less expensive to monitor (i.e., less intensive taxonomy) and less expensive to implement in a regulatory context (i.e., less oversight needed to ensure accuracy of more difficult taxonomic identifications).

4. FURTHER REFINEMENTS OF THE BIOLOGICAL CRITERIA

The data set used here to develop Indices of Biological Integrity (IBIs) and a multivariate RIVPACS-type model can be expanded as additional reference and test sites continue to be added through sampling conducted under the State Water Board’s Surface Water Ambient Monitoring Program (SWAMP). A larger sample size would permit further refinements to biological criteria that can be updated in future reports. The following serves as a short listing of additional information that could be integrated:

1. Test models with validation data sets of new reference and test sites;
2. Use of consistent standardized system for defining reference site selection criteria
3. Include confidence limits in defining assessment certainty (*see* Smith et al. 2005);
4. Contrast with assessments derived from periphyton. (A periphyton IBI is currently under development for the region);
5. Integrate with other Sierra Nevada region data to produce an IBI for the entire extent of the mountain range;

6. Summarize physical and chemical data on habitat quality to serve as an additional indicator of ecological integrity of the surveyed streams.
7. Establish a decision-tree for use in application of IBI assessments for making regulatory decisions regarding compliance, impairment listing, and TMDLs

Appendices to Report:

I. Listing of all metrics screened
II. Listing of core metric calculations
III. Listing of IBI-10 values for streams included in this report (includes repeat years)
IV. Master listing of taxa names and threshold values for invertebrates present in this data set
V. Addendum to Lahontan (Eastern Sierra) IBI Development Report—Simplified taxonomic resolution for bioassessment using a revised Level I standard taxonomic resolution. Includes a color-coded listing of stream / site code names, giving a comparative breakdown of IBI Level II and Level I and O/E scores and condition classes for all 134 site surveys (IBI-10 Level II scores by site and year shown in Appendix III, but this appendix gives a comparative listing by method of assessment and condition class)

LITERATURE CITED

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Kluwer Academic Publishers, New York, New York, USA.

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. U.S. Environmental Protection Agency; Office of Water; Washington, D.C. EPA 481-B-99-002.

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

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16: 1251–1266.

Fore, L.S. 2004. Development and testing of biomonitoring tools for macroinvertebrates in Florida streams. Unpublished report prepared for Florida Department of Environmental Protection.

Hawkins, C.P., R.H. Norris, J.N. Hogue, and J.W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456-1477.

Herbst, D.B. 2002. Quality Assurance Project Plan, Aquatic Invertebrate Bioassessment Monitoring in the Eastern Sierra Nevada: Development of Biological Criteria for Stream Water Quality and Evaluations of Environmental Impacts and Restoration Projects in the Lahontan Region. Unpublished report prepared for the Lahontan Regional Water Quality Control Board, South Lake Tahoe, California. 15 pp. plus appendices. See: http://www.waterboards.ca.gov/lahontan/water_issues/projects/quality_assurance_project_plan/index.shtml

Herbst, D.B. and E.L. Silldorff. 2006. Comparison of the performance of different bioassessment methods: similar evaluations of biological integrity from separate programs and procedures. *Journal of the North American Benthological Society* 25:513-530.

Herbst, D.B., and D.W. Blinn. 2007. Preliminary Index of Biological Integrity (IBI) for Periphyton in the Lahontan Region, Eastern Sierra Nevada, California – Draft Report. Unpublished report submitted to the Lahontan Regional Water Quality Control Board, South Lake Tahoe, California.

Hughes, R. M. 1995. Defining acceptable biological status by comparing with reference conditions. Chapter 4, pp. 31–47, *In:* W. Davis and T. Simon, editors. *Biological assessment and criteria: tools for water resource planning and decision making for rivers and streams.* Lewis, Boca Raton, Florida, USA.

Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10:629–635.

Ode, P.R. 2003 CAMLnet: List of California macroinvertebrate taxa and standard taxonomic effort. Unpublished report of the California Department of Fish and Game, Aquatic Bioassessment Laboratory, Rancho Cordova, California.

Ode, P.R., A.C. Rehn and J.T. May. 2005. A quantitative tool for assessing the integrity of southern coastal California streams. *Environmental Management* 35:493-504.

Rehn, A.C., P.R. Ode, and C.P. Hawkins. 2007. Comparisons of targeted-riffle and reach-wide benthic macroinvertebrate samples: implications for data sharing in stream-condition assessments. *Journal of the North American Benthological Society* 26:332-348.

Smith, J.G., J.J. Beauchamp and A.J. Stewart. 2005. Alternative approach for establishing acceptable thresholds on macroinvertebrate community metrics. *Journal of the North American Benthological Society* 24:428-440.

State of California. 2002. Quality Assurance Management Plan for the State of California's Surface Water Ambient Monitoring Program. California State Water Resources Control Board, Sacramento, California. December 2002 (1st version).

State of California. 2007. Standard Operating Procedures for Collecting Benthic Macroinvertebrate Samples and Associated Physical and Chemical Data for Ambient Bioassessments in California. California State Water Resources Control Board, Sacramento, California. February 2007.

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

Van Sickle, J., C. P. Hawkins, D. P. Larsen, and A. T. Herlihy. 2005. A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society* 24:178–191.

Zar, J.H. 1999. Biostatistical Analysis, 4th edn. Prentice Hall, Upper Saddle River, New Jersey.

Table 1. Coefficient of variation for candidate metrics among reference streams (metric "noise") based on fixed 500-count.

Metric	**Coefficient of Variation**
Richness (total number of taxa)	0.14
Biotic Index (modified Hilsenhoff, composite tolerance)	0.19
Intolerant Percent Richness (% of taxa with TV= 0,1,2)	0.19
Dominance 3 (proportion of 3 most common taxa)	0.20
Predator Richness (number of predator taxa)	0.21
Chironomidae Percent Richness (% of taxa that are midges)	0.23
Ephemeroptera (E) Richness (number of mayfly taxa)	0.25
Tolerant Percent Richness (% of taxa with TV= 7,8,9,10)	0.25
Trichoptera (T) Richness (number of caddisfly taxa)	0.30
EPT abundance (% of total number that are E, P, and T)	0.31
Plecoptera (P) Richness (number of stonefly taxa)	0.35
Acari Richness (number of water mite taxa)	0.40
Percent Shredders (% of total number that are shredders)	1.20

[TV=tolerance values]

Table 2. Ratio of means between reference and test streams for candidate metrics (metric "signal") based on fixed 500-count (ratio inverted for metrics predicted to increase with disturbance*: biotic index, chironomid percent, dominance 3, and tolerant percent).

Metric	**Ratio of Ref/Test Means**
Trichoptera (T) Richness (number of caddisfly taxa)	1.66
Percent Shredders (% of total number that are shredders)	1.55
Plecoptera (P) Richness (number of stonefly taxa)	1.54
Intolerant Percent Richness (% of taxa with TV= 0,1,2)	1.47
EPT abundance (% of total number that are E, P, and T)	1.42
Ephemeroptera (E) Richness (number of mayfly taxa)	1.41
Predator Richness (number of predator taxa)	1.37
Acari Richness (number of water mite taxa)	1.36
Tolerant Percent Richness (% of taxa with TV= 7,8,9,10)*	1.36
Chironomidae Percent Richness (% of taxa that are midges)*	1.26
Richness (total number of taxa)	1.22
Dominance 3 (proportion of 3 most common taxa)*	1.18
Biotic Index (modified Hilsenhoff, composite tolerance)*	1.15

[TV=tolerance values]

Table 3. Standardized difference between reference and test streams (R minus T except for reverse-response metrics (*), divided by the reference standard deviation) for candidate metrics based on fixed 500-count.

Metric	Standardized Difference
Intolerant Percent Richness (% of taxa with TV= 0,1,2)	1.65
Trichoptera (T) Richness (number of caddisfly taxa)	1.32
Richness (total number of taxa)	1.31
Predator Richness (number of predator taxa)	1.27
Chironomidae Percent Richness (% of taxa that are midges)*	1.15
Ephemeroptera (E) Richness (number of mayfly taxa)	1.15
Plecoptera (P) Richness (number of stonefly taxa)	1.00
EPT abundance (% of total number that are E, P, and T)	0.95
Dominance 3 (proportion of 3 most common taxa)*	0.89
Biotic Index (modified Hilsenhoff, composite tolerance)*	0.79
Acari Richness (number of water mite taxa)	0.66
Tolerant Percent Richness (% of taxa with TV= 7,8,9,10)*	0.56
Percent Shredders (% of total number that are shredders)	0.30

[TV=tolerance values]

Table 4. Percent of test streams with scores exceeding a given percentile of reference streams for candidate metrics based on fixed 500-count (metrics marked by * are reverse-scale metrics where the overlap was measured as the percent of streams lower than the 90th, 75th, and 50th percentiles; respectively).

	Percent > Quantile		
Metric	10%	25%	50%
Intolerant Percent Richness (% of taxa with TV= 0,1,2)	39	29	18
Richness (total number of taxa)	42	28	24
Predator Richness (number of predator taxa)	42	29	24
Trichoptera (T) Richness (number of caddisfly taxa)	51	39	7
Ephemeroptera (E) Richness (number of mayfly taxa)	51	39	29
Plecoptera (P) Richness (number of stonefly taxa)	53	38	26
Dominance 3 (proportion of 3 most common taxa)*	53	43	26
Chironomidae Percent Richness (% of taxa that are midges)*	60	43	18
EPT abundance (% of total number that are E, P, and T)	64	40	21
Percent Shredders (% of total number that are shredders)	64	51	39
Biotic Index (modified Hilsenhoff, composite tolerance)*	65	49	28
Acari Richness (number of water mite taxa)	69	56	26
Tolerant Percent Richness (% of taxa with TV= 7,8,9,10)*	71	50	33

[TV=tolerance values]

Table 5. Correlation coefficients (Pearson) for the candidate metrics based on fixed 500-count. Numbers in bold indicate those considered high (>50% variation attributed) suggesting metrics that may be redundant.

Metric	Total Richness	Ephemeroptera Rich.	Plecoptera Rich.	Trichoptera Rich.	EPT % abundance	Acari Rich.	Predator Rich.	Chironomid Rich. %	Dominance-3	Biotic Index	Tolerant Rich. %	Intolerant Rich. %	Shredder %
Total Richness	1												
Ephemeroptera Rich.	**0.72**	1											
Plecoptera Rich.	0.69	0.62	1										
Trichoptera Rich.	0.61	0.40	0.42	1									
EPT % abundance	0.33	0.60	0.52	0.46	1								
Acari Rich.	0.66	0.38	0.40	0.38	0.11	1							
Predator Rich.	**0.90**	0.66	**0.75**	0.57	0.41	**0.73**	1						
Chironomid Rich. %	-0.28	-0.34	-0.49	-0.50	-0.40	-0.38	-0.42	1					
Dominance-3	-0.58	-0.41	-0.43	-0.26	-0.15	-0.36	-0.49	0.03	1				
Biotic Index	-0.29	-0.49	-0.51	-0.39	**-0.79**	-0.17	-0.41	0.39	0.16	1			
Tolerant Rich. %	-0.13	-0.28	-0.21	-0.27	-0.62	-0.09	-0.19	0.30	-0.01	**0.74**	1		
Intolerant Rich. %	0.53	**0.76**	**0.75**	0.57	**0.76**	0.21	0.59	-0.52	-0.28	**-0.71**	-0.39	1	
Shredder %	0.10	0.06	0.21	0.12	0.35	0.03	0.19	-0.11	-0.18	-0.36	-0.18	0.23	1

Table 6. Types of metrics and combinations used in each of final 3 IBI alternatives.

	Metric Type				IBI combination		
Metric	**richness**	**composition**	**tolerance**	**functional**	**12-metric**	**10-metric**	**8-metric**
Total Richness	X				X	X	
Ephemeroptera Rich.	X				X	X	
Plecoptera Rich.	X				X	X	
Trichoptera Rich.	X				X	X	X
EPT % abundance		X			X		
Acari Rich.	X				X	X	X
Predator Rich.	X			X	X		
Chironomid Rich. %		X			X	X	X
Dominance-3		X	X		X	X	X
Biotic Index			X		X	X	X
Tolerant Rich. %		X	X		X	X	
Intolerant Rich. %		X	X				X
Shredder %				X	X	X	X

Table 7. Performance measures of 3 IBI alternatives. Noise measured as coefficient of variation for reference streams; signal measured as ratio of reference mean to test mean; signal:noise ratio measured as difference between reference and test means divided by reference standard deviation; overlap at 25th percentile indicates the empirical proportion of test streams exceeding the 25th percentile of the reference stream distribution.

IBIs >	12-metric	10-metric	8-metric
Noise	0.13	0.13	0.14
Signal	1.55	1.53	1.53
Signal:Noise Ratio	2.74	2.65	2.45
Overlap at 25th%tile	0.21	0.18	0.19

Table 8. Impairment classification cross-comparisons among 3 alternative IBIs.

		REFERENCE SITES				TEST SITES			
			12-METRIC				**12-METRIC**		
	Condition Class	**Supporting Acceptable**	**Supporting Intermediate**	**Not Supporting**	row sum	**Supporting Acceptable**	**Supporting Intermediate**	**Not Supporting**	row sum
10-METRIC	**Supporting Acceptable**	52	0	0	52	16	0	0	16
	Supporting Intermediate	0	6	0	6	1	8	0	9
	Not Supporting	0	0	4	4	0	0	47	47
	column sum	52	6	4	62	17	8	47	72
			12-METRIC				**12-METRIC**		
8-METRIC	**Supporting Acceptable**	51	0	0	51	15	0	0	15
	Supporting Intermediate	1	6	0	7	2	8	4	14
	Not Supporting	0	0	4	4	0	0	43	43
	column sum	52	6	4	62	17	8	47	72
			10-METRIC				**10-METRIC**		
8-METRIC	**Supporting Acceptable**	51	0	0	51	14	1	0	15
	Supporting Intermediate	1	6	0	7	2	8	4	14
	Not Supporting	0	0	4	4	0	0	43	43
	column sum	52	6	4	62	16	9	47	72

Table 9. Changes in taxonomic resolution converting Level II to Level I data set.

Ranked changes of decreased taxa resolution in converting Level II to Level I Data:		
Group collapsed	**Level II**	**Level I**
Chironomidae	100 genus-spp/grp	1 family
Rhyacophila	17 spp groups	1 genus
Drunella	5 spp	1 genus
Brachycentrus	3 spp	1 genus
Tropisternus	3 spp	1 genus
Dolichopodidae	3 genera	1 family
Paraleptophlebia	2 spp	1 genus
Attenella	2 spp	1 genus
Caudatella	2 spp	1 genus
Ephemerella	2 spp	1 genus
Arctopsyche	2 spp	1 genus
Parapsyche	2 spp	1 genus
Hydropsyche	2 genera	1 genus
Lepidostoma	2 spp	1 genus
Optioservus	2 spp	1 genus
Muscidae	2 spp	1 genus
plus some excluded = Nematomorphs, Staphylinidae, Hydrobiidae, Notonectidae		
Sum = 232 at Level I, from 373 total taxa at Level II		

Table 10. Threshold values for condition classes and performance measures contrasting the Level I and Level II IBIs and the Level II RIVPACS O/E scores

L-II IBI-10	L-I IBI-9	O/E	
63.2	62.1	0.839	5th %tile reference = Impairment threshold
80.4	76.5	0.947	25th %tile reference = Full / Partial support limit
84.1	83.5	1.020	reference mean
89.7	89.4	1.060	median of (reference sites > 25th percentile)
42.2	39.9	0.612	median of impaired tests (test < 5th %tile ref)
11.0	11.3	0.116	SD of reference mean
55.0	55.0	0.677	Test Mean
2.65	2.54	2.95	Signal:Noise = (mean R-T)/Ref SD

Tier / Grade	Designations	Ranges
5 / A	supporting	>median of (reference sites > 25th percentile)
4 / B	supporting	25th to (median of (reference sites > 25th percentile))
3 / C	partial supporting	5th-25th percentile reference
2 / D	not supporting	<5th percentile reference (impairment level) & >median of test values in impaired range
1 / F	not supporting	<median of test values in impaired range

Table 11. Condition assessment classification comparison between 10-metric IBI and RIVPACS-type predictive model.

REFERENCE SITES

10-METRIC IBI

	Condition Class	**Supporting Acceptable**	**Supporting Intermediate**	**Not Supporting**	row sum
RIVPACS-type predictive model	**Supporting Acceptable**	44	5	2	51
	Supporting Intermediate	7	0	0	7
	Not Supporting	1	1	2	4
	column sum	52	6	4	62

TEST SITES

10-METRIC IBI

RIVPACS-type predictive model	**Supporting Acceptable**	9	2	0	11
	Supporting Intermediate	3	2	1	6
	Not Supporting	4	5	46	55
	column sum	16	9	47	72

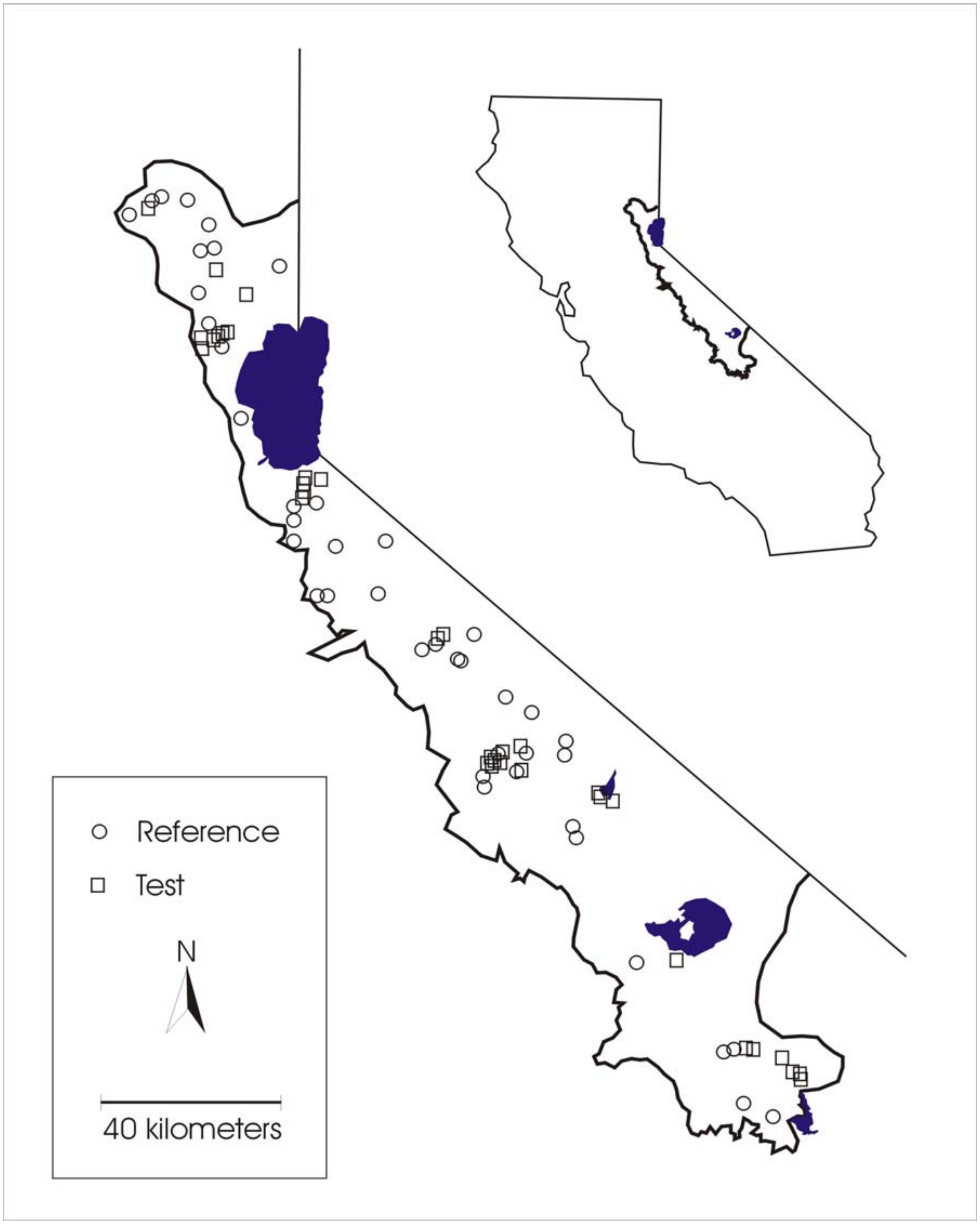
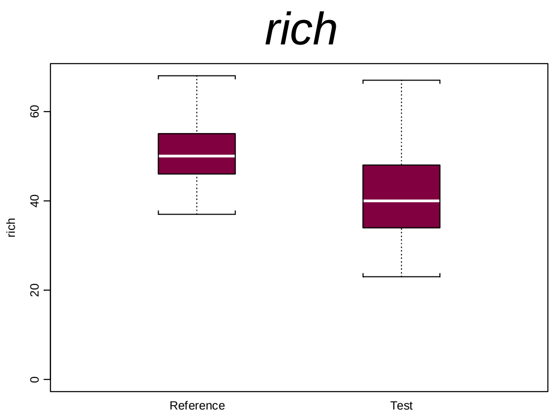
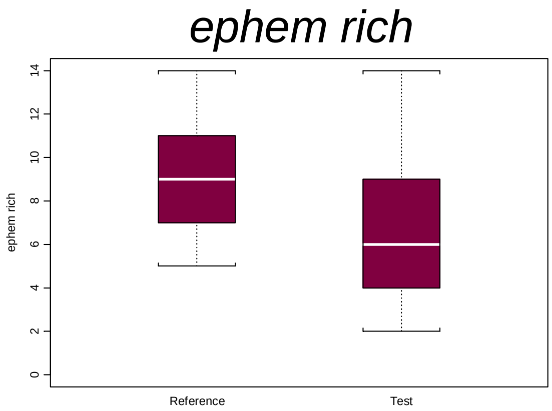
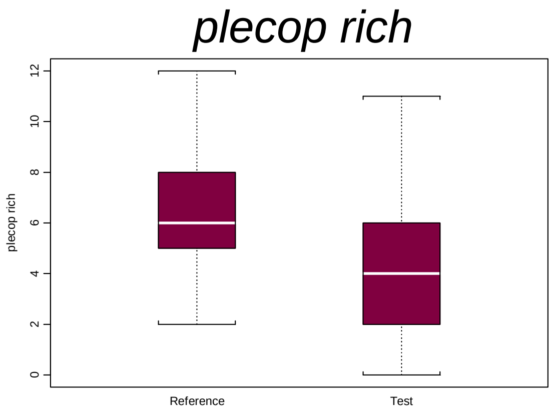


Figure 1. Map of the central Lahontan Region bioassessment sampling locations, consisting of 42 reference sites and 38 test sites. Many of the sites have been repeat sampled in different years over the time period of 1998 to 2003 for a total of 134 site surveys. Replicate sampling at each site along with the between-year data provides a basis for measuring both spatial and temporal components of variability in biological and habitat conditions across sites.

Figure 2. Box-and-whisker plots comparing Reference Streams to Test Streams for 13 Candidate Metrics.







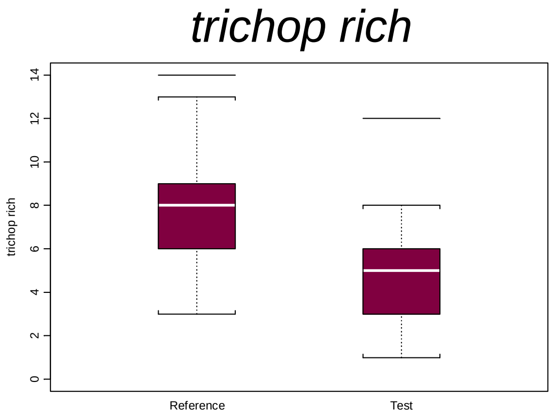
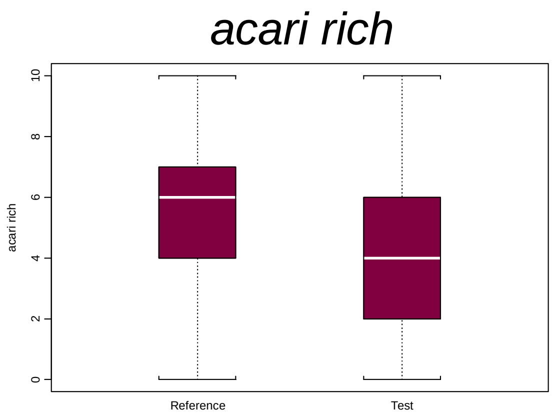
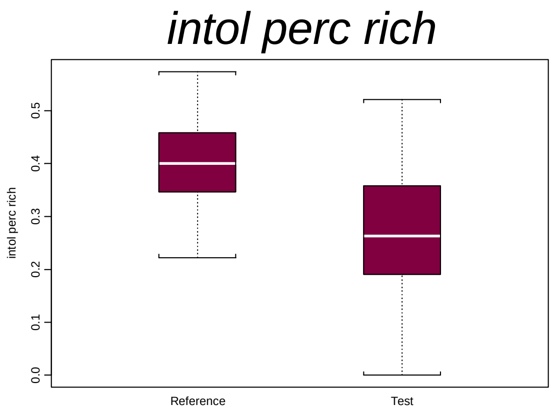


Figure 2. (cont) Box-and-whisker plots comparing Reference Streams to Test Streams for 13 Candidate Metrics.





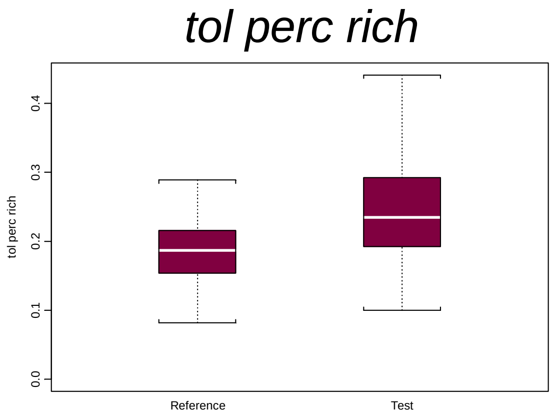
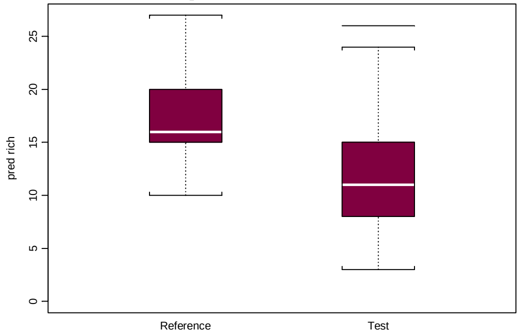
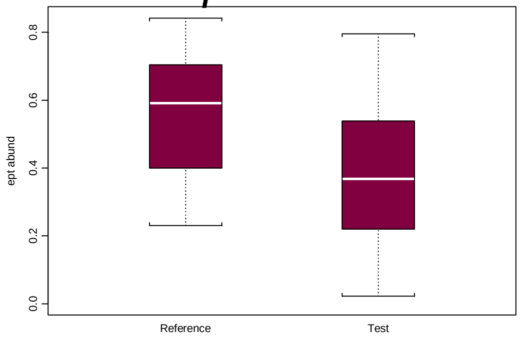


Figure 2. (cont) Box-and-whisker plots comparing Reference Streams to Test Streams for 13 Candidate Metrics.

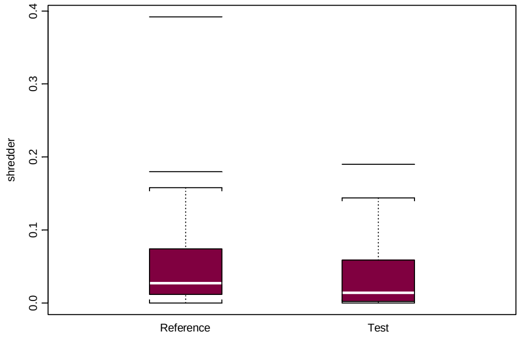
pred rich



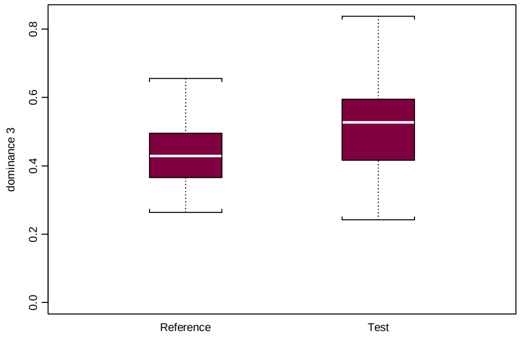
ept abund



shredder



dominance 3



biotic index (bi)

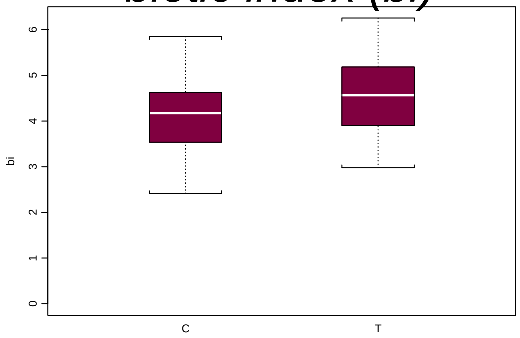
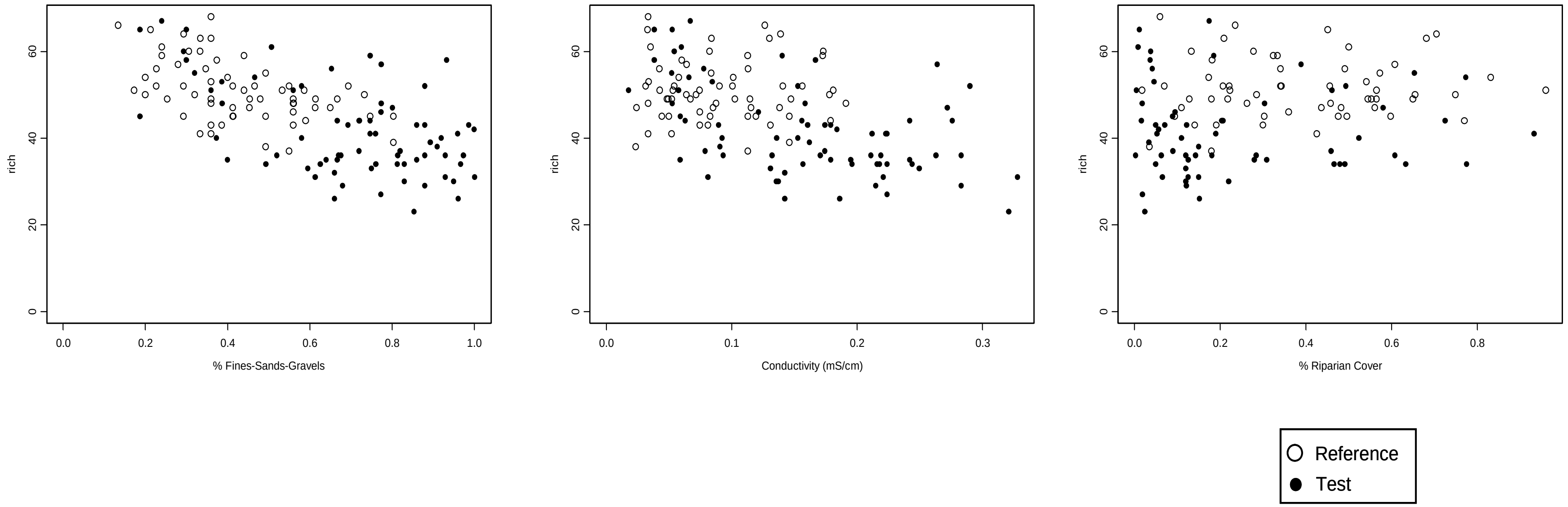


Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance. (solid points represent Test streams, open circles represent Reference streams)

rich



ephem rich

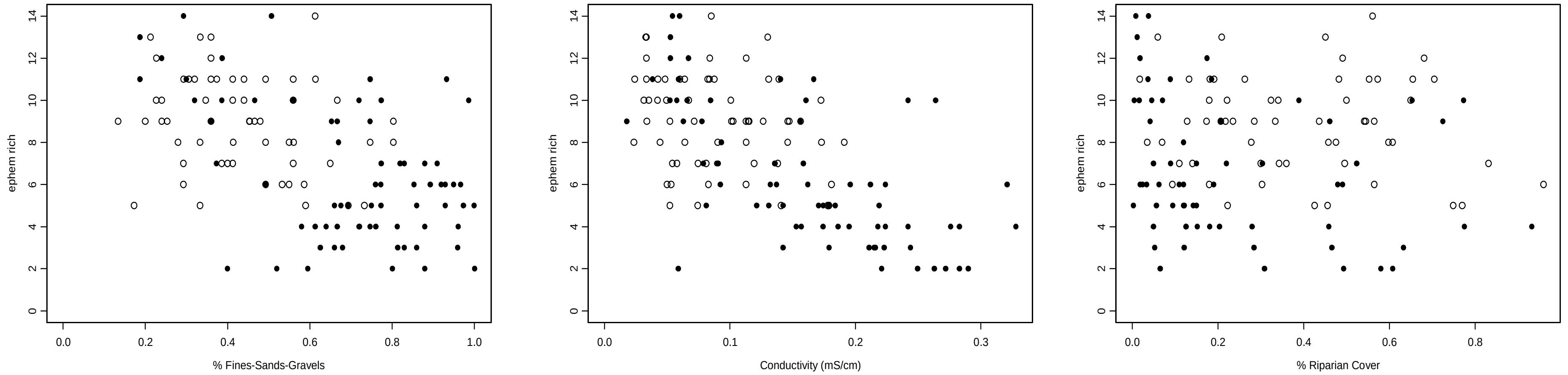
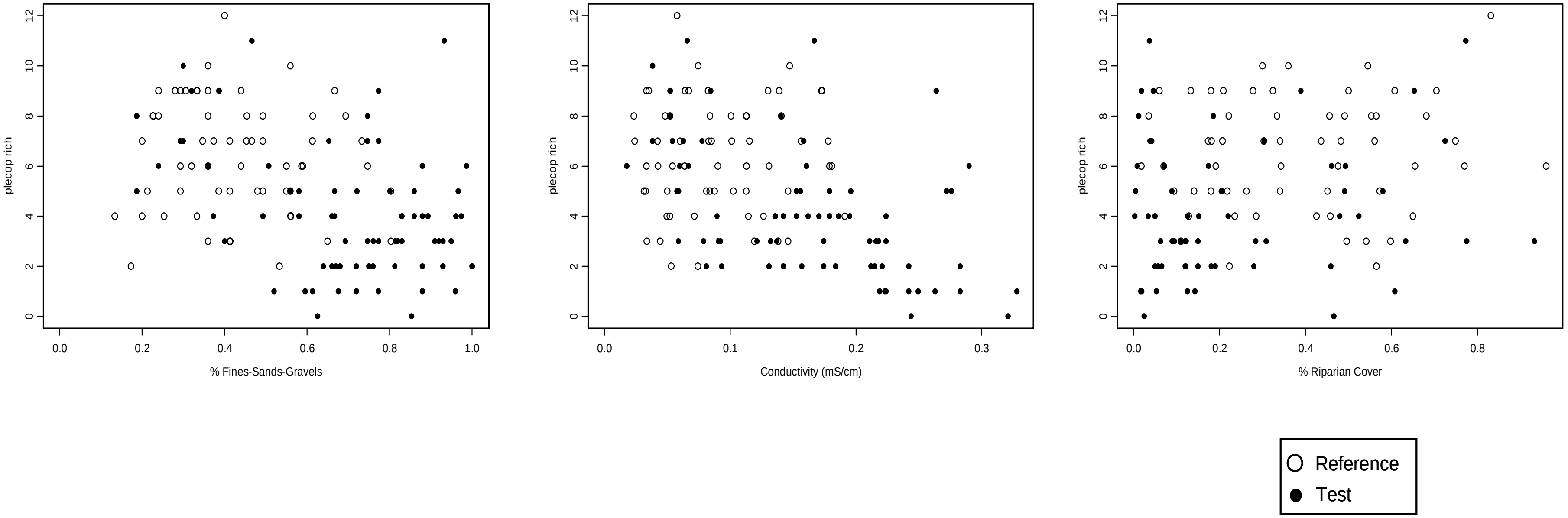


Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)

plecop rich



trichop rich

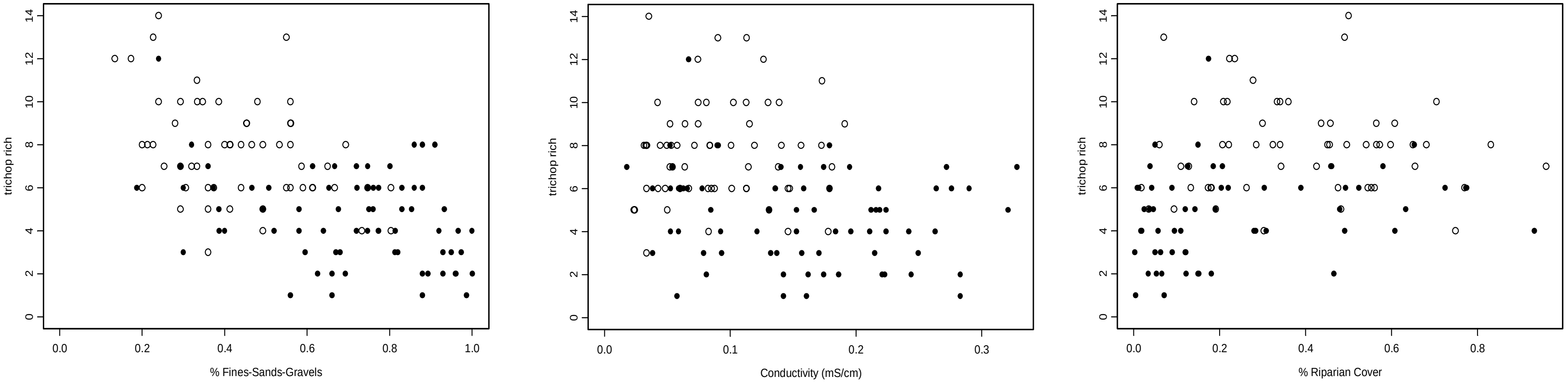
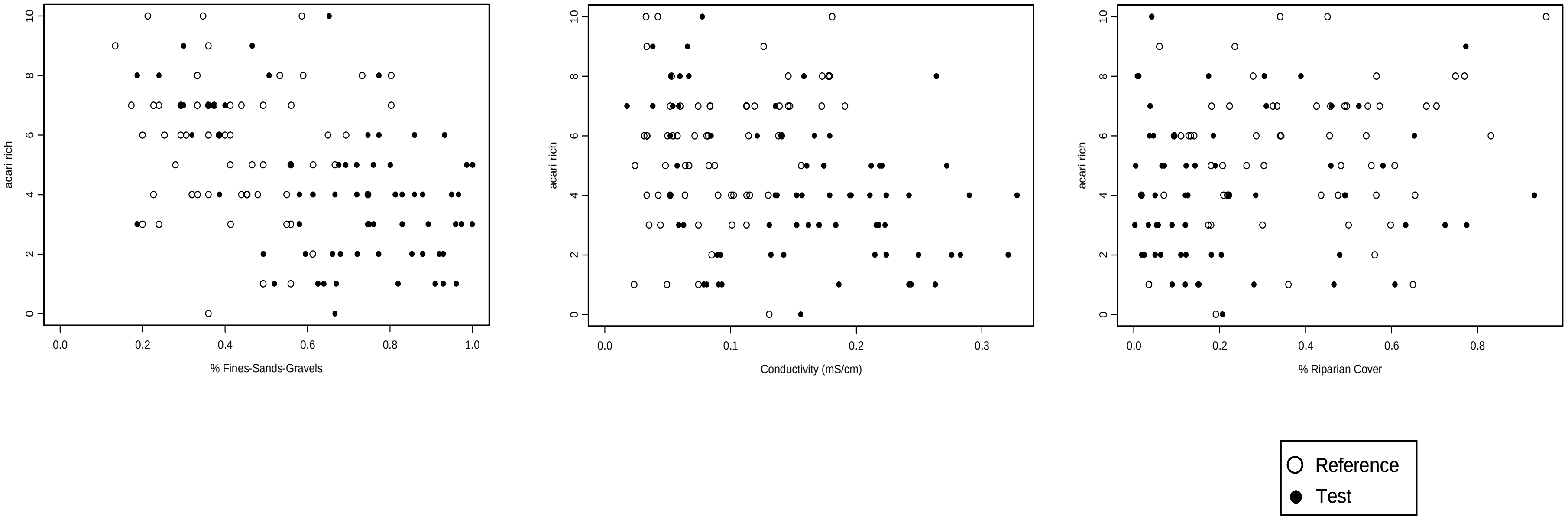


Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)


acari rich

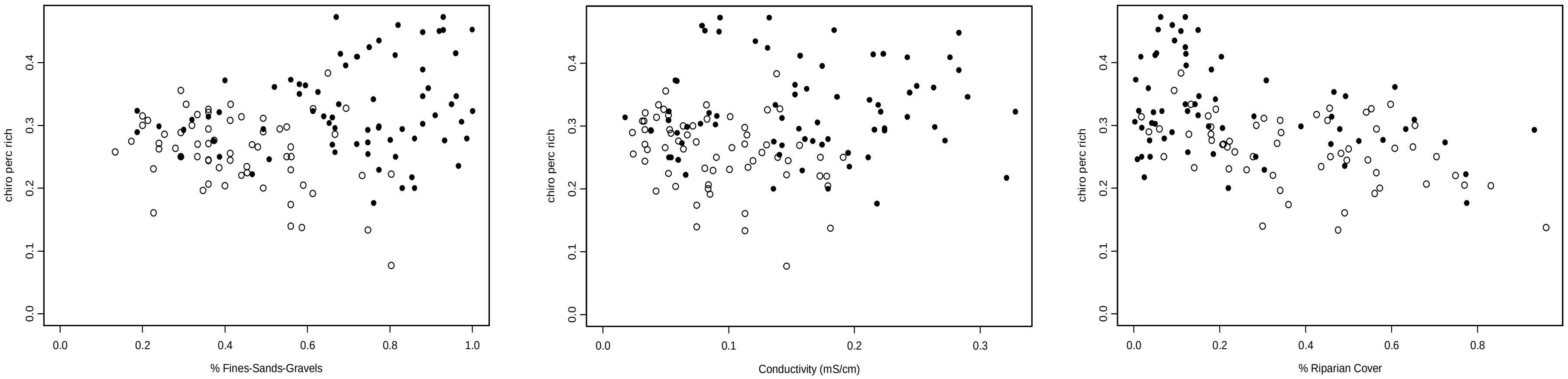
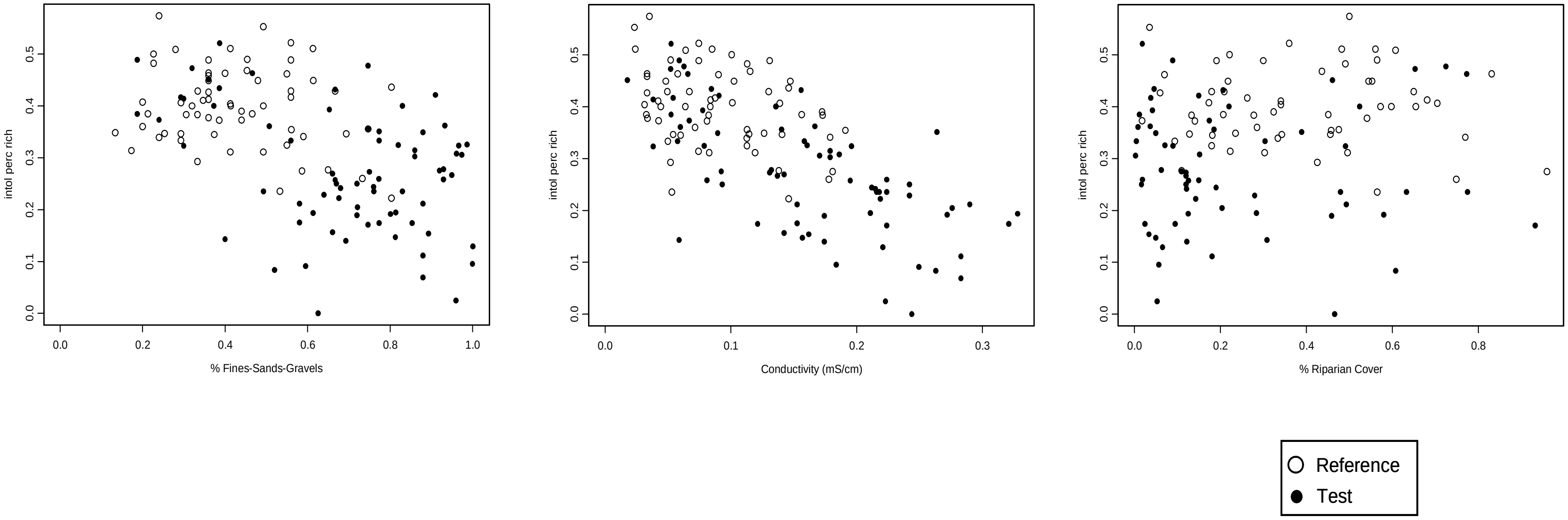

chiro perc rich

Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)

intol perc rich



tol perc rich

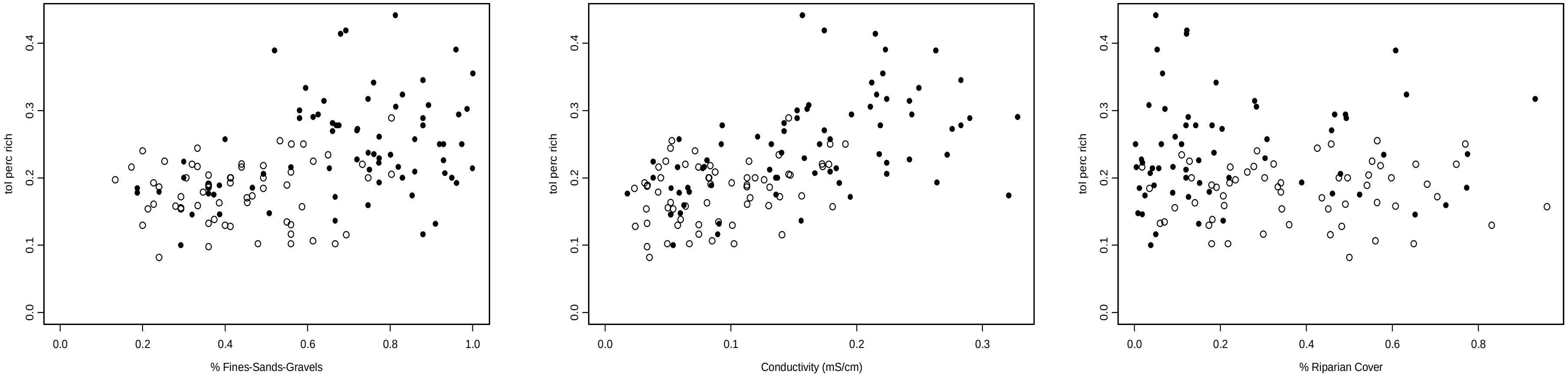
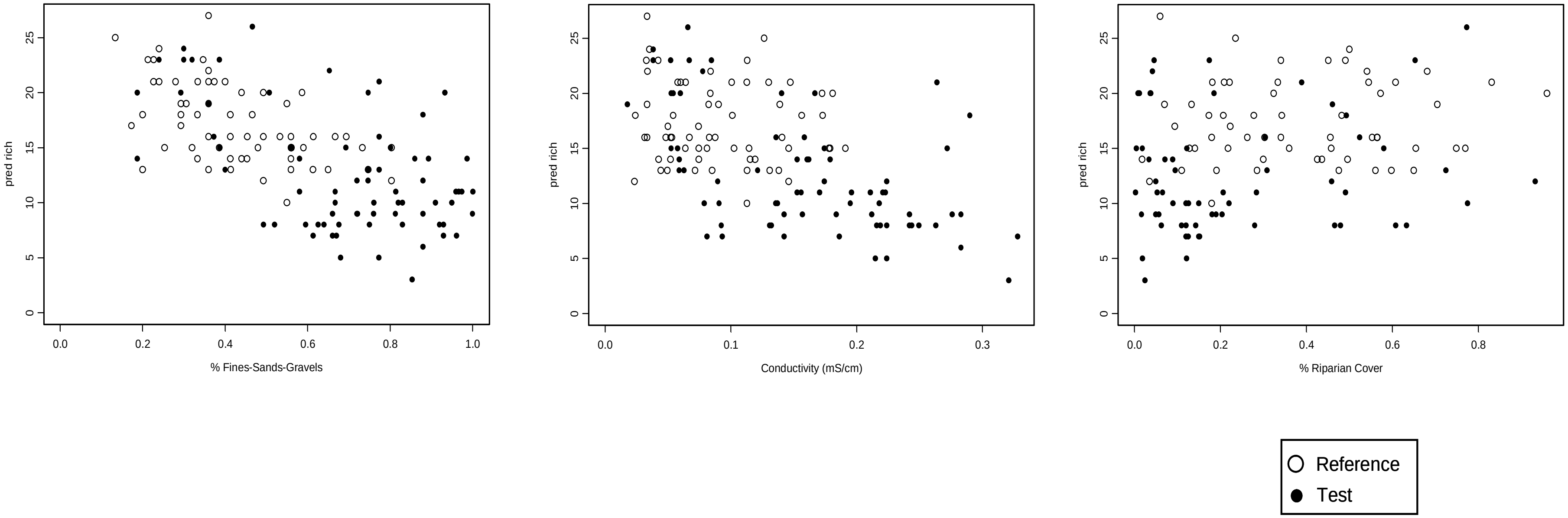


Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)

pred rich



ept abund

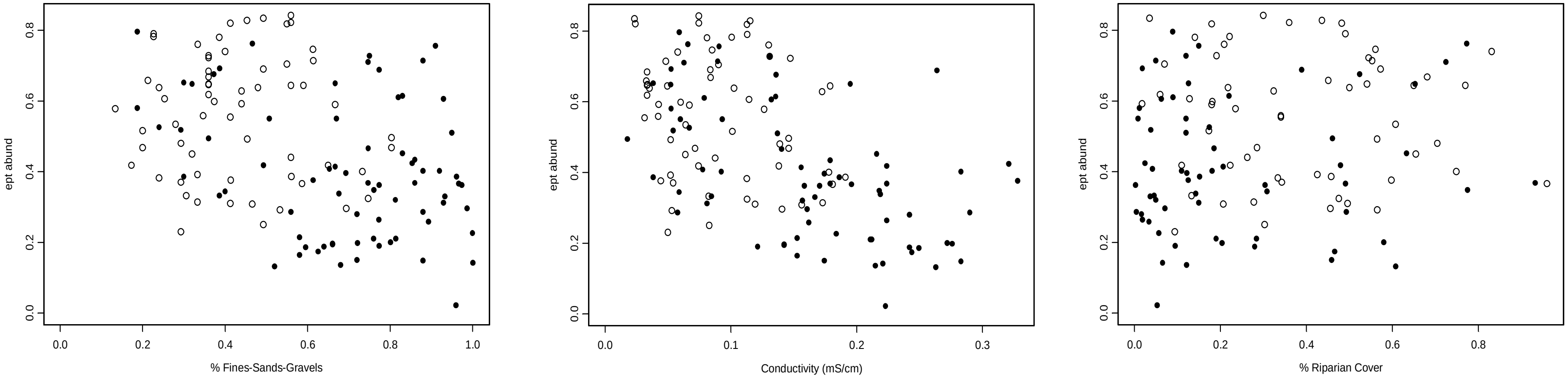
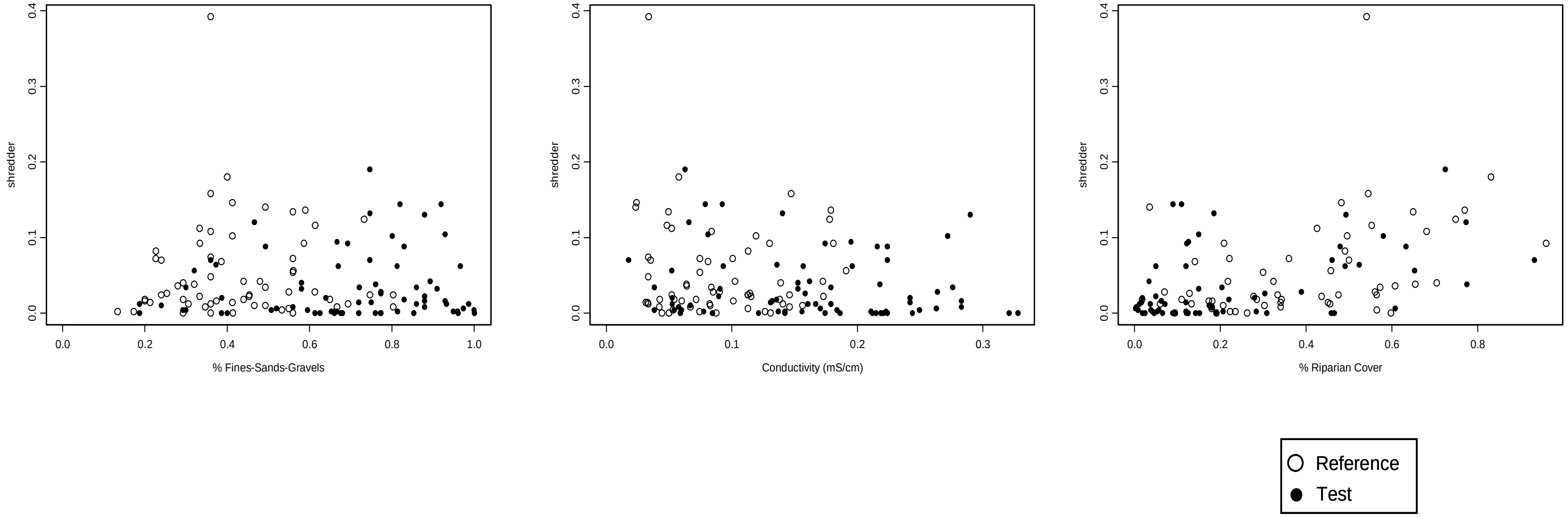


Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)


shredder

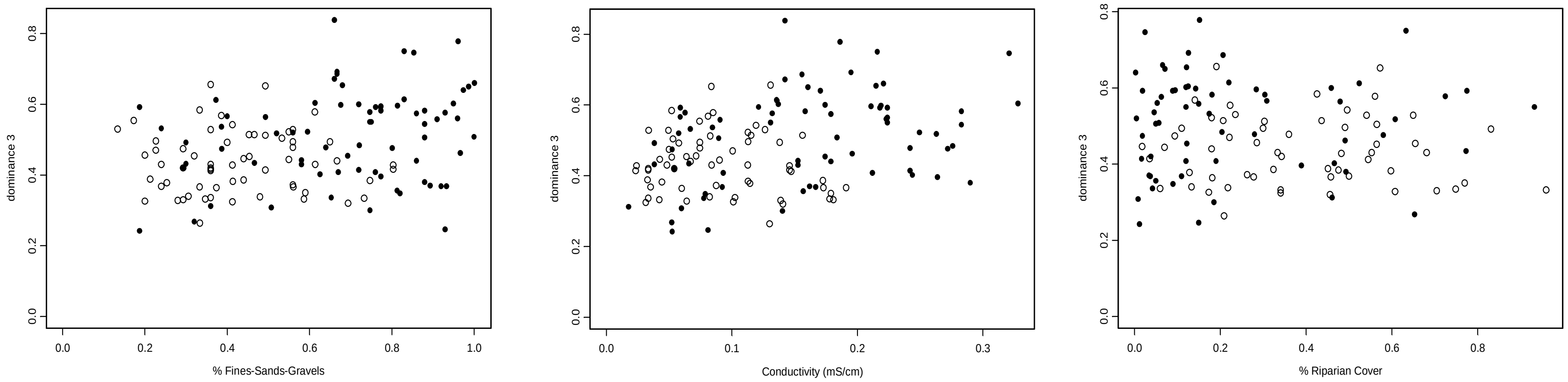

dominance 3

Figure 3. Relationships between 13 Candidate Metrics and 3 Environmental Gradients Related to Human Disturbance.
(cont'd) (solid points represent Test streams, open circles represent Reference streams)

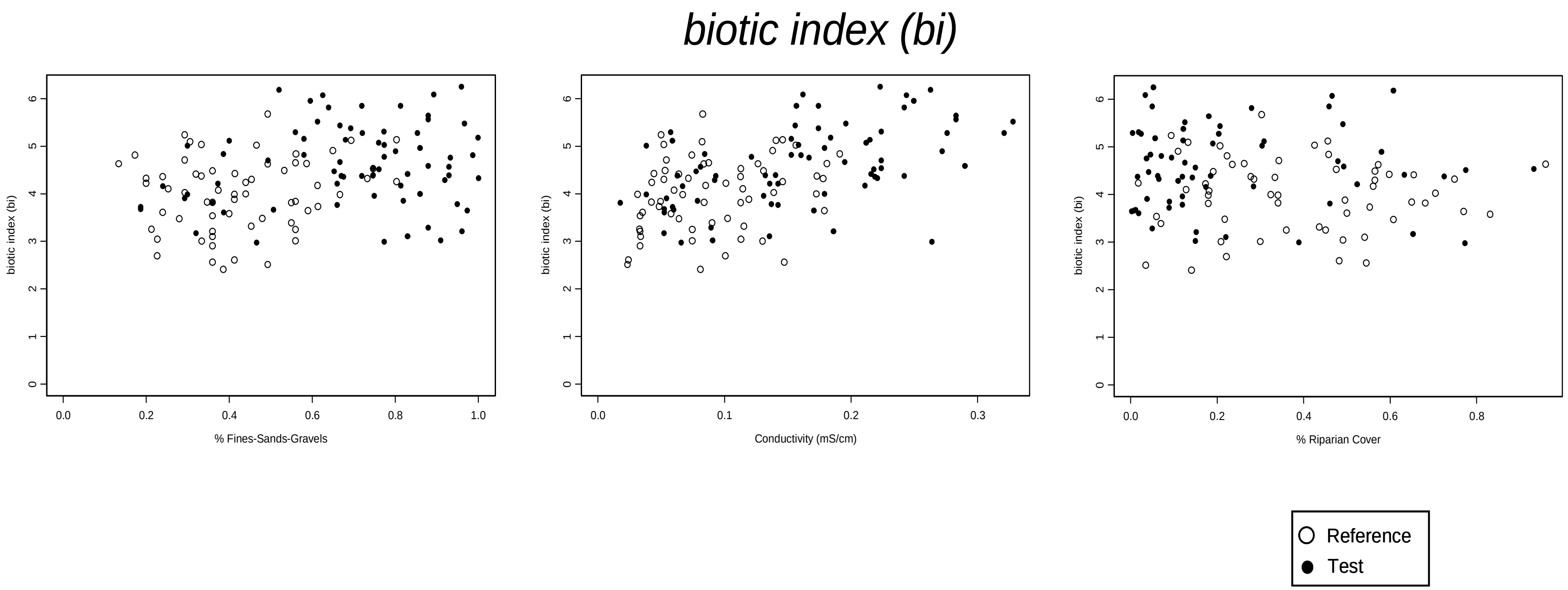
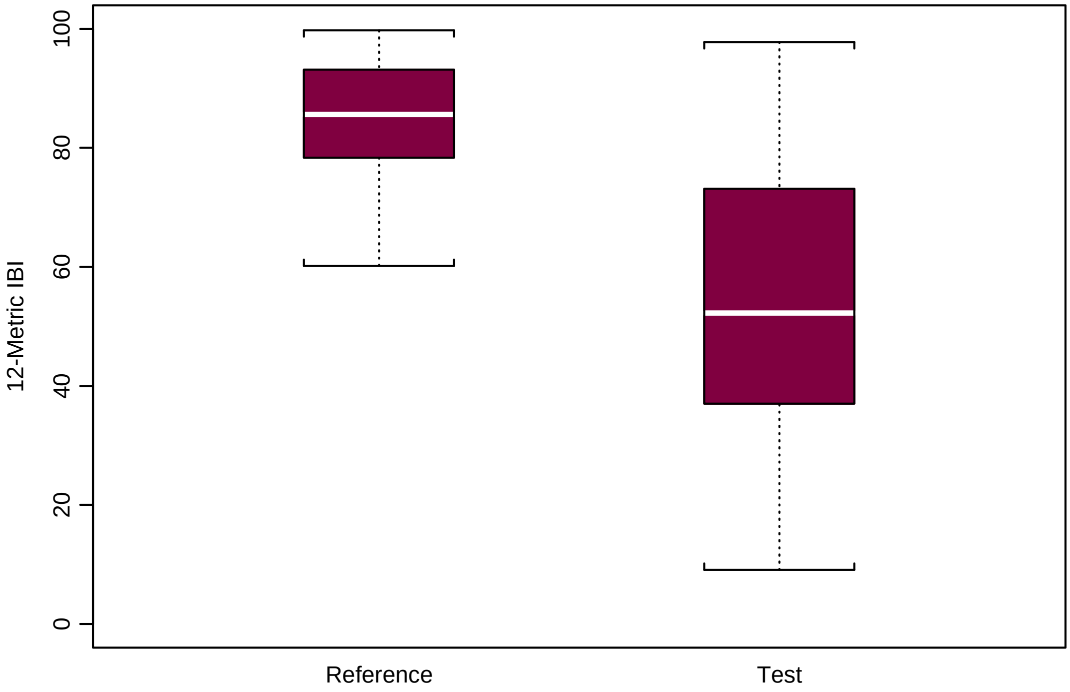


Figure 4. Performance Graphs (box-plot, cumulative distribution plot) for the 12-metric IBI



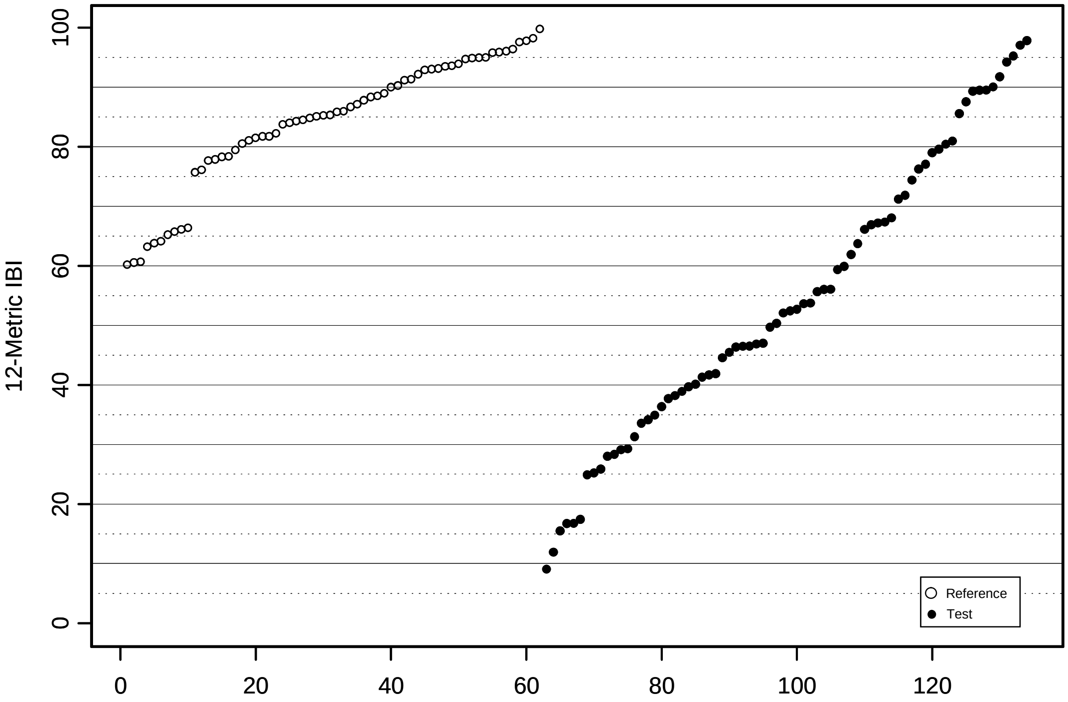


Figure 5. Performance Graphs (box-plot, cumulative distribution plot) for the 10-metric IBI

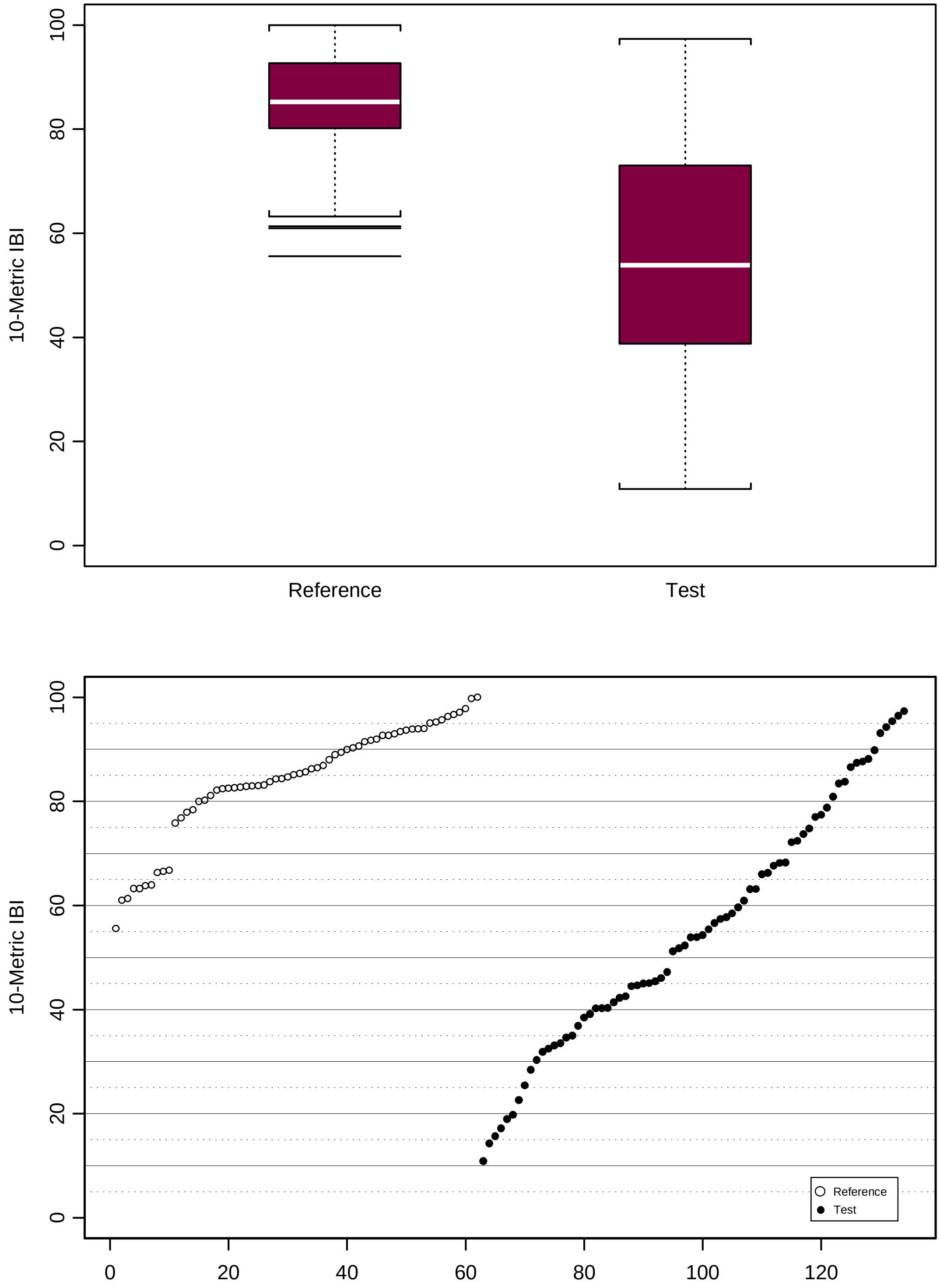
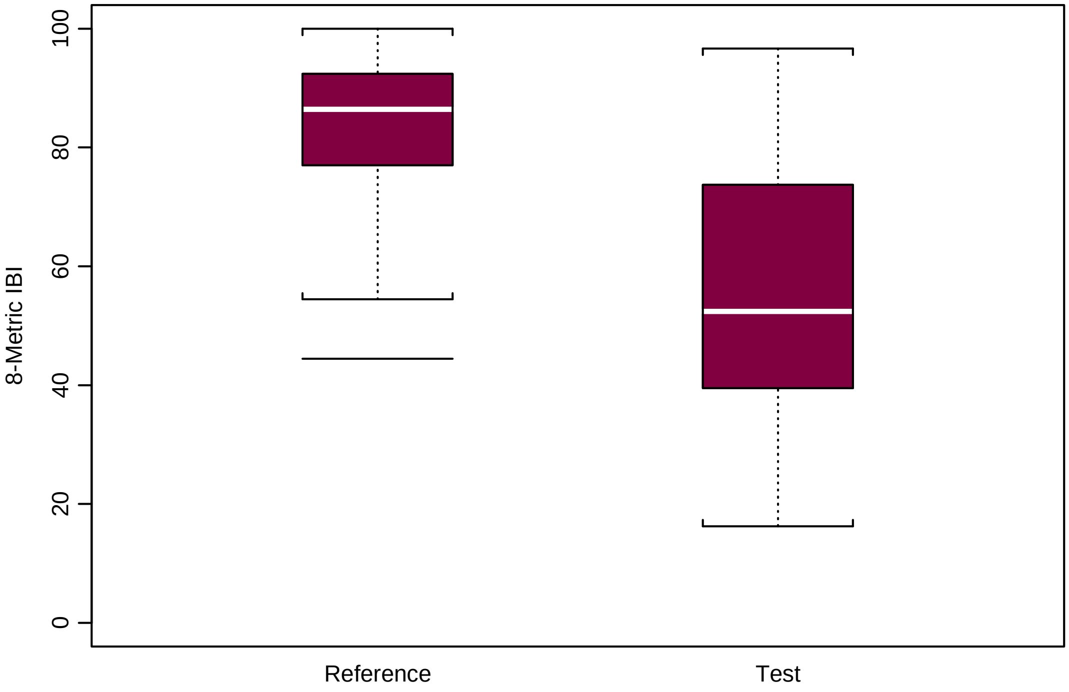
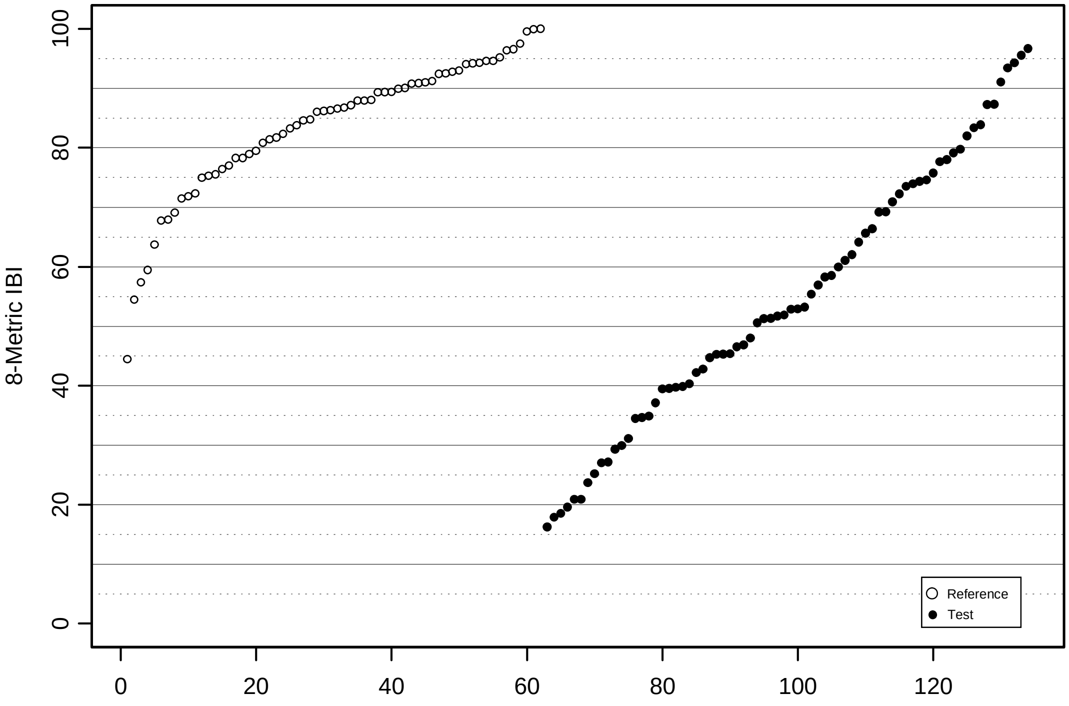


Figure 6. Performance Graphs (box-plot, cumulative distribution plot) for the 8-metric IBI





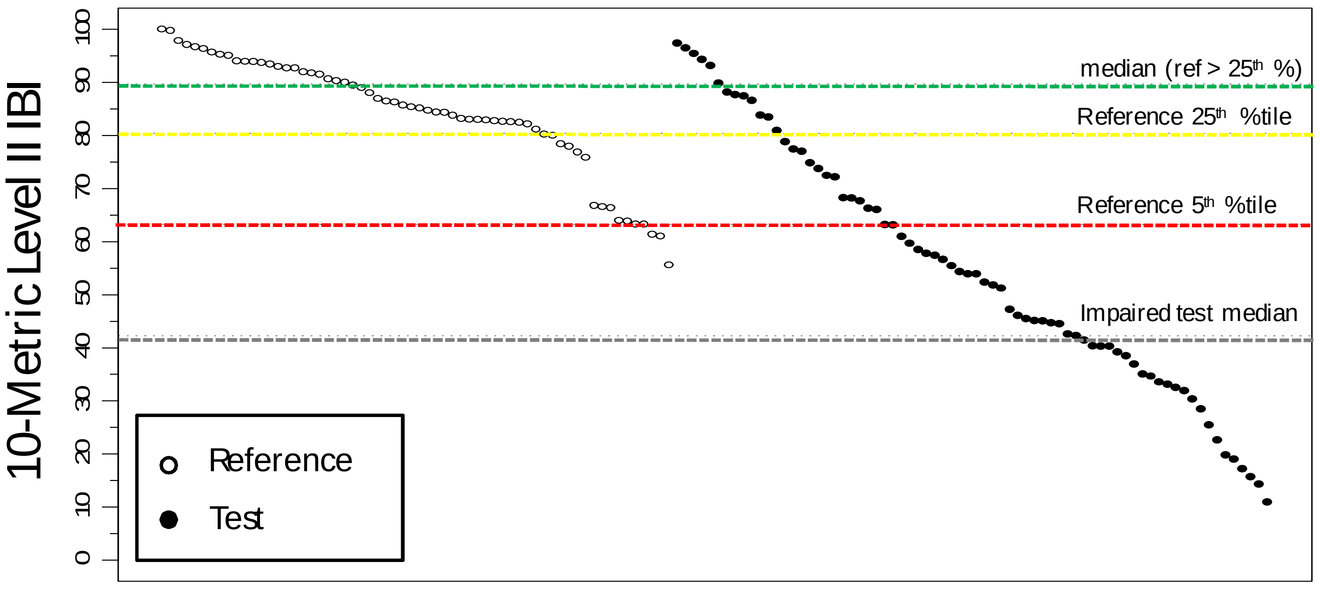


Figure 7. IBI condition or reporting classes forming biological criteria (10-metric IBI). Divisions among these reference and test distributions are based on statistical criteria as outlined in the text. The corresponding descriptive term alternatives are outlined below:

Supporting (Unimpaired)			Impaired	
Acceptable		Intermediate supporting but uncertain	Partially-supporting	Not-supporting
>89.7	80.4 - 89.7	63.2 - 80.4	42.2 - 63.2	<42.2
A	B	C	D	F
Very good	Good	Fair	Poor	Very poor
Good		Fair	Poor	

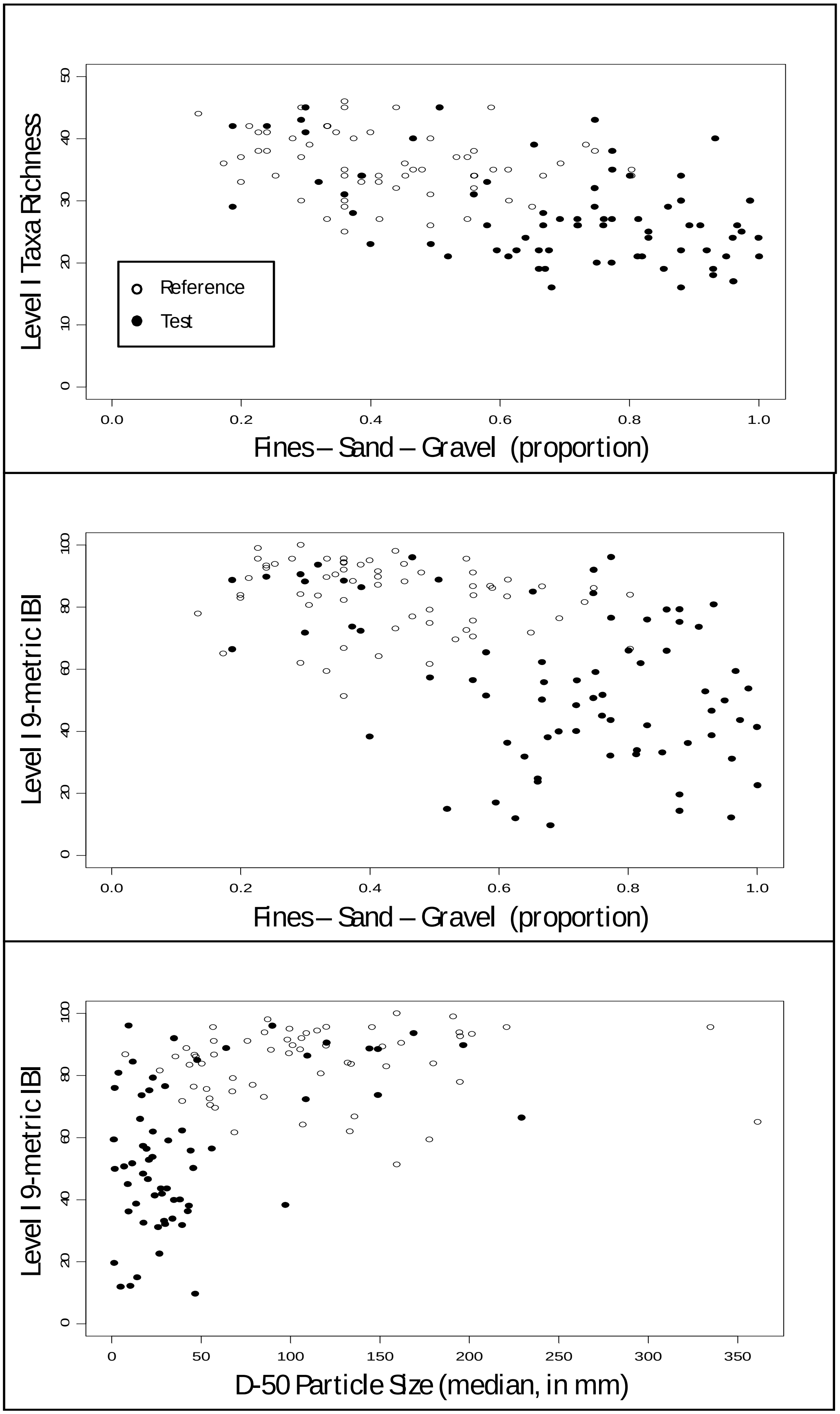


Figure 8. Level I sediment-exposure stressor responses (squares reference, circles test).

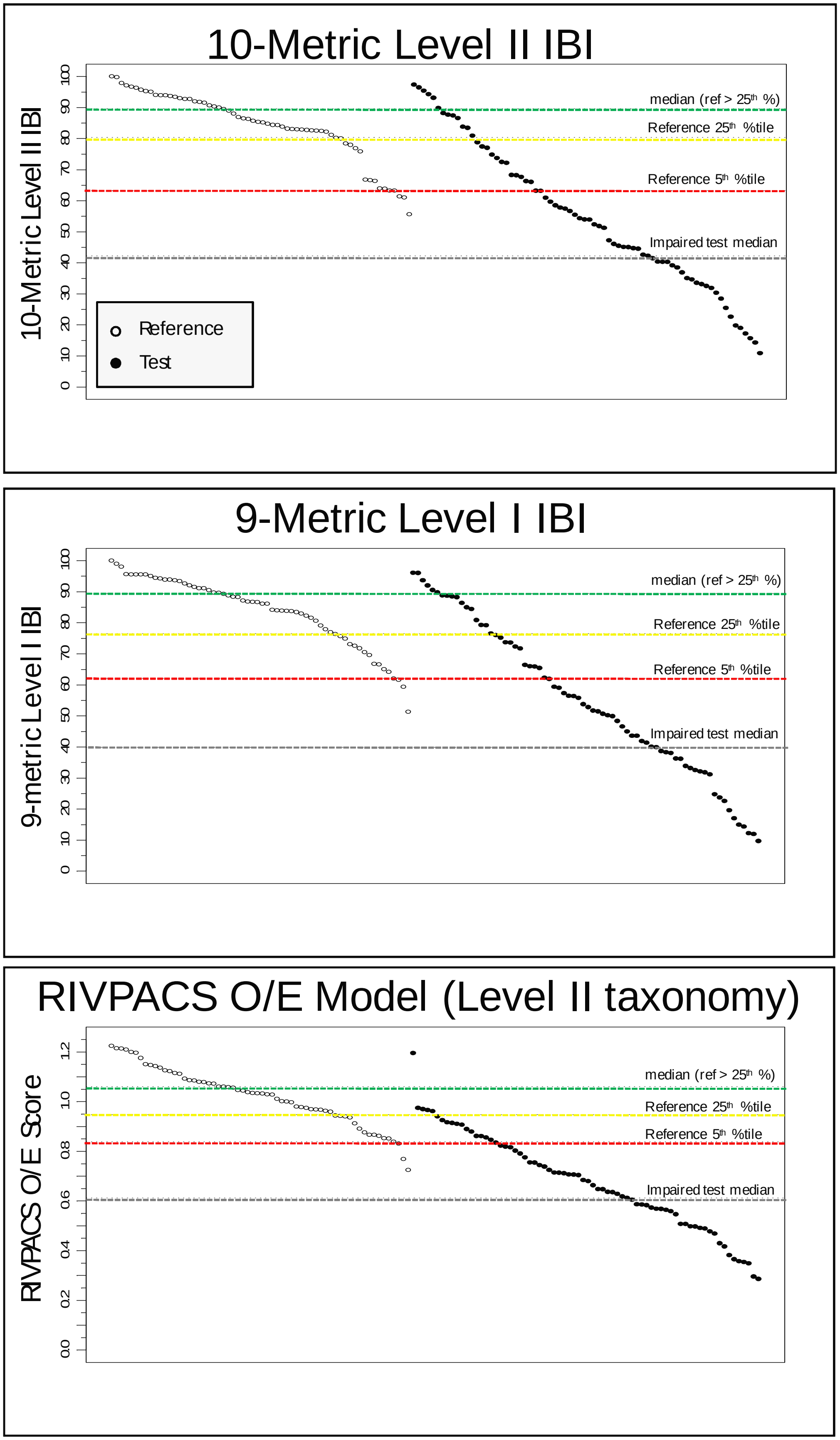


Figure 9. Ranked order distribution of reference followed by test sites.

Appendix I. List of Metrics Tested for IBI Development (**BOLD** indicates metrics included in the recommended 10-metric IBI; * indicates those included among the 30 metrics passing initial screening).

Number of individuals per square meter
Total number of individuals identified per sample
***Raw taxa richness per sample**
Taxa richness per sample using CSBP taxa resolution
Raw taxa richness rarefied to 217 bugs per sample
CSBP taxa richness rarefied to 217 bugs per sample
Raw taxa richness rarefied to 243 bugs per sample (but removed 2 SNARL samples)
CSBP taxa richness rarefied to 243 bugs per sample (but removed 2 SNARL samples)
*Shannon Diversity (H')
Shannon Diversity for samples standardized to CSBP taxonomic resolution
Community Evenness
Community Evenness standardized to CSBP taxa resolution
Simpson's Diversity Measure
Simpson's for CSBP taxa resolution
Hurlbert's "Pi" Diversity Measure
Hurlbert's "Pi" Diversity Measure for CSBP taxa resolution
*Total EPT Taxa per sample
***Total Ephemeroptera taxa per sample**
***Total Plecoptera taxa per sample**
***Total Trichoptera taxa per sample**
*Total Number of *Rhyacophila* species groups per sample
*Total Coleoptera taxa per sample
*Total Number of Elmidae taxa per sample
*Total Dipteran taxa per sample (includes Chironomids)
Total Dipteran taxa per sample but with CSBP taxa resolution for each sample
*Total Number of Chironomidae taxa
Total Number of Chironomidae subfamilies (i.e., CSBP taxa resolution)
*Total Number of Non-Insect Taxa
Total Number of Non-Insect Taxa but with mites just as "mites"
*Percent of Individuals which were Non-Insect Taxa
*Percent of Taxa which were Non-Insect Taxa
***Total Number of Acari (mite) Taxa**
*Percent of Individuals in a Sample which were EPT taxa
*Percent of Taxa which were EPT taxa
Percent of Individuals in a Sample which were EPT taxa but excluding Baetis and Hydropsychidae
Percent of Individuals which were Chironomidae
***Percent of taxa which were Chironomidae**
Ratio of Chironominae abundance to Orthocladiinae abundance
Ratio of EPT Richness to Chironomidae Percent Abundance (ept.rich / perc.chiro.abund)
*The Percent Abundance of the Most Abundant Taxon in a Sample (Dominance)
The Percent Abundance of the Most Abundant Taxon in a Sample but at CSBP taxa resolution
***The Percent Abundance for the 3 most abundant taxa (Dominance 3)**

Appendix I (continued). List of Metrics Tested and Selected for IBI Development

The Percent Abundance for the 3 most abundant taxa but at CSBP taxa resolution
D-50 Dominance = # taxa to get to 50% abundance of sample
D-50 Dominance = # taxa to get to 50% abundance of sample but using CSBP taxa resolution
__*Standard Biotic Index (modified Hilsenhoff)__
Biotic Index but using CSBP taxa resolution
*Number of Taxa which were Intolerant (TV=0,1,2)
*Percent of Taxa which were Intolerant
*Percent Abundance of Taxa which were Intolerant
Number of Taxa which were Intolerant (TV=0,1,2) but using CSBP taxa resolution
Percent of Taxa which were Intolerant but using CSBP taxa resolution
Percent Abundance of Taxa which were Intolerant using CSBP taxa resolution
*Number of Taxa which were Tolerant (TV=7,8,9,10)
__*Percent of Taxa which were Tolerant__
*Percent Abundance of Tolerant Taxa
Percent of Taxa which were Tolerant using CSBP taxa resolution
Percent Abundance of Tolerant Taxa using CSBP taxa resolution
Percentage of Collectors
Percentage of Scrapers
Percentage of Filterers
__*Percentage of Shredders__
*Percentage of Predators
*Number of Taxa which were Predators
Percentage of Piercers
Percentage of Collectors using CSBP taxa resolution
Percentage of Scrapers using CSBP taxa resolution
Percentage of Filterers using CSBP taxa resolution
Percentage of Shredders using CSBP taxa resolution
Percentage of Predators using CSBP taxa resolution
Percentage of Piercers using CSBP taxa resolution

Appendix II. Specific Formulas for Metric Calculations for the 12 Core Metrics:

Metric	Formula	Calculation
1. Richness	rich = $S = \sum_{all\ taxa} I_i$	The sum of non-zero unique taxa in each sample.
2. Ephemeroptera Richness	e rich = $\sum_{all\ E\ taxa} I_i$	The sum of non-zero unique taxa belonging to the order Ephemeroptera (mayflies).
3. Plecoptera Richness	p rich = $\sum_{all\ P\ taxa} I_i$	The sum of non-zero unique taxa belonging to the order Plecoptera (stoneflies).
4. Trichoptera Richness	t rich = $\sum_{all\ T\ taxa} I_i$	The sum of non-zero unique taxa belonging to the order Trichoptera (caddisflies).
5. Acari Richness	acari rich = $\sum_{all\ Acari\ taxa} I_i$	The sum of non-zero unique aquatic mite taxa (Acari).
6. Chironomidae % Richness	chiro perc rich $= \frac{\sum_{all\ Chiro\ taxa} I_i}{S} \cdot 100$	Percent of total Richness composed of Chironomidae taxa (midges)
7. Percent Tolerant Taxa	tol perc rich = $\frac{\sum_{taxa\ with\ TV \geq 7} I_i}{S} \cdot 100$	Percent of total Richness composed of tolerant taxa; Tolerant Taxa defined as the number of non-zero unique taxa whose pollution tolerance score equaled 7, 8, 9, or 10, on a scale of 0 to 10.
8. Predator Richness	pred rich = $\sum_{all\ Predator\ taxa} I_i$	The sum of non-zero unique taxa classified as Predators.
9. EPT % Abundance	ept abund = $\sum_{all\ EPT\ taxa} \frac{N_i}{N} \cdot 100$	Abundance of all EPT taxa divided by the total abundance in the sample (standardized to 500 for this analysis)
10. Shredder % Abundance	shredder = $\sum_{all\ shredder\ taxa} \frac{N_i}{N} \cdot 100$	Percent of total invertebrate abundance composed of shredder individuals.
11. Dominance of 3 Top Taxa	dominance 3 = $\sum_{i=(n),(n-1),(n-2)} \frac{N_i}{N}$	Proportional abundance of the 3 most abundant invertebrate taxa in each sample.
12. Biological Index	BI = $\sum_{all\ taxa} \frac{N_i \cdot TV_i}{N}$	Average of the abundance of each taxon weighted by that taxon's pollution tolerance score.

Where N_i=Abundance of i^{th} taxon; N=Total Abundance across taxa ($\sum N_i$); I_i is an indicator variable which takes a value of 1 when $N_i>0$ and which equals 0 when $N_i=0$; $N_{(n)}$= n^{th} ordered abundance value (i.e., most abundant invertebrate out of n taxa); $N_{(n-1)}$= $(n-1)^{th}$ ordered abundance value (i.e., second most abundant invertebrate out of n taxa); TV_i = Tolerance Value of each taxon, which is a value between 0 and 10 and reflects (for higher numbers) increasing ability to tolerate severe natural or anthropogenic environmental conditions.

Appendix III. 10-Metric IBI Scores Among Years for All Sites Including Multiple Years of Data.

		Year of Invertebrate Survey							
Stream & Site	Test (T) or Reference (R) Site	1998	1999	2000	2001	2002	2003	Mean	Standard Deviation (for sites w/ n ≥ 2)
Owens.abovetun	T			10.9				10.9	
Kirman.upper	T		15.7					15.7	
Bagley.meadow	T		22.6	14.3				18.4	5.9
Bagley.control	T		33.5	17.2		19.8	18.9	22.3	7.5
Poore.below	T		32.5	33.1	28.4	25.4		29.8	3.6
Owens.bridge	T			30.3				30.3	
Owens.417	T			31.9				31.9	
Owens.benton	T			35.0				35.0	
Rush	T			36.9				36.9	
EWalker	T			38.5				38.5	
Owens.power	T			40.2				40.2	
Robinson.below	T			42.5				42.5	
Poore.above	T		44.5			41.4		42.9	2.2
Truck.bart.low	T	40.3	53.8	39.1				44.4	8.2
Owens.belowtun	T			45.1				45.1	
Squaw.upper	T			46.0				46.0	
Truck.bart.up	T	53.9	47.2	44.6				48.6	4.8
Slinkard	T			42.2		55.4	52.3	50.0	6.9
Squaw.lower	T			45.4	56.6			51.0	7.9
Cowcamp	T		54.3	66.3	68.2	40.3		57.3	12.9
Trout	T				57.7			57.7	
Buckeye	T			58.5				58.5	
Squaw.mid	T			34.6	83.4			59.0	34.5
Bagley.restore	T					60.9	57.4	59.2	2.5
Truck.sun.low	T		73.7	45.0				59.3	20.3
Kirman.lower	T		51.8	51.2	63.1	72.2		59.6	10.0
Truck.sun.up	T		66.0	63.1				64.6	2.0
Kirman.upper.alt	T					68.2		68.2	
WWalker.pick.lov	T		59.6			88.1		73.9	20.2
Martis	T				77.0			77.0	
Lit.Walker.cent	T		67.6			86.6		77.1	13.4
Heavenly	T				78.8			78.8	
Perazzo	T				80.9			80.9	
WWalker.settle	T		72.4			89.8		81.1	12.3
WWalker.pick.mi	T		83.8	87.4	77.4	87.6		84.1	4.8
Squaw.sfk	T			74.8	95.4			85.1	14.6
Mill	T		80.0			93.1		86.5	9.3
Squaw.moraine	T			94.3				94.3	
Squaw.nfk	T			96.4	97.4			96.9	0.6
Deadman	R			61.0				61.0	
WCarson.faith	R			61.3				61.3	
WCarson.blm	R			63.8				63.8	
Forestdale	R			63.9				63.9	
Mammoth	R			66.5				66.5	
Lit.Walker.lower	R		55.5			78.4		67.0	16.1
Truck.celio.up	R		84.3	63.2				73.8	14.9
Robinson.camp	R		66.8			86.9		76.8	14.2
Cottonwood	R		76.8	85.7	63.3	83.8		77.4	10.2
Alder	R				77.9			77.9	
Truck.park	R		93.7	66.4				80.0	19.3
Lit.Truck.below	R				80.2			80.2	
Juniper	R				82.2			82.2	
Convict	R			82.4				82.4	
Silver.king	R			82.6				82.6	
Robinson.honey	R			84.4				84.4	
Owens.spring	R			84.7				84.7	
Lit.Truck.perazzo	R			85.1				85.1	
Trib.silver	R			82.9		83.0	90.0	85.3	4.1
Cold	R			85.4				85.4	
Truck.celio.low	R		82.5	89.0				85.7	4.5
Lee	R			86.3				86.3	
Independence	R				86.5			86.5	
WWalker.leavitt	R		75.8	83.1	95.2	92.7		86.7	8.9
Silver.cr	R		82.7			91.8		87.2	6.4
Swauger.above	R		91.9			83.0		87.5	6.3
General	R			88.0				88.0	
ECarson.bagley	R			89.4				89.4	
Willow	R			90.3				90.3	
Swauger.valdez	R		81.1			99.7		90.4	13.1
WWalker.confl	R		91.5			90.6		91.1	0.6
Lacey	R				93.0			93.0	
Prosser.north	R				93.4			93.4	
Wolf	R			93.9				93.9	
Hidden	R				93.9			93.9	
Pole	R			95.1				95.1	
Truck.forest	R		94.0	96.3				95.2	1.6
Sagehen	R			97.8	92.7			95.3	3.6
Prosser.below	R			95.7				95.7	
Bear	R			97.1	96.7			96.9	0.3
Spratt	R			100.0				100.0	

Appendix IV. List of Taxa Collected and Final Taxonomic Resolution Used for Lahontan-Eastern Sierra IBI Database along with Tolerance Values used for each taxon.

Odonata	**Tolerance Value**
Ophiogomphus	4
Cordulegaster dorsalis	3
Argia	7
Coenagrion / Enallagma	9
Ephemeroptera	
Acentrella	4
Baetis	5
Callibaetis	9
Centroptilum	2
Diphetor hageni	5
Fallceon quilleri	5
Procloeon	4
Ameletus	0
Siphlonurus	7
Paraleptophlebia	4
Paraleptophlebia bicornuta	4
Tricorythodes	4
Serratella	2
Attenella delantala	2
Attenella soquele	2
Caudatella heterocaudata	1
Caudatella hystrix	1
Drunella doddsi	0
Drunella flavilinea	0
Drunella grandis	0
Drunella pelosa	0
Drunella spinifera	0
Ephemerella aurivillii	1
Ephemerella infrequens	1
Timpanoga hecuba	7
Cinygma	2
Cinygmula	4
Epeorus	0
Ironodes	3
Leucrocuta / Nixe	3
Rhithrogena	0
Plecoptera	
Malenka	2
Prostoia	2
Visoka cataractae	0
Zapada	2
Capniidae	1
Eucapnopsis brevicauda	1
Despaxia augusta	0
Moselia infuscata	0
Paraleuctra	0
Bisancora	1
Paraperla	0
Plumiperla / Haploperla	1
Suwallia	1
Sweltsa	1
Soliperla	1
Yoraperla	1
Cultus	2
Frisonia picticeps	2

Isoperla	2
Kogotus / Rickera	2
Megarcys	2
Oroperla barbara	2
Perlinodes aureus	2
Skwala	2
Pteronarcys	0
Pteronarcella	0
Calineuria californica	2
Claasenia sabulosa	3
Doroneuria baumanni	1
Hesperoperla pacifica	2
Trichoptera	
Rhyacophila acropedes grp	1
Rhyacophila alberta grp	0
Rhyacophila angelita grp	0
Rhyacophila arnaudi grp	0
Rhyacophila betteni grp	1
Rhyacophila coloradensis grp	2
Rhyacophila hyalinata grp	1
Rhyacophila iranda grp	0
Rhyacophila narvae grp	0
Rhyacophila nevadensis grp	0
Rhyacophila oreta grp	0
Rhyacophila rayneri grp	0
Rhyacophila rotunda grp	0
Rhyacophila sibirica grp	0
Rhyacophila vagrita grp	0
Rhyacophila verrula grp	0
Rhyacophila vofixa grp	0
Agraylea	8
Hydroptila	6
Neotrichia	4
Ochrotrichia	4
Oxyethira	3
Amiocentrus aspilus	3
Brachycentrus americanus	1
Brachycentrus echo	3
Brachycentrus occidentalis	1
Micrasema	1
Mystacides	4
Oecetis	8
Helicopsyche	3
Agapetus	0
Anagapetus	0
Culoptila	2
Glossosoma	1
Protoptila	1
Arctopsyche californica	2
Arctopsyche grandis	1
Parapsyche almota	2
Parapsyche elsis	1
Ceratopsyche	4
Cheumatopsyche	5
Hydropsyche	4
Lepidostoma cascadense	1
Lepidostoma	1
Amphicosmoecus	1
Chyranda centralis	1
Cryptochia	0

Desmona	1
Dicosmoecus	1
Ecclisomyia	2
Hesperophylax	3
Homophylax	0
Limnephilus	3
Onocosmoecus	1
Psychoglypha	2
Yphria californica	1
Dolophilodes	2
Wormaldia	3
Polycentropus	6
Apatania	1
Pedomoecus sierra	0
Neophylax	3
Neothremma	0
Oligophlebodes	1
Gumaga	3
Heteroplectron californicum	1
Lepidoptera	
Petrophila	5
Megaloptera	
Sialis	4
Dysmicohermes	0
Orohermes crepusculus	0
Hemiptera	
Callocorixa audeni	10
Cenocorixa	10
Corisella decolor	8
Sigara washingtonensis	8
Notonecta kirbyi	8
Notonecta undulata	9
Coleoptera	
Amphizoa insoleus	1
Helichus	5
Postelichus	5
Agabinus	8
Agabus	8
Hydroporus	5
Liodessus / Neoclypeodytes	5
Oreodytes	5
Stictotarsus	5
Cleptelmis addenda	4
Heterlimnius corpulentus	4
Lara avara	4
Narpus concolor	4
Optioservus divergens	4
Optioservus quadrimaculatus	4
Zaitzevia parvula	4
Brychius hornii	5
Haliplus	5
Hydraena	5
Ochthebius	5
Helophorus	5
Ametor scabrosus	5
Crenitis	5
Cymbiodyta pacifica	5
Hydrobius	5
Laccobius	5
Tropisternus columbianus	5

	Tropisternus ellipticus	5
	Tropisternus lateralis	5
	Eubrianax edwardsi	4
	Cyphon	5
	Elodes	5
	Staphylinidae	5
Diptera		
	Antocha monticola	3
	Cryptolabis	3
	Dicranota	3
	Erioptera	3
	Hesperoconopa	1
	Hexatoma	2
	Holorusia hespera	6
	Limnophila	3
	Limonia	6
	Molophilus	6
	Ormosia	3
	Pedicia	3
	Rhabdomastix	3
	Tipula	4
	Atherix pachypus	2
	Agathon comstocki	2
	Atrichopogon	6
	Bezzia / Palpomyia	6
	Ceratopogon	6
	Culicoides	6
	Forcipomyia	6
	Monohelea	6
	Culex	8
	Deuterophlebia	0
	Dixa	2
	Dixella	2
	Meringodixa chalonensis	2
	Hydrophorus	6
	Rhaphium	4
	Tachytrechus	6
	Chelifera	6
	Clinocera / Hydrodromia	6
	Hemerodromia	6
	Oreogeton	6
	Weidemannia	6
	Scatella	6
	Limnophora	6
	Muscidae	6
	Glutops	3
	Maruina lanceolata	2
	Pericoma	4
	Ptychoptera	7
	Prosimulium	3
	Simulium	6
	Caloparyphus	7
	Euparyphus	8
	Nemotelus	8
	Odontomyia	8
	Stratiomys	8
	Chrysops	8
	Tabanus	8
	Protanyderus	1
	Thaumalea	3

Diptera - Chironomidae

Monodiamesa	7
Odontomesa	6
Prodiamesa	3
Boreoheptagyia	6
Diamesa	6
Pagastia	1
Potthastia gaedii	2
Potthastia longimana	2
Pseudodiamesa	6
Boreochlus	6
Parochlus kiefferi	6
Apsectrotanypus	7
Brundiniella	7
Meropelopia	6
Pentaneura	6
Trissopelopia	6
Ablabesmyia	8
Larsia	6
Nilotanypus	6
Paramerina	6
Thienemannimyia	6
Zavrelimyia	8
Apedilum	6
Chironomus	10
Cryptochironomus	8
Demicryptochironomus	6
Microtendipes pedellus	6
Microtendipes rydalensis	6
Paracladopelma	7
Paratendipes	8
Phaenopsectra	7
Polypedilum aviceps	6
Polypedilum convictum	6
Polypedilum halterale	6
Polypedilum laetum	6
Polypedilum scalaenum	6
Polypedilum tritum	6
Robackia	6
Stenochironomus	5
Stictochironomus	9
Tribelos	6
Pseudochironomus	5
Cladotanytarsus vanderwulpi	7
Constempellina	4
Micropsectra	7
Paratanytarsus	6
Rheotanytarsus	6
Stempellina	2
Stempellinella	4
Sublettea	4
Tanytarsus	6
Virgatanytarsus	6
Brillia	5
Camptocladius	6
Cardiocladius	5
Chaetocladius dentiforceps	6
Corynoneura	7
Cricotopus / Nostococladius	7
Cricotopus trifascia	7

Cricotopus / Orthocladius	7
Eukiefferiella brehmi	8
Eukiefferiella brevicalar	8
Eukiefferiella claripennis	8
Eukiefferiella coerulescens	6
Eukiefferiella devonica	8
Eukiefferiella gracei	8
Eukiefferiella pseudomontana	8
Eukiefferiella similis	8
Euryhapsis	6
Heleniella	6
Heterotrissocladius marcidus	0
Hydrobaenus	8
Krenosmittia	1
Limnophyes	8
Lopescladius	6
Metriocnemus fuscipes	6
Metriocnemus hygropetricus	6
Nanocladius balticus	3
Nanocladius parvulus	3
Orthocladius / Euorthocladius	6
Parachaetocladius	2
Parakiefferiella	4
Paralimnophyes	5
Parametriocnemus	5
Paraphaenocladius	4
Parorthocladius	6
Psectrocladius psilopterus	8
Psectrocladius sordidellus	8
Pseudorthocladius	5
Pseudosmittia	6
Rheocricotopus	6
Rheosmittia	6
Smittia	6
Symbiocladius	6
Symposiocladius	6
Synorthocladius	2
Thienemanniella fusca	6
Thienemanniella xena	6
Tvetenia bavarica	5
Tvetenia discoloripes	5
Non-Insects (misc)	
Helobdella stagnalis	6
Oligochaeta	5
Tricladida	4
Gordius	4
Paragordius	4
Hydra	5
Ostracoda	8
Pisidium	8
Pacifastacus lenisculus	6
Gammarus lacustris	6
Hyalella azteca	8
Gastropoda	
Hydrobiidae	8
Potamopyrgus	8
Ferrissia	6
Fossaria	8
Radix auricularia	6
Stagnicola	8

	Physa	8
	Gyraulus	8
Acari		
	Arrenurus	5
	Atractides	8
	Aturus	5
	Cheiroseius	5
	Estelloxus	8
	Feltria	5
	Frontipoda	5
	Frontipodopsis	5
	Hydrovolzia	5
	Hydrozetes	5
	Hygrobates	8
	Lebertia	8
	Limnesia	5
	Ljania	5
	Mideopsis	5
	Neumania	5
	Nudomideopsis	5
	Protzia	8
	Sperchon	8
	Sperchonopsis	5
	Stygomomonia	5
	Testudacarus	5
	Thyas	5
	Torrenticola	5
	Tyrrellia	5
	Utaxatax	5
	Wandesia	5

APPENDIX V. Addendum to Lahontan (Eastern Sierra) IBI Development Report

Simplified taxonomic resolution for bioassessment using a revised Level I standard taxonomic resolution

David B. Herbst and Erik L. Silldorff

September 2009

Background

The 10-metric IBI developed for the Lahontan Region was based on taxonomic resolution that identified specimens to what is known as Level II Standard Taxonomic Effort (STE) as defined by SAFIT (Southwestern Association of Freshwater Invertebrate Taxonomists - the successor to CAMLnet). At this level, most taxa are identified at least to genus and many to species where reliable keys are available. The final data set analyzed was based on this STE Level II (refer to the Mater Taxa List for Lahontan Region, and to documentation at http://www.safit.org/ste.html), but as some taxa were not always of certain status ("non-distinct"), some Level II designations were held at Level I so that (a) uncertain identifications were not included in the data set, and (b) no "phantom" taxa were created – those resulting from distinct specimens being recorded separately from non-distinct and being counted as different taxa. In the Level II data set, this was primarily reflected most commonly for species level designations for *Baetis*, *Serratella*, *Zapada*, and many gastropod snails being left at genus (that is, where species identifications were made, they were "backed-off" to genus when entering these data for analysis).

An important consideration in the practical use and application of any IBI is the certainty that identifications are consistent when produced by different laboratories, and that data compilation and computation of metrics and scores by different analysts has no ambiguity. One way to minimize potential discrepancies is to simplify the taxonomy to lower levels of resolution. To achieve this simplification, and compare differences in assessment of biological integrity and impairment, we converted the Level II data set to SAFIT Level I resolution. The main differences in resolution of identifications used are summarized in Table A.1, and show that collapse of the family *Chironomidae* (midges) from 100 genus/species designation to the single family, and *Rhyacophila* from 17 species groups to one genus, account for most of the decrease in resolution in going from Level II to Level I.

Using this revised Level I taxonomic resolution data set, we repeated the process of developing an Index of Biological Integrity (IBI) described in the main report in order to identify a multi-metric IBI that could be used with Level I taxonomic data. Two primary considerations guided the creation of a Level I multi-metric IBI for assessing ecological condition of benthic invertebrate communities. First, comparative performance to the original Level II IBI was sought in order to provide an IBI that yielded similar levels of noise, signal-to-noise ratios, and discrimination efficiency (i.e., the standard performance-based measures advocated by USEPA guidance documents). The second and unique criterion by which the candidate Level I IBIs were evaluated was their ability to match the assessment categories and scores for the Level II IBI. Because the development of this second, coarser taxonomic level IBI will result in two IBIs being utilized within the Lahontan Region, this second criterion was seen as a critical feature to minimize ambiguities in assessments.

Table A.1 Changes in Taxonomic Resolution

Ranked changes of decreased taxa resolution in converting Level II to Level I Data:		
Group collapsed	**Level II**	**Level I**
Chironomidae	100 genus-spp/grp	1 family
Rhyacophila	17 spp groups	1 genus
Drunella	5 spp	1 genus
Brachycentrus	3 spp	1 genus
Tropisternus	3 spp	1 genus
Dolichopodidae	3 genera	1 family
Paraleptophlebia	2 spp	1 genus
Attenella	2 spp	1 genus
Caudatella	2 spp	1 genus
Ephemerella	2 spp	1 genus
Arctopsyche	2 spp	1 genus
Parapsyche	2 spp	1 genus
Hydropsyche	2 genera	1 genus
Lepidostoma	2 spp	1 genus
Optioservus	2 spp	1 genus
Muscidae	2 spp	1 genus
plus some excluded = Nematomorphs, Staphylinidae, Hydrobiidae, Notonectidae		
Sum = 232 at Level I, from 373 total taxa at Level II		

As an additional step in making assessments more consistent, thresholds for defining condition classes were aligned with other IBIs and assessment tools developed in California. Specifically, the limit between the intermediate class of “partially supporting” to “fully supporting” was changed to the 25th percentile of the reference distribution (from the 16th used in the Level II IBI). The impairment threshold was again set at the 5th percentile of reference (approximately 2 standard deviations below the mean), and the upper and lower condition classes (supporting and impaired) were split at the 50th percentile of the reference (median), and the median value of impaired test sites, respectively (Table A.2).

Table A.2. Threshold values for condition classes and performance measures contrasting the Level I and Level II IBIs and the Level II RIVPACS O/E scores

L-II IBI-10	L-I IBI-9	O/E	
63.2	62.1	0.839	5th %tile reference = Impairment threshold
80.4	76.5	0.947	25th %tile reference = Full / Partial support limit
84.1	83.5	1.020	reference mean
89.7	89.4	1.060	median of (reference sites > 25th percentile)
42.2	39.9	0.612	median of impaired tests (test < 5th %tile ref)
11.0	11.3	0.116	SD of reference mean
55.0	55.0	0.677	Test Mean
2.65	2.54	2.95	Signal:Noise = (mean R-T)/Ref SD

Tier / Grade	Designations	Ranges
5 / A	supporting	>median of (reference sites > 25th percentile)
4 / B	supporting	25th to (median of (reference sites > 25th percentile))
3 / C	partial supporting	5th-25th percentile reference
2 / D	not supporting	<5th percentile reference (impairment level) & >median of test values in impaired range
1 / F	not supporting	<median of test values in impaired range

Metric Selection

Metric selection for the Level I IBI started with the same 30 core metrics that provided some degree of performance initially with the Level II IBI and which were rigorously evaluated during the initial IBI creation. Some of these metrics were no longer relevant at the coarser Level I taxonomic resolution (e.g., *Chironomidae* richness or *Rhyacophila* species group richness), and a revised list of 32 metrics was considered. This list of 32 metrics included 4 versions of a weighted “intolerant richness” metric referred to variously as Beck’s Index or the Florida Index (Beck 1954, Barbour et al. 1999). A number of states have considered or incorporated some variation of this weighted intolerant richness metric for their ecological assessments of benthic invertebrate data. The full suite of metrics evaluated for the Level I IBI is presented in Table A.3.

Performance evaluations for the Level I metrics followed the procedures used for the Level II metrics (see Section 2.5 of the main report), with emphasis placed on variation (noise) at reference sites, the signal-to-noise ratio between reference and test sites, and the discrimination efficiency at various threshold of the reference distribution. Like with the Level II IBI, some of these 32 metrics showed little or no ability to discriminate between reference sites and sites with human-caused disturbance, and nine metrics with poor individual performance were eliminated from further consideration in building a Level I multi-metric IBI.

The remaining 23 core metrics (highlighted in Table A.3) were re-scaled to a continuous range between 0 and 10 in the same manner as the Type II metrics, with the median reference site providing the threshold for a 10-score, the 10th percentile of the test distribution used as the threshold for a 0-score, and linear interpolation between these two thresholds (note: 90th percentile of test distribution was used for reverse-scale metrics instead of 10th percentile). The rationale for these thresholds is discussed in detail in Section 2.5 of the main report.

Various combinations of these 23 core metrics were then added together to create candidate multi-metric IBI scores. Like with the original Level II candidate IBIs, the combination of metrics selected for Level I candidate IBIs involved both quantitative information about the performance of each metric, the redundancy and correlations among metrics, as well as subjective evaluations of the performance and utility of the particular ecological measurements. In total, 27 Level I candidate IBIs were created sequentially in an attempt to optimize performance as well as classification consistency with the Level II IBI and to evaluate different emphases in a multi-metric scoring system. These 27 candidate IBIs contained between 6 metrics and 11 metrics, with emphasis directed at IBIs containing fewer metrics. Following each individual IBI creation, the performance and classification were evaluated and the information gleaned from its performance was used in selecting subsequent metric combinations for candidate IBIs.

Comparison of Three Highest Ranked IBIs

All 27 IBIs considered at the Level I taxonomic resolution had high performance characteristics that were broadly suitable for assessment decisions (Table A.4). Specifically, the estimate of noise (the coefficient of variation for the reference sites) had a narrow range between 10.6 and 14.7. Similarly, the discrimination efficiency as measured by the proportion of test sites overlapping with the upper 75% of reference sites (i.e., greater than the 25th percentile) ranged only from 19% to 28% across the 27 IBIs considered. Other performance measures likewise varied in a narrow range. Thus, although there were some differences among the IBIs based on traditional performance measures, all IBIs performed suitably well for further consideration.

Rankings of the 27 candidate IBIs therefore were developed to a large degree based on one qualitative measure of functional traits, one quantitative measure of redundancy, and two quantitative measures of classification match to the Level II IBI. First, the distribution of metrics across the four main metric categories (richness, composition, tolerance, functional) was considered as a qualitative criterion, with a more even distribution among categories being desirable. Second, the extent of redundancy and overlap among component metrics was evaluated using simple pairwise correlation coefficients, with a goal to allow some redundancy but to eliminate IBIs where component metrics provided largely overlapping information (i.e., $r > 0.8$). Third, the agreement for IBI scores between the candidate Level I IBI and the established Level II IBI was measured as the R^2 of the regression relationship. Finally, the agreement for classification decisions between the candidate Level I IBI and the established

Level II IBI was measured for reference sites and test sites separately (see Table A.4; note that the thresholds for the 10-metric Level II IBI were changed to 25th percentile reference data set to consistently define limits between the intermediate level of "partial" to "full" support).

Based on these evaluations, three candidate IBIs emerged as having strong performance and superior classification matches with the original Level II IBI while also minimizing redundancy and capturing a breadth of the functional categories for metrics. These three candidate IBIs are presented in detail to both demonstrate their performance and to illustrate the trade-offs with selecting among them as the Level I IBI for ecological assessment decisions in the Lahontan Region.

First, the component metrics to these three IBIs are presented in Table A.5. The 9 metrics used in these three IBIs are the same as those used in the 10-metric Level II IBI, with the exception that percent richness composed of *Chironomidae* taxa is no longer considered since this family of dipterans is not identified in detail in the Level I taxonomy. Although metrics beyond those in Table A.5 were included in various forms of the candidate IBIs, no such IBIs demonstrated exceedingly high levels of performance or strong enough classification matches to the Level II IBI to merit further consideration.

The performance scores for the three highlighted IBIs, their maximum correlation between metrics (a measure of metric redundancy), and their numerical and classification performance compared to the Level II IBI are presented in Table A.6 along with the summary measures of performance for all 27 candidate IBIs. For redundancy, only the 9-metric IBI has correlations between metrics exceeding 0.707 (R^2=0.5), and the 0.78 correlation (R^2=0.61) is similar to the maximum correlation included in the Level II IBI (r=0.74, R^2=0.55). For performance measures, all three IBIs show they perform suitably well, frequently scoring in the top 25% across the 27 IBIs considered. Thus, there is little distinction between these three IBIs based on these performance measures. The strong performance for each of these three IBIs is reinforced by the boxplots of reference vs. test sites (Figure A.1). In these graphs, clear separation is demonstrated between the reference and test sites, with tight clustering of reference sites near the high end of the IBI scale. Figure A.2 shows these properties in the form of cumulative distribution functions (CDFs), whose portrayal of the data also reveals the shape of the data distributions. Like with the boxplots, these graphs suggest the three IBIs are comparable in performance.

Among the classification measures, the highest numerical correspondence with the Level II IBI comes through the use of the 9-metric Level I IBI (Table A.6). This result is not surprising given that the 9 metrics used in this Level I are also used in the 10-metric Level II IBI, with one additional metric (percent richness composed of *Chironomidae*) being the only metric included in the Level II that is not included in the 9-metric Level I. The strength of this relationship is reinforced in the classification plot between these two IBIs (Figure A.3). Although the correspondence between the Level I and Level II IBIs is also strong for the 7-metric and 6-metric versions (Table A.6, Figures A.4 and A.5), there is increased variation in the relationship, particularly for the 6-metric IBI.

The non-correspondence rate between the three Level I IBIs and the Level II IBI suggests similar performance in consistency of classification among the three Level I IBIs (Table A.6). These rates are among the lowest seen for the 27 IBIs considered in this study. In addition, a property not demonstrated in Table A.6 is that other candidate IBIs often yielded excellent classification matches for reference sites but not test sites, or vice versa. Thus, the relatively low non-correspondence rates for both reference and test sites in these three candidate Level I IBIs is unusual among the IBIs evaluated in this study. Yet it is also informative to compare these disagreements in assessment seen among the Level II IBIs considered in the main report. In that original component of the study (see Table 8 in main report), the non-correspondence rates range from 0% to 5.6% when comparing the candidate 8-metric, 10-metric, and 12-metric IBIs, indicating that sites would be assigned to a different primary assessment category less than 1-in-17 times regardless of which Level II IBI was used.

The poorer classification rate for the Level I IBIs seen in Table A.6 is expected, to some degree, since the structure of the underlying data has been substantially altered in using a coarser level of taxonomic resolution. Yet non-correspondence rates between the candidate Level I IBIs and the Level II IBI frequently exceeded a 1-in-5 level (20%). Such high rates of disagreement in the primary assessment category between methods could lead to substantial uncertainty in the status of many sites, and could even lead to the selection of a taxonomic resolution (and thus an IBI) that would benefit the organization conducting a study. Thus, high non-correspondence rates were seen as unacceptable for candidate Level I IBIs since they may be used concurrently with the established Level II IBI. Among the three primary Level I IBIs presented here for consideration, the non-correspondence rates are typically in the range of 10% (1-in-10), among the lowest rates among any IBIs considered, with similar rates among these three IBIs for the current data set (Table A.6).

Predicting the performance of the three candidate IBIs for future sites and data sets is an additional important consideration in the final selection of the Level I IBI. Because the selection of the current three IBIs has been partly based on their ability to classify the current set of “test” sites in a manner similar to the Level II IBI, these analyses do not present an objective evaluation of how each IBI would perform on new sites. With a large enough data set, a subset of sites could have been removed from the original analysis in order to use them for later “validation” after the IBIs were developed and tested. In the present study, no separation into “calibration” and “validation” data sets was possible given the modest sample sizes available. As mentioned in the main report, evaluation of such a “validation” data set is one of the recommendations for future work.

Even though a validation data set was not available in the present study, some measure of future performance can be ascertained from the strength of the relationship in raw IBI scores between Level I and Level II IBIs. With an R^2 of 0.96, the 9-metric Level I IBI match to the Level II IBI is extremely strong, and the consistently low variation around this relationship (seen in Figure A.3) reflects the close correspondence in scoring across various levels of impairment. The 7-metric IBI comparison to the Level II IBI is nearly as strong with an R^2 of 0.94, but the pattern in Figure A.4 reflects some weakening in this relationship. Finally, considerable scatter is observed in the relationship between the 6-metric IBI when compared to the Level II IBI (Figure A.5) even with a strong R^2 of 0.90. These patterns suggest that the 9-metric Level I IBI will likely lead to more consistent scorings compared to the Level II IBI for future sites, and these more consistent scores would be expected to lead to fewer non-correspondences. The increased variance seen in the 7-metric and the 6-metric Level I IBIs therefore poses a somewhat greater risk for ambiguity in site assessments in the future.

Recommended Level I IBI

Some advantage can be derived from simplifying the Level I IBI not only through a coarser taxonomic resolution but also through a reduction in the number of metrics needed to make an ecological assessment. As demonstrated above, such simplification leads to few changes in IBI performance as measured through traditional statistics. Yet the simultaneous use of different IBIs in a geographic area for making ecological assessments demands that such assessments be as consistent as possible in order to avoid ambiguity in site classifications. Ideally, the non-correspondence rate would be so low that disagreements would be rare and little consideration would be given to which IBI was utilized in making an assessment.

The current analyses demonstrate both quantitatively and qualitatively that the most consistent scoring and classification is obtained by using a larger number of metrics, with the 9-metric IBI most closely matching the 10-metric Level II IBI (R^2 of 0.96). Given that the only two differences between these Level I and Level II IBIs are the taxonomic resolution and the inclusion or exclusion of one metric (percent richness composed of *Chironomidae*), such an outcome is not unexpected. In addition, the 9-metric IBI performs equally well to the 7-metric and 6-metric IBIs when compared using other parameters.

Because of the similar general performance and the superior matching of the 9-metric IBI to the Level II IBI, the current research therefore indicates that the 9-metric Level I IBI is the most suitable IBI for both accurate assessments of ecological condition as well as consistent assessments when compared to the alternative Level II IBI. Although some trade-offs exist with such a selection, on balance the expected consistency in future scorings and classifications warrants the selection of this 9-metric IBI, even with the greater number of metrics compared to the alternatives.

Therefore, the 9-metric Level I IBI is recommended for adoption as an alternative assessment endpoint in the Lahontan Region when coarser taxonomic resolution is desired or required by the data set (see Table A.7 for calculation details).

Table A.7. Recommended 9-metric Level I IBI and scoring thresholds for each metric.

Metric	**10-score threshold**	**0-score threshold**
Total Richness	35	19.1
Ephemeroptera Richness	8	3
Plecoptera Richness	6	1
Trichoptera Richness	6	2
Acari Richness	6	1
Dominance-3 (top 3 taxa %)	56 %	81 %
Biotic Index	3.90	5.49
Tolerant Rich. %	11 %	27 %
Shredder %	2.6 %	0 %

Response to Stressors and Site-by-Site Comparison for Recommended Level I IBI

The primary stress gradient that occurred in test sites of the eastern Sierra was related to the impacts of erosion and sediments produced by livestock grazing. Using either the percent of smallest particle sizes (fines, sand, gravel), or the D-50 median particle size as measures of sedimentation exposure, the Level I richness and IBI showed clear responses to this stressor (Figure A.6abc). Effect thresholds of biological response were apparent at %FSG > 50-60%, and at a D-50 value < 50 mm, and are consistent with those observed for the Level II IBI.

To compare site assessments across the methods, the thresholds for the 10-metric Level II IBI and RIVPACS O/E were changed to the 25th percentile of the reference data set to define limits between the intermediate level of "partial" to "full" support. Assessments for references sites (Table A.8a , Figure A.7) were in broad agreement among methods, particularly when comparing the Level I and Level II IBIs. The RIVPACS O/E model had a more mixed correspondence to both IBIs reflecting the distinct nature of this predictive-model assessment relative to the two multi-metric assessments. Generally, mixed classifications between the two IBIs occurred for borderline sites where a site fell marginally above or below a threshold for the second IBI. Stronger differences between the IBI scores reflected changes in dominance scoring as a result of the different taxonomic levels, particularly the combining of *Chironomidae* midges into a single taxon.

For defining impaired condition of test sites, biological condition assessments showed close matching among the differing approaches (Table A.8b). The 47 test sites judged as being impaired (red color) by the Level II IBI agreed in all but three cases with the Level I IBI, and these again reflected borderline sites. On the other hand, there were no cases where the Level I IBI showed impairment at a site but Level II did not. These relatively rare discrepancies of judged impairment show that the results obtained between the Level I and II IBIs are usually in accordance, with most differences resulting from borderline assessment decisions. Test sites judged unimpaired were also typically in agreement between the Level I and Level II IBIs, but O/E scores resulted in more variable assessments across the range of Level I and Level II scores, without a clear pattern for the differences in classifications. (See Section 3.5 of the main report for a discussion of the differences between the RIVPACS-type O/E scores and the multi-metric assessments.)

References Cited

Beck, W.M., Jr. 1954. A simplified ecological classification of organisms. Quart. J. Florida Acad. Sci. 17(4): 211-227.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers (2nd Ed). EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Table A.3 List of Metrics Tested for IBI Development Based on Level I Taxonomic Resolution (* denotes core metrics considered for candidate IBIs**; BOLD** indicates metrics included in the recommended 9-metric IBI).

***Raw taxa richness per sample**
*Taxa richness per sample using rarefaction to standardize at 250 individuals per sample
*Shannon Diversity (H')
Community Evenness
*Total EPT Taxa per sample
***Total Ephemeroptera taxa per sample**
***Total Plecoptera taxa per sample**
***Total Trichoptera taxa per sample**
Total Coleoptera taxa per sample
Total Number of Elmidae taxa per sample
*Total Dipteran taxa per sample (includes Chironomids)
Total Number of Non-Insect Taxa
Percent of Individuals which were Non-Insect Taxa
Percent of Taxa which were Non-Insect Taxa
***Total Number of Acari (mite) Taxa**
*Percent of Individuals in a Sample which were EPT taxa
Percent of Taxa which were EPT taxa
Percent Abundance of the Most Abundant Taxon in a Sample (Dominance)
***Percent Abundance for the 3 most abundant taxa (Dominance 3)**
***Standard Biotic Index (modified Hilsenhoff)**
*Number of Taxa which were Intolerant (TV=0,1,2)
*Percent of Taxa which were Intolerant
*Percent Abundance of Taxa which were Intolerant
Number of Taxa which were Tolerant (TV=7,8,9,10)
***Percent of Taxa which were Tolerant**
***Percentage of Shredders**
*Percentage of Predators
*Number of Taxa which were Predators
Weighted Intolerant Richness v1 (aka, Beck's, Florida Index): 2(# Taxa w/ TV = 0, 1, or 2) + 1*(# Taxa w/ TV = 3 or 4)
Weighted Intolerant Richness v2: 2(TV=0, 1) + 1*(TV=2,3,4)
Weighted Intolerant Richness v3: 3(TV=0) + 2*(TV=1) + 1*(TV=2)
Weighted Intolerant Richness v4: 5(TV=0)+4*(TV=1)+3*(TV=3)+2*TV=4)

Table A.4. Summary Statistics (minimum, 25th percentile, median, 75th percentile, and maximum) of Performance, Redundancy, and Comparability for 27 candidate IBIs considered for the Lahontan Level I IBI. **IBI Performance:** noise measured as coefficient of variation for reference streams; signal measured as ratio of reference mean to test mean; signal:noise ratio measured as difference between reference and test means divided by reference standard deviation; overlap at 25th percentile indicates the empirical proportion of test streams exceeding the 25th percentile of the reference stream distribution. **Redundancy:** Maximum of correlation coefficient among component metrics included in the IBI. **Comparison to Level II IBI:** R^2 given for pairwise regression between Level I and Level II IBIs; classification match indicates the percent of sites not assigned to same primary impairment category ("supporting," "partially supporting," or "not supporting")

	Min	**25th %**	**Median**	**75th %**	**Max**
IBI Performance					
Noise	10.6	11.8	12.6	13.6	14.7
Signal	1.49	1.53	1.55	1.60	1.62
Signal:Noise Ratio	1.99	2.18	2.44	2.56	2.83
Overlap at 25th Percentile	0.19	0.23	0.25	0.25	0.28
Redundancy - Max. Corr (r)	0.64	0.67	0.78	0.83	0.90
Comparison to Level II IBI					
R^2 - Level I vs Level II	0.85	0.89	0.90	0.92	0.96
Classification Match (ref)	10%	13%	17%	23%	29%
Classification Match (test)	6%	13%	14%	17%	28%

Table A.5. Types of metrics and combinations used in each of final three Level I alternatives.

	Metric Type				**IBI combination**		
Metric	**richness**	**composition**	**tolerance**	**functional**	**9-metric**	**7-metric**	**6-metric**
Total Richness	X				X	X	X
Ephemeroptera Rich.	X				X	X	X
Plecoptera Rich.	X				X		
Trichoptera Rich.	X				X		
Acari Rich.	X				X	X	X
Dominance-3		X	X		X	X	X
Biotic Index			X		X	X	X
Tolerant Rich. %		X	X		X	X	X
Shredder %				X	X	X	

Table A.6. Performance of the 3 Primary IBI Candidates at the Level I taxonomic resolution (summary statistics of performance measures for 27 candidate IBIs presented for comparison). See Table A.4 for definition of terms.

	Primary IBI Candidates			**Summary Statistics Across All 27-IBIs Considered**				
	9-Metric	**7-Metric**	**6-Metric**	**Min**	**25th %**	**Median**	**75th %**	**Max**
IBI Performance								
Noise	11.3	11.8	10.8	10.6	11.8	12.6	13.6	14.7
Signal	1.52	1.51	1.53	1.49	1.53	1.55	1.60	1.62
Signal:Noise Ratio	2.54	2.39	2.74	1.99	2.18	2.44	2.56	2.83
Overlap at 25th Percentile	0.22	0.22	0.19	0.19	0.23	0.25	0.25	0.28
Redundancy - Max Corr (r)	0.78	0.67	0.67	0.64	0.67	0.78	0.83	0.90
Comparison to Level II IBI								
R^2 - Level I vs Level II	0.96	0.94	0.90	0.85	0.89	0.90	0.92	0.96
Classification Match (ref)	13%	10%	16%	10%	13%	17%	23%	29%
Classification Match (test)	8%	13%	6%	6%	13%	14%	17%	28%

TABLE A.8a. REFERENCE SITES COMPARED ACROSS BIOCRITERIA
(L-II = Level II; L-I = Level I; O/E = RIVPACS O/E)
Green = Supporting; Yellow = Supporting Intermediate; Red = Not Supporting

Stream.site	Stream	Site	Year	L-II (IBI-10)	L-I (IBI-9)	O/E
Spratt	Spratt Cr	above road xing	2000	100.0	100.0	1.18
Swauger.valdez	Swauger Cr	lower Valdez	2002	99.7	95.6	1.22
Sagehen	Sagehen Ck	below field station	2000	97.8	98.0	1.06
Bear	Bear Cr	lower	2000	97.1	99.0	1.03
Bear	Bear Cr	lower	2001	96.7	92.6	1.03
Truck.forest	Upper Truckee R	above Christmas Valley (forest)	2000	96.3	95.6	0.96
Prosser.below	Prosser Cr	below confluence	2000	95.7	95.6	1.13
WWalker.leavitt	West Walker R	upper Leavitt	2001	95.2	94.2	1.01
Pole	Pole Cr	tributary ref	2000	95.1	94.4	0.85
Truck.forest	Upper Truckee R	above Christmas Valley (forest)	1999	94.0	93.3	1.08
Hidden	Hidden Valley Cr	above confluence	2001	93.9	95.0	0.96
Wolf	Wolf Cr	above trailhead	2000	93.9	95.6	1.14
Truck.park	Upper Truckee R	at State Park	1999	93.7	95.6	1.21
Prosser.north	North Prosser Cr	below USFS boundary	2001	93.4	91.1	1.03
Lacey	Lacey Canyon Cr	confined section	2001	93.0	90.4	1.00
Sagehen	Sagehen Ck	below field station	2001	92.7	89.6	1.11
WWalker.leavitt	West Walker R	upper Leavitt	2002	92.7	92.0	1.21
Swauger.above	Swauger Cr	above East Fork	1999	91.9	88.2	0.87
Silver.cr	Silver Cr	above fence	2002	91.8	88.4	1.21
WWalker.confl	West Walker R	upper confluence	1999	91.5	91.5	1.06
WWalker.confl	West Walker R	upper confluence	2002	90.6	89.3	1.20
Willow	Willow Cr	lower	2000	90.3	87.1	1.20
Trib.silver	Trib 1 - Silver King Cr	above SKC	2003	90.0	86.8	1.03
ECarson.bagley	East Carson R	above Bagley Valley	2000	89.4	93.8	0.94
Truck.celio.low	Upper Truckee R	Celio lower	2000	89.0	93.8	1.15
General	General Cr	below loop road	2000	88.0	88.8	0.89
Robinson.camp	Robinson Cr	below Robinson Cr campground	2002	86.9	83.7	1.12
Independence	Independence Cr	below road	2001	86.5	84.1	1.03
Lee	Lee Vining Cr	Moraine campground	2000	86.3	66.7	1.11
Cottonwood	Cottonwood Cr	Sweetwater Meadow site	2000	85.7	83.9	1.06
Cold	Coldstream Cr	upper gravel pit	2000	85.4	86.6	0.98
Lit.Truck.perazzo	Little Truckee R	at upper Perazzo Mdw	2000	85.1	93.6	0.97
Owens.spring	Upper Owens R	below Big Springs	2000	84.7	83.8	0.84
Robinson.honey	Robinson Cr	Honeymoon Flat	2000	84.4	82.9	1.15
Truck.celio.up	Upper Truckee R	Celio upper	1999	84.3	91.1	1.00
Cottonwood	Cottonwood Cr	Sweetwater Meadow site	2002	83.8	86.1	1.07
WWalker.leavitt	West Walker R	upper Leavitt	2000	83.1	82.2	0.94
Swauger.above	Swauger Cr	above East Fork	2002	83.0	79.1	1.00
Trib.silver	Trib 1 - Silver King Cr	above SKC	2002	83.0	81.6	0.88
Trib.silver	Trib 1 - Silver King Cr	above SKC	2000	82.9	86.1	0.97
Silver.cr	Silver Cr	above fence	1999	82.7	73.0	1.04
Silver.king	Silver King Cr	above valley	2000	82.6	83.9	1.09
Truck.celio.low	Upper Truckee R	Celio lower	1999	82.5	86.7	1.07
Convict	Convict Cr	lower SNARL	2000	82.4	89.7	0.98
Juniper	Juniper Cr	above rd xing	2001	82.2	76.9	0.97
Swauger.valdez	Swauger Cr	lower Valdez	1999	81.1	75.6	0.86
Lit.Truck.below	Little Truckee R	below Coldstream	2001	80.2	80.6	1.06
Mill	Mill Cr	central	1999	80.0	83.4	0.87
Lit.Walker.lower	Little Walker R	lower	2002	78.4	77.9	0.94
Alder	Alder Cr	meadow	2001	77.9	76.4	1.08
Cottonwood	Cottonwood Cr	Sweetwater Meadow site	1999	76.8	70.5	1.08
WWalker.leavitt	West Walker R	upper Leavitt	1999	75.8	74.8	0.85
Robinson.camp	Robinson Cr	below Robinson Cr campground	1999	66.8	64.1	1.09
Mammoth	Mammoth Cr	substation	2000	66.5	69.5	0.97
Truck.park	Upper Truckee R	at State Park	2000	66.4	71.7	1.05
Forestdale	Forestdale Cr	upper	2000	63.9	59.3	0.91
WCarson.blm	West Carson R	lower BLM	2000	63.8	65.0	0.72
Cottonwood	Cottonwood Cr	Sweetwater Meadow site	2001	63.3	66.5	1.14
Truck.celio.up	Upper Truckee R	Celio upper	2000	63.2	72.5	1.04
WCarson.faith	West Carson R	Upper Faith	2000	61.3	62.0	0.94
Deadman	Deadman Cr	above Big Springs campground	2000	61.0	61.6	0.83
Lit.Walker.lower	Little Walker R	lower	1999	55.5	51.3	0.77

TABLE A.8b. TEST SITE ASSESSMENTS USING DIFFERING BIOCRITERIA
(L-II = Level II; L-I = Level I; O/E = RIVPACS O/E)
Green = Supporting;   Yellow = Supporting Intermediate; Red = Not Supporting

Stream.site	Stream	Site	Year	L-II (IBI-10)	L-I (IBI-9)	O/E
Squaw.nfk	Squaw Cr, N Fork	below Silverado	2001	97.4	93.6	0.85
Squaw.nfk	Squaw Cr, N Fork	below Silverado	2000	96.4	96.0	0.82
Squaw.sfk	Squaw Cr, S Fork	below lower Headwall	2001	95.4	88.5	0.71
Squaw.moraine	Squaw Cr	moraine	2000	94.3	96.1	0.91
Mill	Mill Cr	central	2002	93.1	92.0	0.97
WWalker.settle	West Walker R	Settlemeyer	2002	89.8	90.5	0.92
WWalker.pick.low	West Walker R	lower Pickel	2002	88.1	88.8	0.96
WWalker.pick.mid	West Walker R	middle Pickel	2002	87.6	88.7	0.89
WWalker.pick.mid	West Walker R	middle Pickel	2000	87.4	88.2	0.85
Lit.Walker.cent	Little Walker R	central	2002	86.6	89.8	1.19
WWalker.pick.mid	West Walker R	middle Pickel	1999	83.8	86.4	0.74
Squaw.mid	Squaw Cr	middle meadow	2001	83.4	80.8	0.80
Perazzo	Perazzo Cr	meadow	2001	80.9	85.0	0.94
Heavenly	Heavenly Valley Cr	above powerline	2001	78.8	84.4	0.97
WWalker.pick.mid	West Walker R	middle Pickel	2001	77.4	71.7	0.92
Martis	Martis Cr	above confluence	2001	77.0	76.5	0.91
Squaw.sfk	Squaw Cr, S Fork	below lower Headwall	2000	74.8	73.7	0.71
Truck.sun.low	Upper Truckee R	Sunset Stable lower	1999	73.7	79.2	0.97
WWalker.settle	West Walker R	Settlemeyer	1999	72.4	72.3	0.91
Kirman.lower	Kirman Cr	lower	2002	72.2	65.9	0.74
Cowcamp	Cowcamp Cr	lower Schoettler	2001	68.2	65.9	0.82
Kirman.upper.alt	Kirman Cr	alternate upper	2002	68.2	75.2	0.75
Lit.Walker.cent	Little Walker R	central	1999	67.6	66.4	0.86
Cowcamp	Cowcamp Cr	lower Schoettler	2000	66.3	79.1	0.60
Truck.sun.up	Upper Truckee R	Sunset Stable upper	1999	66.0	73.6	0.88
Kirman.lower	Kirman Cr	lower	2001	63.1	65.4	0.72
Truck.sun.up	Upper Truckee R	Sunset Stable upper	2000	63.1	76.0	0.78
Bagley.restore	Bagley Valley Cr	restoration project	2002	60.9	62.2	0.57
WWalker.pick.low	West Walker R	lower Pickel	1999	59.6	56.4	0.83
Buckeye	Buckeye Cr	below WRID	2000	58.5	48.3	0.57
Trout	Trout Cr	Bennett Flat	2001	57.7	59.3	0.70
Bagley.restore	Bagley Valley Cr	restoration project	2003	57.4	57.3	0.56
Squaw.lower	Squaw Cr	lower meadow	2001	56.6	53.7	0.68
Slinkard	Slinkard Cr	restoration area	2002	55.4	51.7	0.65
Cowcamp	Cowcamp Cr	lower Schoettler	1999	54.3	50.2	0.86
Truck.bart.up	Upper Truckee R	Barton upper	1998	53.9	61.9	0.49
Truck.bart.low	Upper Truckee R	Barton lower	1999	53.8	52.8	0.57
Slinkard	Slinkard Cr	restoration area	2003	52.3	50.7	0.65
Kirman.lower	Kirman Cr	lower	1999	51.8	56.3	0.82
Kirman.lower	Kirman Cr	lower	2000	51.2	51.4	0.59
Truck.bart.up	Upper Truckee R	Barton upper	1999	47.2	55.8	0.64
Squaw.upper	Squaw Cr	upper mdw	2000	46.0	36.2	0.36
Squaw.lower	Squaw Cr	lower meadow	2000	45.4	39.9	0.35
Owens.belowtun	Upper Owens R	below Mono tunnel	2000	45.1	45.0	0.58
Truck.sun.low	Upper Truckee R	Sunset Stable lower	2000	45.0	49.9	0.79
Truck.bart.up	Upper Truckee R	Barton upper	2000	44.6	59.0	0.71
Poore.above	Poore Cr	above fence	1999	44.5	43.6	0.61
Robinson.below	Robinson Cr	below WRID	2000	42.5	43.5	0.62
Slinkard	Slinkard Cr	restoration area	2000	42.2	41.9	0.71
Poore.above	Poore Cr	above fence	2002	41.4	33.8	0.71
Truck.bart.low	Upper Truckee R	Barton lower	1998	40.3	38.6	0.43
Cowcamp	Cowcamp Cr	lower Schoettler	2002	40.3	40.0	0.75
Owens.power	Upper Owens R	Ebasco Powerline	2000	40.2	38.0	0.48
Truck.bart.low	Upper Truckee R	Barton lower	2000	39.1	46.5	0.63
EWalker	East Walker R	below WRID fence	2000	38.5	41.3	0.42
Rush	Rush Cr	bottomlands	2000	36.9	38.2	0.64
Owens.benton	Upper Owens R	below Benton xing	2000	35.0	33.1	0.29
Squaw.mid	Squaw Cr	middle meadow	2000	34.6	32.5	0.30
Bagley.control	Bagley Valley Cr	control	1999	33.5	31.8	0.55
Poore.below	Poore Cr	below fence	2000	33.1	24.7	0.36
Poore.below	Poore Cr	below fence	1999	32.5	31.1	0.50
Owens.417	Upper Owens R	Ebasco 417s	2000	31.9	32.1	0.38
Owens.bridge	Upper Owens R	above bridge	2000	30.3	36.2	0.35
Poore.below	Poore Cr	below fence	2001	28.4	23.7	0.50
Poore.below	Poore Cr	below fence	2002	25.4	22.5	0.66
Bagley.meadow	Bagley Valley Cr	lower meadow	1999	22.6	19.6	0.56
Bagley.control	Bagley Valley Cr	control	2002	19.8	11.9	0.51
Bagley.control	Bagley Valley Cr	control	2003	18.9	14.9	0.59
Bagley.control	Bagley Valley Cr	control	2000	17.2	17.0	0.51
Kirman.upper	Kirman Cr	upper	1999	15.7	12.2	0.68
Bagley.meadow	Bagley Valley Cr	lower meadow	2000	14.3	14.3	0.49
Owens.abovetun	Upper Owens R	above Mono tunnel	2000	10.9	9.6	0.47

Figure A.1. Box-and-Whisker Plot showing the comparison between reference and test sites for each of the final 3 candidate IBIs for the Level I taxonomic resolution.

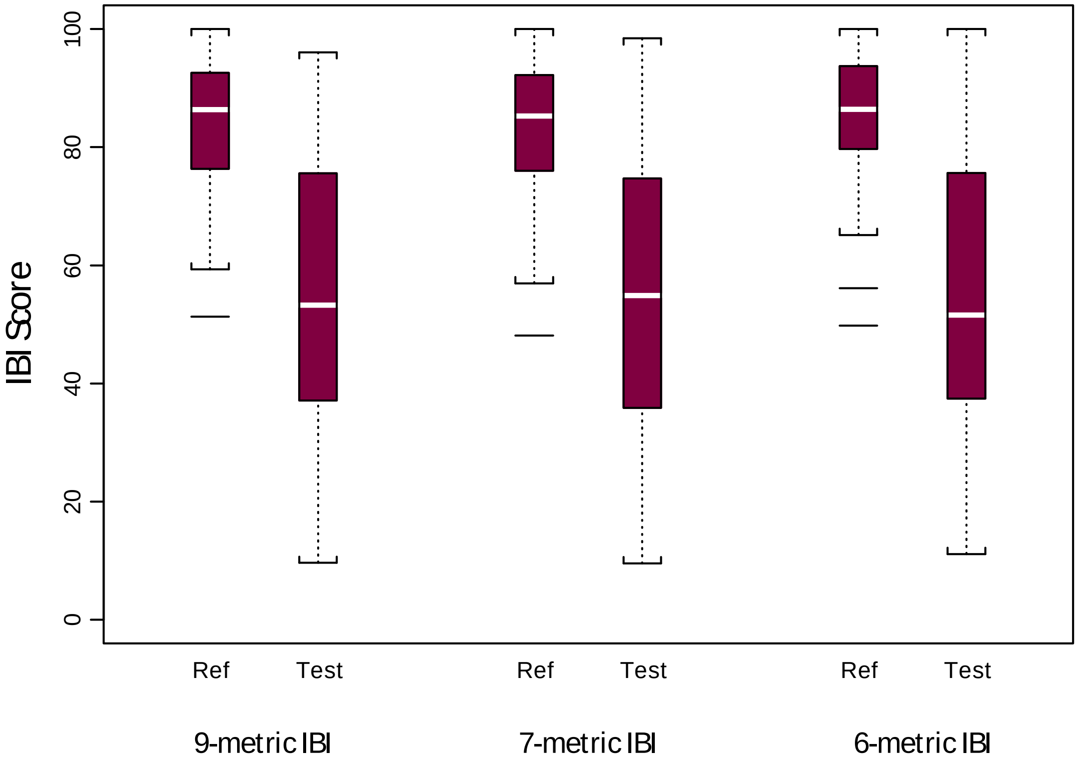


Figure A.2. Cumulative Distribution Function plots of IBI scores for Reference and Test Sites for the three final IBIs considered at the Level I resolution.

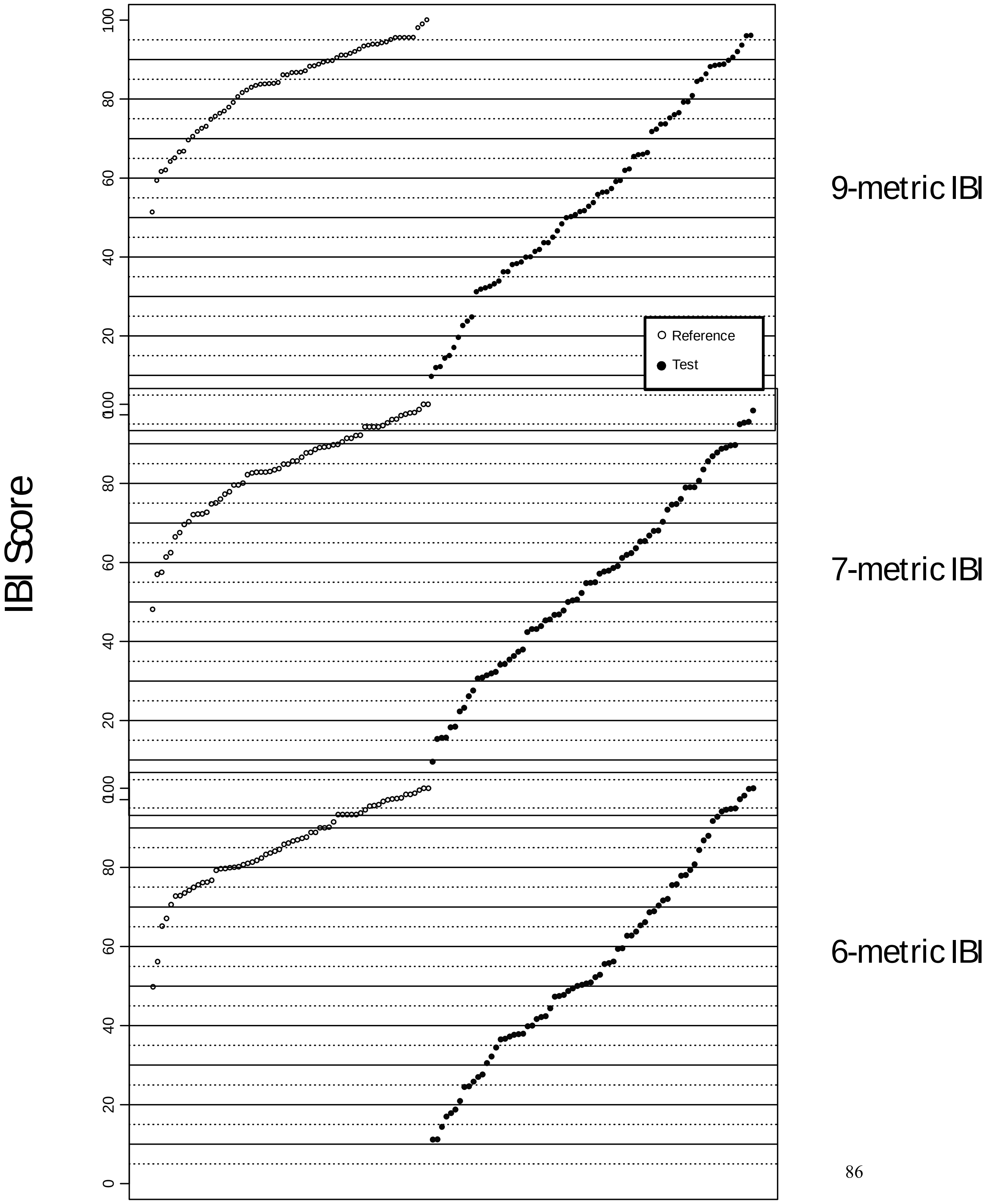


Figure A.3. Scoring and Classification of Level II 10-metric IBI against candidate Level I 9-metric IBI. (red points indicate impair/attain non-correspondences; blue points indicate levels of attainment non-correspondence; dotted lines indicate 5th and 25th percentiles of reference sites)

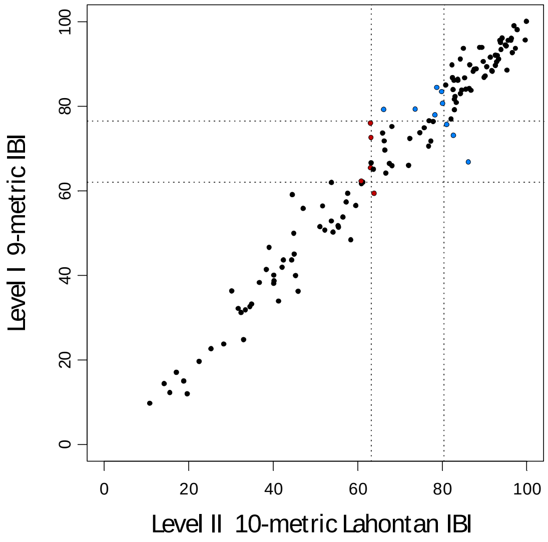


Figure A.4. Scoring and Classification of Level II 10-metric IBI against candidate Level I 7-metric IBI. (symbols as in Fig. A.3)

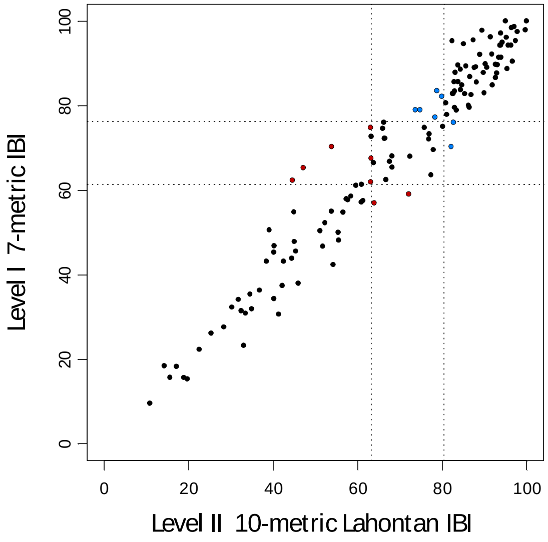


Figure A.5. Scoring and Classification of Level II 10-metric IBI against candidate Level I 6-metric IBI. (symbols as in Fig. A.3)

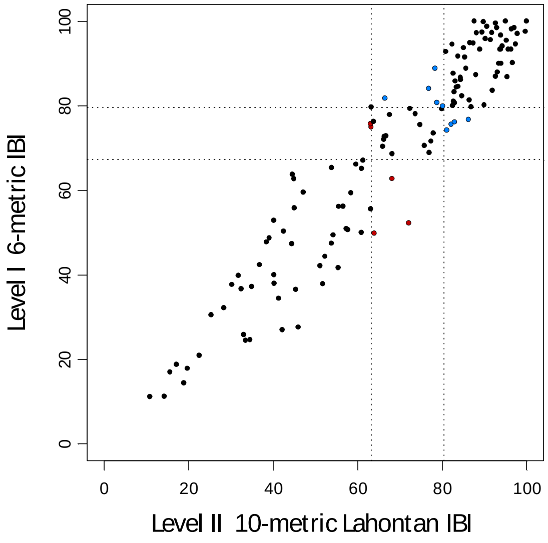


Figure A.6 abc. Sediment-exposure stressor responses among assessment methods.

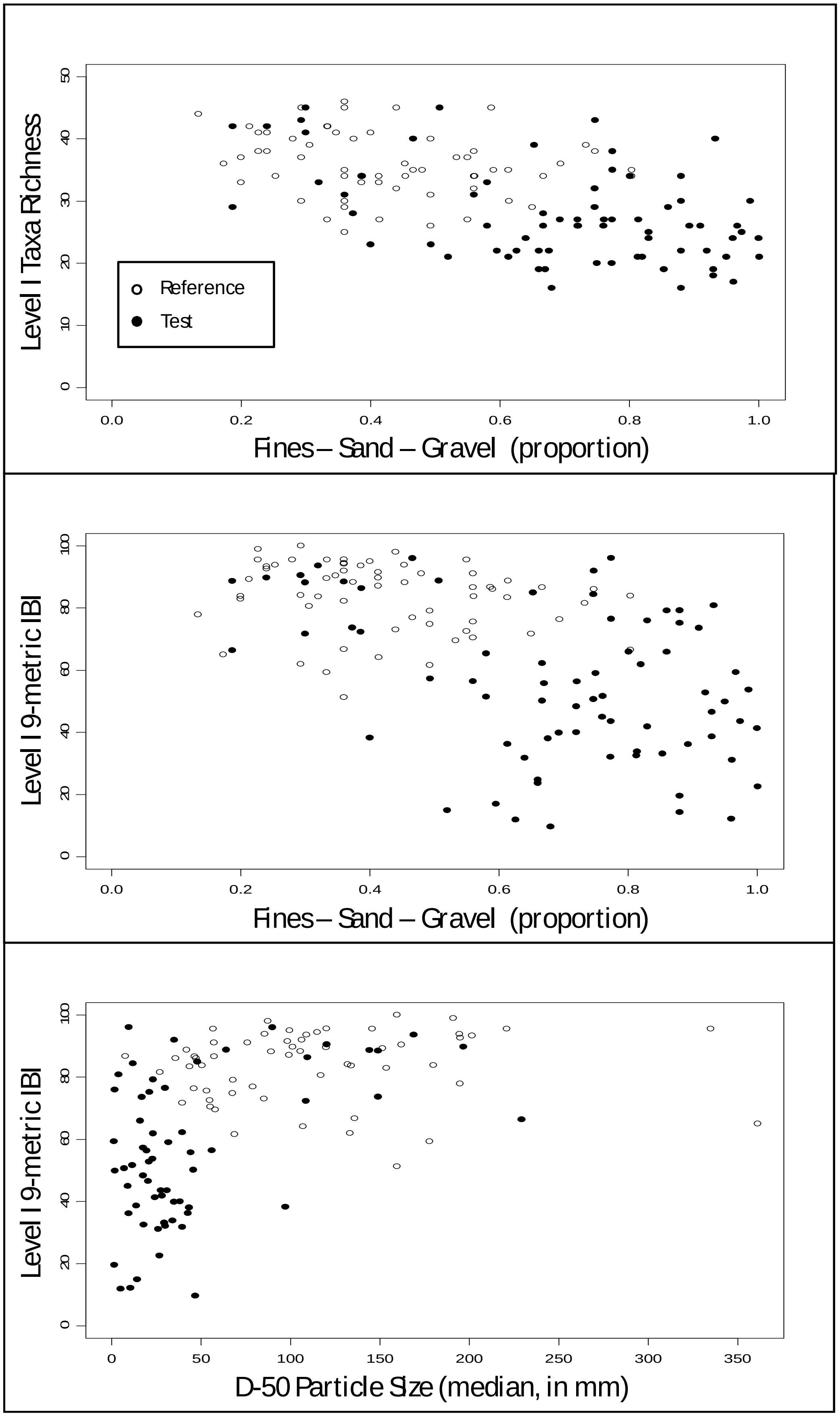


Figure A.7. Ranked order distribution of reference followed by test sites.