

DEVELOPMENT AND TESTING OF A BIOLOGICAL INDEX FOR COLDWATER STREAMS OF ARIZONA

Prepared for

Arizona Department of Environmental Quality
Water Quality Division
3033 North Central Avenue
Phoenix, Arizona 85012

Work Assignment Manager
Patti Spindler, Arizona DEQ

Prepared by

Erik W. Leppo
Jeroen Gerritsen

Tetra Tech, Inc.
10045 Red Run Blvd., Suite 110
Owings Mills, Maryland 21117

July 2000 (Revised March 2006)

ABSTRACT

Objectives of the analysis were to develop a biological index for coldwater Arizona streams. We analyzed benthic macroinvertebrate data collected by Arizona DEQ from 1992 - 1998. Arizona currently divides their streams into warmwater and coldwater (generally defined by elevation; warmwater below 5,000 feet and coldwater above 5,000 feet). This report is to develop a multimetric index for coldwater streams similar to that developed for warmwater streams (Gerritsen and Leppo 1998).

Seven benthic metrics were selected for a biological index for coldwater Arizona streams: total taxa, Diptera taxa, intolerant taxa, Hilsenhoff Biotic Index (HBI), percent Plecoptera, percent scrapers, and scraper taxa. The resultant index was able to fully discriminate between reference sites and stressed sites. The spring index period was used as there were too few fall samples on which to base our analysis. To correspond with the warmwater sampling, only riffle samples were examined. Identifications to a split level of taxonomy were used (Chironomids to family, all other organisms to genus-level), and rarefaction was not used.

TABLE OF CONTENTS

ABSTRACT..... ii
LIST OF FIGURES iv
LIST OF TABLES iv

1 INTRODUCTION..... 1
 1.1 Background..... 1
 1.2 Objective 2

2 METHODS 3
 2.1 Sample Collection 3
 2.2 Data Management 4
 2.3 Metric Calculation and Evaluation 4
 2.4 Index Development..... 5

3 RESULTS AND DISCUSSION..... 9
 3.1 Metric Evaluation: Season and Habitat 9
 3.2 Metric Selection..... 10
 3.3 Index Calculation 11
 3.4 Proposed Assessment Classes 12

4 RECOMMENDATIONS..... 14
 4.1 Future Sampling..... 14

LITERATURE CITED..... 15

APPENDIX A: PERFORMANCE OF CANDIDATE METRICS (Spring Riffle, Split-level taxonomy)

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1 Diagram of relationships of tables in Access database of Arizona data..... | 5 |
| 2 Working index for Arizona coldwater streams..... | 12 |

LIST OF TABLES

| Table | Page |
|---|------|
| 1 Distribution of coldwater samples by sample classes..... | 3 |
| 2 Attributes of benthic macroinvertebrates used as candidate metrics..... | 6 |
| 3 Example of metric standardization, using values from the sample at site SR2LCBRB000.18..... | 8 |
| 4 Discriminatory ability of candidate metrics considered for Arizona=s coldwater streams. WW and CW refer to warmwater and coldwater, respectively. Comments and use of index are relevant to riffle samples, only.. .. | 9 |
| 5 Statistics of selected metrics for all Spring Riffle samples | 10 |
| 6 Pearson correlations of potential metrics used for the Coldwater index | 11 |
| 7 Proposed narrative categories assigned to Arizona coldwater index scores..... | 13 |

1 INTRODUCTION

1.1 Background

Multimetric, macroinvertebrate indices of biotic integrity, variously called RBP (Rapid Bioassessment Protocol; Plafkin et al. 1989); ICI (Invertebrate Condition Index; Ohio EPA 1989); B-IBI (Benthic IBI, Kerans and Karr 1994) have been developed for many regions of North America and are generally accepted for biological assessment of aquatic resource quality (e.g., Gibson et al. 1996; Southerland and Stribling 1995; Karr 1991). The framework of bioassessment consists of characterizing reference conditions upon which comparisons can be made, and identifying appropriate biological attributes with which to measure the condition. Reference conditions are “best available” conditions where biological potential is at its highest for the particular region or area. These reference conditions are representative of sustainable ecosystem health.

The Arizona Department of Environmental Quality (DEQ) is developing an Index of Biotic Integrity (IBI) for use in assessing the quality of Arizona's streams, and for eventual development of biological criteria for stream resource quality. The Arizona index is being developed for benthic macroinvertebrates (Spindler, in press).

Biological integrity is commonly defined as “the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity and functional organization comparable to that of the natural habitat of the regions” (Karr and Dudley 1981; Gibson et al. 1996). This concept refers to the natural assemblage of indigenous organisms that would inhabit a particular area if it had not been affected by human activities. This integrity or naturally occurring diversity becomes the primary reference condition used to measure and assess waterbodies in a particular region.

Careful measurement of key attributes of the natural aquatic ecosystem and its constituent biological communities can determine the condition of biological integrity. These key attributes or biological endpoints indicate the quality of the waters of concern. They are established by biosurveys based on sampling of fish, invertebrates, plants, and other flora and fauna. Such biosurveys establish the endpoints or measures used to summarize several community characteristics such as taxa richness, number of individuals, sensitive or insensitive species, observed pathologies, and the presence or absence of essential habitat elements.

A study to classify macroinvertebrate communities in reference streams of Arizona showed that streams could be classified into two groups at an elevation cutoff of 5,000 feet: montane or coldwater streams, and desert or warmwater streams (Spindler, in press). This classification provides the basis for further development of a multimetric index in this report.

The biological attributes to be measured represent elements of the structure and function of the bottom-dwelling macroinvertebrate assemblage and are called *metrics*. A metric is defined as a characteristic of the biota that changes in some predictable way with increased human influence (Barbour et al. 1996). These metrics are specific measures of diversity, composition, and functional feeding group representation and include ecological information on pollution tolerance.

A warmwater index was developed for AZ DEQ using the same methods that will be employed in developing the coldwater index. The warmwater index was developed for sites of elevation less than 5,000 feet and consisted of 9 metrics: total taxa, Hilsenhoff Biotic Index (HBI), % 1 dominant taxon, % Ephemeroptera, Ephemeroptera taxa, Trichoptera taxa, Diptera taxa, % scrapers, and scraper taxa. The final index was based on a 0-100 scale where sites most like reference conditions scored closest to 100 and as perturbation increased the scores decreased.

1.2 Objective

The objective of this report is to develop and test a multimetric biological index for coldwater Arizona streams, using data collected by Arizona DEQ throughout Arizona from 1992 - 1998.

2 METHODS

2.1 Sample Collection

Arizona DEQ has sampled candidate reference and impaired streams throughout the state since 1992, in two biological index periods, spring and fall. Samples were taken in two habitats, pool and riffle, and analyzed separately. Benthic macroinvertebrates in the samples were identified to the lowest practical taxonomic level with Chironomidae (midges) only identified to family level.

Arizona DEQ collected macroinvertebrate samples statewide from 1992 through 1998 in both riffle and pool habitats using modified Rapid Bioassessment Protocols (RBP; Plafkin et al. 1989). The Arizona sampling protocol (Meyerhoff and Spindler, 1994) for a riffle sample consists of a three minute timed composite sample from three riffle habitats in the same reach, using a D-frame dipnet. Pool samples were collected similarly. Microhabitats were sampled separately by hand-picking organisms from microenvironments for 30 minutes. A minimum of 500 macroinvertebrates (300 for 1992-1994) were randomly subsampled and identified to the lowest practical taxonomic level, usually to genus (Meyerhoff and Spindler 1994). Chironomidae taxa were then lumped at the family level.

Individual samples were examined in this study rather than sites. A sample was defined as a sampling event in a particular place at a particular time. Thus, sites can have more than one sample.

Sites were grouped by Arizona DEQ into 5 sample classes: reference, potential reference, non-reference, suspected impacted, and effluent dependent waters (EDW). For analysis these 5 site classes were regrouped into 3 site classes: reference (reference and potential reference), non-reference (non-reference) and stressed (suspected impacted and effluent dependent waters). The distribution of samples among seasons, habitats, and sample classes is shown in Table 1.

Table 1. Distribution of coldwater samples by sample classes.

| Season | Habitat | Reference | Non-Reference | Stressed | Total |
|--------------|---------|-----------|---------------|----------|-------|
| Fall | Pool | 0 | 0 | 0 | 0 |
| | Riffle | 2 | 0 | 0 | 2 |
| Spring | Pool | 54 | 27 | 4 | 85 |
| | Riffle | 102 | 64 | 14 | 180 |
| Total | | | | | 267 |

2.2 Data Management

Data were received from Arizona DEQ as Quattro Pro7 spreadsheet files. The data were then transferred to a database (MS-Access7), and checked for data integrity:

- § Inconsistent taxonomic names were corrected through discussions with Arizona DEQ (e.g., Baetidae-larvae was changed to Baetidae).
- § Attributes of taxa (genus or family-level), such as functional feeding groups and tolerance values, were supplied by Arizona DEQ or were taken from published reference works when not supplied (Merritt and Cummins, 1994, US EPA, draft 1990).

The structure of a database is designed to store information in a compact manner that is also flexible enough to extract needed information. The Quattro Pro7 spreadsheets were transformed into lists of information and then imported into Access. These lists were made into tables that contained only a single type of information (e.g., sample taxon, and number counted) but had identifiers to link to data in other tables. The tables were linked to form relationships that reflect the underlying structure of the data and how they are collected (Figure 1). For example, to extract information for a particular site one would pick a StationID in the Stations table. The StationID is linked to samples for that station in the BenSamps, HabSamps, and ChemSamps tables. The StationID is in turn linked to the taxa and number found at the site in the Benthics table. The Benthics_Master_Taxa table contains only information on the taxa, i.e., tolerance values, functional feeding groups, and complete phylogenetic names. The taxa information in the Benthic_Master_Taxa table is linked by FinalID to the numbers of organisms in the Benthics table.

2.3 Metric Calculation and Evaluation

Attributes (candidate metrics) that were calculated are listed in Table 2. These attributes were selected based upon known or suspected ability to discriminate impairment and are the same as those used for the development of Arizona's warmwater index (Gerritsen and Leppo 1998). The attributes in Table 2 were calculated for each habitat (pool and riffle).

Box-and-whisker plots were used to evaluate differences in value ranges of the metrics among the stream classes (reference, non-reference, and stressed). This type of plot displays the statistics (i.e., median value, minimum, maximum, 25th, and 75th percentile) of a population of sites. Decisions regarding the distinction of the populations of sites were made based on the degree of similarity or dissimilarity among the plots (i.e., the discrimination ability of each metric).

Table 2. Attributes of benthic macroinvertebrates used as candidate metrics.

| Category | Metric | Definition | Expected response to increasing perturbation |
|----------------------|---------------------------|--|---|
| Richness measures | Total number of taxa | Measures the overall variety of the macroinvertebrate assemblage. | Decrease |
| | No. of EPT taxa | Sum of the number of taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies). | Decrease |
| | No. of Ephemeroptera taxa | Number of taxa of mayfly nymphs. | Decrease |
| | No. of Plecoptera taxa | Number of taxa of stonefly naiads. | Decrease |
| | No. of Trichoptera taxa | Number of taxa of caddisfly larvae. | Decrease |
| | No. of Chironomidae taxa | Number of taxa of chironomid (midge) larvae. | Decrease |
| | No. of Diptera taxa | Number of taxa of truefly larvae | Decrease |
| Composition measures | % Dominant taxon | The percent abundance of the single most abundant taxon. | Increase |
| | % EPT | Percent abundance of mayfly nymphs, stonefly naiads, and caddisfly larvae. | Decrease |
| | % Ephemeroptera | Percent abundance of mayfly nymphs. | Decrease |
| | % Plecoptera | Percent abundance of stonefly naiads. | Decrease |
| | % Trichoptera | Percent abundance of caddisfly larvae. | Decrease |
| | % Chironomidae | Percent abundance of chironomid larvae. | Increase |
| | % Diptera | Percent abundance of individuals classed as dipterans. | Increase |
| | % Oligochaeta | Percent abundance of aquatic worms. | Increase |
| Tolerance measures | Hilsenhoff Biotic Index | Abundance-weighted average tolerance of assemblage. | Increase |
| | % Tolerant organisms | Percent abundance of organisms with a Tolerance Value ≥ 7 . | Increase |
| | No. of Intolerant taxa | Number of taxa with a Tolerance Value ≤ 3 . | Decrease |

Table 2. Continued.

| Category | Metric | Definition | Expected response to increasing perturbation |
|-----------------|-----------------------|-------------------------------------|---|
| Trophic | % Collector-Filterers | Percent abundance of the collector- | Increase |

| | | | |
|----------------|--------------------------------|---|----------|
| measures | | filterer functional feeding group. | |
| | No. of Collector-Filterer taxa | Number of taxa of the collector-filterer functional feeding group. | Decrease |
| | % Collector-Gatherers | Percent abundance of the collector-gatherer functional feeding group. | Decrease |
| | No. of Collector-Gatherer taxa | Number of taxa of the collector-gatherer functional feeding group. | Decrease |
| | % Predators | Percent abundance of the predator functional feeding group. | Decrease |
| | No. of Predator taxa | Number of taxa of the predator functional feeding group. | Decrease |
| | % Scrapers | Percent abundance of the scraper functional feeding group. | Decrease |
| | No. of Scraper taxa | Number of taxa of the scraper functional feeding group. | Decrease |
| | % Shredders | Percent abundance of the shredder functional feeding group. | Decrease |
| | No. of Shredder taxa | Number of taxa of the shredder functional feeding group. | Decrease |
| Habit measures | % Clingers | Percent abundance of the clinger habit group. | Decrease |
| | No. of Clinger taxa | Number of taxa of the clinger habit group. | Decrease |

Selected metrics were standardized on a “best” value (Table 3). The best value was the 95th percentile of the total distribution, for those metrics that decrease in value with perturbation (e.g., total taxa, intolerant taxa, % Plecoptera, % scrapers). The 95th percentile was used to eliminate unusual outliers as a scoring standard, to avoid skewing the resultant scores. All scores greater than or equal to this standard were given a score of 100, those scores less than the

$$\text{score} = \left(\frac{x}{x_{95} - x_{\min}} \right) \times 100 \text{ for } x \leq x_{95};$$

standard were scored as a percentage of the standard:

where x is the metric value; x_{95} is the 95th percentile value; and x_{\min} is the minimum possible value (usually 0).

For those selected metrics that increase in value with an increase in perturbation (e.g., HBI) the 5th percentile of reference sites was used as the standard. All values less than or equal to this standard were given a score of 100; and values greater than the standard were given scores as the

$$\text{score} = \left(\frac{x_{\max} - x}{x_{\max} - x_5} \right) \times 100 \text{ for } x \geq x_5;$$

percentage of the range from to the maximum (worst) value to the 5th percentile (best) value: where x_5 is the 5th percentile value; and x_{\max} is the maximum possible value (e.g., 100% for percentage metrics; 10 for HBI; maximum observed value for taxa richness metrics).

Table 3. Example of metric standardization, using values from the sample at site SR2LCBRB000.18.

| Metric | Direction | Percentile Used | Standard (best value) | Measured Metric Value | Standardized Metric Score |
|-----------------|------------------|------------------------|------------------------------|------------------------------|----------------------------------|
| Total Taxa | Decrease | 95 th | 37 | 21 | 56.8 |
| Diptera Taxa | Decrease | 95 th | 11 | 10 | 90.9 |
| Intolerant Taxa | Decrease | 95 th | 5 | 1 | 20.0 |
| HBI | Increase | 5 th | 4.2 (max 10) | 5.3 | 81.7 |
| % Plecoptera | Decrease | 95 th | 18.6 | 0.84 | 4.5 |
| % Scrapers | Decrease | 95 th | 49.3 | 0.28 | 0.6 |
| Scraper Taxa | Decrease | 95 th | 11 | 2 | 18.2 |
| Total IBI Score | | | | | 39.0 |

The individual metrics are then averaged to obtain the total IBI score for a sample. The closer the score to 100 the closer the sample is to reference conditions. Only sites that exceed reference conditions for each metric will receive a total score of 100.

3 RESULTS AND DISCUSSION

3.1 Metric Evaluation: Season and Habitat

Metrics were evaluated on their ability to discriminate impairment among site classes and consistent with previous investigations (e.g., Barbour et al. 1996, Karr and Chu 1997, Fore et al. 1996). Table 4 shows the discriminatory rating of 30 metrics by season and habitat. Discrimination efficiency was also calculated for each metric. The numbers reported are the percentage of stressed sites that were assessed correctly (i.e., falling below the 25th percentile for metrics that decrease with impairment and above the 75th percentile for metrics that increase with impairment). A higher number indicates better performance at distinguishing between reference and stressed samples.

Table 4. Discriminatory ability of candidate metrics considered for Arizona's coldwater streams. WW and CW refer to warmwater and coldwater, respectively. Comments and use of index are relevant to riffle samples, only.

| Spring Metrics | Expected Response | Quartile Used | Quartile Value | | Discrimination Efficiency | | Comment (Riffle) | Index Use (Riffle) | |
|-------------------------|-------------------|---------------|----------------|------|---------------------------|------|------------------|--------------------|-----|
| | | | Riffle | Pool | Riffle | Pool | | WW | CW |
| Total Taxa | D | 25 | 23.0 | 21.0 | 75.0 | 50 | strong | yes | yes |
| HBI | I | 75 | 5.5 | 6.0 | 68.8 | 50 | intermediate | yes | yes |
| % 1 Dominant Taxon | I | 75 | 43.6 | 63.7 | 50.0 | 75 | weak | yes | |
| % Tolerant | I | 75 | 12.3 | 23.9 | 37.5 | 25 | weak | | |
| Intolerant Taxa | D | 25 | 4.0 | 1.0 | 100.0 | 50 | strong | | yes |
| % EPT | D | 25 | 29.6 | 22.3 | 43.8 | 75 | weak | | |
| EPT Taxa | D | 25 | 10.0 | 10.0 | 68.8 | 75 | intermediate | | |
| % Ephemeroptera | D | 25 | 18.4 | 10.7 | 50.0 | 50 | weak | yes | |
| Ephemeroptera Taxa | D | 25 | 4.0 | 4.0 | 56.3 | 0 | intermediate | yes | |
| % Plecoptera | D | 25 | 1.3 | 0.4 | 62.5 | 75 | intermediate | | yes |
| Plecoptera Taxa | D | 25 | 1.0 | 1.0 | 43.8 | 75 | too few | | |
| % Trichoptera | D | 25 | 2.5 | 2.2 | 50.0 | 75 | weak | | |
| Trichoptera Taxa | D | 25 | 4.0 | 3.0 | 43.8 | 50 | weak | yes | |
| % Chironomidae | I | 75 | 25.3 | 63.0 | 43.8 | 75 | weak | | |
| Chironomid Taxa | D | 25 | 1.0 | 1.0 | 0.0 | 0 | too few | | |
| % Diptera | I | 75 | 47.3 | 64.5 | 56.3 | 75 | intermediate | | |
| Diptera Taxa | D | 25 | 5.0 | 4.0 | 62.5 | 25 | intermediate | yes | yes |
| % Oligochaeta | I | 75 | 1.7 | 3.3 | 37.5 | 0 | weak | | |
| % Filterers | I | 75 | 18.4 | 0.7 | 68.8 | 50 | intermediate | | |
| Filterer Taxa | I | 75 | 2.0 | 1.0 | 50.0 | 50 | weak | | |
| % Collector-Gatherers | D | 25 | 12.7 | 14.1 | 25.0 | 25 | weak | | |
| Collector-Gatherer Taxa | D | 25 | 4.0 | 5.0 | 31.3 | 0 | weak | | |
| % Predators | D | 25 | 2.8 | 2.4 | 62.5 | 50 | too few | | |
| Predator Taxa | D | 25 | 5.0 | 4.0 | 56.3 | 25 | intermediate | | |
| % Scrapers | D | 25 | 8.8 | 2.2 | 81.3 | 50 | strong | yes | yes |
| Scraper Taxa | D | 25 | 5.0 | 3.0 | 68.8 | 75 | intermediate | yes | yes |
| % Shredders | D | 25 | 1.0 | 2.2 | 68.8 | 75 | too few | | |
| Shredder Taxa | D | 25 | 1.0 | 2.0 | 37.5 | 75 | too few | | |
| % Clingers | D | 25 | 0.0 | 0.0 | 0.0 | 0 | too few | | |

| | | | | | | | | | |
|--------------|---|----|-----|-----|-----|---|---------|--|--|
| Clinger Taxa | D | 25 | 0.0 | 0.0 | 0.0 | 0 | too few | | |
|--------------|---|----|-----|-----|-----|---|---------|--|--|

A preliminary index was calculated comprising seven metrics: total taxa, Diptera taxa, intolerant taxa, HBI, percent Plecoptera, % scrapers, and scraper taxa. These metrics showed a good ability to discriminate impairment over other metrics that were evaluated. Each of the selected metrics was able to score greater than 60% of the stressed sites correctly. Statistics of the selected metrics are shown in Table 5.

Table 5. Statistics of selected metrics for all Spring Riffle samples.

| Metric | Minimum | 5th percentile | 25th percentile | Median | 75th percentile | 95th percentile | Maximum |
|-----------------|---------|----------------|-----------------|--------|-----------------|-----------------|---------|
| Total taxa | 3 | 12 | 20 | 25 | 30 | 37 | 45 |
| Diptera taxa | 2 | 3 | 5 | 6 | 8 | 11 | 13 |
| Intolerant taxa | 0 | 0 | 1 | 2 | 4 | 5 | 8 |
| HBI | 3.46 | 4.16 | 4.77 | 5.20 | 5.64 | 6.21 | 7.34 |
| % Plecoptera | 0 | 0 | 0.19 | 2.07 | 7.32 | 18.63 | 46.86 |
| % Scrapers | 0 | 0.31 | 3.78 | 16.48 | 29.46 | 49.30 | 87.18 |
| Scraper taxa | 0 | 1 | 3 | 6 | 7 | 11 | 15 |

The fall season was not considered for the index as there were too few samples (2) to analyze. In the spring, there were only 4 stressed samples taken in pools, which made it difficult to make a determination of responsiveness to perturbation. Therefore, only spring riffle samples were considered for the index.

3.2 Metric Selection

Discrimination efficiency was used to select metrics that were best (> 50%) at differentiating stressed sites from reference sites. Metrics that were unresponsive or weak were excluded from further analysis. This left only 12 metrics to be considered for the index.

Minimization of redundancy among metrics was the next selection criteria; for example, % EPT and % Trichoptera would be redundant because Trichoptera are the most abundant component of the % EPT metric. Pearson correlations were also run on all metrics such that no two metrics that were highly correlated ($r > 0.9$) with each other were part of the final index. Table 6 shows the Pearson correlations of the 12 potential metrics for the final index.

Metrics then were selected that were balanced among the different metric categories (richness, composition, tolerance, and trophic function; Table 2). To select metrics from only one category would be to bias the final index towards the potential stressors that affect that category.

Table 6. Pearson correlations of potential metrics used for the coldwater index.

| | % Plecoptera | % Diptera | % Filterer | % Scraper | Scraper Taxa | Predator Taxa | Total Taxa | EPT Taxa | Ephemeroptera Taxa | Diptera Taxa | Intolerant Taxa | HBI |
|--------------------|--------------|-----------|------------|-----------|--------------|---------------|------------|----------|--------------------|--------------|-----------------|-------|
| % Plecoptera | 1.00 | -0.34 | -0.22 | 0.17 | 0.12 | 0.16 | 0.16 | 0.21 | 0.09 | 0.08 | 0.34 | -0.39 |
| % Diptera | -0.34 | 1.00 | 0.61 | -0.52 | -0.30 | -0.06 | -0.24 | -0.29 | -0.22 | 0.04 | -0.12 | 0.52 |
| % Filterer | -0.22 | 0.61 | 1.00 | -0.33 | -0.14 | -0.12 | -0.20 | -0.17 | -0.05 | -0.04 | 0.00 | 0.19 |
| % Scraper | 0.17 | -0.52 | -0.33 | 1.00 | 0.51 | 0.16 | 0.26 | 0.33 | 0.24 | 0.03 | 0.29 | -0.42 |
| Scraper Taxa | 0.12 | -0.30 | -0.14 | 0.51 | 1.00 | 0.36 | 0.75 | 0.79 | 0.62 | 0.32 | 0.43 | -0.25 |
| Predator Taxa | 0.16 | -0.06 | -0.12 | 0.16 | 0.36 | 1.00 | 0.73 | 0.54 | 0.32 | 0.69 | 0.30 | 0.01 |
| Total Taxa | 0.16 | -0.24 | -0.20 | 0.26 | 0.75 | 0.73 | 1.00 | 0.88 | 0.64 | 0.66 | 0.40 | -0.06 |
| EPT Taxa | 0.21 | -0.29 | -0.17 | 0.33 | 0.79 | 0.54 | 0.88 | 1.00 | 0.80 | 0.38 | 0.53 | -0.28 |
| Ephemeroptera Taxa | 0.09 | -0.22 | -0.05 | 0.24 | 0.62 | 0.32 | 0.64 | 0.80 | 1.00 | 0.23 | 0.47 | -0.29 |
| Diptera Taxa | 0.08 | 0.04 | -0.04 | 0.03 | 0.32 | 0.69 | 0.66 | 0.38 | 0.23 | 1.00 | 0.24 | 0.08 |
| Intolerant Taxa | 0.34 | -0.12 | 0.00 | 0.29 | 0.43 | 0.30 | 0.40 | 0.53 | 0.47 | 0.24 | 1.00 | -0.42 |
| HBI | -0.39 | 0.52 | 0.19 | -0.42 | -0.25 | 0.01 | -0.06 | -0.28 | -0.29 | 0.08 | -0.42 | 1.00 |

Table 4 summarizes the discrimination ability of 30 candidate metrics calculated from the statewide data set. Box and whisker plots of each of the 30 candidate metrics by site class are shown in Appendix A.

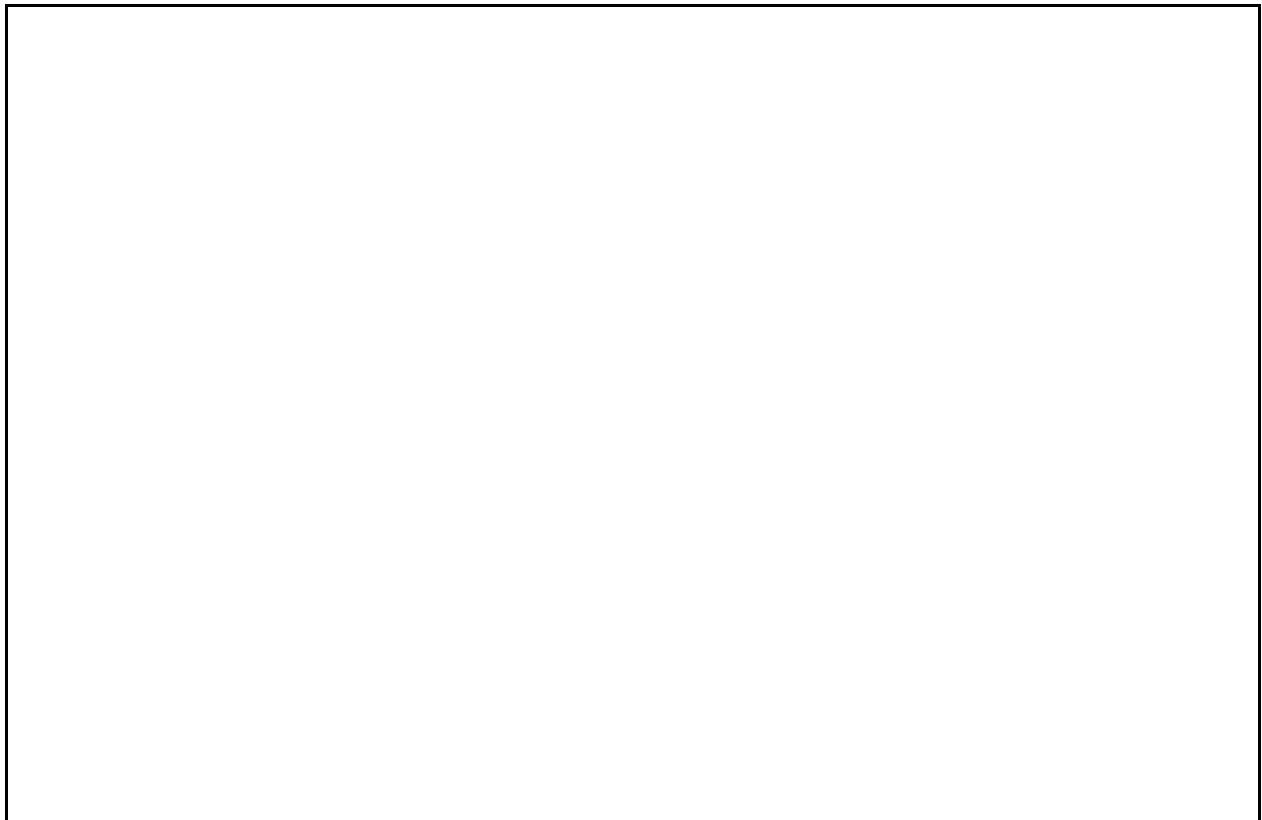
The selection process yielded 7 metrics that are responsive and satisfy the selection criteria:

- \$ Total Taxa
- \$ Diptera Taxa
- \$ Intolerant Taxa
- \$ HBI
- \$ Percent Plecoptera
- \$ Percent Scrapers
- \$ Scraper Taxa

Statistics for each of the selected metrics are shown in Table 5.

3.3 Index Calculation

The scoring standards for the preliminary index were taken as the 5th or 95th percentiles of the metrics for all sites, as described in Section 2.4, Table 3. The scores for each of the six metrics were then averaged for a final index of score with a range of 0 - 100. The resultant index discriminates both reference and stressed sites as shown in Figure 2. Index scores are expected to decrease with increasing perturbation. Scores of 100 are most like reference conditions and as scores decrease the sites are less like the reference conditions. The discrimination ability of the index as a whole was 87.5%.



3.4 Proposed Assessment Classes

To further classify samples, narrative categories can be assigned to the numeric scores of the coldwater index (Table 7). It has already been stated that index scores decrease with perturbation and become less similar to reference conditions. So the narrative categories would communicate a level of similarity between a particular sample and reference conditions. The highest category would be defined as being greater than to the 75th percentile of the reference samples and would be described as “exceptional”. The next category, “good”, would also be

comparable to reference conditions and would be those sites scoring between the 75th percentile and the 25th percentile of reference conditions. The remaining area below the 25th percentile would be bisected. The sites falling within the upper half of the bisection would be described as “fair” while those falling in the lower half of the bisection would be described as “poor”. This scoring method is the same as used for the Arizona warmwater index (Gerritsen and Leppo, 1998).

Table 7. Proposed narrative categories assigned to Arizona coldwater index scores.

| Index Score | Comparison to Reference | Narrative Description |
|--------------------|---|------------------------------|
| 60.3 - 100 | above | very good |
| 50.8 – 60.3 | between the 75 th and 25 th percentiles | good |
| 40.5 – 50.8 | upper half of bisection below 25 th percentile | fair |
| 0 – 40.5 | lower half of bisection below 25 th percentile | poor |

4 RECOMMENDATIONS

The working coldwater index for Arizona streams appears to be robust enough for statewide assessment of stream condition in coldwater streams. Further refinement maybe necessary as the status of reference sites are reviewed and as more stressed sites are identified. Narrative categories may also be refined as more stressed sites are sampled.

4.1 Future Sampling

Arizona Department of Environmental Quality (DEQ) should continue to sample a small set of reference sites for monitoring trends and seasonal variation. However, the majority of sampling effort should be directed at sampling independent sites over a wide range of impairment conditions and sources of stressors, from reference sites to severely impaired.

As new data are collected, it is advisable to refine the index, by testing the index with a new, independent data set (say, a year=s sampling). This will allow Arizona DEQ to examine the response of the index and its component metrics to defined stressors; and to reexamine site classification (e.g., by ecoregions, elevation, watershed area) to determine if additional classes (e.g., ecoregions) or covariates (e.g., area) would improve performance of the index. We did not examine classification of reference sites in this project but it is often found that refinements are worthwhile in a large geographic area, such as Arizona. The refinements become possible when enough sites have been monitored that patterns begin to emerge. We caution that the mere existence of patterns is not sufficient: the patterns should be relevant to the performance of the index in discriminating impairment.

LITERATURE CITED

- Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M.L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *J. N. Am. Benthol. Soc.* 15(2):185-211.
- Fore, L.S., J.R. Karr, and R.W. Wisseman. 1996. Assessing invertebrate responses to human activities: Evaluating alternative approaches. *J. N. Am. Benthol. Soc.* 15:212-231.
- Gerritsen, J., and E.W. Leppo. 1998. Development and testing of a biological index for warmwater streams of Arizona. Prepared by Tetra Tech, Inc., Owings Mills, MD, for Arizona Department of Environmental Quality, Water Quality Division, Phoenix, Arizona.
- Gibson, G.A., M.T. Barbour, J.B. Stribling, J. Gerritsen, and J.R. Karr. 1996. Biological criteria: Technical guidance for streams and rivers. EPA/822-B-94-001. U.S. Environmental Protection Agency (US EPA), Office of Science and Technology, Washington, D.C.
- Hilsenhoff, W.L. 1982. Using a Biotic Index to Evaluate Water Quality in Streams. Wisconsin Department of Natural Resources, Madison, Wisconsin. Technical Bulletin No. 132.
- Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577-586.
- Karr, J.R. 1991. Biological integrity: A long-neglected aspect of water resource management. *Ecol. Applic.* 1:66-84.
- Karr, J.R. and E.W. Chu. 1997. Biological monitoring: Essential foundation for ecological risk assessment. *Hum. Ecol. Risk Assess.* 3:993-1004.
- Karr, J.R., and D.R. Dudley. 1981. Ecological perspectives on water quality goals. *Environ. Manage.* 5:55-68.
- Kerans, B.L., and J.R. Karr. 1994. Development and testing of a benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecol. Applic.* 4(4):768-785.
- Merritt, R.W., and K.W. Cummins, editors. 1994. *An Introduction to the Aquatic Insects of North America*, 2nd edition. Kendall/Hunt Publishing Company, Dubuque, Iowa.
- Meyerhoff, R.D. and P.H. Spindler. 1994. Biological sampling protocols: Reference site selection and sampling methods. Arizona Department of Environmental Quality, Phoenix, Arizona.

Ohio Environmental Protection Agency (EPA). 1989. Addendum to biological criteria for the protection of aquatic life, volume II: users manual for biological field assessment of Ohio surface water. Ohio EPA, Division of Water Quality Planning and Assessment, Ecological Assessment Section, Columbus, Ohio.

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

Southerland, M.T., and J.B. Stribling. 1995. Status of biological criteria development and implementation. Pages 81-96 *in* W.S. Davis and T.P. Simon (editors). Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Spindler, P.H. In press. A classification scheme for macroinvertebrate community distribution among Arizona reference sites, 1992-1994. Arizona Department of Environmental Quality, Phoenix, Arizona.

U.S. Environmental Protection Agency (US EPA). 1990 (DRAFT). Freshwater Macroinvertebrate Species List Including Tolerance Values and Functional Feeding Group Designations for Use in Rapid Bioassessment Protocols. EA Report No. 11075.05. Prepared by EA Engineering, Science, and Technology for US EPA, Office of Water, Washington, D.C.

APPENDIX A

PERFORMANCE OF CANDIDATE METRICS (Spring Riffle, Split-level taxonomy)