Aquatic Life Use Attainment

Methodology to Determine Use Attainment
for
Rivers and Streams

Policy Statement 10-1

Colorado Department of Public Health and Environment
Water Quality Control Commission
4300 Cherry Creek Dr South
Denver Colorado 80246-1530

Approved: October 12, 2010
Expires: August 31, 2017

This page intentionally left blank

Table of Contents

I. INTRODUCTION 1

II. COLORADO'S AQUATIC LIFE USES 2

III. BACKGROUND 2

IV. BIOASSESSMENT AND BIOLOGICAL THRESHOLDS 3

V. METHODOLOGY FOR DETERMINING CURRENT BIOLOGICAL CONDITION 5

VI. HISTORY OF USE ATTAINMENT DECISIONS BASED ON BIOLOGICAL INFORMATION 5

VII. ASSOCIATION BETWEEN AQUATIC LIFE POLICY AND OTHER WQCC ACTIONS……7

VIII. AQUATIC LIFE USE ATTAINMENT: BASIC CONCEPTS 8

X. AQUATIC LIFE USE ATTAINMENT: NUMERICAL THRESHOLDS 9

X. REFERENCES……14

XI. APPENDICES

A. MMI FINAL REPORT - RECALIBRATION OF THE MACROINVERTEBRATE MULTI-METRIC INDEX FOR COLORADO

B. STANDARD OPERATING PROCEDURES FOR THE COLLECTION AND PRESERVATION OF BENTHIC MACROINVERTEBRATES

C. STANDARD OPERATING PROCEDURES FOR THE IDENTIFICATION AND ENUMERATION OF BENTHIC MACROINVERTEBRATES

D. METHODOLOGY FOR DETERMINING BIOLOGICAL CONDITION

E. THRESHOLD DEVELOPMENT - TECHNICAL UNDERPINNINGS

List of Tables

Table 1 Aquatic Life Use Thresholds 10
Table 2 Auxiliary Metric Thresholds for Class 1 Waters with MMI Scores Between the Attainment and Impairment Thresholds.....12
Table 3 Criteria for Protection of High Scoring Waters 12

List of Figures

Figure 1 Flowchart: Comparison with Aquatic Life Thresholds 11
Figure 2 Flowchart: Protection of High Scoring Waters.....13

List of Acronyms and Abbreviations

AAH	Administrative Action Hearing
Commission	Colorado Water Quality Control Commission
CDOW	Colorado Division of Wildlife
CWQCA	Colorado Water Quality Control Act
CWA	Clean Water Act
Division	Colorado Water Quality Control Division
EPA	U.S. Environmental Protection Agency
EPT Taxa	The number of Ephemeroptera, Plecoptera and Trichoptera genera represented in a macroinvertebrate sample. These genera are generally very sensitive to pollution.
MMI	Multi-metric Index
WQCC	Water Quality Control Commission
WQCD	Water Quality Control Division

Definitions

Aquatic Assemblage means a group of interacting species in a given water body, united by habitat and ecology. For example, "fish assemblage" or "benthic macroinvertebrate assemblage."

Benthic Macroinvertebrates (also macrobenthos) means the animals (without backbones) living on or in the streambed. These animals generally are large enough to be seen by the unaided eye; the size of the smallest organisms in this community is determined operationally by what can be retained by a U.S. Standard No. 30 sieve (28 openings/inch, 0.595-mm openings).

Biological Assessment (or bioassessment) means an evaluation of the biological condition of a water body using biological survey data and other direct measurements of resident biota in surface waters.

Biological Thresholds means numeric values or narrative expressions that describe the expected biological quality of aquatic communities in waters of a given classified aquatic life use. Biological thresholds are based on direct measurement of the aquatic community, including its structure and functions.

Macroinvertebrates see benthic macroinvertebrate.

Metric means a calculated term or enumeration representing some aspect of biological assemblage, function, or other measurable aspect. A multimetric approach combines metrics to provide an integrative assessment of the status of aquatic resources.

AQUATIC LIFE USE ATTAINMENT
METHODOLOGY TO DETERMINE USE ATTAINMENT
FOR RIVERS AND STREAMS

I. INTRODUCTION

Direct measurement and characterization of the health and well being of the aquatic community in surface water are critical to determining how effectively Colorado is achieving the ultimate goals of the Colorado Water Quality Control Act ("CWQCA") and the federal Clean Water Act ("CWA"), which are to restore, maintain, protect, and improve the quality of our water resources.[1] To this end, one of the specific policy goals of the CWQCA is the "protection and propagation of wildlife and aquatic life.[2]" The federal CWA similarly speaks to the importance of healthy aquatic ecosystems in its often-quoted Congressional declaration, which sets forth the CWA's key objective - to "restore and maintain the chemical, physical and *biological integrity* of the Nations waters.[3]" These legislative policy statements highlight the importance of having tools and techniques available to evaluate the current aquatic-life conditions in our surface waters, and to set goals for the future.

This policy document provides the Colorado Water Quality Control Commission's ("Commission") methodology for determining whether the aquatic life use is attained in rivers and streams. The procedures detailed in this guidance rely upon direct measurement of the aquatic life use rather than on comparing existing water quality to numeric standards for individual pollutants (e.g. copper, zinc, ammonia) set to protect aquatic life.

The Commission's approach to making aquatic life use-attainment decisions has evolved over time. The intent of this policy is to describe the cumulative result of regulatory actions taken by the Commission. It is envisioned that as the Commission gains experience and takes action on individual use-attainment decisions in the context of Regulation No. 93 "Colorado's Section 303(d) List of Impaired Waters and Monitoring and Evaluation List" and in considering re-classification of water bodies in rulemaking hearings, this policy document will be revised to reflect those new policy decisions.

The contents of this document have no regulatory effect, serving instead to summarize the Commission's thinking and actions in a single public document. In other words, as opposed to a rule or regulation, this policy statement has no binding effect on the Commission, the Water Quality Control Division ("Division"), or the regulated community. Moreover, this policy is not intended, and should not be interpreted, to limit any options that may be considered or adopted by the Commission in future rulemaking proceedings. Therefore, this policy statement can, and will, be modified over time as warranted by future rulemaking decisions.

---

[1] *See* Colorado Water Quality Control Act, § 25-8-102(2), C.R.S.; Federal Clean Water Act, § 101(a), 33 U.S.C. § 1251(a).
[2] Colorado Water Quality Control Act at section 25-8-101(2) and Clean Water Act at section 101(a) (2).
[3] Federal Clean Water Act at section 101(a), emphasis added.

II. COLORADO'S AQUATIC LIFE USES

Colorado classifies water bodies for beneficial uses that, along with chemical and physical standards, assure their protection. Along with water supply, recreation and agriculture, Colorado classifies waters for "aquatic life" use. Colorado's aquatic life use classifications are assigned to individual water body segments based upon the potential to support that use according to a narrative description in Regulation No. 31 "Basic Standards and Methodologies for Surface Water" at section 31.13:

> **31.13 STATE USE CLASSIFICATIONS**
>
> Waters are classified according to the uses for which they are presently suitable or intended to become suitable. In addition to the classifications, one or more of the qualifying designations described in section 31.13(2), may be appended. Classifications may be established for any state surface waters, except that water in ditches and other manmade conveyance structures shall not be classified.
>
> (1) Classifications
>
> . . .
>
> (c) Aquatic Life These surface waters presently support aquatic life uses as described below, or such uses may reasonably be expected in the future due to the suitability of present conditions, or the waters are intended to become suitable for such uses as a goal:
>
> (i) Class I - Cold Water Aquatic Life These are waters that (1) currently are capable of sustaining a wide variety of cold water biota, including sensitive species, or (2) could sustain such biota but for correctable water quality conditions. Waters shall be considered capable of sustaining such biota where physical habitat, water flows or levels, and water quality conditions result in no substantial impairment of the abundance and diversity of species.
>
> (ii) Class 1 - Warm Water Aquatic Life These are waters that (1) currently are capable of sustaining a wide variety of warm water biota, including sensitive species, or (2) could sustain such biota but for correctable water quality conditions. Waters shall be considered capable of sustaining such biota where physical habitat, water flows or levels, and water quality conditions result in no substantial impairment of the abundance and diversity of specifies.
>
> (iii) Class 2- Cold and Warm Water Aquatic Life These are waters that are not capable of sustaining a wide variety of cold or warm water biota, including sensitive species, due to physical habitat, water flows or levels, or uncorrectable water quality conditions that result in substantial impairment of the abundance and diversity of species.

III. BACKGROUND

Historically, Colorado's aquatic life use classifications have sometimes been difficult to determine for a given water body. The broad categories and the generalized terms in the

classification descriptions have led to disagreements and sometimes contentious hearings about the appropriate classification of specific water bodies.

The current effort to refine the aquatic life use classifications and provide a more consistent basis for determining the condition of aquatic life arose from specific issues in Colorado, as well as an increased national focus on biological thresholds. In the late 1990's, a contentious hearing about the proper classification of the Lower Arkansas raised awareness of problems with the existing descriptions of the aquatic life use classifications. This coincided with EPA encouragement for States to incorporate biological assessments and biological thresholds as major components of the standards and monitoring programs. In June 2000, the Division established the Aquatic Life Work Group to develop biological assessment methods and biological thresholds, and propose refinements of the aquatic life use classifications.

In 2002, an amendment of the CWQCA (§ 25-8-309) directed the Division to study (among other things) the need for refined aquatic life use classifications and a better method with a uniform approach to determine whether a water body had attained its aquatic life use. These subjects were explored by the Aquatic Life Work Group. The result of the "309" study was to flesh out the expected condition concept and identify a "Strawman" classification system which contained nine primary aquatic life use categories and 27 possible subcategories and called for voluntary pilot studies to characterize the "expected condition" for candidate sub-categories. However, the pilot studies did not proceed as intended due to a lack of interest, funding, and study sites.

In November of 2004, the Commission reviewed the Aquatic Life Work Group's "expected condition concept" during the Basic Standards Issues Formulation Hearing. At that time, the Commission decided not to incorporate the expected condition concept in the 2005 Basic Standards and Methodologies for Surface Water (Regulation #31) Rulemaking Hearing. Instead, the Commission tasked the Aquatic Life Work Group with strengthening the underlying biological data from which the Strawman's nine principle use classification categories and sub-classifications would be defined. At that time the Commission targeted the 2010 Rulemaking Hearing for Basic Standards for the next Strawman iteration. Coincidentally, the Division began to work with EPA and their contractors to use Colorado's data to develop bioassessment tools. This effort is described below in Section IV.

Rather than restructuring the existing classification system with potentially numerous ramifications, the Division decided to proceed with the more modest step of utilizing bioassessment tools and data to refine how existing classifications are interpreted and applied. This policy represents the first official step to add more specificity to Colorado's aquatic life use classifications.

IV. BIOASSESSMENT AND BIOLOGICAL THRESHOLDS

Bioassessment is a quantitative measure of the biological condition of the resident aquatic community. A biological threshold is the establishment of numeric criteria against which the current biological condition can be evaluated. Each of these regulatory concepts – biological assessment and biological thresholds – is defined briefly, and then discussed in more detail in Section IV subsection A.

Biological Assessment (or bioassessment) means an evaluation of the biological condition of a water body using biological survey data and other direct measurements of resident biota.

Biological Thresholds means numeric values or narrative expressions that describe the expected biological quality of aquatic communities in waters of a given classified aquatic life use, and are based on direct measurement of the aquatic community, its structure and functions.

A. Bioassessment Tools

Aquatic communities are an association of interacting populations of aquatic organisms in a given waterbody or habitat (EPA 1991). They provide valuable information regarding the overall biological integrity of a given waterbody. These organisms are capable of integrating the effects of stressors over time; thus, they can provide a broad ecological measure of fluctuating environmental conditions (EPA 1999). Bioassessments provide the means to evaluate biological condition in relation to the goal or expectation, such as designated use support.

Several biological assemblages are available for study including periphyton (attached algae), benthic macroinvertebrates and fish, each of which responds somewhat differently to the suite of anthropogenic stressors. The decision to rely exclusively on the benthic macroinvertebrates assemblage was based on community characteristics and practical considerations. The combination of relatively long life spans, limited mobility, representation in most Colorado habitats, and ease of collection make macroinvertebrates the best single assemblage for bioassessment.

Within the benthic macroinvertebrate assemblage, metrics are selected that represent some measureable aspect of the community structure and function. These measurements are grouped into metric categories such as taxa richness, composition, pollution tolerance, functional feeding groups, and habit. Combining metrics from these categories into a multi-metric index transforms taxonomic identifications and individual counts into a unitless score that ranges from 0-100.

Choosing metrics that are sensitive to stressors yields an index that defines the health, or the biological condition, of the aquatic community. A bioassessment tool of this kind is a powerful basis for a direct assessment of aquatic communities that can be used to reach conclusions about attainment of the aquatic life use.

B. Colorado's Multi-Metric Index (MMI) Bioassessment Tool

The Division, with assistance from EPA and its contractors, has developed a multi-metric bioassessment tool for Colorado that is composed of separate indices calibrated to respond to stressors affecting aquatic communities in one of the three analytically defined biotypes[4].

[4] A biotype is an aggregation of macrobenthos sites that have similar community composition. Environmental attributes (elevation, stream slope and ecoregion) common to sites within each of the biotypes can be used to predict membership of a new site.

Multi-metric indices were then calibrated for each biotype. Each index is composed of several metrics selected to represent categories of community characteristics including richness, composition, functional feeding group, mode of locomotion, and pollution tolerance. Metrics were chosen on the basis of their ability to discriminate between reference and stressed sites, represent multiple metric categories, are ecologically meaningful, and are not redundant with other metrics in the index. See Appendix A for details.

Because of this metric selection process, the metrics used in the MMI vary in their ability to detect or diagnose specific types of stress. The MMI is designed to detect environmental stresses that result in alteration of the biological community. No specific stressors are identified because the intent is to have a generalized tool that responds to a wide range of potential stressors. In other words, the MMI tool cannot determine if the stressor is a specific pollutant, pollution or habitat limitation (including flow). Once impairment is identified, however, other tools are available to identify the likely cause of impairment.

V. METHODOLOGY FOR DETERMINING CURRENT BIOLOGICAL CONDITION

Determining biological condition involves the steps of calculating the MMI for the appropriate biotype and comparing the MMI score to a threshold.

To calculate an MMI a representative macroinvertebrate sample is collected from the appropriate habitat(s). Standardized procedures for collecting macroinvertebrates are in Appendix B. Field sampling is typically followed by preservation, laboratory identification and enumeration, and entry into a biological database capable of calculating the MMI. See Appendices C and D for details.

VI. HISTORY OF USE ATTAINMENT DECISIONS BASED ON BIOLOGICAL INFORMATION

Direct measurement and characterization of the health and well being of the aquatic community in surface water is critical to determine whether or not the aquatic life use is attained at a given site. Historically, Colorado has used biological information to make attainment decisions, even though no specific methods of assessment or thresholds were set. This section describes the history of such decisions.

A. Impaired Waters Listing Decisions

In 1998, the Colorado List of Impaired Waters included three segments found to be impaired for aquatic life. The associated listing methodology explained the protocol for reaching those decisions:

> Biological assessments by the Colorado Division of Wildlife (CDOW) were utilized in developing the List. These consist of fish surveys performed by CDOW staff using both seining and electrofishing. The results of these assessments were compared with the Standards and Classification System in the following manner. For segments that are

> designated as Aquatic Life Class 1, evidence of a decline over time from a healthy and diverse fish community or the absence of a Species of Critical Concern[5] constitutes an impairment of the use. For segments that are designated as Aquatic Life Class 2, evidence of significant reduction of the species composition of a fish community over time constitutes an impairment of the use. The Division limited the time frame for comparison of fish communities as shown by fish surveys, to only the late 1970's (when aquatic life classes were established) through more recent conditions.

The requirement for a Section 303(d) List in 2000 was suspended by EPA, and no List of Impaired Waters was compiled. The 2002 303(d) List included six waters with an impaired aquatic life use. Decisions about impairment were based on the following narrative that appeared in the 2002 Listing Methodology:

> B. IMPAIRMENT OF NARRATIVE STANDARDS AND CLASSIFIED USES
>
> Impairment of narrative standards and classified uses may be supported by chemical data and/or information generated by biological and/or physical assessments. In instances where a determination of impairment is based solely upon biological and/or physical assessments, such assessments must provide clear and convincing evidence of non-attainment. For aquatic life uses, as previously referenced, the WQCD will generally consider impairment of narrative standards and classified uses to be demonstrated when either the physical/habitat data or biological community metrics reflect a condition that is significantly less than reference condition. When such data do not indicate specific pollutant(s) causing non-attainment, the WQCD may place the segment on the M&E List for further study.

In 2004, the Section 303(d) List was promulgated as Regulation No. 93, and the Monitoring and Evaluation List became Regulation No. 94. As in 2002, a separate Listing Methodology document was developed. The 2004 Section 303(d) List included six waters with an impaired aquatic life use. The 2004 Listing Methodology section regarding decisions of impairment based on biological information is essentially the same as 2002, but with the addition of the concept of "expected condition," as provided below.

> 3. Impairment of Narrative Standards and Classified Uses
>
> Impairment of narrative standards and classified uses may be supported by chemical data and/or information generated by biological and/or physical assessments. In instances where a determination of impairment is based solely upon biological and/or physical assessments, such assessments must provide clear and convincing evidence of non-attainment. For aquatic life uses, as previously referenced, the Division will generally consider impairment of narrative standards and classified uses to be demonstrated when either the physical/habitat data or biological community metrics reflect a condition that is significantly less than *the expected or* reference condition. When such data do not indicate specific pollutant(s) causing non-attainment, the Division will recommend placing the segment on the M&E List for further study. (*emphasis added*).

[5] Species of Critical Concern includes native fish species observed to be in decline and rare in abundance or limited in distribution (as identified by CDOW in the Inventory and Status of South Platte River Native Fishes in Colorado, CDOW, 1997).

The 2006 Section 303(d) List included three segments determined to be impaired for the aquatic life use. The 2006 Listing Methodology considerations were the same as 2004. In 2008, the structure of this portion of the Listing Methodology was modified; however, the level of proof (clear and convincing) and decline from reference condition (significantly less than) remained unchanged. The 2008 Section 303(d) List contained four segments with aquatic life use impairment (not including Fish Consumption Advisories). The 2008 Listing Methodology explained as follows:

> 3. Impairment of Narrative Standards and Classified Uses
>
> Impairment of narrative standards and classified uses may be supported by chemical data and/or information generated by biological and/or physical assessments. In instances where a determination of impairment is based solely upon biological and/or physical assessments, such assessments must provide clear and convincing evidence of non-attainment.
>
> a. Aquatic Life Use: For aquatic life uses, as previously referenced, the Division will generally consider impairment of narrative standards and classified uses to be demonstrated when either the physical/habitat data or biological community metrics reflect a condition that is significantly less than the expected or reference condition.

The 2010 Section 303(d) List contained three segments for aquatic life use impairment (not including Fish Consumption Advisories). The 2010 Listing Methodology has the same provisions as the 2008 Methodology; however, Regulation Nos. 93 and 94 were combined as Regulation No. 93.

VII. ASSOCIATION BETWEEN AQUATIC LIFE POLICY AND OTHER WQCC ACTIONS

Aquatic Life Use Policy 10-1 provides guidance for determining whether sites on rivers and streams with watersheds lesser than 2700 square miles are attaining their aquatic life use. It offers a systematic assessment method for comparing the biological condition of a test site in a given biotype, as characterized by an MMI score, against an expectation of biological condition (threshold) for a similar biotype. If biological conditions fall below expectations, the site may be deemed "impaired".

Once a determination of impairment is reached, however, there is a range of possible regulatory outcomes that may be reached. One outcome is inclusion in the biennial Regulation No. 93 "Colorado's Section 303(d) List of Impaired Waters and Monitoring and Evaluation List". Another is through the consideration of re-classification of water bodies in rulemaking hearings. These Commission actions are guided by their own policy considerations, which should be consulted separately as the need arises.

It is envisioned that as the Commission gains experience and takes action on individual use-attainment decisions in these or other contexts, this policy document will be revised to reflect those evolving policy decisions. Further, if major changes are made in related policy documents, (e.g., the section regarding "Impairments where the Pollutant is Unknown" in the Section 303(d)

Listing Methodology), those policy changes may necessitate changes to Aquatic Life Use Policy 10-1.

VIII. AQUATIC LIFE USE ATTAINMENT: BASIC CONCEPTS

In order to protect and maintain Colorado's aquatic life, it is the policy of the Commission to apply a defined method and a uniform approach to determine whether a water body has attained its aquatic life use classification. The method and approach are based on the following tenets:

1. Quantitative Assessment of the Biological Condition**:** The Commission has determined that the Colorado Multi-Metric Index is an appropriate tool for quantitative bioassessment because:

- The condition of the macroinvertebrate community is an appropriate surrogate for the entire aquatic community.
- The macroinvertebrate community is an appropriate indicator of stream quality because of limited migration patterns and short life spans. As such, they are capable of integrating the effects of physical and chemical stressors over short time periods (weeks to months).
- The Multi-Metric Index tool was developed exclusively with macroinvertebrate data collected in Colorado.
- Supporting data represent the variability expected within the state and accounts for the broad environmental gradients through spatial classification.

2. Comparison with Expected Condition: The Commission continues to support the use of "reference condition" or "expected condition" as the basis for characterizing use support. It is important to note that this concept of use support embraces considerable variation in the biological community. This variability must be acknowledged in developing the biological thresholds and implementation approach.

3. Impairment is a Significant Departure from Expected Condition: At this time, the Commission affirms the position taken in prior decisions made in the context of the Section 303(d) Listing Methodology -- that clear and convincing evidence is needed to show impairment, and that the status of non-attainment represents a significant departure from reference or expected condition.

4. Recognition of Warm / Cold Distinction: The Commission believes it is important to preserve the existing distinction between warm-water and cold-water biota that is set forth in the Basic Standards. Accordingly, the assessment methodology should accommodate the different thermal tolerances that are expected within each subclass.

5. Recognition of Class 1 / Class 2 Distinction: The Commission believes it is important to preserve the existing distinction between Class 1 and Class 2 aquatic life classifications that is set forth in the Basic Standards, i.e. that Class 1 waters are to be capable of sustaining a wide

variety of biota including sensitive species.[6] Accordingly, showing attainment of the Class 1 aquatic life use may require additional demonstration that diversity and sensitivity characteristics are maintained.

6. Statistical Methods: Consistent with CWQCA at section 25-8-204(5), the Commission requires that statistical methodologies be based on assumptions that are compatible with the water quality data. Application of those methodologies should be transparent with respect to uncertainty and risk of mistaken conclusions.

7. Protection Against Large Decline: Natural variability in biological condition at reference sites leads to thresholds that could inadvertently expose high scoring sites to the risk of undetected degradation. It is the Commission's policy to provide additional protection by defining an allowable change in biological condition that is independent of biological thresholds.

IX. AQUATIC LIFE USE ATTAINMENT: NUMERICAL THRESHOLDS

It is the Commission's intent that there should be predictable, transparent, and understandable techniques for evaluating the aquatic life use, and that the methods for assessing attainment of the aquatic life use should be consistent with the classifications and standards system previously adopted by the Commission.

A. Use Attainment Thresholds for Streams and Small Rivers

For Colorado's streams and rivers with a watershed area less than 2,700 mi^2, the Commission has established aquatic life use attainment thresholds that were derived using a statistical methodology that relies on a normal operating range to define use support and interval and equivalence tests.

1. Quantitative Assessment of Biological Condition: Implementation of this policy relies upon use of the Colorado MMI (version 3, 2010, see Appendix A). The following documents describe the procedures for collecting data and determining MMI scores:

- Standard Operating Procedures for the Collection and Preservation of Benthic Macroinvertebrates (Appendix B)
- Standard Operating Procedures for the Identification and Enumeration of Benthic Macroinvertebrates (Appendix C)
- Methodology for Determining Biological Condition (Appendix D)

2. Recognition of Warm / Cold Distinction: Use of the predictor variables for establishing biotype membership provides appropriate refined recognition of the thermal tolerances that are expected within the warm water and cold water biota subclasses. Sites are to be compared against the attainment and impairment thresholds for their predicted biotype.

[6] Regulation No. 31, section 31.13(1)(c)(i).

3. Comparison with Aquatic Life Thresholds: Thresholds are established based on analysis of the biological condition at reference sites in each of three biotypes. Figure 1 provides a flowchart of the comparison process. Derivation of thresholds is described in Appendix E.

Table 1. Aquatic Life Use Thresholds

Biotype		Attainment Threshold	Impairment Threshold
1	Transition	52	42
2	Mountains	50	42
3	Plains & Xeric	37	22

- Class 2 water bodies: When the MMI score falls within the gray zone (between the attainment and impairment thresholds discussed above) and the site is from a Class 2 stream, the stream is considered to be in support of the use.
- Class 1 water bodies: An MMI score in the gray zone for site on a Class 1 water requires examination of additional information – see Subsection 4 below.

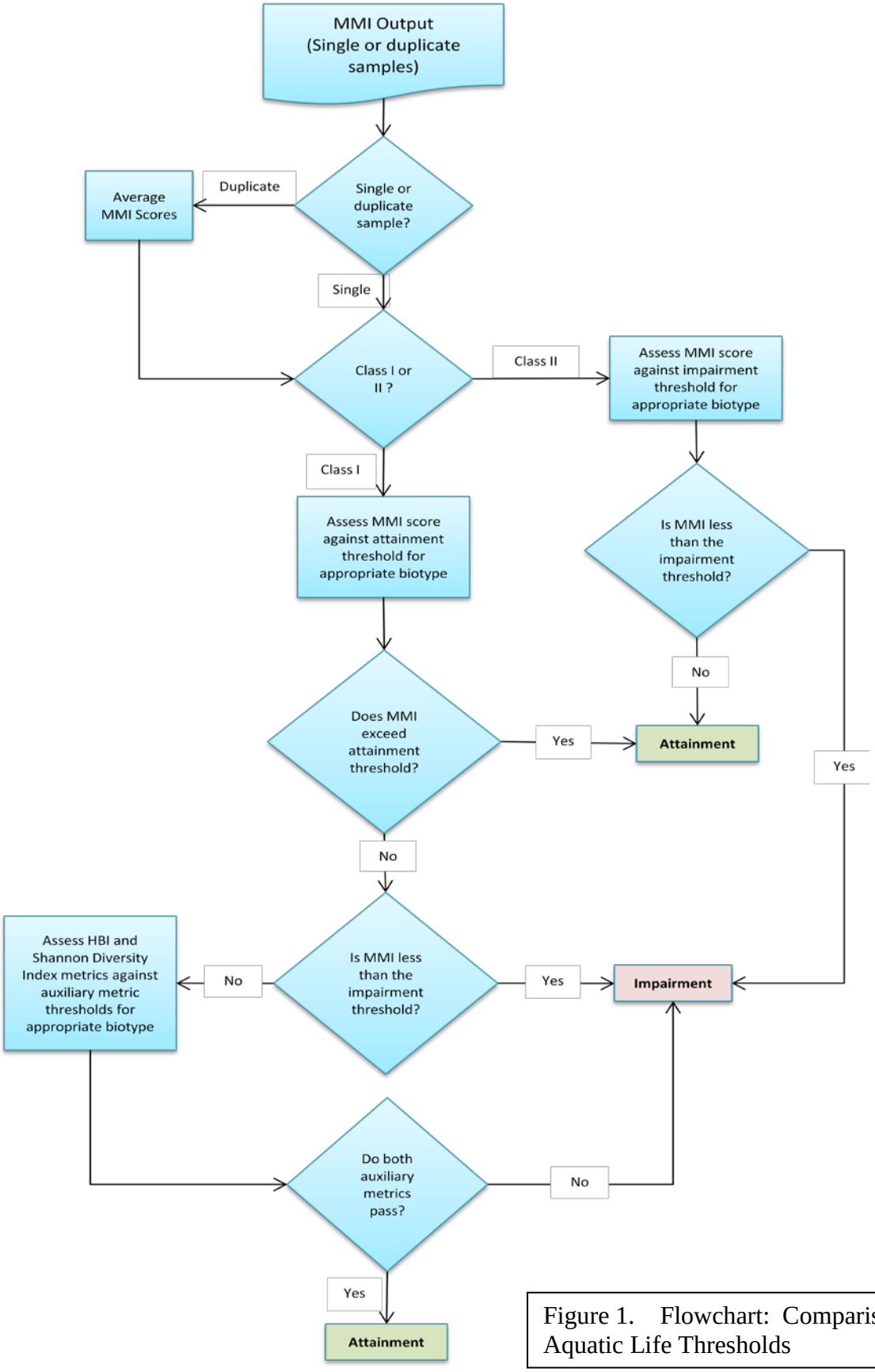


Figure 1. Flowchart: Comparison with Aquatic Life Thresholds

4. Recognition of Class 1 / Class 2 Distinction: The Hilsenhoff Biotic Index and the Shannon Diversity Index are used as auxiliary metrics that supplement the MMI for Class 1 waters with MMI scores between the attainment and impairment thresholds. If a Class 1 site fails to meet the criteria shown in Table 2 for either auxiliary metric, the site will be considered impaired. For Class 2 waters, the premise that these "waters are not capable of sustaining a wide variety of cold or warm biota, including sensitive species" mitigates the need to further confirm attainment or impairment through use of auxiliary metrics. Accordingly, for Class 2 waters, a single threshold provides the basic determinant between attainment and impairment (see Table 1). Auxiliary metrics are not applicable to Class 2 waters. Derivation of Auxiliary Metric thresholds is described in Appendix E.

Table 2. Auxiliary Metric Thresholds For Class 1 Waters with MMI Scores Between the Attainment and Impairment Thresholds

Biotype		Hilsenhoff Biotic Index	Shannon Diversity Index
1	Transition	<5.4	>2.4
2	Mountains	<5.1	>3.0
3	Plains & Xeric	<7.7	>2.5

5. Protection Against a Large Decline: Thresholds are established for a decline in biological condition that is acceptable over time. Samples must be representative of normal conditions and be separated by more than 12 months. Conditions under which factors of representativeness apply may be found in Section IX, subsection C of this document. Table 3 presents the thresholds for "high scoring waters[7]" for each biotype. A 22 point decline in the MMI score results in a conclusion of impairment. Figure 2 provides a flowchart for determining protection of high scoring waters. Derivation of High Scoring Waters thresholds is described in Appendix E.

Table 3. Criteria for Protection of High Scoring Waters [7]

Biotype		High Scoring Water (MMI score)	Allowable MMI decline
1	Transition	>64	-22
2	Mountains	>64	-22
3	Plains & Xeric	>44	-22

[7] In concept, this threshold applies to all sites; in practice, it affects only high-scoring sites since a large decline for a moderate scoring site would bring it below the attainment threshold. The MMI threshold that constitutes a high score is indicated in Table 3.

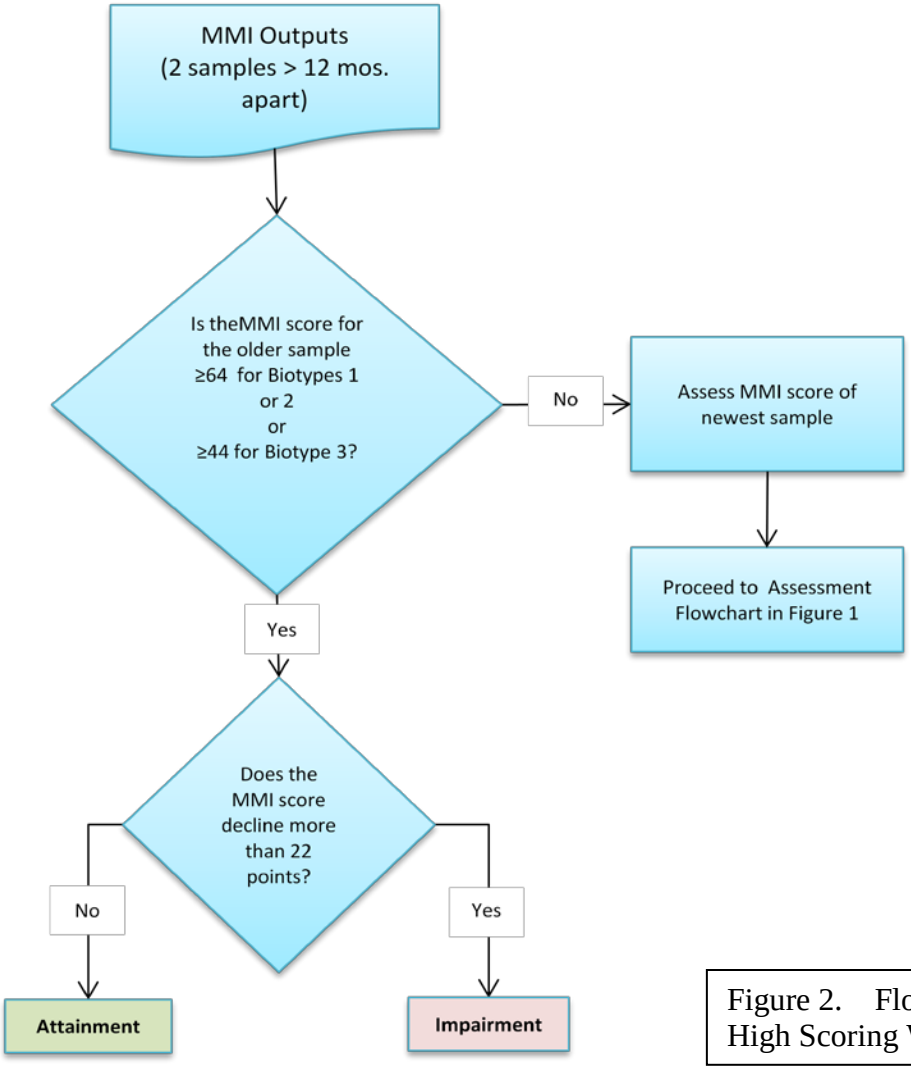


Figure 2. Flowchart: Protection of High Scoring Waters

B. Use Attainment Thresholds for Bigger Rivers: *RESERVED*

[The Commission will not apply the 2010 MMI-based Impairment Thresholds to Big Rivers (watershed areas >2700 mi^2) as a single indicator.]

C. Considerations of Representative Data

The 2010 Section 303(d) Listing Methodology contains appropriate coverage of the representative data issue: It clearly provides: "Data that are not representative of normal conditions shall typically be discounted in making the listing decision." The 2010 Listing Methodology explains "representativeness" as follows: "[T]he Monitoring and Evaluation List identifies water bodies where there is reason to suspect water quality problems, but there is also uncertainty regarding one or more factors, such as the representative nature of the data."

1. The representativeness of an individual sample may be affected by antecedent physical or chemical events within the generation time of the macroinvertebrate community. Events might include scouring flood flows or accidental spills of toxic chemicals.

2. The representativeness of an individual sample may be inferred from spatial or temporal patterns where multiple samples exist. For example, an anomalously low MMI score at one site in a spatial group or on one date in a time series might be evidence that a site is not representative. However, there must be no evidence of perturbation that could explain the low score.

3. The length of stream that a single sample can fairly represent depends on the site-specific considerations of the location. Generally, the length would be limited to a portion that was hydrologically consistent with the sample location (i.e., no intervening dams, diversions, tributaries or discharges).

D. Alternative Thresholds

In order to set an alternative threshold, a party would need to establish that the situation is "natural or man-induced irreversible" (i.e., there are no sources to remediate) and determine the central tendency of the MMI scores.

X. REFERENCES

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

U.S. Environmental Protection Agency. 1991. Biological Criteria: State Development and Implementation Efforts. EPA 440-5-91-003. Office of Water; Washington, D.C.

Colorado Division of Wildlife, 1997. Inventory and Status of South Platte River Native Fishes in Colorado.

APPENDIX A

MMI FINAL REPORT

Recalibration
of the
Macroinvertebrate Multi-Metric Index
for
Colorado

Prepared by:

Ben Jessup
Tetra Tech, Inc.
Center for Ecological Sciences
15 State Street, Suite 301
Montpelier, VT 05602

Prepared for:

Colorado Department of Public Health and Environment
Water Quality Control Division – Monitoring Unit
Denver, CO

and

EPA Region 8

Final Report
January, 2010

This page intentionally blank.

Abstract

The Colorado Department of Public Health and Environment (CDPHE) monitors streams throughout the state for assessment of water resource quality. The benthic macroinvertebrate assemblage can be associated with known gradients of natural conditions and of human disturbance. Association of biological communities with natural gradients allows classification by expected community characteristics, which should be recognized before trying to find relationships between biology and the disturbance gradient. Site classes were conceptualized by grouping sites with similar taxonomic composition and identifying distinctive environmental characteristics of the groups. Cluster analysis and ordination were used to define three site classes. Predicting site class membership was accomplished by developing classification rules with three environmental variables: ecoregion, site elevation, and stream slope. Within each site class, responses of 145 biological measures (metrics) were tested to find those that were indicative of differences among sites with minimal degradation (reference) and those with greater degrees of degradation (stressed). Several mathematical combinations of metrics were evaluated to find indices that were responsive to stress, ecologically meaningful, and both diverse and non-redundant in metric information. Three indices were identified, because metrics were not equally responsive to stress in each site class. The recommended MMIs apply to three specific site classes, contain 5-6 metrics each, and discriminate reference from stressed sites with minimal error (overall Discrimination Efficiency = 96.7%).

This page intentionally blank

Table of Contents

Abstract i
Table of Contents iii
List of Tables iv
List of Figures v
1 Introduction 1
2 Data Description 3
3 Analytical Methods 6
3.1 Site Classification 6
Biological Groupings 7
Discerning Environmental Factors 8
Predicting Site Classes 10
3.2 Metric Calculations 11
Metric Calculation and Evaluation 11
Indicator taxa 12
Metric adjustment to environmental conditions 12
3.3 Index Composition 13
4 Results 14
4.1 Site Classification 14
4.2 Metric Development 19
4.3 Index Performance 20
4.4 Index Application 23
6 Conclusions and Recommendations 23
7 References cited 24

Appendices
A Reference and Stressed Site Selection Process
B Reference and Stressed Sites Used in Analysis
C Metric Discrimination Efficiencies
D Index Alternatives

List of Tables

Table 1. Reference and stressed site tally by data source. 4
Table 2. Data treatments for consistency in analysis. 5
Table 3. Predictor variables listed categorically. This list does not include all the possible variations of each variable, which included statistical summaries by watershed or at the sampling point. 9
Table 4. Site designations in three clusters. 14
Table 5. Classification rules. 18
Table 6. Indicator taxa for the low elevation sites (groups 1 and 3).. 19
Table 7. Metrics included in the Biotype 1 MMI with adjustment and scoring formulae. 20
Table 8. Metrics included in the Biotype 2 MMI with scoring formulae. 22
Table 9. Metrics included in the Biotype 3 MMI with scoring formulae. 22
Table 10. MMI performance statistics 23

List of Figures

Figure 1. Colorado map showing sample sites within ecoregions. Some stressed sites are obscured. Ecoregions are as follows: 18 = Wyoming Basin, 20 = Colorado Plateaus, 21 = Southern Rockies, 22 = Arizona/ New Mexico Plateau, 25 = High Plains, 26 = Southwestern Tablelands. 4
Figure 2. Typical inputs and results in a discrete classification process. Inputs include natural environmental variables and biological similarities among sites. Resulting site classes are discrete bins, where a site can be placed in only one bin..... .7
Figure 3. Illustration of metric discrimination efficiency (DE) between reference and stressed sites. 12
Figure 4. Cluster dendrogram showing reference sites grouped by similarities in the presence (and absence) of taxa in the samples..... 15
Figure 5. Distributions of EPT taxa richness in clusters. 16
Figure 6. NMS ordination of taxa presence-absence with sites marked by groups resulting from cluster analysis. 16
Figure 7. Distributions of environmental variables (elevation, precipitation, maximum air temperature, and stream slope) in the biologically-defined groups..... 17
Figure 8. Distribution of index scores in reference and stressed sites in Biotype 1..... 21
Figure 9. Distribution of index scores in reference and stressed sites in Biotype 2..... 21
Figure 10. Distribution of index scores in reference and stressed sites in Biotype 3..... 22

1 Introduction

The State of Colorado is well-known for its natural beauty, often epitomized in images of clear mountain streams. These streams and all surface waters in Colorado should support a diverse and functional fauna. Land use activities such as mining, agriculture, silviculture, industrialization, and urbanization have contributed to water pollution and have often degraded the quality of the surface waters. It is the responsibility of Colorado's Department of Public Health and Environment (CDPHE) to maintain and protect the quality of the State's waters. In keeping with the Clean Water Act of 1972 (CWA, PL-92-500, and revisions of 1977, 1987) and technical guidance from the U. S. Environmental Protection Agency (USEPA), CDPHE has developed water quality standards for the protection of human and ecosystem health.

The Clean Water Act has as one of its primary goals the maintenance and restoration of biological integrity, which incorporates biological, physical, and chemical quality. This concept of biological integrity refers to the natural assemblage of indigenous organisms that inhabit a particular area that has not been affected by human activities (Frey 1977, Karr et al. 1986). The measurable definition of biological integrity is the *reference condition* (Barbour et al. 1995, 1999), which is characterized using data from minimally-disturbed sites within a region.

> ***Biological integrity*** *is commonly defined as "the capability of supporting and maintaining a balanced, integrated, adaptive community or 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)

Biological metrics observed in the undisturbed reference sites become the standard to which biological conditions in other sites are compared. Therefore, proper identification of reference sites, characterization of the biological reference conditions, recognition of natural variability, and selection of the metrics that are most responsive to prevalent stressors are essential for making accurate comparisons and valid assessments.

Through the 303(d) and Total Maximum Daily Load (TMDL) framework outlined in the CWA, waters considered impaired and threatened must be identified and improved to meet their designated uses. Such uses almost always include aquatic life uses. *Impairment*, as defined by natural resource management or regulatory agencies in accordance with CWA provisions, is typically based on attainment or non-attainment of water quality criteria associated with a waterbody's designated use. If those criteria are not met (or attained) then the waterbody is considered to be impaired.

Resident aquatic biota in a stream can indicate environmental quality, responding to the effects of both episodic as well as cumulative pollution and habitat alteration. Thus, the assessment of ecosystem integrity is greatly enhanced through biomonitoring, or direct

survey of aquatic biota and comparison of meaningful and responsive measures of the samples to the regional ecological potential. In this exercise, we use the benthic macroinvertebrate assemblage to represent biological conditions.

The current investigation was framed in the context of previous indicator development processes, which were described by Paul et al. (2005). After applying the 2005 Multi-Metric Index (MMI), CDPHE recognized a need for recalibration of the index. The recalibration was expected to address perceived inadequacies of the 2005 indicator by incorporating new data, newly defined reference and stressed sites, and refinement of the stream classifications. The perceived inadequacies pertained to site classification, which did not address transition zones between ecoregion-defined classes.

Biomonitoring in Colorado

Colorado DPHE has established a biological monitoring program for streams throughout the State. The information collected in the biomonitoring program has been incorporated into Colorado's 305(b) report and its 303(d) list of impaired streams through application of indictors developed by Paul et al. (2005). The objective of this project was to recalibrate the multimetric biological index for interpreting the condition of Colorado streams. Recalibration is a logical improvement based on the addition of sample data since the last index development. The additional biological, chemical, and physical habitat data filled gaps noted in the spatial coverage during the previous indicator development effort. The specific questions investigated in this study were:

- Can site classification for assessing ecosystem condition across the diverse landscape and physiographic regions of Colorado be improved by considering additional data?
- Can site classification be developed so that it is sensitive to transitional areas that have characteristics of multiple site classes?
- Which metrics are most appropriate for use in a Colorado multimetric macroinvertebrate stream condition index?

Concurrent Index Development Efforts

As during the 2005 MMI calibration, a concurrent predictive model (Observed/Expected taxa) development process was performed. The two efforts were coordinated so that identical reference samples and environmental data would be considered. Otherwise, the indicator development processes were independent and we do not report on predictive model results here.

Application Considerations

The multimetric index development process was intended to recognize the highest degrees of sensitivity, accuracy, and precision in a biological indicator. This should allow CDPHE to apply the index with a high degree of confidence that they are assessing stream resources correctly. While we expect the MMI to indicate relative biological integrity, and can recommend possible application scenarios, we think the CDPHE is best qualified to decide on final index application protocols.

Index Development Process
The MMI development process includes a series of data preparation and analytical steps.

1. Collect, Compile, & QC Data
2. Identify Reference and Stressed Conditions
3. Identify Natural Variability (Site Classification)
4. Calculate Metrics
5. Find Discriminating Metrics
6. Score Metrics
7. Combine Scores in an Index
8. Apply within the development context

These steps are generally linear, though reiteration can occur when a later step prompts revisions of earlier assumptions or questioning of earlier results. Most of the effort for steps 1 and 2 were accomplished by CDPHE and results were transferred to Tetra Tech for further analysis.

2 Data Description

Colorado DPHE provided a robust dataset in the Ecological Data Application System (EDAS; an Access database) for use in this study. This dataset included benthic macroinvertebrate data. The dataset included data points obtained from other agencies. These additional datasets were of high quality and employed similar methods. The datasets used were Western EMAP and Southern Rockies Regional EMAP (US EPA Environmental Monitoring and Assessment Program), USU-STAR (Utah State University Western Center for Monitoring and Assessment of Freshwater Ecosystems Science to Achieve Results program), USFS (U.S. Forest Service), and NAWQA (US Geological Survey National Water Quality Assessment program).

Although different programs do not use the exact same methods, the protocols used by each program were relatively comparable (Paul et al. 2005). All the programs sampled in riffles, used kick or D-frame nets with similar mesh (425 - 600 µm), sampled a similar size area (0.25 – 1.0 m^2) using comparable kick methods, and identified organisms to a similar taxonomic level. The principal difference among the programs was the subsample size. Most programs enumerated the whole sample, while CDPHE identified a 300 count subsample. To standardize sample size, large samples were randomly subsampled to a 300 organism count before analysis.

Reference and Stressed Site Identification
The basis of biological assessment is comparison of conditions observed in a test site to conditions that are found in sites of the highest quality, the reference condition. CDPHE provided a list of reference sites, stressed sites, and sites that almost met the criteria for those categories. Sites that were not categorized were generally not used in index calibrations, except to illustrate final index distributions. The selection process for reference and stressed sites included remote sensing of land uses in the immediate and entire catchments of the sites. Sites were screened using Geographic Information System

(GIS) analysis and confirmed through inspection of aerial photography. The process is described in detail in **Appendix A**. As a result, 133 reference sites and 74 stressed sites were identified throughout Colorado (**Table 1, Figure 1**). In the stressed sites, 14 sites were not included because the samples had less than 150 organisms (resulting in 60 final stressed sites).

Table 1. Reference and stressed site tally by data source.

	CDPHE	WEMAP	REMAP	STAR	USFS	NAWQA
Reference	63	18	18	18	15	1
Stressed	48	5	1	0	0	6

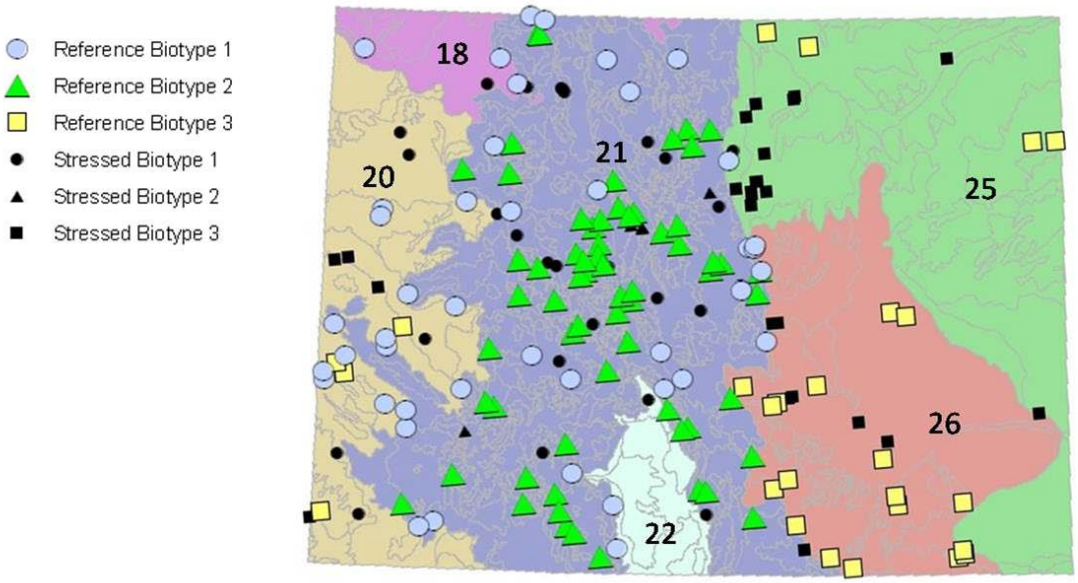


Figure 1. Colorado map showing sample sites within ecoregions. Some stressed sites are obscured. Ecoregions are as follows: 18 = Wyoming Basin, 20 = Colorado Plateaus, 21 = Southern Rockies, 22 = Arizona/New Mexico Plateau, 25 = High Plains, 26 = Southwestern Tablelands.

Data Treatments

For certain statistical analyses, the biological data were summarized, reduced, or filtered to conform to analytical requirements. These data treatments are listed in **Table 2** and described below. The final list of reference and stressed samples used in the analysis are in **Appendix B**.

Table 2. Data treatments for consistency in analysis.

Data treatment	**Analytical justification**
Elimination of rare taxa	Eliminates potential over-weighting of uncommon taxa in site classification
Condensing taxa to Operational Taxonomic Units (OTU's)	Eliminates potential confusion of identical taxa as unique, for site classification
Re-sampling samples to 300 organisms	Eliminates potential over-estimation of taxa richness in site classification and metric calculations
Elimination of samples with incomplete midge taxonomy	Standardizes taxonomy on genus level for midges
Elimination of samples with less than 150 individuals	Eliminates under-estimation of richness in index calibration
Selection of one sample per site	Reduces potential bias associated with pseudo-replication

Rare taxa can have undue influence on analyses that are used to detect biological site types. They were discounted from samples during cluster and ordination analyses only. Rare taxa were defined as those that occurred in less than 7 of the 133 reference samples.

The taxon identifications used for classification were called the operational taxonomic units (OTUs) and were the most common and confident taxonomic identification levels in the data set. Ambiguous taxa are those that are not identified to the targeted taxonomic levels because of damaged or undeveloped specimens. The site-by-taxon matrix was reduced to retain as much information as possible while excluding ambiguous taxa.

As mentioned above, the principal difference among the multiple sample collection procedures was the subsample size. Larger subsamples tend to contain more taxa than smaller subsamples. This uneven taxa representation can be misinterpreted in analyses of site classification or metric responsiveness to stress. Therefore, a standard subsample size was established as the smallest size of all the data sets, 300 organisms. All large samples were randomly reduced to a 300 organism count prior to classification or metric analysis.

The elimination of samples with incomplete midge taxonomy and less than 150 individuals were measures taken to standardize comparisons among samples. Some samples had midges (*Diptera: Chironomidae*) identified to the sub-family level. This level of identification was uncommon and the information from genus-level identifications could help in discriminating stressor effects. Therefore, the few samples with midges identified to sub-family were not used.

Samples with fewer than 150 organisms were not used because metrics and taxa lists are sensitive to such sample size. Interpretations of metrics from small sample are not always

straightforward, likely indicating stressed conditions, but not always. Since the target subsample size is 300 individuals, small samples were not used in index calibration.

Several samples were collected at some sites. However, including more than one sample from some sites and only one from other sites would over-weight the importance of those sites with multiple samples. Therefore, one sample per site was selected. Preferred samples were those that were collected recently, during the standard sampling period, and met all other analytical requirements. When several samples were of equal preference, selection was random.

3 Analytical Methods

3.1 Site Classification

Reference conditions are expected to vary across the many natural settings in Colorado, due to environmental limitations or opportunities affecting the macroinvertebrates. When comparing conditions in test sites to the reference condition, the test sites must be of a similar site class, so that the biological community has the same potential regardless of stresses. Recognizing natural variability and evaluating sites according to their natural potential is extremely important to accurate bioassessment. In the site classification analysis we attempt to establish the distinctions among reference-quality macroinvertebrate community types across the state and to identify the environmental variables that are most important in determining the site classes.

Site classes can be defined using the reference biological samples to identify compositional similarity. Environmental variables are then used to determine the basis for the biological structure. Site classes can be conceptualized in terms of both biological and environmental characteristics (e.g., high elevation, cold water streams have more oxygen sensitive organisms, such as mayflies, stoneflies and caddisflies). For MMI development, sites are typically categorized into one of a set of possible classes in a discrete manner. While each site is unique, the site characteristics among site of a distinct site class are somewhat homogenous in those natural environmental factors that affect biological potential (**Figure 2**). For multimetric indices, classification simplifies biological index development, allowing an analytical focus on biological responses within one site type at a time (Gerritsen and Barbour 2000).

Ecoregions are commonly used as the primary classifying variable because they incorporate multiple environmental characteristics. They are often used to classify sites for bioassessment (Paul et al. 2005, Stribling et al. 2000). The ecoregion concept is a trusted method for identifying general ecological categories of site characteristics (Omernik and Bailey 1997).

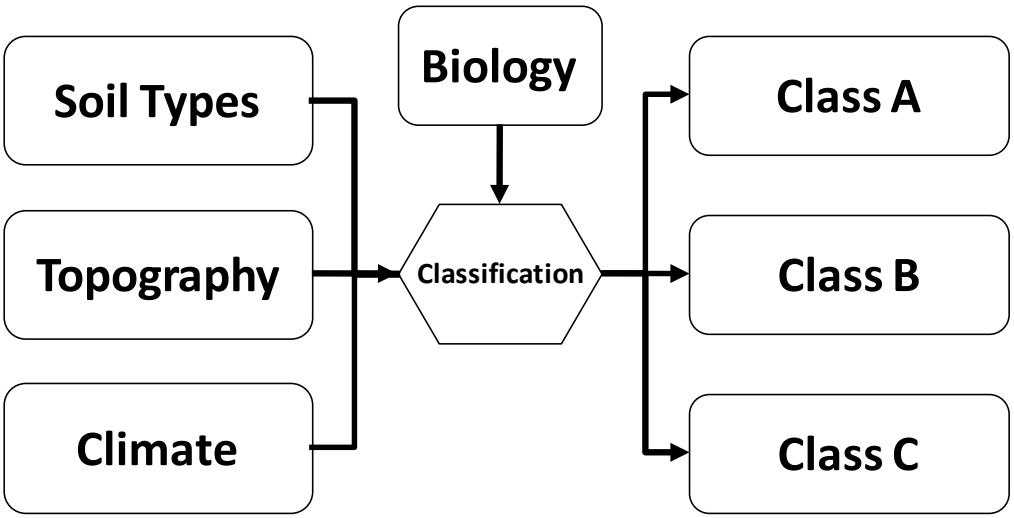


Figure 2. Typical inputs and results in a discrete classification process. Inputs include natural environmental variables and biological similarities among sites. Resulting site classes are discrete bins, where a site can be placed in only one bin.

Biological Groupings

Cluster Analysis

In cluster analysis, sites are arranged in a branched hierarchy of similar groups by first calculating the taxonomic similarities between samples and then clustering sites with the flexible-beta algorithm (McCune and Grace 2002). Biologically distinct groups can be defined by breaking the dendrogram into several branches. Each group can then be associated with environmental factors through discriminant analysis and through display of the groups in the ordination diagram. We use the terms groups and clusters interchangeably.

A site-by-taxon matrix was compiled to reflect presence of each taxon in each site. Similarity among reference biological samples was determined using the Bray-Curtis similarity measure. The Bray-Curtis (BC) formula is sometimes written in shorthand as

$$BC = 1-2W/(A+B)$$

where W is the sum of shared abundances and A and B are the sums of abundances in individual samples. The analytical software (PC-Ord, McCune and Mefford 1999) calculates a site-by-site matrix of BC similarity from which the arrangement of samples in the ordination diagram is derived. Multiple dimensions are compressed into two or three dimensions that we can perceive. Taxon presence does not take abundance into account, so that a single individual or 50 individuals would be counted equally in this exercise. Clusters were also derived using relative abundance of taxa and calculated metrics.

Ordination diagrams and metric distributions are used to confirm the biological groupings indicated in the cluster analysis. When clusters have unique taxonomic composition, they

will appear in separate areas of the ordination diagram (see ordination details below). Significant overlap in the ordination diagram would suggest that the taxonomic composition in one group is similar to the overlapping group.

A biological metric is a numerical expression of a biological community attribute that responds to human disturbance in a predictable fashion. Because metrics are the building blocks of MMI's, it is important that the distributions of metric values among clusters are distinct. Metrics can be insensitive to the clustered groups depending on the constancy of taxonomic characteristics among taxa that are substituted across groups. In other words, the metrics may measure an assemblage characteristic or function that is independent of the specific taxa. Therefore, we used metric distributions in potential site groupings to confirm the proposed groupings and assure that the groupings are meaningful in the assessment context. In our analysis, those groups that show consistently overlapping metrics were considered for combination as a single group.

Metric distributions can also indicate when groupings are inadequate. Metric distributions that show considerable variability (broad ranges of values) within a single group suggest that the natural variability of the site grouping has not been accounted for. In these cases, we would look into the possibility of multiple refined groupings within the variable group or metric adjustments using linear regressions.

Discerning Environmental Factors

The natural environmental settings of the sites were described using remotely sensed, sampled, observed, estimated, or calculated environmental data. Variables that might indicate site classes (predictor variables) were compiled among seven categories, including ecoregions, site location, site and watershed physical characteristics (including topography), geology, soil characteristics, climate, and flow (**Table 3**). Variables that were subject to alteration through human activity were relegated to lower analytical importance because we were seeking natural biological differences, not stressor induced differences. Stress levels should be minimal in the reference sites, but classifying sites based on stressor sensitive variables could possibly introduce errors in classification of non-reference sites.

We discerned the most important environmental determinants of biological groups using three techniques: 1) ordination overlays, 2) variable distributions, and 3) discriminant function analysis. These three techniques were used simultaneously in a weight of evidence approach, where results from one technique were used to inform and validate results from the others.

Table 3. Predictor variables listed categorically. This list does not include all the possible variations of each variable, which included statistical summaries by watershed or at the sampling point.

Site Location	Latitude Longitude Level 8 Hydrologic Unit Major Basin
Physical Characteristics	Catchment Area Relief Ratio Drainage Density Watershed Shape Factor Site Elevation Average Catchment Elevation Stream Slope[1]
Ecoregion	Dominant Level 3 Ecoregion in the site catchment Dominant Level 4 Ecoregion in the site catchment Site Level 3 Ecoregion Site Level 4 Ecoregion
Geology	Dominant Geology Site Geology Percent Granitic Geology Percent Sedimentary Geology Percent Volcanic Geology Geologic Chemical Activity Geologic Nutrient Activity Geologic Physical Activity Predicted Alkalinity Predicted Conductivity
Soil Characteristics	Soil Water Capacity Soil Density Soil Erodibility Soil Organic Content Soil Permeability Soil Depth Water Table Depth
Climate	Precipitation, Watershed Site Precipitation Number of Wet Days Average Temperature Maximum Summer Temperature Number of Frost Free Days Date of First Frost
Flow	Hydrologic Stability

[1] Stream slope was calculated the same way as USGS National Hydrography Dataset (NHD) slope but the distances of the stream segments were shorter resulting in differing values. Spatially mapped slope predictions will be made available to the public for estimating slope values.

Ordination Overlays
Biological groups were explored using non-metric multidimensional scaling (NMS), a comparison of taxa within each sample (using a pair-wise Bray-Curtis similarity measure) and an arrangement of the samples so that similar samples plot closer together than dissimilar samples in multiple dimensions. Natural environmental variables can be associated with the biological groups through visual inspection of the ordination diagrams and correlations with the biologically defined axes of the NMS diagram. NMS is a robust method for detecting similarity and differences among ecological community samples and works as well with presence/absence data as with relative abundance data (McCune and Mefford 1999, Reynoldson et al. 1995).

The ordination software (PC-Ord; McCune and Mefford 1999) allows for the display of environmental variables associated with the biologically arranged samples. The display was used to discern those variables that were correlated with (continuous variables) or biased on (categorical variables) the ordination axes. Through rotation of the ordination axes, it is possible to show correlation or orthogonality among the environmental variables.

Variable Distributions
Environmental variable distributions and comparisons are used to identify those variables that differ among groups. Variables with disparate distributions among biological groups are candidate determinants of the groupings. Box and whisker diagrams were used to identify breakpoints of the variable values among groups.

Discriminant Function Analysis
Discriminant function analysis (DFA) is commonly used in predictive assessment tools to assign probabilities of class membership to sites. A discriminant function is created as a linear combination of independent variables. If DFA is effective, the model predicts correct groups for a high percentage of sites. The environmental variables that repeatedly appear as significant determinants in the DFA models are the ones that affect the natural biological community and may be valuable as classifying variables.

Predicting Site Classes

We observed patterns in the environmental data distributions using graphic displays of environmental variables and cluster assignments. Mapping exercises displaying sites coded by potential groupings and overlain on landscape characteristics (such as ecoregion) were useful to conceptualize biological types (Biotypes). Rules for assigning sites to Biotypes were derived through an iterative process which included establishing preliminary rules, checking for correct assignment of sites to biological clusters, refining the rules through examination of assignment errors, and checking correct assignments after refining the rules. Rule establishment and refinement were based on biological-environmental relationships recognized in the ordination overlays, variable distributions, and DFA.

3.2 Metric Calculations

Metric Calculation and Evaluation

Metrics were calculated as counts of taxa, relative percentages of individuals, or complex algorithms derived from the sample taxa lists. Several commonly applied, empirically proven, and theoretically responsive metrics were calculated for possible inclusion in a multimetric index. The calculations were made in the EDAS database. The metrics were organized into five categories: richness, composition, pollution tolerance, functional feeding group, and habit (mode of locomotion).

Discrimination efficiency (DE) is the capacity of the biological metric or index to detect stressed conditions. It is measured as the percentage of stressed sites that have values lower than the 25th percentile of reference values (Stribling et al. 2000, **Figure 3**). For metrics that increase with increasing stress, DE is the percentage of stressed sites that have values higher than the 75th percentile of reference values. DE can be visualized on box plots of reference and stressed metric or index values with the inter-quartile range plotted as the box (**Figure 5**). When there is no overlap of boxes representing reference and stressed sites, the DE is greater than 75%. A metric with a high DE thus has a greater ability to detect stress than metrics with low DEs. Metrics with DEs <25% do not discriminate and were not considered for inclusion in the index.

Metrics are based on some attribute of the taxa, as suggested by the metric categories. Several metrics are based on the sensitive insect orders Ephemeroptera, Plecoptera, and Trichoptera (EPT), commonly known as mayflies, stoneflies and caddisflies. Other metrics are based on pollution tolerance, as in the Hilsenhoff Biotic Index (HBI, Hilsenhoff 1987). The HBI was developed with organic pollution as the primary stressor, and the values may not be applicable in certain stressors in Colorado. For that reason, we analyzed indicator taxa for the stressors and samples in the current data set.

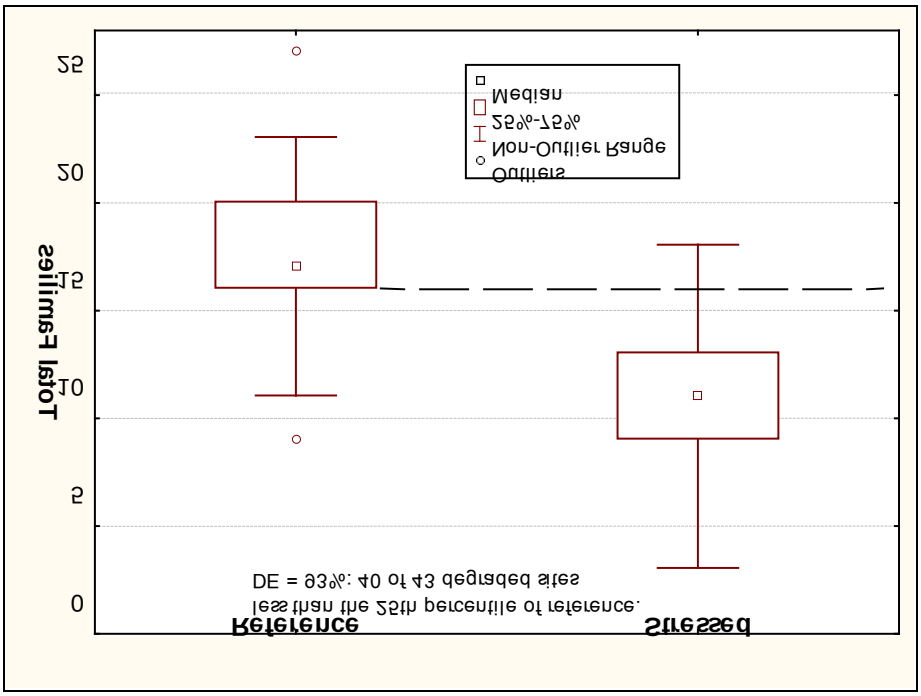


Figure 3. Illustration of metric discrimination efficiency (DE) between reference and stressed sites.

Indicator Taxa

Indicator taxa are those that differ markedly in frequency of occurrence and relative abundance between groups of samples (Dufresne and Legendre 1997), such as those from reference and stressed sites. Frequency of occurrence was calculated as the percentage of sites in which each taxon occurred (by site category; reference or stressed). Relative abundance was the percentage of all individuals in a taxon that occurred in either reference or stressed samples. We grouped taxa at the family taxonomic level to simplify the analysis and resulting metrics. We considered responses in two site groupings, high elevation and low elevation, because of perceived differences in taxonomic distributions, sensitivities, and environmental characteristics between the two categories.

The indicator analysis included an empirical step and a subjective review step. During review, the indicator taxa that were more or less prevalent by reference category were scrutinized by the index development workgroup, which included biologists who were familiar with benthic taxa sensitivities. The workgroup removed some taxa from the lists of those empirically identified as indicators of stress. No taxa were added to the lists during the review process.

Metric Adjustment to Environmental Conditions

Site classification with discrete class assignments can result in groupings of sites with similar biological characteristics. However, high biological metric variability within a class can persist. This can be controlled if relationships between metrics and environmental variables are recognized and adjustments are applied. Within reference

sites, correlations among metrics and environmental variables were analyzed. Where high correlations existed, metrics were adjusted to the environmental variable through linear regression. The ultimate metric used in metric testing and index development was then the residual of the metric value predicted by the regression equation.

3.3 Index Composition

A multimetric index is a combination of metric scores that indicates a degree of biological stress in the stream community (Barbour et al. 1999). Individual metrics are candidate for inclusion in the index if they:

- discriminate well between reference and stressed sites;
- are ecologically meaningful (mechanisms of responses can be explained);
- represent diverse types of community information (multiple metric categories); and
- are not redundant with other metrics in the index.

Metrics are scored on a common scale prior to combination in an index. The scale ranges from 0 to 100 and the optimal score is determined by the distribution of data. For metrics that decrease with increasing stress, the 95th percentile of all high gradient data was considered optimal and scored as 100 points. All other metric values were scored as a percentage of the 95th percentile value except those that exceeded 100, which were assigned a score of 100. Metrics that increase with stress are scored using the 5th percentile as the optimum. The 5th and 95th percentile values were selected as optimal instead of the maximum so that outlying values would not skew the scoring scale.

Other Metric Considerations
Index alternatives were evaluated based on index DE, a *z-score* to quantify reference and stressed differences, variability of reference scores, the number of metric categories included, and ecological meaningfulness of the included metrics. The DE for indices was calculated as described above for individual metrics. The *z-score* was calculated as the difference in mean index values between reference and stressed sites, divided by the standard deviation of reference index scores.

Ecologically meaningful metrics are those for which the assemblage response mechanisms are understandable and are represented by the calculated value. Ecological meaningfulness is a professional judgment based on theoretical or observed response mechanisms. Those metrics that respond according to expectations established in other studies are more defensible.

Metrics contribute information representative of integrity if they are from diverse metric categories. As many metric categories as practical should be represented in an index so that signals of various stressors can be integrated into the index. While several metrics should be included to represent biological integrity, those that are included should not be redundant with each other. Redundancy was evaluated using a Pearson Product-Moment correlation analysis.

For metrics to discriminate on a gradient of stress, they must have a sufficient range of values. Metrics with limited ranges (e.g., richness of taxa poor groups or percentages of rare taxa) may have good discrimination efficiency. However, small metric value changes will result in large and perhaps meaningless metric scoring changes.

4 Results

4.1 Site Classification

Cluster Analysis
For our analytical data set, presence-absence data were more robust than relative abundance or metrics for identifying sample similarities. The test of robustness entailed iterative removal of a portion of sites during successive cluster analyses and checking on the consistency of assignments of sites to clusters in each run. We ultimately relied on the clusters indicated using presence and absence of taxa in all reference sites because the cluster assignments were relatively robust among iterations compared to methods using relative abundance or metrics. We suspect that any inconsistencies introduced through multiple sampling techniques would have less effect on taxa presence/absence than it would on taxa relative abundance. In addition, presence-absence data are commonly used in ordination and cluster analyses (e.g., RIVPACS analyses use presence-absence data exclusively).

In each of 20 iterations of the cluster analysis with presence/absence data, 10% of the sites were randomly removed. Site dendrograms and site cluster assignments were examined for each iteration. When more than 3 clusters were distinguished, sites moved among clusters in successive iterations. Based on this variability and results of other preliminary cluster analyses, we used only 3 cluster designations.

The cluster analysis of presence-absence data with flexible beta clustering technique resulted in a cluster dendrogram with minimal chaining when beta was set at -0.5. Chaining is the sequential addition of small groups to one or a few large groups. Dendrograms that are highly chained are undesirable as they are generally not helpful in defining subgroups. The beta value of -0.5 was selected based on visual inspection of alternative cluster dendrograms and consideration of the percent chaining statistic. Three groups were defined when using a cut line with 30% information remaining (**Figure 4**). The three groups were biologically distinct (**Figure 5**) and could be associated with conceptual sites types (**Table 4**).

Table 4. Site designations in three clusters.

Cluster number and conceptual description	Number of Sites
1 (mid-level, intermediate temperatures)	43
2 (high elevation, cold water)	61
3 (low elevation, warm water)	29

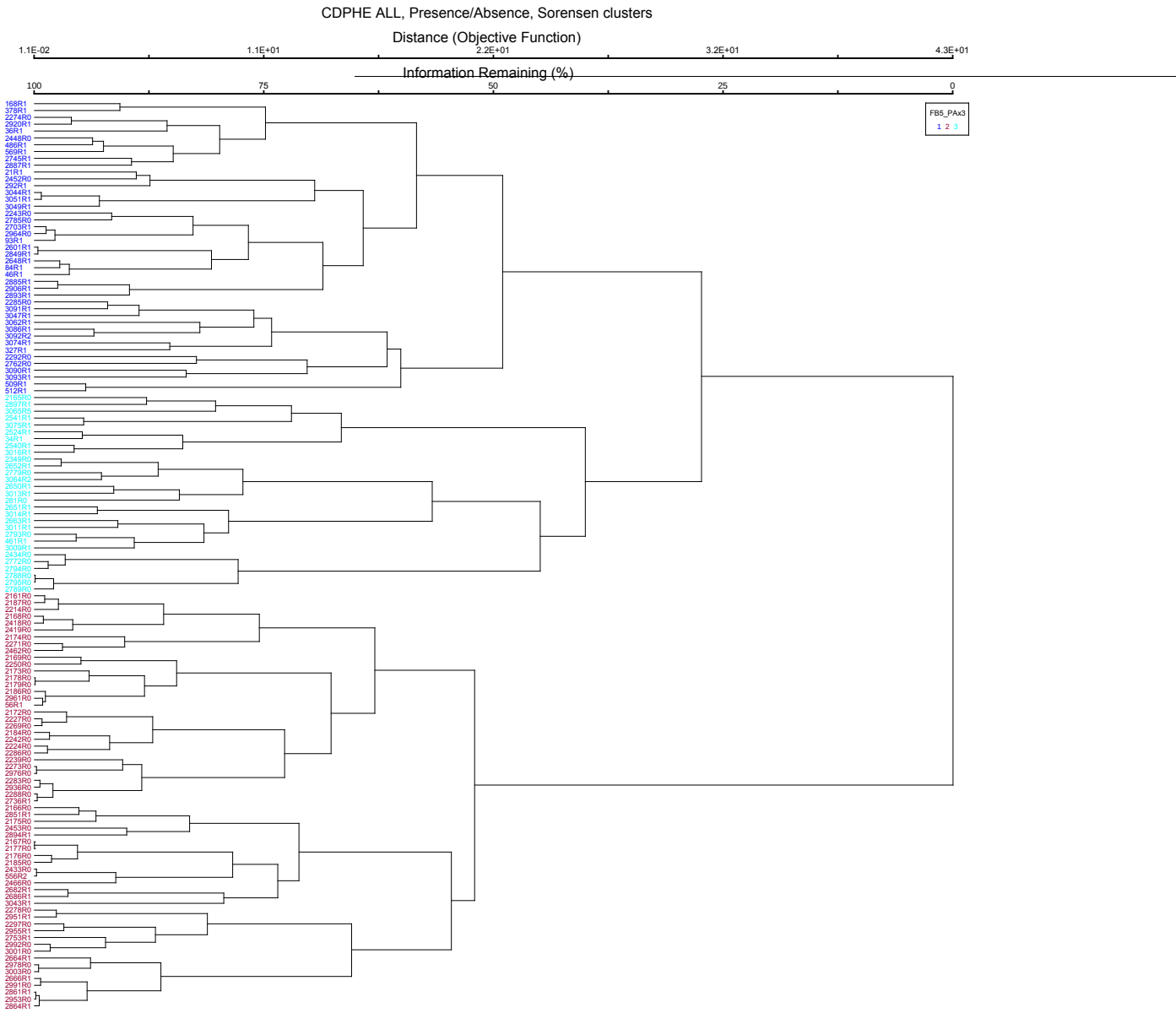


Figure 4. Cluster dendrogram showing reference sites grouped by similarities in the presence (and absence) of taxa in the samples.

Ordination Analysis

Ordination showed that groups defined through cluster analysis also showed separation in non-metric multidimensional scaling (**Figure 6**). The primary axis of the ordination diagram divided the mountainous sites on the left and the plains sites on the right. This axis was also associated with temperature, elevation, and precipitation. The second axis was strongly related to longitude, but this variable did not prove to be useful in separating the groups. Ecoregions were well separated on the ordination diagram, though some overlap among the clustered groups was obvious. The Southern Rockies did not overlap at all with the High Plains and Southwestern Tablelands ecoregions.

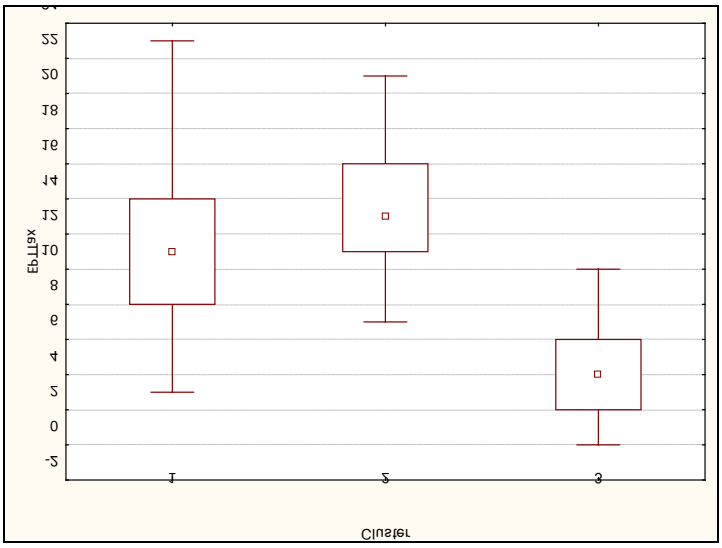


Figure 5. Distributions of EPT taxa richness in clusters.

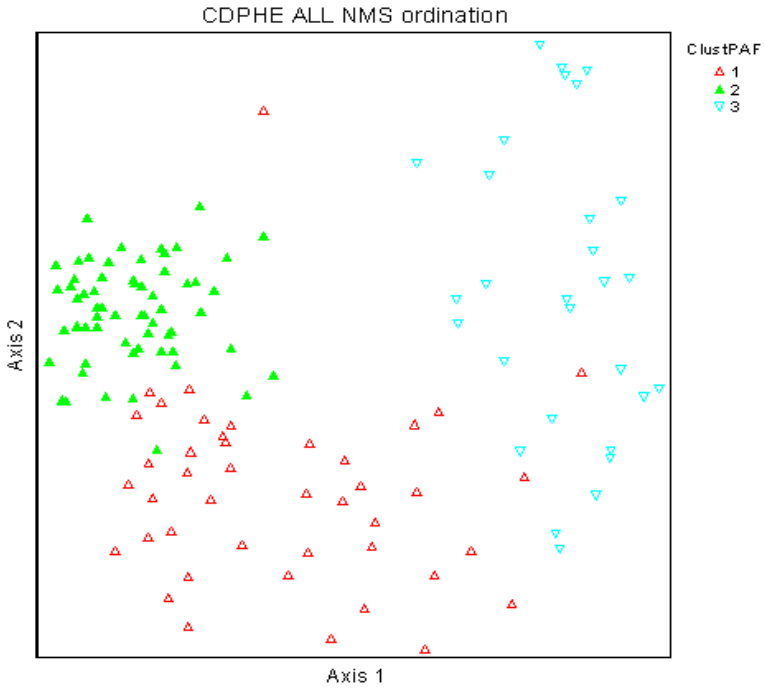


Figure 6. NMS ordination of taxa presence-absence with sites marked by groups resulting from cluster analysis.

Only two axes were needed in the ordination diagram to achieve a suitable arrangement (stress = 20.6). Ordination stress is a measure of the departure from the linear relationship between the dissimilarity (distance) in the original multi-dimensional space and distance in the reduced-dimensional ordination space. A target stress value for deciding on the number of ordination dimensions is 20. These ordination results were based on presence-

absence data. Ordinations using taxa relative abundance and metrics showed relationships that verified those described with presence-absence data.

Variable Distributions

The distributions of environmental variables in biological groups show that the environmental characteristics of the site clusters are somewhat distinctive (**Figure 7**). While groups 2 and 3 could be distinguished on elevation and temperature, overlapping ranges of values were evident in precipitation and stream slope and with the intermediate group, group 1. This figure includes only a subset of the environmental variables; those which showed the greatest distinctions among groups and were not redundant with each other. Other variables, such as catchment area and longitude, also showed distinctions among groups.

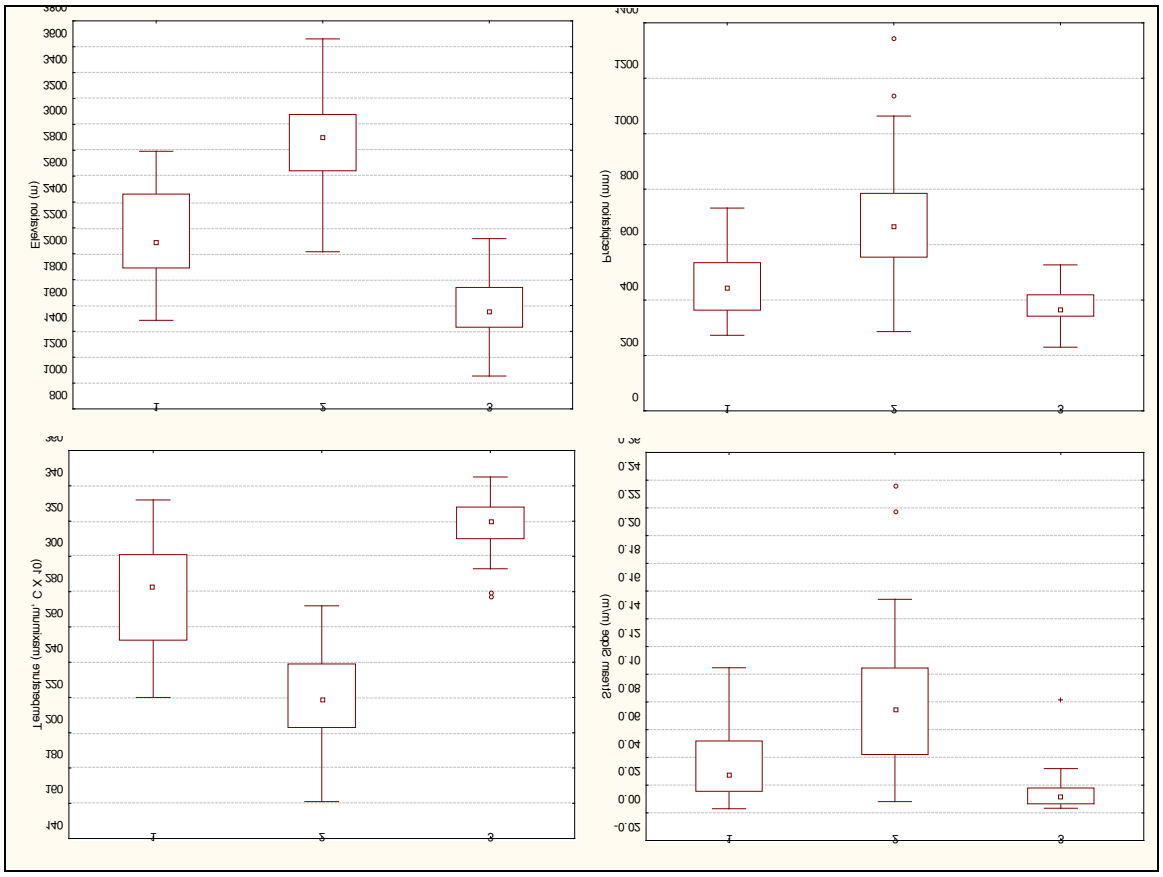


Figure 7. Distributions of environmental variables (elevation, precipitation, maximum air temperature, and stream slope) in the biologically-defined groups.

Discriminant Analysis

A forward stepwise DFA was conducted using the environmental variables. The DFA did not consider categorical variables such as ecoregion. The most accurate prediction of group membership used three variables in the model: elevation, slope, and precipitation. These were not the most significant variables to enter the model at each step, but they were forced into the model after eliminating variables that were relatively difficult to derive (e.g., average maximum summer temperature or average precipitation in the watershed) or that were conceptually non-linear (e.g., longitude). Elevation was the most significant determinant when summer temperature was removed, and the two were highly correlated. The model placed 76% of reference sites in the correct groups given equal probability of membership and 83% correct given probability weighted by group size.

Conceptual Site Classes

The groups defined by cluster analysis were indicative of three site types, which are conceptually described as:

1. Mid-elevation, semi-cold, low gradient, moist (Transitional)
2. High elevation, cold-water sites (Mountains)
3. Low elevation, warm-water, low gradient, dry (Plains)

The site types were in agreement with (and more refined than) the existing MMI assessment regions and represented groupings that could be conceptually associated with biological differences.

Predicting Site Classes

The process for assigning sites to classes included review of all analytical results, development and application of decision rules, and iteratively testing the results. The three variables used in the classification rules (**Table 5**) include site ecoregion, site elevation, and stream slope. These were decided upon because of corroborating results from ordination analysis, variable distribution evaluation, and DFA. Of the 133 sites used to derive the classification rules, 120 (90.2%) had predicted classes matching the biologically defined clusters. Class assignments are listed in **Appendix B**.

Table 5. Classification Rules.

Group 3 prediction: If ecoregion = 25 or 26 (except sub-ecoregion 25l and 26i), then Predicted Class = 3 If elevation at the site < 1550, the Predicted Class = 3
Group 2 prediction: If sub-ecoregion (level 4) = 21a, 21b, 21e, or 21g, then Predicted Class = 2 If sub-ecoregion (level 4) = 21c and slope > 0.04, then Class = 2 (1 if shallower) If sub-ecoregion (level 4) = 21f and elevation >2500, then Class = 2 (1 if lower)
Group 1 prediction: If site is not predicted in classes 2 or 3, then Predicted Class = 1 (e.g., If sub-ecoregion (level 4) = 25l, 26i, 21d, 21h, 21i, or 21j, then Predicted Class = 1)
Sub-ecoregions used in the rules Used in rules for group 3. 26i = Pine-Oak Woodlands

Table 5 (continued)

25l = Front Range Fans

Used in rules for group 2.

21a = Alpine Zone
21b = Crystalline Subalpine Forests
21c = Crystalline Mid-Elevation Forests (northern Front Range)
21e = Sedimentary Subalpine Forests
21f = Sedimentary Mid-Elevation Forests (primarily western slopes)
21g = Volcanic Subalpine Forests

Used in group 1 rules:

21d = Foothill Shrublands
21h = Volcanic Mid-Elevation Forests
21i = Sagebrush Parks
21j = Grassland Parks

4.2 Metric Development

Indicator Taxa

Indicator taxa analysis using presence/absence family-level information revealed the taxa that were more common to reference or stressed samples. This analysis was performed separately in two sets of site groups: high elevation and low elevation. Because other tolerance measures were responsive in high elevation sites, indicator taxa were not as thoroughly pursued there compared to the low elevation sites. During a working meeting held in 2008, Colorado stakeholders scrutinized the analytical results for the low elevation sites and reduced the length of the lists based on professional experience (**Table 6**). These taxa were used to calculate metrics in the pollution tolerance metric category.

Table 6. Indicator taxa for the low elevation sites (groups 1 and 3).

Families Sensitive to Stress		Families Tolerant of Stress
Elmidae	Naucoridae	Tubificidae
Hydrophilidae	Gomphidae	Corixidae
Dytiscidae	Nemouridae	Naididae
Tipulidae	Limnephilidae	Erpobdellidae
Caenidae	Hydropsychidae	Enchytraeidae
Leptophlebiidae	Leptoceridae	

Me*trics Adjustments to Environmental Variables*

In Biotype 1 reference sites, several metrics were correlated with elevation and maximum temperatures. For instance, the EPT taxa metric was correlated with elevation (Spearman r = 0.62) and summer temperature (Spearman r = -0.68). Our expectations were that EPT taxa would be more numerous in cooler and higher elevation reference sites. We adjusted metrics as residuals of regressions to eliminate the bias. Other relationships were noted in Biotype 1 and in the other biotypes. However, the relationships within other Biotypes were not strong. Metric variability was mostly accounted for by the classification scheme. Metric adjustment did not always improve the DE of the metrics.

Metric Performance
A total of 145 metrics were calculated using EDAS queries of the macroinvertebrate data with identifications reduced to the standard OTU level. This included 56 metrics calculated using environmental adjustments for elevation or summer temperature. The DE of each metric was calculated within the *predicted* site classes (**Appendix C**). Thirty-eight (38) metrics showed a similar discriminating trend in all three biotypes.

Metrics with limited ranges of values were not included in index trials. Metrics that quantify rare taxa can have limited ranges of values. Metrics with ranges of less than 5 taxa or percentages less than 10% were tested for responsiveness to stress. However, they were generally not included in index trials because small changes in metric values could result in large changes in metric scores and increased index variability.

In Biotype 1, metrics from each metric category had DEs greater than 50% and metrics in the richness, composition, and tolerance categories had DEs ranging to above 75%. The metrics with the highest DEs were Plecoptera taxa (adjusted for elevation) and percent Plecoptera.

Metrics from each category in Biotype 2 had DEs greater than 85%, including seven metrics in the richness composition and habit categories with DEs of 100%. In Biotype 3, richness and specialized tolerance metrics performed best. Tolerance values ranged upwards to 92%. Except for % Coleoptera, composition metrics did not have DEs greater than 72%, and feeding group metrics did not have any DEs greater than 64%.

4.3 Index Performance

One hundred and four (104) index alternatives were calculated and tested using reference and stressed sites predicted in the three site classes (**Appendix D**). The index alternatives that performed best were reconsidered and the following indices were recommended.

In Biotype 1, 18 of 59 alternatives had DEs greater than 85%, including three with the highest DE of 92.9%. The alternative selected as the MMI for 1 (**Table 7, Figure 8**) included six metrics from all five metric categories and had the lowest standard deviation in reference sites. Two of the metrics were sensitive to site elevation and required adjustment before scoring.

Table 7. Metrics included in the Biotype 1 MMI with adjustment and scoring formulae.

Metric (X)	**Metric Adjustment (Xa[2])**	**Scoring Formulae[3]**
% non-insect taxa		100*(28.09375-X)/ 28.0937
EP Taxa	(X-(-4.803+ (0.00583)*Elev))	100*(Xa+6.481)/ (2.870+6.481)
% Chironomidae		100*(66.839- X)/ (66.839-1.308)
% Sens. Fam. Plains		100*X/ 61.812
Predator, Shredder taxa		100*X/ 14
Clinger taxa	(X -(-3.262+ (0.00694)*Elev))	100*(Xa +8.615)/ (3.136+8.615)

[2] Xa = Elevation adjusted metric
[3] Scoring formulae are rounded

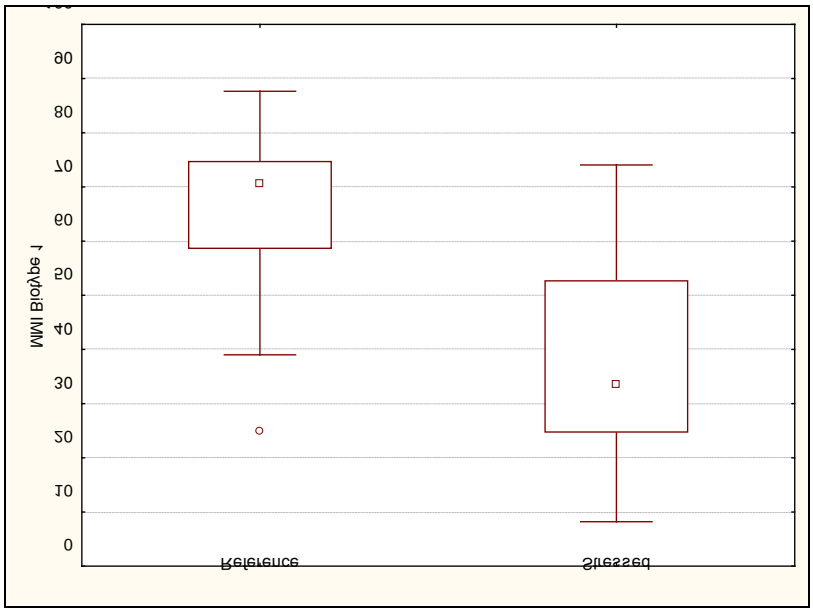


Figure 8. Distribution of index scores in reference and stressed sites in Biotype 1.

In Biotype 2, 16 of 19 alternatives had DEs of 100%. Therefore, other index evaluation factors were important in selecting the final MMI. The alternative selected as the MMI for Biotype 2 (**Table 8, Figure 9**) included five metrics, one from each of the five metric categories. It had the lowest variability in reference sites of those indices that had 100% DE.

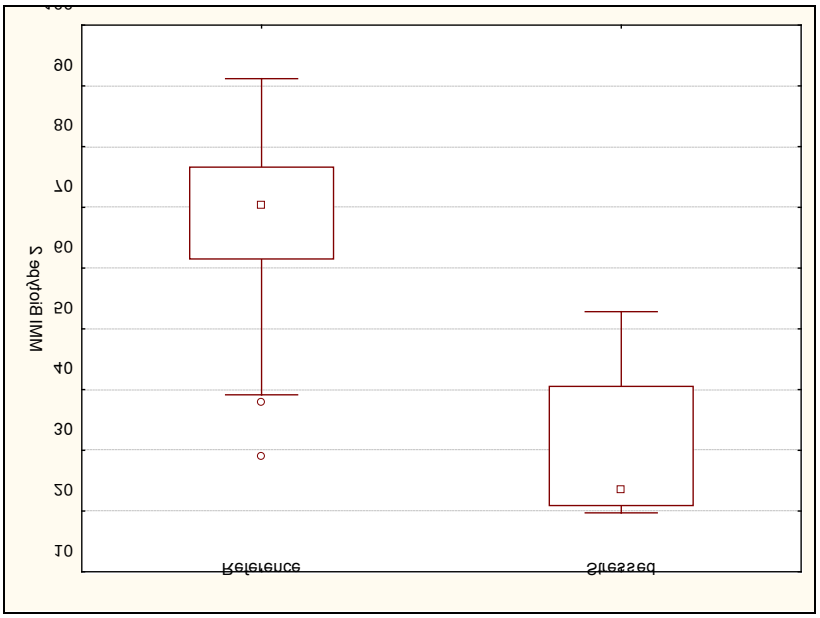


Figure 9. Distribution of index scores in reference and stressed sites in Biotype 2.

Table 8. Metrics included in the Biotype 2 MMI with scoring formulae.

Metric (X)	**Scoring Formulae**
Total taxa	100* X / 36
% Ephemeroptera	100* X / 72.090
Predator, Shredder taxa	100* X / 14
Beck's Biotic Index	100* X / 33
Clinger taxa	100* X / 17

In Biotype 3, 16 of 26 alternatives had DEs of 100%. Therefore, other index evaluation factors were important in selecting the final MMI. The alternative selected as the MMI for Biotype 3 (**Table 9, Figure 10**) included six metrics, representing each of the five metric categories. Some index alternatives included metrics with low ranges of values (e.g., Coleoptera taxa). These alternatives were not selected because such metrics could introduce high variability to the index.

Table 9. Metrics included in the Biotype 3 MMI with scoring formulae.

Metric (X)	**Scoring Formulae**
Insect taxa	100* X / 26
% non-insect taxa	100*(44.444- X)/ (44.444-1.977)
% Sens. Fam. Plains	100* X / 57.739
% dominant taxon	100*(79.590- X)/ (79.590-18.138)
Predator, Shredder taxa	100* X / 10
% sprawler	100* X / 50.625

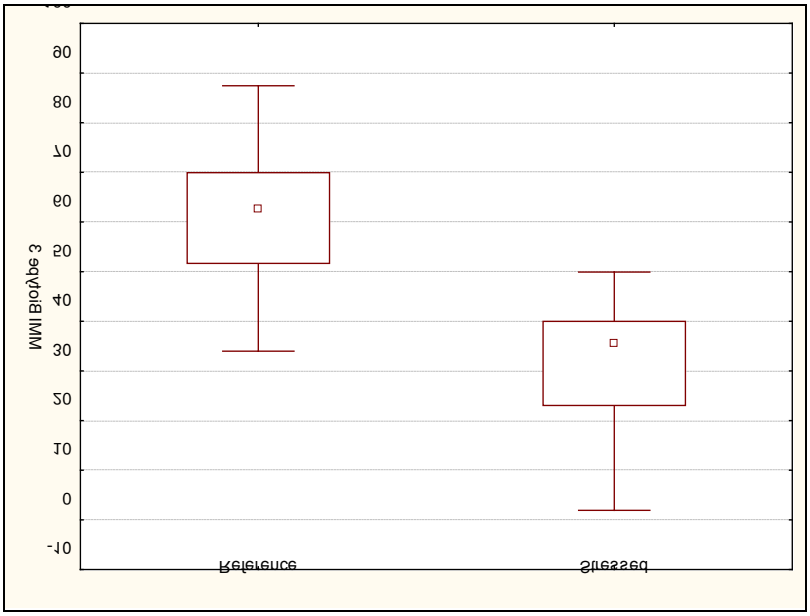


Figure 10. Distribution of index scores in reference and stressed sites in Biotype 3.

4.4 Index Application

Application of a typical MMI is straightforward; calculate the index for the appropriate site class and compare it to an established threshold. Thresholds will be determined by CDPHE.

Index application should proceed as follows:

1) Enter the biological and predictor variable data into the analytical database.
2) Determine the site class of the site using the classification rules.
3) Calculate index metrics and scores (**Table 7, 8, or 9**) and average as an MMI.
4) Compare the resulting MMI value to impairment thresholds for the appropriate site class.
5) Include MMI performance statistics when reporting assessments (**Table 10**).

Table 10. MMI performance statistics.

Site Class	DE	25th %ile of reference scores[4]
Biotype 1, Transitional	92.9	58.7
Biotype 2, Mountains	100	61.5
Biotype 3, Plains	100	51.7
All Biotypes	96.7	

6 Conclusions and Recommendations

The proposed classification scheme and multi-metric indices gives CDPHE a system in which sites can be recognized in three unique site classes and assessed with specific index formulations. Sites were grouped by biological similarities. Cluster analysis, ordination analysis, and metric distributions corroborated the biological groups. Environmental variables, that helped to classify sites, included ecoregions, elevation, and stream slope. These determinants and the way they were applied in classification rules were derived through multiple analyses, including DFA. One measure of error associated with site classification is the percentage of reference sites in which the classification rules assign a class that does not match the biological cluster. That error rate is 9.8%.

The MMIs were selected among several alternatives. All final MMIs included metrics from each of five metric categories and had a combined DE of 96.7%. This high DE and low classification error indicate that assessments of Colorado stream benthic condition, made using the MMIs, will be accurate.

[4] In EPA guidelines, the 25th percentile of reference is often used as an impairment threshold

7 References Cited

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

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

Dufrene, M. and P. Legendre. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67:345-366.

Frey, D. G. 1977. Biological integrity of water – an historical approach. Pages 127-140 in R.K. Ballantine and L. J. Guarraia (editors). The Integrity of Water. Proceedings of a Symposium, March 10-12, 1975, U. S. Environmental Protection Agency, Washington, DC.

Gerritsen, J. and M. Barbour. 2000. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. Journal of the North American Benthological Society 19(3): 487–496.

Gibson, G.R., M.T. Barbour, J.B. Stribling, J. Gerritsen, J.R. Karr (1996) Biological Criteria and Technical Guidance for Streams and Small Rivers. EPA 822-B-96-001. USEPA, Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC

Hilsenhoff, W. L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist 20:31-39.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessment of biological integrity in running waters: A method and its rationale. Illinois Natural History Survey, Champaign, Illinois. Special Publication 5.

Karr, JR, and D.R. Dudley. 1981. Ecological Perspective on Water Quality Goals. Environmental Management 5:11.

McCune, B., and J.B. Grace. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, OR.

McCune, B. and M.J. Mefford. 1999. Multivariate Analysis of Ecological Data, Version 4.10. MJM Software, Gleneden Beach, Oregon.

Omernik, J.M., and R.G. Bailey. 1997. Distinguishing between watersheds and ecoregions. Journal of the American Water Resources Association 33(5):935-949.

Paul M.J., J. Gerritsen, C. Hawkins, and E. Leppo. 2005. Development of Biological Assessment Tools for Colorado. Prepared for the Colorado Department of Public Health and Environment, Water Quality Control Division – Monitoring Unit, Denver, CO.

Reynoldson, T.B., R.C. Bailey, K.E. Day, and R.H. Norris. 1995. Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using multivariate approach for predicting biological state. Australian Journal of Ecology (1995) 20: 198-219.

Stribling, J. B., B. K. Jessup, J. Gerritsen. 2000. Development of Biological and Physical Habitat Criteria for Wyoming Streams and Their Use in the TMDL Process. Prepared for U. S. EPA Region 8, Denver, CO.

Appendix A

Reference and Stressed Site Selection Process

Appendix A: Reference and Stressed Site Selection Process

Reference sites were selected using a GIS based approach that focused on using mapped human disturbance categories, delineated watershed area polygons and delineated 5 km polygons upstream from each station. GIS was then used to calculate percent (%) disturbed area or count within the total watershed area ("far field") or the 5 km radius "near field" area. Candidate reference sites were further screened by reviewing Global Explorer images at EPA Region VIII to confirm reference status.

As recommended by EPA, the WQCD ("Division") decided to avoid issues of circularity further down the road during nutrient criteria development by dropping water chemistry and habitat criteria all together. Nitrogen and phosphorus parameters were dropped outright due to concern over circularity (i.e. nutrient criteria development conflict). The Division plans on establishing nutrient thresholds by plotting MMI scores versus nutrient values and then finding a cutoff where MMI scores are low when nutrient values are high.

Other water chemistry criteria were dropped due to three reasons:

- Lacked screening power
- Lacked sufficient amount of data across all candidate sites
- Difficulty setting thresholds (what values are appropriate for distinguishing between ideal and unacceptable?)

Habitat parameters were also dropped for the following reasons:

- Lacked sufficient amount of data across all candidate sites
- Difficulty setting thresholds (what values are appropriate for distinguishing between ideal and unacceptable?)

Using the GIS-based approach allowed each candidate site to be considered in the reference screening process. In other words, all sites started off on equal footing in this process. The issue of data gaps could be avoided all together.

GIS Based Approach

The Division used the following human disturbance parameters in the screening process:

- Irrigated agriculture (%)
- Dryland agriculture (%)
- Urban (%)
- Permitted Point Sources - outfalls (count)
- Diversions – headgated only (count)
- Road density (average)
- Abandoned mines (count)
- Oil and gas facilities (count)
- CAFO's (count)

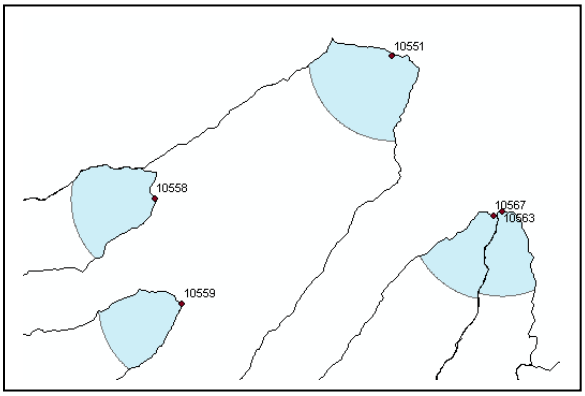
Sources of GIS Information

Irrigated agriculture	CO Division of Wildlife Aquatic GAP land use/land cover
Dryland agriculture	CO Division of Wildlife Aquatic GAP land use/land cover
Urban	CO Division of Wildlife Aquatic GAP land use/land cover
Permitted point sources	WQCD GIS shapefile (updated summer 2007)
Headgated diversions	CO Division of Water Resources (updated fall 2007)
Roads	U.S. Census Bureau 2007 TIGER/Line shapefile
Abandoned mines	U.S. Geological Survey Abandoned Mines Land Initiative
Oil and Gas facilities	CO Oil and Gas Conservation Commission GIS Online
CAFO's	WQCD GIS shapefile

Total Watershed Area vs. 5 km Radius

Total watershed area is a calculation of the total area that drains to a station. These areas were calculated using Utah State University's Multi-Watershed Delineation Tool.

The 5 km radius was created by first setting a 5 km buffer around each station. Then a buffer polygon was intersected with a total watershed area for each given station. The area where the two intersect is considered the 5 km radius area above the station. The intersections are illustrated in blue below.



GIS Calculation

Irrigated agriculture	km^2 per total watershed area (km^2)
Dryland agriculture	km^2 per total watershed area (km^2)
Urban	km^2 per total watershed area (km^2)
Permitted point sources	# of PPS's within 5 km radius
Headgated diversions	# of diversions within 5 km radius
Roads	linear miles (mi.) per total 5 km radius area ($mi.^2$)
Abandoned mines	# of abandoned mines within total watershed area
Oil and Gas facilities	# of oil and gas facilities within 5 km radius area
CAFO's	# of CAFO's within 5 km radius area

Human disturbance values for 1032 stations in CO-EDAS were calculated on GIS or on Excel spreadsheets. Results were stored on a master spreadsheet and then moved to a final spreadsheet that could provide a platform for selecting thresholds or cutoffs.

The process for selecting thresholds was iterative and involved the best professional judgment of WQCD Standards Unit staff. For the mountains bioregion, values for each human disturbance parameter were plotted. Each plot was reviewed and a threshold was selected that best partitioned values that were deemed ideal and those deemed marginally ideal for a reference status in the mountains. Once this was completed the threshold values were manually input into the same spreadsheet where a formula calculated the number of mountain sites that would qualify for reference status given the thresholds provided. The iterative process involved striking a balance between an adequate number of candidate reference sites and maintaining a sensible partition between ideal and marginally ideal. This process continued until the Division felt there was an adequate set of candidates that could then continue forward in the screening process.

The same process was used in selecting candidate reference sites in the plains and xeric bioregions. This was much more difficult since more human activity occurs in these bioregions. The partition between ideal and marginally ideal shifted to a partition between acceptable and undesirable. The threshold result often granted an allowance for some minimum amount of human disturbance.

The thresholds selected for each bioregion are found in the following table:

	Irrigated Agriculture	Dryland Agriculture	Urban	Permitted Pt. Sources	Diversions	Road Density	Abandoned Mines	O & G	CAFO's
	% of WSA	% of WSA	% of WSA	#/km^2	#/km^2	mi/mi^2	#/km^2	#/km^2	#/km^2
Mnts	0.0	-	0.0	0	1	0.6	0.03	-	-
Plains	6	50	1	2	30	3	0.002	3	2
Xeric	5	2	0.05	0	10	2	0.05	0	-

The outcome of this process yielded 110 candidate reference sites in the mountains bioregion, 60 in the plains bioregion and 64 in the xeric bioregion.

Site Duplication Removal

The next phase removed multiple sites from the same waterbody. For instance, the Division did not want to have 6 reference sites on one given stream. The Division established a "rule of thumb" for removing duplication sites. If adjacent sites on the same waterbody had a greater than 20% increase in watershed area from the upstream site

to the downstream site then both sites were retained for further analysis. If the difference was <20% then the Division selected the site with the most recent sample year. 2005 would supersede 2004 and so forth. However, there were two exceptions to the rule. The first exception was if the data was collected in 2002. This was a drought year and this data was avoided all-together. For example, if data were collected in 2001 and 2002, then 2001 would supersede 2002. The second exception was if one of the two stations was below a significant hydrological modification(s), such as below an impoundment or clustering of upstream diversion. The minimally impacted site would supersede the most impacted site.

The outcome of this process yielded 100 candidate reference sites in the mountains bioregion, 48 in the plains bioregion and 60 in the xeric bioregion.

Aerial Imagery Review

This phase of the reference screening process involved reviewing aerial images on Global Explorer at EPA Region VIII – Denver. EPA staff assisted Division staff in queuing up images on Global Explorer through a GIS routine. Then one at a time each station and a 3 km distance upstream was scrutinized for human activity and scored on a 0 to 10 scale. The source of the scoring scale is the Rapid Screening Disturbance Scoring (RSDS) instruction manual (draft 12/10/03) developed by EPA. Assignment of numeric disturbance scores was based upon visual interpretation of aerial orthophotos. Stations with obvious signs of human perturbation were removed from further consideration. Images were converted to PDF files and saved.

The outcome of this process yielded 88 reference sites in the mountains bioregion, 34 in the plains bioregion and 28 in the xeric bioregion.

After the aerial image review the Division mapped the final list of candidate reference points on GIS and examined the spatial distribution across the state. It was determined that it was necessary to fill spatial gaps particularly in the NE plains, the “core” sections of the xeric bioregion and the North Platte basin. Sites that were removed during the aerial image review were re-examined and further scrutinized. For the mountains bioregion, two sites were added to the reference list. These two sites were initially screened out simply because they fell too close to the mountains and plains bioregion interface. For the plains bioregion, five sites were added to the reference list. These sites primarily filled spatial gaps in the NE quadrant of the Colorado plains. For the xeric bioregion, 11 sites were added to the reference list. These sites primarily filled spatial gaps on big rivers and “core” portions of the xeric bioregion.

The outcome of this process yielded 90 reference sites in the mountains bioregion, 39 in the plains bioregion and 39 in the xeric bioregion. The total number of reference sites was 168. These reference sites were forwarded to Utah State to kick start the recalibration of the RIVPACS style OE macroinvertebrate model. During the stressed site screening process 3 xeric reference sites were removed thereby knocking the total number of reference sites in the xeric bioregion down to 36 and the grand total to 165. These sites will be provided to Tetra Tech for the recalibration of the Colorado MMI.

WQCD Stressed Site Screening Process

The stressed site screening process involved using the same human disturbance values derived during the reference site screening process and aerial image review on Google Earth. The goal was to take the remaining EDAS stations not selected as reference sites and run these sites through a screening process that would result in true, "bookend" stressed sites. To this end, the Division focused on avoiding any stressed sites that might be considered marginally stressed, an issue that likely impacted the first version of the CO multimetric index in 2005.

For each bioregion, 85th percentile values were calculated for each human disturbance parameter (i.e. irrigated agriculture, CAFO's, etc). Eighty-fifth percentile values varied across all human disturbance parameters and across all three bioregions. This was expected as human disturbance increases significantly in both the plains and xeric bioregions. The goal in this exercise was to flag values that exceeded acceptable conditions and could be best described as "unacceptable". Values that exceeded the 85th percentile were assigned a score of "1". Values that fell below the 85th percentile were assigned a score of "0". Scores were tallied on a spreadsheet for each site. The aggregated scores were stored in a spreadsheet and then sorted highest to lowest thus revealing sites that had "stressed" potential. The maximum aggregated score potential was 9 and the lowest 0. However, the highest score turned out to be 5, which occurred on several occasions. Scores that ranged from 3-5 were further examined in the aerial image phase.

To supplement the stressed candidate list, each human disturbance parameter was sorted from highest to lowest, thus revealing the upper echelon of values for each disturbance type. This was done in the event that one human disturbance parameter value was extremely high but other parameter values were negligible for any given site. For instance, at a particular site an irrigated agriculture value was 95% but all other parameters were okay. This site would have an aggregated score of 1 and would not be flagged as having cumulative disturbance. However, it would be flagged in this supplemental screening and reviewed further. There was no set system for how many sites were selected from each human disturbance parameter, but typically it ranged from the top 10 to the top 20.

Sites with cumulative high scores or with one obviously high value were reviewed by examining Google Earth aerial images on a desktop PC with internet access.
Each site and an upstream section of the waterbody (approximately 2-3 km) were examined for clear characteristics of stressed condition, such as high concentration of diversions, impoundments, road crossings, near stream grazing, mining, adjacent agriculture and riparian instability, to name a few. Only one person performed this initial image review. This provided a consistent and repetitive approach. Images were saved to PDF files and stored on a network drive.

On occasion, when a stressed determination was difficult to make from aerial orthophotos alone, the reviewer would examine certain tolerance and diversity metrics to help make sense of things, sort of a reality check. This was used sparingly and only as backstop to

prevent true, “bookend” stressed sites from being unnecessarily removed from the process. The initial set of candidate stressed sites were not selected using metric values, only human disturbance values and aerial orthophotos. Metric values were only used to substantiate a handful of candidate sites in the xeric bioregion where finding stressed sites was fairly onerous. While the Division knew that examining biological data is frowned upon in bioassessment tool circles, the inability to utilize water chemistry and habitat data as criteria left the Division with few screening options.

Once a list of candidate stressed sites was assembled the reviewer stepped away from the process for 3 weeks. After 3 weeks the reviewer took a fresh look at each candidate stressed site. Google images and human disturbance values were re-examined in order to fine-tune the final list of stressed sites. This was an iterative and image-intensive venture.

This process yielded 31 stressed sites in the mountains bioregion, 23 sites in the plains bioregion and 20 sites in the xeric bioregion. These sites will be used for the recalibration of the Colorado MMI.

Appendix B

Reference and Stressed Sites Used in Analysis

Appendix B: Reference and Stressed Samples

Station	Site Name	Sample_ Replicate	Collection Date	Latitude	Longitude	Ref-Biotype
Reference sites						
WCOP99-0502	Adams Fork Conejos River	2418_0	8/5/2000	37.3296	-106.6896	R-2
EPA01-247	Agate Creek	2173_0	7/17/2001	38.4510	-106.3410	R-2
USFS Avalanche	Avalanche Creek	2936_0	9/7/2005	39.2311	-107.1992	R-2
CO151M	Badger Creek	2285_0	8/22/1995	38.5942	-105.8364	R-1
7630	Bear Creek	3016_1	5/2/2007	37.6138	-104.7738	R-3
CO125M	Beaver Creek	2274_0	8/21/1995	37.9997	-108.1939	R-1
USFSPIKE4	Beaver Creek	2666_1	11/6/2003	39.0287	-104.9448	R-2
10262	Beaver Creek	3044_1	8/1/2007	38.5525	-107.0530	R-1
EPA01-246	Big Blue Creek	2172_0	7/16/2001	38.1760	-107.3870	R-2
10666	Billy Creek	3051_1	7/16/2007	38.2948	-107.7020	R-1
7285	Boggs Creek	327_1	3/21/2000	38.2294	-104.7462	R-3
9716A	Box Canyon Creek	2851_1	7/11/2006	37.4489	-108.2078	R-2
7164	Browns Creek	2682_1	7/27/2005	38.6690	-106.1610	R-2
5341	Cache La Poudre River	486_1	5/9/2002	40.7083	-105.7266	R-1
11206	Carr Creek	2885_1	9/13/2006	39.5665	-108.5017	R-1
USFS Cataract	Cataract Creek	2951_1	9/26/2006	39.8285	-106.3272	R-2
WCOP01-0777	Chacuaco Creek	2772_0	4/22/2003	37.4942	-103.6313	R-3
7512	Chico Creek	3013_1	4/30/2007	38.3577	-104.3873	R-3
WCOP03-R003	Chief Creek	2779_0	7/10/2003	40.1033	-102.3225	R-3
5775	Cook Creek	3064_2	8/16/2007	39.1817	-104.8968	R-1
EPA01-248	Cottonwood Creek	2174_0	7/18/2001	38.0567	-105.5790	R-2
7997b	Cottonwood Creek	2651_1	4/7/2004	37.1109	-103.0744	R-3
USFS Cross 4	Cross Creek	2953_0	10/2/2006	39.4841	-106.5039	R-2
USFS Cross 1	Cross Creek	2992_0	9/15/2005	39.5413	-106.4334	R-2
WCOP99-0633	Crystal River	2762_0	6/18/2003	38.7239	-106.6725	R-2
Cuchara	Cucharas River	461_1	4/5/2000	37.6797	-104.6520	R-3
EPA01-259	Deeds Creek	2184_0	8/8/2001	39.2979	-106.6590	R-2
7595	Del Agua Arroyo	3014_1	5/29/2007	37.3472	-104.5742	R-3
10700	Dolores River	34_1	10/19/1994	38.3711	-108.8020	R-3
WCOP99-0597	Dyer Creek	2453_0	5/21/2002	38.5928	-107.4472	R-2
7997a	East Carrizo Creek	2650_1	4/7/2004	37.1354	-103.0157	R-3
7999a	East Carrizo Creek	2652_1	4/7/2004	37.1685	-103.0345	R-3
USFS EFH Upper	East Fork Homestake Creek	2955_1	9/18/2006	39.3559	-106.4540	R-2
WCOP03-R009	East Fork Piedra River	2785_0	8/8/2003	37.4817	-107.0971	R-2
11765	East Miller Creek	2906_1	8/7/2006	39.8768	-107.7649	R-2
5779B	East Plum Creek	3043_1	8/14/2007	39.1843	-104.9307	R-2
WCOP99-0595	East Rifle Creek	2452_0	8/2/2001	39.6431	-107.7104	R-1
EPA01-242	El Rito Azul	2169_0	7/11/2001	37.2743	-106.6190	R-2
10551	Escalante Creek	2524_1	9/28/2004	38.7178	-108.2686	R-3

Station	Site Name	Sample_ Replicate	Collection Date	Latitude	Longitude	Ref-Biotype
EPA01-168	Fawn Creek	2161_0	8/11/2001	39.8525	-107.3150	R-2
CO162M	Fernleaf Gulch	2292_0	8/25/1995	38.4069	-105.6389	R-1
CO153M	Fourmile Creek	2286_0	8/24/1995	38.9661	-106.1397	R-2
5772A	Garber Creek	3062_1	8/16/2007	39.3552	-105.0272	R-1
CO116M	Garner Creek	2271_0	8/28/1995	38.1897	-105.7764	R-2
EPA01-250	Goose Creek	2176_0	7/22/2001	39.1856	-105.3910	R-2
USFS Gore 1	Gore Creek	3001_0	9/27/2004	39.6277	-106.2710	R-2
7226	Hardscrabble Creek	3009_1	5/30/2007	38.3434	-105.0682	R-3
7130	Hayden Creek	2736_1	9/14/2005	38.3356	-105.8022	R-1
CO03RS	Hope Creek	2250_0	9/21/1995	37.5531	-106.8022	R-2
EPA01-241	Huerfano River	2168_0	7/10/2001	37.6138	-105.4720	R-2
12961	Illinois River	84_1	9/23/1997	40.4624	-106.1770	R-1
5771	Jackson Creek	509_1	5/30/2003	39.3457	-104.9812	R-1
EPA01-262	Jefferson Fork	2187_0	8/11/2001	39.4626	-105.8580	R-2
CO072M	Junction Creek	2243_0	8/1/1994	37.3336	-107.9094	R-1
10570	Kannah Creek	3049_1	7/19/2007	38.9612	-108.2297	R-1
8337	La Jara Creek	2745_1	9/28/2005	37.1775	-106.2119	R-1
8715	La Manga Creek	2753_1	9/29/2005	37.1164	-106.3778	R-2
9680	La Plata River	2601_1	3/30/2005	37.2920	-108.0416	R-1
10906	La Sal Creek	3091_1	9/12/2007	38.3205	-108.9770	R-1
EPA01-261	Lake Fork	2186_0	8/10/2001	39.2853	-106.4490	R-2
10415	Leroux Creek	21_1	11/5/1996	38.8811	-107.7858	R-1
CO070M	Lime Creek	2242_0	8/4/1994	37.6770	-107.7509	R-2
12832	Little Bear Creek	2920_1	8/16/2006	40.6888	-107.4345	R-1
CO04RS	Little Cimarron Creek	2224_0	8/16/1994	38.2100	-107.4636	R-2
7385	Little Fountain Creek	2849_1	4/12/2006	38.6817	-104.8581	R-1
11408	Little Snake River	46_1	10/7/1998	40.9970	-107.1508	R-1
6753400	Lonetree Creek	2349_0	7/20/1994	40.8983	-104.8680	R-3
11535	Lost Creek	292_1	6/7/2000	40.0506	-107.4687	R-1
WCOP99-0503	Lost Man Creek	2419_0	7/27/2000	39.1595	-106.5718	R-2
12785A	Lost Man Creek	2864_1	9/10/2006	39.1292	-106.6180	R-2
CO051M	Lottis Creek	2227_0	8/17/1994	38.7725	-106.6225	R-2
USFS Meadow	Meadow Creek	3003_0	9/11/2003	39.5954	-106.1237	R-2
CO155M	Middle Fork Brush Creek	2288_0	9/6/1995	38.9542	-106.8583	R-2
10559	Middle Fork Escalante Creek	3075_1	7/18/2007	38.5812	-108.4059	R-1
WCOP99-0578	Middle Fork Little Snake R.	2448_0	8/14/2001	40.9715	-107.0190	R-1
CO038M	Middle Fork North Crestone	2214_0	8/11/1994	38.0361	-105.6425	R-2
10834	Naturita Creek	3086_1	9/11/2007	38.1591	-108.4031	R-1
EPA01-249	Newlin Creek	2175_0	7/19/2001	38.2662	-105.1898	R-2
WCOP99-0649	Noname Creek	2466_0	7/23/2002	39.5817	-107.2881	R-1
10558	North Fork Escalante Creek	3074_1	7/18/2007	38.6369	-108.4272	R-1
CO142M	North Anthracite Creek	2283_0	9/13/1995	38.9817	-107.1911	R-2

Station	Site Name	Sample_ Replicate	Collection Date	Latitude	Longitude	Ref-Biotype
EPA01-253	North Cottonwood Creek	2179_0	7/27/2001	38.8708	-106.2660	R-2
10917	North Fork Mesa Creek	3065_5	9/10/2007	38.5032	-108.7904	R-1
NPlatte	North Platte River	378_1	11/4/1997	40.6973	-106.4125	R-1
CO113M	North St. Vrain Creek	2269_0	9/14/1995	40.2047	-105.4061	R-2
USFS N Tenmile	North Tenmile Creek	2961_0	9/8/2005	39.5730	-106.1722	R-2
11462	Oliver Creek	2893_1	10/3/2006	40.8771	-107.0871	R-2
11463	Oliver Creek	2894_1	10/4/2006	40.8641	-107.0674	R-2
WCOP99-0518	Ouzel Creek	2433_0	7/10/2000	40.1998	-105.6258	R-2
USFSPIKE1	Pine Creek	2663_1	11/3/2003	39.2406	-105.2826	R-2
7170	Pine Creek	2686_1	7/26/2005	38.9988	-106.2318	R-2
USFS Piney 2	Piney River	2964_0	9/28/2005	39.7504	-106.4713	R-1
EPA01-238	Purgatoire River	2165_0	7/6/2001	37.5584	-103.6520	R-3
10350	Razor Creek	3047_1	8/2/2007	38.3846	-106.6733	R-1
USFS Ripple	Ripple Creek	2976_0	8/12/2004	40.0746	-107.3028	R-2
11208	Roan Creek	2887_1	9/13/2006	39.5096	-108.5248	R-1
12182	Roaring Fork Arapaho Creek	56_1	9/9/1997	40.1308	-105.7675	R-2
10980	Roc Creek	3093_1	9/10/2007	38.4418	-108.8774	R-3
EPA01-252	Rock Creek	2178_0	7/26/2001	39.3746	-105.6840	R-2
EPA01-260	Rock Creek	2185_0	8/9/2001	39.2169	-106.3980	R-2
7284	Rock Creek	3011_1	5/1/2007	38.2080	-104.7931	R-3
8395	Rock Creek	93_1	4/15/1993	37.4903	-106.2589	R-1
USFSPIKE2	Rough and Tumbling Creek	2664_1	11/4/2003	39.0360	-106.1159	R-2
5971c1	Rule Creek	2648_1	10/25/2001	39.0477	-105.0983	R-1
WCOP04-R008	Rush Creek	2794_0	6/9/2004	40.0994	-102.0974	R-3
7571	San Francisco Creek	2541_1	8/26/2004	37.1206	-104.2614	R-3
10805	San Miguel River	36_1	10/19/1994	38.1260	-108.2077	R-1
CO174M	Scott Gomer Creek	2297_0	9/11/1995	39.5081	-105.7047	R-2
CO067M	Silver Creek	2239_0	8/1/1994	37.4289	-106.7589	R-2
USFS Snowmass	Snowmass Creek	2978_0	9/8/2004	39.1801	-107.0220	R-2
EPA01-239	South Apache Creek	2166_0	7/7/2001	37.8522	-104.9850	R-2
5588	South Boulder Creek	569_1	6/24/2003	39.9805	-105.2208	R-1
12759	South Fork Fryingpan River	2861_1	9/12/2006	39.2372	-106.5900	R-2
CO122M	South Fork Saguache Creek	2273_0	8/23/1995	37.9192	-106.7153	R-2
WCOP99-0563	South Rush Creek	2434_0	5/17/2001	38.8867	-103.6830	R-3
AR0097	South Rush Creek	281_0	12/19/2000	38.8507	-103.5394	R-3
5548	South St. Vrain Creek	556_2	6/24/2003	40.0788	-105.5778	R-2
WCOP04-R009	Timpas Creek	2795_0	7/6/2004	37.8271	-103.7730	R-3
7560	Trinchera Creek	2540_1	8/26/2004	37.0446	-104.0510	R-3
CO133M	Trout Creek	2278_0	8/17/1995	37.6661	-107.0700	R-2
WCOP04-R003	Two Butte Creek	2789_0	5/19/2004	37.5134	-103.0267	R-3
WCOP99-0634	Ute Creek	2462_0	8/13/2002	37.5955	-105.3989	R-2
10922	Ute Creek	3092_2	9/13/2007	38.7237	-108.9097	R-1

Station	Site Name	Sample_ Replicate	Collection Date	Latitude	Longitude	Ref-Biotype
11485	Vermillion Creek	2897_1	8/15/2006	40.7197	-108.7518	R-1
EPA01-240	Wahatoya Creek	2167_0	7/9/2001	37.4123	-104.9640	R-2
8116	West Alder Creek	2703_1	8/22/2005	37.7050	-106.6478	R-1
USFS West Lake	West Lake Creek	2991_0	9/27/2006	39.5395	-106.6160	R-2
10905G	West Paradox Creek	3090_1	9/12/2007	38.3827	-108.9960	R-1
5777A	West Plum Creek	512_1	5/30/2003	39.3692	-104.9655	R-1
EPA01-251	Wigwam Creek	2177_0	7/25/2001	39.2470	-105.3580	R-2
WCOP04-R001	Willow Creek	2788_0	7/29/2004	40.8003	-104.4593	R-3
12802	Yampa River	168_1	8/28/2001	40.5013	-107.2643	R-1
WCOP04-R007	Yellow Jacket Creek	2793_0	6/3/2004	37.3642	-108.9511	R-3
Stressed Sites						
7701	Apishapa River	477_1	6/11/2002	38.0958	-103.9847	S-3
7186A	Arkansas River	2828_1	4/22/2006	39.2024	-106.3530	S-1
5762	Bear Creek	3082_1	9/17/2007	39.6432	-105.3072	S-1
WCOP01-0756	Big Dry Creek	2787_0	7/15/2003	40.0321	-104.8729	S-3
5417	Big Thompson River	275_0	5/21/2001	40.3943	-104.9643	S-3
12304b	Blue River	2570_1	5/12/2004	39.4865	-106.0463	S-2
5575	Boulder Creek	2692_1	8/12/2005	40.0511	-105.1810	S-1
12761B	Brush Creek	2862_1	9/7/2006	39.2247	-106.9210	S-1
AR0018	Buffalo Creek	287_0	12/20/2000	38.1487	-102.3110	S-3
12893	Burgess Creek	436_1	8/29/2001	40.4517	-106.8102	S-1
5306	Cache La Poudre River	359_1	6/17/2002	40.4227	-104.5980	S-3
12715	Castle Creek	388_1	10/23/2001	39.1958	-106.8329	S-1
6713500	Cherry Creek	2320_0	7/18/2002	39.7425	-105.0000	S-3
34	Clear Creek	2563_1	5/10/2005	39.8276	-104.9521	S-3
5605	Clear Creek	2587_1	5/10/2005	39.7742	-105.1413	S-3
10320	Cochetopa Creek	2520_1	10/13/2004	38.5171	-106.7862	S-1
12102	Colorado River	3094_1	10/15/2007	40.1083	-106.0036	S-1
WCOP99-0594	Colorado River	2451_0	8/9/2001	39.1817	-108.9331	S-3
12731	Crystal River	64_1	8/18/1998	39.4081	-107.2183	S-1
10908B	Dolores River	2720_1	9/28/2005	37.7938	-108.8181	S-1
12852	Dry Creek	518_0	4/16/2003	40.4942	-107.2675	S-1
9149480	Dry Creek	2394_0	8/5/1998	38.6458	-108.0490	S-1
12870	Fish Creek	298_0	8/29/2001	40.4665	-106.8247	S-1
7390	Fountain Creek	2850_1	4/13/2006	38.2853	-104.6034	S-3
12191	Fraser River	3030_1	5/31/2007	39.9903	-105.8297	S-1
9046530	French Gulch	2371_0	8/4/1998	39.4930	-106.0447	S-2
WCOP99-0648	Gunnison River	2465_0	8/26/2002	39.0050	-108.5064	S-3
WCOP99-0622	Hartman Draw	2456_0	6/19/2002	37.3659	-108.5933	S-1
8650	Kerber Creek	410_1	3/24/2000	38.2500	-105.9498	S-1
7187G	Lake Fork	2836_1	4/22/2006	39.2512	-106.3754	S-2

Station	Site Name	Sample_ Replicate	Collection Date	Latitude	Longitude	Ref-Biotype
5454	Little Thompson River	2612_1	6/14/2004	40.2945	-105.0588	S-3
6753990	Lonetree Creek	2354_0	8/6/2002	40.4425	-104.5888	S-3
CO050M	Noname Creek	2226_0	8/18/1994	38.7831	-106.4822	S-1
133	North Fork Clear Creek	2575_1	5/27/2004	39.7479	-105.3993	S-2
11601	Piceance Creek	2898_1	8/8/2006	39.9643	-108.2682	S-1
7580	Purgatoire River	154_1	6/24/1999	37.1842	-104.4882	S-3
9153290	Reed Wash	2398_0	11/12/1997	39.2114	-108.8037	S-3
12851	Sage Creek	498_1	4/16/2003	40.4833	-107.1705	S-1
7309	Sand Creek	2840_1	4/12/2006	38.8101	-104.7480	S-3
5212a	Sand Creek	2582_1	7/7/2004	39.7611	-104.8576	S-3
8375	Sangre de Cristo Creek	2751_1	9/26/2005	37.4282	-105.3949	S-1
11191b	South Canyon Creek	226_1	3/8/2000	39.5619	-107.4081	S-1
5965	South Fork South Platte River	2819_1	7/28/2006	38.9835	-105.8834	S-1
5009	South Platte River	552_1	5/7/2003	40.7161	-103.1257	S-3
5164	South Platte River	2579_1	7/7/2004	39.8313	-104.9480	S-3
6714000	South Platte River	2325_0	9/9/1998	39.7597	-105.0033	S-3
5951a	South Platte River	2815_1	7/29/2006	38.9001	-105.4668	S-1
WCOP99-0671	South Platte River	2766_0	5/13/2003	39.6631	-105.0038	S-3
7336	Spring Creek	2845_1	4/11/2006	38.8078	-104.7954	S-3
12336	Tenmile Creek	60_1	9/23/1998	39.5028	-106.1403	S-2
7501	Timpas Creek	2695_1	8/8/2005	37.9606	-103.7185	S-3
5971a3	Trout Creek	2639_1	10/28/2003	39.0881	-105.0969	S-1
10609	Uncompahgre River	27_1	8/21/1996	37.9897	-107.6475	S-2
12337	West Tenmile Creek	492_1	5/8/2003	39.5020	-106.1467	S-2
7293	Wildhorse Creek	372_1	3/21/2000	38.2732	-104.6365	S-3
8105	Willow Creek	406_1	3/23/2000	37.8484	-106.9245	S-1
12804	Yampa River	420_1	8/28/2001	40.4993	-107.5532	S-1
12806F	Yampa River	455_1	8/29/2001	40.4832	-106.8323	S-1
11732A	Yellow Creek	2904_1	8/9/2006	40.1181	-108.3609	S-1
9830	Yellow Jacket Creek	2625_1	3/17/2005	37.3280	-109.0447	S-3

Appendix C

Metric Descriptions and Discrimination Efficiencies

Appendix C: Metric Descriptions and Discrimination Efficiencies

Metric Category	Metric	Biotype 1	Biotype 2	Biotype 3	Trend*
Richness	ChiroTax	UnResp	71.4	76.0	--
Richness	ChiroTaxaSmr	UnResp	NA	NA	0
Richness	ColeoTax	64.3	71.4	68.0	---
Richness	ColeoTax_aSmr	NA	71.4	NA	-
Richness	CrMolTax	53.6	UnResp	40.0	+/-
Richness	DipTax	50.0	85.7	72.0	---
Richness	DipTaxaSmr	UnResp	NA	NA	0
Richness	EphemTax	53.6	100.0	28.0	---
Richness	EphemTax_aElv	75.0	NA	NA	-
Richness	EphemTaxaSmr	71.4	NA	NA	-
Richness	EPtax	67.9	71.4	28.0	---
Richness	EPtax_aElv	78.6	NA	NA	-
Richness	EPTTax	64.3	71.4	36.0	---
Richness	EPTTax_aElv	67.9	NA	NA	-
Richness	EPTTaxaSmr	60.7	NA	NA	-
Richness	InsectTax	78.6	100.0	92.0	---
Richness	InsectTax_aElv	75.0	NA	NA	-
Richness	InsectTaxaSmr	67.9	NA	NA	-
Richness	NonInsPT	78.6	57.1	64.0	+++
Richness	OligoTax	35.7	28.6	52.0	+++
Richness	OligoTax_aSmr	NA	NA	76.0	+
Richness	PlecoTax	78.6	71.4	UnResp	--
Richness	PlecoTax_aElv	82.1	NA	NA	-
Richness	PlecoTaxaSmr	75.0	NA	40.0	-
Richness	TanytTax	UnResp	85.7	56.0	--
Richness	TotalTax	64.3	100.0	88.0	---
Richness	TotalTax_aElv	64.3	NA	NA	-
Richness	TotalTaxaSmr	50.0	NA	NA	-
Richness	TrichTax	46.4	71.4	UnResp	--
Richness	TrichTax_aElv	57.1	NA	NA	-
Richness	TrichTaxaSmr	50.0	NA	NA	-
Richness	WormlikeTax	50.0	UnResp	UnResp	+
Composition	AmphPct	42.9	UnResp	UnResp	+
Composition	AmphPct_aElv	NA	NA	52.0	-
Composition	BivalPct	UnResp	UnResp	UnResp	0
Composition	ChironominiPct	50.0	UnResp	44.0	++
Composition	ChironominiPct_aSmr	NA	NA	52.0	+
Composition	ChiroPct	50.0	57.1	UnResp	++
Composition	ColeoPct	53.6	71.4	96.0	---
Composition	ColeoPct_aSmr	NA	71.4	NA	-
Composition	CorbPct	UnResp	UnResp	UnResp	0

Metric Category	Metric	Biotype 1	Biotype 2	Biotype 3	Trend*
Composition	CrCh2ChiPct	UnResp	UnResp	44.0	+
Composition	CrMolPct	67.9	UnResp	40.0	+/-
Composition	D_Mg	71.4	100.0	68.0	---
Composition	DiamesPct	35.7	42.9	UnResp	++
Composition	DiamesPctaSmr	32.1	NA	40.0	++
Composition	DipPct	46.4	57.1	UnResp	++
Composition	EphemPct	50.0	71.4	44.0	---
Composition	EPTPct	57.1	57.1	44.0	---
Composition	Evenness	67.9	85.7	52.0	---
Composition	GastrPct	53.6	UnResp	UnResp	+
Composition	GastrPct_aElv	NA	NA	40.0	-
Composition	GastrPct_aSmr	NA	NA	36.0	-
Composition	IsoPct	UnResp	UnResp	UnResp	0
Composition	NonInPct	64.3	57.1	52.0	+/-
Composition	NonInPct_aElv	NA	42.9	NA	-
Composition	OdonPct	UnResp	UnResp	UnResp	0
Composition	OdonPctaSmr	35.7	NA	NA	-
Composition	OligoPct	53.6	42.9	56.0	+++
Composition	Orth2ChiPct	32.1	42.9	28.0	+/-
Composition	Orth2ChiPct_aSmr	NA	NA	UnResp	0
Composition	PentanPct	UnResp	UnResp	68.0	-
Composition	PercDom01	50.0	42.9	64.0	+++
Composition	PlecoPct	82.1	57.1	UnResp	--
Composition	PlecoPct_aElv	NA	NA	40.0	+
Composition	PlecoPct_aSmr	NA	NA	44.0	-
Composition	Shan_2	67.9	100.0	56.0	---
Composition	SimpsonD	50.0	71.4	72.0	+++
Composition	TanytPct	UnResp	100.0	56.0	--
Composition	TanytPct_aSmr	NA	NA	52.0	-
Composition	Tnyt2ChiPct	UnResp	85.7	56.0	--
Composition	Tnyt2ChiPct_aElv	NA	NA	56.0	-
Composition	Tnyt2ChiPct_aSmr	NA	NA	52.0	-
Composition	TrichPct	53.6	57.1	UnResp	--
Composition	WCSpct	71.4	42.9	48.0	+++
Feeding	CllctPct	UnResp	UnResp	UnResp	0
Feeding	CllctTax	39.3	UnResp	UnResp	-
Feeding	FiltrPct	50.0	42.9	56.0	+/-
Feeding	FiltrPct_aElv	NA	42.9	NA	+
Feeding	FiltrPctaSmr	39.3	UnResp	NA	-
Feeding	FiltrTax	UnResp	UnResp	48.0	-
Feeding	FiltrTax_aSmr	NA	42.9	NA	-
Feeding	PredPct	39.3	57.1	64.0	---
Feeding	PredTax	53.6	85.7	64.0	---

Metric Category	Metric	Biotype 1	Biotype 2	Biotype 3	Trend*
Feeding	PreShrTax	60.7	85.7	64.0	---
Feeding	PreShrTaxaSmr	75.0	NA	NA	-
Feeding	ScrapPct	46.4	57.1	60.0	---
Feeding	ScrapTax	UnResp	57.1	56.0	--
Feeding	ScrapTax_aElv	53.6	NA	NA	-
Feeding	ShredPct	50.0	57.1	44.0	+/-
Feeding	ShredPctaSmr	UnResp	NA	NA	0
Feeding	ShredTax	57.1	42.9	36.0	---
Feeding	ShredTaxaSmr	71.4	NA	NA	-
Habit	BrrwrPct	42.9	57.1	32.0	---
Habit	BrrwrTax	UnResp	42.9	44.0	--
Habit	ClmbrPct	35.7	42.9	36.0	+/-
Habit	ClmbrPctaSmr	32.1	NA	NA	-
Habit	ClmbrTax	UnResp	UnResp	UnResp	0
Habit	ClmbrTaxaSmr	46.4	NA	NA	-
Habit	ClngrPct	57.1	71.4	56.0	---
Habit	ClngrPct_aSmr	UnResp	NA	60.0	-
Habit	ClngrTax	67.9	100.0	40.0	---
Habit	ClngrTax_aElv	75.0	NA	NA	-
Habit	ClngrTaxaSmr	60.7	NA	NA	-
Habit	ClnSprPct	50.0	UnResp	76.0	--
Habit	SprwlPct	35.7	71.4	72.0	+/-
Habit	SprwlTax	42.9	28.6	76.0	---
Habit	SwmmrPct	60.7	UnResp	56.0	--
Habit	SwmmrPctaSmr	UnResp	NA	NA	0
Habit	SwmmrTax	UnResp	42.9	UnResp	+/-
Tolerance	BeckBI	64.3	85.7	28.0	---
Tolerance	BeckBI_aElv	78.6	NA	NA	-
Tolerance	BeckBIaSmr	67.9	NA	NA	-
Tolerance	GeneraSensChir	28.6	57.1	88.0	---
Tolerance	GeneraSensChirPLNS	UnResp	42.9	76.0	--
Tolerance	GeneraSensFamPlns	67.9	42.9	84.0	---
Tolerance	GeneraTolChir	35.7	42.9	28.0	+/-
Tolerance	GeneraTolChir_aElv	UnResp	NA	NA	0
Tolerance	GeneraTolChiraSmr	39.3	NA	NA	+
Tolerance	GeneraTolChirPLNS	32.1	UnResp	48.0	++
Tolerance	GenTolerFamPLNS	32.1	UnResp	60.0	++
Tolerance	IntolTax	60.7	71.4	36.0	---
Tolerance	IntolTax_aElv	78.6	NA	NA	-
Tolerance	IntolTaxaSmr	64.3	NA	NA	-
Tolerance	newHBI	71.4	71.4	56.0	+++
Tolerance	PctSensChir	53.6	71.4	92.0	---
Tolerance	PctSensChirPLNS	39.3	57.1	88.0	---

Metric Category	Metric	Biotype 1	Biotype 2	Biotype 3	Trend*
Tolerance	PctSensFamPlns	64.3	42.9	88.0	+/-
Tolerance	PctTolChir	46.4	57.1	56.0	+++
Tolerance	PctTolChiraSmr	UnResp	NA	NA	0
Tolerance	PctTolChirPLNS	32.1	UnResp	72.0	++
Tolerance	PctTolChirPLNS_aSmr	NA	NA	52.0	+
Tolerance	PctTolerFamsPLNS	53.6	42.9	60.0	+++
Tolerance	PercBaetEph	39.3	42.9	32.0	+/-
Tolerance	PercBaetEph_aElv	NA	NA	32.0	+
Tolerance	PercBaetEph_aSmr	NA	NA	32.0	+
Tolerance	PercHydr2Tri	39.3	42.9	UnResp	+/-
Tolerance	PercHydr2TriaSmr	UnResp	NA	NA	0
Tolerance	PercIntol	71.4	71.4	72.0	---
Tolerance	PercTol	64.3	UnResp	48.0	++
Tolerance	PercTol_aSmr	NA	NA	40.0	+
Tolerance	TolerPT	71.4	57.1	52.0	+/-
Tolerance	TolerPT_aSmr	NA	NA	56.0	+
Tolerance	TolerTax	46.4	57.1	UnResp	+/-

*Trend symbols:

+++ = metric value increases with increasing stress in all three Biotypes
++ = metric value increases with increasing stress in two of three Biotypes
+ = metric value increases with increasing stress in one of three Biotypes
+/- = mixed metric responsiveness, by Biotype
0 = not responsive in any Biotype where the metric was tested
- = metric value decreases with increasing stress in one of three Biotypes
-- = metric value decreases with increasing stress in two of three Biotypes
--- = metric value decreases with increasing stress in all three Biotypes

Appendix D

Index Alternatives

Appendix D: Index Alternatives

Biotype 1

	DE_1	64.3	82.1	85.7	85.7	78.6	78.6	82.1	78.6	78.6	82.1
	Z-score	1.55	1.67	1.79	1.91	2.00	1.97	1.87	2.02	1.96	1.95
	Ref SD	16.94	15.52	17.95	18.36	16.57	15.40	14.67	14.19	14.35	14.08
	# of Categories	1 of 5	5 of 5	3 of 5	3 of 5	3 of 5	3 of 5	4 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	<0.80	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.86
	Remarks										
	Metric	Ix1-1	Ix1-2	Ix1-3	Ix1-4	Ix1-5	Ix1-6	Ix1-7	Ix1-8	Ix1-9	Ix1-10
Rch	InsectTax	1	2	3							
Rch	InsectTax_aElv				4	5	6	7	8	9	10
Rch	NonInsPT	1		3	4	5	6	7	8	9	10
Rch	EPTTax	1									
Rch	EPTTax_aElv		2	3	4	5	6	7	8	9	10
Cmp	EPTPct					5	6	7	8	9	10
Cmp	PlecoPct		2	3	4						
Cmp	ChiroPct		2				6	7	8	9	10
Tol	BeckBI_aElv		2	3	4	5	6	7	8		
Tol	IntolTax_aElv									9	
Tol	newHBI										10
FFG	ScrapPct		2					7	8	9	10
Hab	ClngrTax		2								
Hab	ClngrTax_aElv								8	9	10

	DE_1	78.6	75.0	78.6	78.6	78.6	78.6	78.6	78.6	78.6	78.6
	Z-score	2.05	1.87	1.81	1.76	1.81	1.90	1.77	2.00	2.02	1.99
	Ref SD	14.63	15.40	14.88	14.80	14.88	15.39	16.39	16.57	14.48	15.46
	# of Categories	3 of 5	3 of 5	4 of 5	4 of 5	4 of 5	4 of 5	4 of 5	3 of 5	5 of 5	5 of 5
	Redundancy	0.86	<0.80	0.93	<0.80	0.93	0.93	0.93	0.93	<0.85	<0.85
	Remarks										
	Metric	Ix1-11	Ix1-12	Ix1-13	Ix1-14	Ix1-15	Ix1-16	Ix1-17	Ix1-18	Ix1-19	Ix1-20
Rch	InsectTax_aElv	11	12	13	14	15	16	17	18	19	20
Rch	NonInsPT	11	12	13	14	15	16		18	19	20
Rch	EPTTax_aElv	11	12	13	14	15	16	17	18		
Rch	EphemTax_aElv									19	20
Rch	PlecoTax_aElv									19	20
Cmp	EPTPct	11	12	13	14	15	16	17	18	19	20
Cmp	ChiroPct	11	12	13	14	15				19	
Tol	BeckBI_aElv						16	17	18	19	20
Tol	IntolTax_aElv			13		15					
Tol	newHBI	11	12		14						
FFG	ScrapPct			13	14	15	16	17		19	20
Hab	ClngrTax_aElv	11								19	20

Biotype 1 cont.

	DE_1	85.7	78.6	78.6	82.1	82.1	78.6	78.6	89.3	75.0	82.1
	Z-score	1.99	2.09	1.97	2.15	2.06	1.98	1.91	2.09	2.02	1.99
	Ref SD	15.53	15.14	14.26	14.85	15.05	15.75	15.87	15.50	15.96	15.98
	# of Categories	3 of 5	4 of 5	5 of 5	4 of 5	5 of 5	3 of 5	4 of 5	2 of 5	3 of 5	2 of 5
	Redundancy	<0.85	<0.85	<0.85	0.86	0.86	0.86	0.86	<0.80	<0.80	<0.80
	Remarks										
	Metric	Ix1-21	Ix1-22	Ix1-23	Ix1-24	Ix1-25	Ix1-26	Ix1-27	Ix1-28	Ix1-29	Ix1-30
Rch	InsectTax_aElv	21	22	23	24	25	26	27	28	29	30
Rch	NonInsPT	21	22	23	24	25	26	27	28	29	30
Rch	EPTax_aElv				24	25	26	27	28	29	
Rch	EphemTax_aElv	21	22	23							30
Rch	PlecoTax_aElv	21	22	23							30
Cmp	EPTPct	21	22	23	24	25	26	27	28	29	30
Cmp	ChiroPct	21	22	23	24		26	27			
Tol	BeckBI_aElv	21	22		24	25					
Tol	newHBI			23			26	27		29	
FFG	ScrapPct			23		25		27			
Hab	ClngrTax_aElv		22	23	24	25					

	DE_1	82.1	78.6	82.1	85.7	92.9	89.3	82.1	78.6	89.3	82.1
	Z-score	1.97	1.89	2.16	2.21	2.23	2.10	2.04	2.20	2.33	2.44
	Ref SD	16.94	15.85	15.20	14.85	16.02	17.14	17.31	15.47	14.40	14.61
	# of Categories	3 of 5	4 of 5	5 of 5	4 of 5	4 of 5	5 of 5	5 of 5	5 of 5	5 of 5	4 of 5
	Redundancy	<0.85	<0.85	<0.85	<0.85	0.93	0.93	0.88	<0.80	<0.85	<0.85
	Remarks			Summer Adjust						Concept Redund.	Concept Redund.
	Metric	Ix1-31	Ix1-32	Ix1-33	Ix1-34	Ix1-35	Ix1-36	Ix1-37	Ix1-38	Ix1-39	Ix1-40
Rch	InsectTax_aElv	31	32	33	34	35					40
Rch	NonInsPT	31	32	33	34	35				39	40
Rch	EPTTax_aElv					35					
Rch	EPTax_aElv						36	37	38		
Rch	EphemTax_aElv	31	32								
Rch	PlecoTax_aElv	31	32								
Cmp	EPTPct	31	32								
Cmp	PlecoPct			33	34	35					
Cmp	D_Mg			33	34						
Cmp	NonInPct						36	37	38	39	40
Cmp	WCSpct			33	34	35					
Tol	BeckBI_aElv	31	32	33	34	35	36			39	40
Tol	IntolTax_aElv							37			
Tol	newHBI								38		
FFG	ScrapPct		32								
FFG	PreShrTax						36	37	38	39	
FFG	PreShrTaxaSmr			33							
Hab	ClngrTax_aElv			33	34	35	36	37	38	39	40

Biotype 1 cont.

	DE_1	85.7	89.3	89.3	78.6	82.1	82.1	85.7	78.6	85.7	82.1
	Z-score	2.13	2.24	2.26	2.02	2.16	2.04	2.15	2.34	2.06	2.11
	Ref SD	16.23	15.66	16.34	17.00	17.63	15.07	15.57	13.81	13.21	15.25
	# of Categories	5 of 5	5 of 5	4 of 5	4 of 5	4 of 5	5 of 5	5 of 5	5 of 5	4 of 5	5 of 5
	Redundancy	<0.85	0.86	0.86	0.86	0.86	<0.85	0.86	<0.80	<0.80	0.86
	Remarks		Concept Redund.	Concept Redund.				Summer Adjust			
	Metric	Ix1-41	Ix1-42	Ix1-43	Ix1-44	Ix1-45	Ix1-46	Ix1-47	Ix1-48	Ix1-49	Ix1-50
Rch	InsectTax_aElv	41				45					
Rch	NonInsPT		42	43	44		46	47	48	49	50
Rch	EPTax_aElv		42	43	44	45		47	48		50
Cmp	EPTPct				44		46				
Cmp	ChiroPct							47	48	49	50
Cmp	NonInPct	41	42	43							
Tol	BeckBI_aElv	41	42	43	44	45	46	47			50
Tol	PercTol								48		
FFG	PreShrTax	41	42				46		48	49	50
FFG	PreShrTaxaSmr							47			
Hab	ClngrTax_aElv	41	42	43	44	45	46	47	48	49	50

	DE_1	78.6	78.6	85.7	82.1	92.9	92.9	85.7	78.6	85.7
	Z-score	1.90	1.93	2.12	2.06	2.19	2.25	2.24	2.29	2.14
	Ref SD	17.09	16.52	14.17	14.16	14.04	13.21	13.55	14.40	14.66
	# of Categories	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	0.86	0.86	<0.85	<0.80	<0.80	<0.80	<0.80	<0.80	<0.80
	Remarks						**CDPHE choice**			
	Metric	Ix1-51	Ix1-52	Ix1-53	Ix1-54	Ix1-55	Ix1-56	Ix1-57	Ix1-58	Ix1-59
Rch	InsectTax				54					
Rch	NonInsPT	51	52	53	54	55	56	57	58	59
Rch	EPTax_aElv	51	52		54	55	56	57	58	59
Cmp	Shan_2		52							
Cmp	EPTPct				54			57	58	59
Cmp	ChiroPct			53	54	55	56			
Cmp	D_Mg	51								
Tol	BeckBI_aElv	51	52	53	54					
Tol	PercTol								58	
Tol	GenSensFamPlns					55				59
Tol	PctSensFamPlns						56	57		
FFG	ScrapPct				54					
FFG	PreShrTax	51	52	53		55	56	57	58	59
Hab	ClngrTax_aElv	51	52	53	54	55	56	57	58	59

Biotype 2

	DE_1	100	100	100	85.71	85.71	100	100	100	100	100
	Z-score	3.27	2.83	2.94	2.97	2.91	2.89	2.92	2.93	2.88	2.79
	Ref SD	13.46	14.21	14.28	13.05	13.11	14.55	13.45	13.19	12.86	13.13
	# of Categories	3 of 5	4 of 5	4 of 5	3 of 5	4 of 5	4 of 5	5 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	<0.80	<0.80	0.92	0.92	0.92	<0.80	<0.80	0.92	<0.80	<0.80
	Remarks										
	Metric	Ix2-1	Ix2-2	Ix2-3	Ix2-4	Ix2-5	Ix2-6	Ix2-7	Ix2-8	Ix2-9	Ix2-10
Rch	TotalTax	1	2	3			6	7	8	9	10
Rch	InsectTax				4	5					
Rch	NonInsPT				4	5					
Rch	EPTTax			3	4	5			8		
Rch	EphemTax		2				6	7		9	10
Rch	TrichTax		2				6	7		9	10
Rch	PlecoTax		2								
Cmp	EphemPct							7	8	9	10
FFG	PreShrTax		2	3		5	6	7	8	9	10
Tol	BeckBI	1	2	3	4	5	6	7	8		10
Tol	newHBI									9	
Hab	ClngrTax	1	2	3	4	5	6	7	8	9	
Hab	ClngrPct										10

	DE_1	100	100	100	100	100	100	100	85.7	100
	Z-score	2.99	3.13	3.06	2.98	2.95	3.20	3.08	2.78	3.20
	Ref SD	14.01	13.91	12.86	12.75	12.88	12.90	13.77	12.58	12.84
	# of Categories	4 of 5	3 of 5	4 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	<0.80	0.92	0.92	<0.80	<0.80	<0.80	<0.80	<0.85	<0.80
	Remarks				**CDPHE choice**					Concept Redund.
	Metric	Ix2-11	Ix2-12	Ix2-13	Ix2-14	Ix2-15	Ix2-16	Ix2-17	Ix2-18	Ix2-19
Rch	TotalTax	11	12	13	14					19
Rch	InsectTax					15	16	17	18	
Rch	EPTTax		12	13						
Cmp	EphemPct			13	14	15			18	
Cmp	Shan_2*						16			19
Cmp	D_Mg*							17		
FFG	PreShrTax	11			14	15	16	17	18	19
Tol	BeckBI	11	12	13	14	15	16	17		19
Tol	newHBI								18	
Hab	ClngrTax	11	12	13	14	15	16	17	18	19

Biotype 3

	DE_1	96.0	100.0	96.0	100.0	96.0	100.0	96.0	96.0	96.0	96.0
	Z-score	1.62	2.09	2.20	2.11	2.29	2.18	2.11	2.25	2.36	2.46
	Ref SD	14.34	12.09	11.09	12.16	11.37	12.13	12.77	11.13	11.04	10.49
	# of Categories	3 of 5	4 of 5	4 of 5	5 of 5	4 of 5	4 of 5	4 of 5	4 of 5	4 of 5	4 of 5
	Redundancy	<.8	<.8	<.8	<.8	<.8	<.8	<.8	0.84	0.84	0.84
	Remarks				Col metric						
	Metric	Ix3-1	Ix3-2	Ix3-3	Ix3-4	Ix3-5	Ix3-6	Ix3-7	Ix3-8	Ix3-9	Ix3-10
Rch	TotalTax	1	2	3	4				8		
Rch	InsectTax					5	6	7		9	10
Rch	NonInsPT					5					10
Rch	DipTax								8	9	10
Tol	PctSensFamPlns	1	2	3	4	5	6	7	8	9	10
Tol	TolerPT			3							
Cmp	ColeoPct				4						
FFG	PredPct	1	2	3	4	5	6	7	8	9	10
Hab	SprwlPct		2	3	4	5	6		8	9	10
Hab	ClnSprPct							7			

	DE_1	96.0	100.0	100.0	100.0	96.0	100.0	100.0	100.0	100.0	100.0
	Z-score	2.75	3.23	3.16	3.01	2.38	2.51	3.21	2.63	2.14	2.56
	Ref SD	9.51	8.75	9.15	9.34	12.06	10.18	8.75	10.05	14.12	10.47
	# of Categories	5 of 5	4 of 5	4 of 5	4 of 5	4 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	<.8	<.8
	Remarks		Col metric	Col metric	Col metric		% Pred	Col metric	Redund.		% Pred
	Metric	Ix3-11	Ix3-12	Ix3-13	Ix3-14	Ix3-15	Ix3-16	Ix3-17	Ix3-18	Ix3-19	Ix3-20
Rch	TotalTax				14	15	16	17			
Rch	InsectTax	11	12	13					18	19	20
Rch	NonInsPT	11	12								
Rch	DipTax	11	12	13	14	15	16	17	18		
Rch	ColeoTax		12	13	14			17			
Tol	PctSensFamPlns	11	12	13	14		16	17	18	19	20
Tol	GenSensFamPlns					15					
Cmp	PercDom01	11					16	17	18	19	20
FFG	PredPct	11	12	13	14	15	16	17	18		20
FFG	PreShrTax									19	
Hab	SprwlPct	11	12	13	14	15	16	17	18	19	20

Biotype 3 cont.

	DE_1	100.0	92.0	100.0	100.0	100.0	100.0
	Z-score	2.33	2.13	2.28	2.68	2.36	2.25
	Ref SD	13.12	13.97	12.84	9.86	12.53	14.36
	# of Categories	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5	5 of 5
	Redundancy	<.8	0.81	<.8	<.8	<.8	<.8
	Remarks			**CDPHE choice**	% Pred		
	Metric	Ix3-21	Ix3-22	Ix3-23	Ix3-24	Ix3-25	Ix3-26
Rch	InsectTax	21	22	23	24	25	26
Rch	NonInsPT			23	24	25	26
Tol	PctSensFamPlns	21	22	23	24	25	
Tol	GenSensFamPlns						26
Cmp	PercDom01	21	22	23	24	25	26
FFG	PredPct				24		
FFG	PreShrTax	21	22	23		25	26
Hab	SprwlPct			23	24		26
Hab	SprwlTax		22				
Hab	ClnSprPct	21				25	

APPENDIX B

Benthic Macroinvertebrate Sampling Protocols

Water Quality Control Division
Standard Operation Procedure
WQCDSOP-001
May 2010

1. Overview

The use of benthic macroinvertebrates for assessing and monitoring the condition of lotic systems has become increasingly widespread and acceptable in the domain of Colorado's water quality standards setting. Macroinvertebrates are particularly suitable indicators of the condition of lotic systems as they are found in almost all freshwater environments, are easy to sample and identify, and different taxa show varying degrees of sensitivity to pollution and other impacts (Boothroyd & Stark 2000). The recent advent of statewide multimetric indices or the "bioassessment indicator tool" necessitates supplementary macroinvertebrate data from which to support the use of this indicator tool within the Water Quality Control Division's assessment methodology.

2. Scope and Applicability

This standard operation procedure describes semi-quantitative methods for collecting a single aquatic benthic macroinvertebrate sample from perennial, wadeable streams.

Perennial is defined as a well-defined channel that contains water year round during a year of normal rainfall with the aquatic bed located below the water table for most of the year. A perennial stream exhibits the typical biological, hydrological, and physical characteristics commonly associated with the continuous conveyance of water.

Wadeable is defined as a waterbody that can be safely traversed when collecting samples. Separate protocols are provided for hard-bottomed and soft-bottomed streams. Benthic macroinvertebrate data collected on unwadeable large rivers or intermittent type streams are beyond the scope of this procedure.

2.1 Index Period

The index period is the period of time that samples should be collected to minimize seasonal variation. The standard index period utilized by the Water Quality Control Division is summer to early fall, namely July 1 to October 1. This period is congruent with the central tendency of sample dates of macroinvertebrate replicates used to regionally calibrate the multimetric indices.

Eastern plains, as defined by the boundaries of EPA Level III Ecoregions Western High Plains and Southwestern Tablelands within Colorado, and Western xeric streams, as defined by those streams

with elevations less than 1500 meters within EPA Level III Ecoregions Colorado Plateaus and Arizona/New Mexico Plateau within Colorado, may be sampled from May 1 to October 1.

Unforeseen severe runoff or summer drought conditions may predispose WQCD planners to employ alternative sampling schedules that deviate from the standard index period listed above (see Section 2.2).

2.2 Sampling Frequency

This protocol recommends that benthic macroinvertebrate samples be collected once per year and within the standardized index period provided in Section 2.1.

If a second sample is taken, it should be collected before runoff (May) in the mountains streams and late summer (August-September) for plains/xeric streams. The spring sample in the mountains will provide a clearer representation of the stonefly fauna and the late summer sample in the plains/xeric streams will supplement the annual picture especially for some mayflies and midges. The fauna of plains/xeric streams may appear depauperate by late fall.

2.3 Site Selection

The study reach length should be one of the following: 1) 20 times the bankfull width of the wadeable waterbody or 2) long enough to encompass multiple riffles/runs (for hard- bottomed sites) or glides/pools/microhabitats (for soft-bottomed sites) from which to produce a single, representative sample from the predominant habitat type. The study reach should be representative of the typical habitat conditions that occur at or immediately above and below the greater stream segment.

Riffle habitat refers to the portions of the stream where moderate velocities and substrate roughness produce turbulent conditions which break the surface tension of the water and may produce whitewater (Bain and Stevenson, eds. 1999). Run habitat refers to the portions of the stream where there are moderate velocities, but lack the turbulent conditions that break the surface tension of the water (Bain and Stevenson, eds. 1999). A glide generally refers to a calm stretch of shallow water flowing smoothly.

Although riffle/run areas with hard-bottomed substrates are generally the most diverse and productive habitat type in mountain streams, these may not be entirely representative of the overall types of habitat present within the study reach. Alternately, although glide/pool areas with soft-bottomed substrates are generally the most diverse and productive habitat type in plains and plateau streams, these may not be entirely representative of the overall types of habitat present within the study reach.

There are some advantages to taking samples in or near the thalweg. Especially in small streams, the thalweg portion of the riffle usually has larger and cleaner substrate, better food supply and more reliable flow. When the thalweg is not or cannot be sampled, attention must be paid to the recent history of flow. Many Colorado streams are subject to flow variation on a short time scale due to flow regulation. Substrate that has been inundated only recently or that is inundated only occasionally should not be sampled because it is unlikely to support many specimens.

2.3.1 Hard-Bottomed Streams

A hard-bottomed stream is one where the stream substrate is dominated by particles gravel size or larger. Riffle/run habitats are common in these high to moderate gradient streams. Gravel, cobble and boulder sized substrate are frequent in these streams. These types of streams are conducive to the single habitat approach described in Section 2.4.

2.3.2 Soft-Bottomed Streams

Soft-bottomed streams are usually low gradient, often found in the Eastern Plains and in the far western xeric plateaus of Colorado, and are dominated by glide/pool habitats. The dominant substrate is sand, silt, clay or mud. Gravel, cobble and boulder sized substrate are naturally rare or entirely absent in these low gradient streams. These types of streams are conducive to the multihabitat approach described in Section 2.4.

2.4 Sample Collection Information

There are two macroinvertebrate sampling procedures used by the WQCD to collect the required single representative sample within a stream reach:

1) Semi-quantitative sample collection of hard-bottomed streams that focuses collecting macroinvertebrates from riffle/run habitats. These samples are collected using a modified kick net.

2) Semi-quantitative sample collection of soft-bottomed streams that focuses collecting macroinvertebrates from non-riffle/run habitats, such as vegetated bank margins, submerged woody debris or snags and aquatic macrophytes. These samples are generally collected using a jab or sweeping technique that utilizes the same modified kick net.

Semi-quantitative sample collection methods are designed to collect the widest variety of aquatic macroinvertebrates available at the study reach. For these methods, it is not necessary to know the exact area sampled. Both procedures are suitable for use with both relative abundance and fixed count processing protocols from which a variety of species richness and relative abundance metrics (Stark et al 2001) and multimetric predictive model analysis can be calculated.

In hard-bottomed streams or those streams predominated by substrate greater than gravel size, a single sample shall consist of a one-minute timed sample collected over an area of one square meter (1 m^2). The investigator shall select a single riffle or run from within the study reach that represents the predominant velocity and substrate type.

In soft-bottomed streams or those streams predominated by substrate smaller than sand size, a single sample shall consist of several individual sweeps or jabs collected from a fixed area of approximately 1 m^2. The multihabitat sampling effort is limited to 1 minute. Time spent traversing from one habitat type to another is not included in the total time.

If the predominant habitat type expected to occur at the site does not occur along the defined study reach, then the investigator should specify some other stable and productive habitat type to sample. In circumstances where hard-bottomed or cobble substrates represent less than 20% of the study reach, multihabitat(s) will need to be sampled. Alternatively, in circumstances where hard-bottomed or cobble substrates represent greater than 80% of the study reach in plains/xeric streams, then a riffle/run sample will need to be collected.

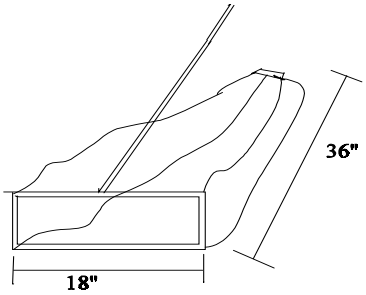
2.5 Equipment and Supplies

- Sampler with 500 to 600-µm mesh collection bag. Suggested: Kick-net with long handle, rectangular frame (18" x 8") dip-net, and 500-µm mesh nylon bag.
- Sieve dolphin bucket (504 µm mesh)
- 1-liter wide-mouth sample jars with screw tight lids
- 95% ethanol stored in sealed and labeled polyurethane carboys or bottles
- 1-liter rinse bottle
- 5-gallon bucket with handle
- Dissecting tray
- Rectangular "*Rite in the Rain*" labels
- Standard #2 pencil(s)
- Fine-tip forceps
- Waterproof stop watch
- Number 30 (600 µm) **or** 35 (500 µm) standard sieve
- Tape measurers (100 ft)
- 48 quart or larger ice chests or sealable Rubbermaid totes
- Geographic Positioning System (GPS) unit set to NAD 83 or WGS 84
- De Lorme Gazetteer map of Colorado
- Safety glasses
- Rubber gloves
- Hip or chest waders with wading boots
- Pertinent USGS topographic quad maps
- Applicable field sheets

2.5.1. Kick Net Specifications

The kick net is comprised of the following components:

1) 18" x 8" rectangular frame
2) 500 to 600 µm mesh nylon bag with canvas reinforced bottom and shroud reinforced opening
3) 1 or 2 piece long handle (≈ 70 inches long)
4) Sieve dolphin bucket (504 µm mesh)
5) Sieve bucket adaptor



3. Sample Collection

The following section discusses procedures used to collect benthic macroinvertebrate samples in perennial, wadeable streams.

3.1 Riffle or Run Habitat Method

1) Ensure that the sampling net and sieve bucket are clean prior to usage.
2) Select the dominant riffle or run habitat within the study reach according to Section 2.4.
3) Place the net frame flush to the streambed with the frame open to the upstream flow. Check that the nylon bag and sieve bucket are freely floating immediately downstream of the net frame. This will ensure that once the substrate is disturbed that specimens will be directed through the nylon bag and into the capture sieve bucket.
4) Carefully lower the handle forward in an upstream direction until the sampling net is nearly horizontal to the water surface but the net frame is still flush to the streambed. The point at which the tip of the handle extends along the streambed is the point at which the kicking activity will cease. This distance multiplied by the width of the net frame equals one square meter. Return the handle to its vertical position.
5) Position yourself next to sampling net and begin to disturb the substrate immediately upstream of the net. Disturb the substrate using the heel of your boot or entire foot by kicking to dislodge the upper layer of cobbles or gravel and to scrape the underlying bed. The area disturbed should extend no further than the point delineated in Step 4 and not exceed 1 minute. Approximately 0.25 meters should be disturbed for every 15 seconds.
6) Larger cobble may be scraped by hand, if necessary, to remove specimens. Cobble should be scraped clean quickly and efficiently as the scraping is counted within the one minute time frame.
7) Transfer material (matrix of specimens and insubstantial amount of stream substrate/detritus) from the interior of the net and sieve bucket into the sample jar and wash or pick all specimens off the net interior. Specimens that cling to the exterior of the net are not considered part of the sample. They may be removed and placed back into the stream.
8) Release back into the stream any fish, amphibians, reptiles or crayfish/rusty crayfish caught in the net.
9) If excessive or large debris items are present refer to Section 3.4.2.
10) The kick-net should be rinsed clean by backwashing with site water before collecting additional samples.
11) At this point refer to Section 3.4 Sample Processing.

3.2 Multihabitat Method

1) Ensure that the sampling net and sieve bucket are clean prior to usage.
2) Sample multiple habitats, as defined below, using the following procedures. The design is to sample an equivalent of a 1 meter sweep across multiple non-riffle/run habitats. Avoid dredging the kick net through mud or silt and clumps of leafy detritus or algal material. Also avoid hard-bottomed substrates as those habitats will be sampled separately according to Section 2.4.

Woody Debris or Snags – Jab the kick net into an area of submerged and partially decayed woody debris to dislodge specimens, followed by 1-2 "cleaning" sweeps through the water column to capture specimens in the water column. Scrub larger debris by hand over the opening of the kick net. The area of the larger debris should be included in the one meter unit effort.

Bank Margins – Locate an area of bank within the study reach. Jab the kick net vigorously into the bank for a distance of 1 meter to dislodge specimens, followed by 1 to 2 "cleaning" sweeps to collect specimens in the water column.

Aquatic Macrophytes – Sweep the kick net through submerged or emergent vegetation for a distance of 1 meter to loosen and capture specimens, followed by 1 to 2 "cleaning" sweeps to collect specimens in the water column.

3) Transfer material (matrix of specimens and insubstantial amount of stream substrate/detritus) from the interior of the net and sieve bucket into the sample jar and wash or pick all specimens off the net interior. Specimens that cling to the exterior of the net are not considered part of the sample. They may be removed and placed back into the stream.
4) Release back into the stream any fish, amphibians, reptiles or crayfish/rusty crayfish caught in the net.
5) If excessive or large debris items are present refer to Section 3.4.2.
6) The kick-net should be rinsed clean by backwashing with site water before collecting additional samples.
7) At this point refer to Section 3.4 Sample Processing.

3.3 Field Duplicates

One out of ten (10%) sample events shall include a duplicate field sample to ensure quality control (QC). For example, when a biosurvey consists of collecting benthic macroinvertebrates at 10 stations, then 1 out of the 10 stations shall include a duplicate field sample. It is acceptable to increase the rate of duplicate field samples (QC>10%). However, it is unacceptable for the rate to fall below 10%.

The duplicate field sample shall be collected within the same habitat type and in close proximity to the standard field sample and in a manner consistent with procedures set forth in Sections 3.1 or 3.2.

3.4 Sample Processing (On-site)

Sample processing is characteristically conducted in the field, generally on the bank of the stream being sampled. Sample processing consists of excessive material or large debris item removal and rinsing, elutriation (if necessary), preservation and storage.

3.4.1 Removing Excessive and Large Debris Items

Picking and rinsing should be performed in a Number 30 (600 µm) or 35 (500 µm) standard sieve. Rinse off and remove any excessive debris such as algal clumps or large debris items such as leaves,

sticks, or rocks that will not fit into a 1-liter sample jar or will lessen the effectiveness of the preservative. Calmly rinse the debris with stream water over the sieve opening using care not to cause unnecessary splattering of material. Examine larger debris to ensure that all specimens have been thoroughly rinsed or scraped into the sieve. Discard the material.

Transfer the remaining sample matrix in the sieve to a 1-liter wide-mouth polyethylene sample jar. Each sample jar should be no more than 1/2 full of sample material. Consequently, splitting the sample into two or more sample jars is acceptable. See Section 3.4.4 for labeling split samples.

3.4.2 Elutriation

Elutriation is a technique used to extract specimens from excessive substrate that has been captured during the sample collection process. This technique works best when the substrate is comprised of fines, sands and pebbles and should be used in circumstances when the amount of substrate is disproportionate to the amount of the detritus/specimen matrix. This step follows the removal of large debris items detailed in Section 3.4.1.

Keeping the sample in the 5-gallon bucket, add stream water to the bucket. Gently swish the sample around in the bottom of the bucket to liberate organic material and macroinvertebrates from the substrate. Pour the water and all floating material and specimens into a Number 30 (600 µm) or 35 (500 µm) standard sieve. This process may not work for heavy invertebrates such as snails, larger annelids or case-building caddis flies that use sand. Continue rinsing in similar fashion 2-3 more times to maximize retention of specimens collected. If it appears that the heavy invertebrates are not being separated from the substrate, pour the remaining sample in the bucket into a tray and spread the sample homogenously across the bottom of the tray. Use forceps to remove remaining specimens and place them into the sieve.

Transfer the remaining sample matrix in the sieve to a 1-liter wide-mouth polyethylene sample jar. Each sample jar should be no more than 1/2 full of sample material. Consequently, splitting the sample into two or more sample jars is acceptable. See Section 3.4.4 for labeling split samples.

3.4.3 Sample Preservation

Sample preservation is very important to ensure the integrity of the benthic organisms collected from the site. The sample is preserved by decanting as much remaining water as possible and completely filling the sample container with 95% ethanol (ETOH). Gently invert the sample jar several times to thoroughly homogenize the sample and preservative. This will make certain that the entire sample is preserved. Poorly preserved specimens can impede the identification and enumeration process. Any liquid leaking from the jar lid with the bottle inverted indicates an incomplete seal.

Allowing for dilution with water remaining in the sample container, the minimum ethanol concentration should always be greater than 70%. If in doubt, or with samples containing a large amount of organic material, the ethanol should be decanted after initial preservation and replaced with fresh 95% ethanol. In general, the volume of the container should contain no more than 50% of the sample.

3.4.4 Labels

Add pre-printed, moisture resistant labels to both the inside and outside of the sample container. Affix the label to the outside using transparent packaging tape. The following information should be recorded with a pencil on each label and placed in each sample container:

- Site number
- Stream name
- Stream description
- Date
- Indicate if kicknet was used
- Habitat type sampled
- Collector's initials
- Indicate if sample is a duplicate
- Indicate if sample is split

If splitting the sample among several containers, label appropriately to indicate that the sample has been split (e.g., Sample 1 of 2 and Sample 2 of 2). If not, simply label sample as 1 of 1.

3.4.5 Storage

Place the sample jars in a hard-cased ice chest or equivalent container for transport to the laboratory. Ensure that jar lids are thoroughly tightened to eliminate leakage and fumes from developing inside vehicle cargo holds or truck camper shells.

4. Invasive Species

Precautions must be taken to avoid collecting invasive species from areas where the Colorado Division of Wildlife ("CDOW") has issued urgent closures to the taking of those species. Those species captured incidentally and prohibited under order of the Director of the Colorado Department of Natural Resources must either be immediately returned to the waterbody, where it was captured, or immediately killed.

Further precautions must be taken to avoid inadvertently transferring invasive species from one waterbody to another waterbody. This is best accomplished by following appropriate disinfection procedures of personal gear (i.e. waders/boots) as prescribed by CDOW.

5. References

Bain, M. B., and Stevenson, N. J., eds. (1999). Aquatic habitat assessment: Common methods. American Fisheries Society, Bethesda, MD.

Boothroyd, I. K. G; Stark, J. D. 2000: Use of invertebrates in Monitoring In: New Zealand Stream Invertebrates: Ecology and implications for management. Collier K. & Winterbourn, M. J. eds. New Zealand Limnological Society, Hamilton. Pp. 344-373.

Stark, J. D.; Boothroyd, I. K. G; Harding, J. S.; Maxted, J. R.; Scarsbrook, M. R. 2001: Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment. Sustainable Management Fund Project No. 5103. 57 p.

APPENDIX C

STANDARD OPERATING PROCEDURE FOR LABORATORY IDENTIFICATION AND ENUMERATION

Preserved macroinvertebrate samples and chain of custody records are delivered to the laboratory whereby all samples received are verified against the chain of custody upon arrival. The laboratory randomly chooses 10% of the received samples for quality assurance/quality control (QA/QC).

Samples are prepared for identification by sorting the macroinvertebrates from sediment and debris, and placing the organisms into labeled vials containing ethanol. Samples are rinsed to remove ethanol preservative and placed into processing trays. The processing trays are marked with a grid pattern totaling 15 cells that are the same size and are numbered. The sample is agitated in the tray to evenly distribute the material among the grids.

All 15 grids are processed (a whole sample) unless excessive numbers of organisms are present. If excessive numbers are found, one of the grids chosen at random will be picked (1/15 subsample fraction). If at least 300 organisms are found after picking the entire grid subsample, picking is considered complete. If less than 300 organisms are found, processing will continue by picking another randomly chosen grid subsample. Additional grids will be picked, if necessary, until the 300-organism count is reached. If the 300-organism count is reached during the sorting of any single grid subsample, picking is continued until that entire fraction has been picked. All material from each randomly selected grid is removed from the tray and transferred to a large picking tray for processing.

All organisms processed in the subsample are placed in a separate vial and labeled according to the number of grids picked. When sub-sampling, the entire sample is also processed for any large or rare organisms, which are placed in a separate vial labeled 100% Large/Rare. Vials are filled with an ethanol preservative.

The taxonomist empties a vial into a glass dish and places it under a dissecting microscope. All individuals are counted and identified to the lowest possible taxonomic level using the most current taxonomic literature and returned to the vial containing the ethanol preservative. Both larvae and pupae Chironomids (midges) are identified to genus or species group using the appropriate taxonomic references. In certain instances, early instars or specimens in poor condition may only be identified to subfamily or tribe. Oligochaetes (worms) are identified to the lowest possible taxonomic level using the appropriate taxonomic references. Identifications and counts are recorded on data benchsheets.

A QA/QC check is performed on 10% of the picked/sorted samples. This involves a second person checking over the processed/picked samples to ensure that no more than 5% of the total organisms originally picked from a sample are missed. If more than 5% of the total organisms that were originally picked are found, samples are then re-picked until less than 5% of the original total are missed. Identifications and organism counts are verified by another taxonomist

to ensure data accuracy. When necessary, specimens of a particular taxonomic group are sent to recognized experts for taxonomic verification purposes.

APPENDIX D

METHODOLOGY FOR DETERMINING BIOLOGICAL CONDITION

Determining biological condition involves calculating the MMI for the appropriate site class and comparing the MMI score to a threshold. While this application is relatively straightforward, the collection, preservation and enumeration of a macroinvertebrate sample and the prediction of the biological site class from which that sample was collected is a meticulous process.

The process of using macroinvertebrate data in the application of the MMI first begins with collecting a representative sample from a single or multi-habitat as a means to standardize assessments among waterbodies having similar of habitats. Field sampling is typically followed by preservation, transporting to a laboratory for identification and enumeration analysis, returning data spreadsheets, resolving taxonomic inconsistencies, and lastly electronically uploading the data to a biological database capable of running the MMI.

The process of predicting site class involves extracting three environmental variables (elevation, stream slope and Level III or IV Ecoregion) from maps or ArcMap GIS feature layers for each site. The values are then evaluated against a set of pre-defined classification rules, which determines the site class and the appropriate index to calculate a score. The following sections explain the step-wise procedures necessary to compute a site class and a MMI score from which to assess biological condition.

A. Macroinvertebrate Data Collection

Aquatic benthic macroinvertebrates are animals without backbones that live on submerged rocks, logs, sediment, debris and aquatic plants during some period in their life. They are an important part of the food chain, especially as a food source for fish. The primary advantage of using macroinvertebrates to monitor water quality is that they cannot move around much, so they are less able to evade the deleterious effects of pollutants that diminish water quality. The abundance and species composition of macroinvertebrates can provide water quality scientists with reliable and valuable information on stream and river water quality.

Macroinvertebrate samples are primarily collected using a semi-quantitative method that varies depending on the type of substrate. This sampling method targets riffle and run habitats in rocky bottom streams. These habitats are characteristic of high to intermediate elevation streams in Colorado. However, large portions of eastern Colorado and several regions abutting Utah have streams with sandy or silty substrates. In these low gradient streams a multi-habitat approach is taken. This method targets an abundance of representative habitat types, such as snags, submerged macrophytes and vegetated banks.

The equipment used to collect macroinvertebrate samples, regardless of habitat or stream type, is a modified kick net (or dip net), which includes a long handle, an 18” x 9” rectangular frame and a 500 to 600 µm mesh net that measures no less than 36” long (**Figure 1**).

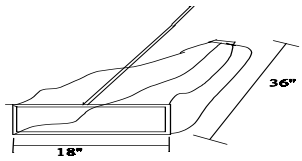


Figure 1. Modified Kick Net

In rocky bottom streams, a riffle or run is selected for sampling. Sampling begins at the downstream portion of the sampling reach and proceeds in an upstream direction. A "kick" is accomplished by positioning the net frame flush against the stream substrate and disturbing one square meter of bottom substrate upstream of the net for a period of one minute. The heel or toe of the samplers' boot dislodges the upper layer of cobble and organisms are washed into the net. Sampling in low gradient streams, lacking riffle/run habitats, involves systematically disturbing multiple habitat types that result in sampling one square meter of mixed habitats for a period of one minute.

The entire content of the kick net is deposited into a one liter polyurethane jar, which is labeled, 95% ethyl alcohol (EtOH) or ethanol preservative is added and the jar is sealed. If excessive detritus or material prohibits sufficient preservation then the sampler will either split the sample into multiple jars or try to remove some larger rocks and organic matter/litter from the sample to leave enough space to add ethanol preservative.

These methods and equipment have been utilized by Division from 1996 through 2009. Further information related to the Division's Standard Operating Procedure for the collection of benthic macroinvertebrates, including more details germane to sample processing and preservation may be found in Appendix B of Water Quality Control Commission ("WQCC") Policy Statement 2010-1.

B. Laboratory Processing and Identification

Preserved samples are delivered to a qualified laboratory, which performs benthic macroinvertebrate sample analysis including taxonomic identification, enumeration and preparation of project reference collections. Taxonomic identification and abundance results are reported to the Division. Further information related to the Standard Operating Procedures for laboratory processing and identification may be found in Appendix C of WQCC Policy Statement 2010-1.

After the taxonomic identifications are completed, the data is transferred from the benchsheets into an Excel spreadsheet that is structured specifically for uploading data directly into Colorado's Ecological Data Application System ("EDAS"). These spreadsheets are known as "Benthic Import Sheets". After entry is completed, the data sets are printed out and proofed against the bench sheets in order to ensure accurate data entry. Spreadsheets are checked so that all spelling and calculations are correct.

C. Data Management

Electronic copies and printed versions of the Benthic Import Sheets are submitted to the Division. The Benthic Import Sheets are checked against the authenticated final identifications in EDAS to identify potential taxonomic nomenclature inconsistencies and the subsequent actions necessary to remedy those inconsistencies (see Table 1). The authenticated final identifications are those listed under the "FinalID" column in the Benthics Master Taxa table.

Table 1. Common Taxonomic Nomenclature Inconsistencies

Inconsistencies	Action
Taxonomic spelling errors	Check correct spelling in EDAS and fix error on Benthic Import Sheets
Taxonomic suffixes (such as “species”, “adults” and “groups”)	Check correct FinalID in EDAS and remove suffixes on Benthic Import Sheets
Taxonomic authorities	Remove authorities on Benthic Import Sheets
New taxa are identified	Verify taxa and add to EDAS’ Benthics Master Taxa table

Final identifications in EDAS are streamlined so nomenclature is based upon the taxonomic hierarchy system, with few exceptions. Thus, final identifications do not include suffixes nor do they include taxonomic references to “authorities”.

New taxa shall be verified on the Integrated Taxonomic Information System (“ITIS”) website[1]. This website shall be considered an authoritative source on taxonomic information. In some instances, an invalid taxonomic name is reported by the laboratory, but the valid taxonomic name exists in EDAS. In these instances, it is appropriate to change the reported taxonomic name to the valid EDAS equivalent on the Benthic Import Sheets.

New taxa are input into EDAS’ Benthics Master Taxa database table. When new taxa are added to this table, ancillary taxonomic information related to that new taxon must be added as well. The generation of metrics and index scores is dependent upon the most updated information available for each new taxon. For instance, many species richness and composition metrics are dependent upon a complete taxonomic hierarchy. Also, many metrics are dependent upon the species tolerance value, functional feeding group, and habit preference type.

Taxonomic hierarchies can be verified on ITIS and added to the Benthics Master Taxa table.

Regional tolerance values, functional feeding groups and habit types shall be verified using the following reliable sources:

- 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. United States Environmental Protection Agency, Office of Water, Washington, D.C.
- Merritt, R. W., and K. W. Cummins (eds.) 1984. An introduction to the aquatic insects of North America. Second edition. Kendall/Hunt, Dubuque, Iowa.

Additional literature sources may include:

[1] http://www.itis.gov/

- Bode, R. W., M.A. Novak, L.A. Abele. 1996. Quality assurance work plan for biological stream monitoring in New York State. NYS Department of Environmental Protection; Division of Water; Bureau of Monitoring and Assessment; Stream Biomonitoring Unit; Albany, NY.

Once verified, regional tolerance values, functional feeding groups and habit types, along with their respective sources shall be added to the Benthics Master Taxa table.

At this point, the reported taxonomic names on the Benthic Import Sheets will be in full synchronization with the "FinalID" column in the Benthics Master Taxa table. The Benthic Import Sheets are now ready for importing into EDAS.

D. Generating Predictor Variables

A prerequisite to importing the Benthic Import Sheets into EDAS is the "Predictor Variable Import Sheet". The function of this spreadsheet is to populate EDAS with a nominal amount of new station information from which to run the multi-metric indices ("MMI"). This spreadsheet includes general site information, such as the station ID, waterbody name, station description and latitude and longitude geographic coordinates in decimal degrees (i.e. 37.3923 and -104.5687). It also includes the set of MMI predictor variables that will determine both the station's biotype and the MMI to be used in calculating an index score.

The general site information is gathered while in the field and recorded to field sheets. Station coordinates are determined in the field with a hand-held GPS unit and verified using desktop tools such as Google Earth or Google Maps to ensure an accurate location. This information is appended to the Predictor Variable Import Sheet.

The verified station coordinates are converted to "point" layers on ArcMap GIS. From these points, the MMI predictor variables are extracted from feature or raster layers. The MMI predictor variables and brief descriptions are listed below:

- **Elevation -** a National Elevation Dataset raster layer from the U.S. Geological Survey. The unit is meters or a resolution of 1 arc-second (about 30 meters).

- **Slope -** a feature stream layer developed by Utah State University. Slope was calculated by identifying segments of channel that were between upstream and downstream tributaries. The range of elevation values, extracted from a 30-meter Digital Elevation Model, was divided by the length of the segment to obtain the slope of the segment. The slope value was then applied to the entire segment.

- **Level IV Ecological Subregions -** a feature layer that was developed through a collaborative effort between the U.S. Environmental Protection Agency, National Health and Environmental Effects Laboratory, U.S. Forest Service, Natural Resources Conservation Service and various other state and federal agencies. The subregions

denote areas of general similarity in ecosystems and in the type, quality, and quantity of environmental resources.

The MMI predictor variables are appended to the Predictor Variable Import Sheet. It is not necessary to include any of the River Invertebrate Prediction and Classification System ("RIVPACS") Observed over Expected ("O/E") predictor variables to the Predictor Variable Import Sheet, since they are not relevant to the generation of MMI scores. The Predictor Variable Import Sheet will upload correctly without the O/E predictor variables.

The Predictor Variable Import Sheet can be uploaded into EDAS by selecting the "Multimetric and Predictive Models" button from the EDAS home page and opening the "Model Analysis" form. The user clicks the "Import/Update Predictor Variables for MMI and Predictive Model" box. A new window opens. The user navigates to the Predictor Variable Import Sheet location and selects the file to upload. After uploading, check the "TTImportPredictors" table to confirm that all the data was imported into EDAS. The "TTImportPredictors" table can be accessed from the home page of EDAS by clicking the "Access Database Structure" button. If the "TTImportPredictors" table is empty the import was successful. This table temporarily stores stations and predictor variable data that EDAS is not programmed to recognize and therefore did not properly upload to their assigned tables. Errors in the import process are often due to incorrect spreadsheet formatting. For instance, the date format is something other than "short date" or a Station ID is formatted as a number rather than generally unformatted. These errors can be remedied within the Predictor Variable Import Sheet. First, clear the contents of the "TTImportPredictors" table. Then re-attempt to import the Predictor Variable Import Sheet as described above. Repeat until the upload results in an empty "TTImportPredictors" table.

If the Predictor Variable Import Sheet import was successful, proceed to upload the Benthic Import Sheet through the Model Analysis form. The Model Analysis form can be accessed by selecting the "Multimetric and Predictive Models" button from the home page of EDAS. The user clicks the "Import Benthic Data" box. A new window opens and the user selects the file to upload. After the upload is completed, check the "TTImport" table to check for taxonomic identifications that were not imported into EDAS. The "TTImport" table can be accessed from the home page of EDAS by clicking the "Access Database Structure" button. If the "TTImport" table is empty the import was successful. If not, re-check those taxonomic identifications and apply these changes to the Benthic Import Sheet. Clear the contents of the "TTImport" table. It is not necessary to remove the taxonomic data that did successfully imported on the first attempt. Re-attempt to import the entire Benthic Import Sheet. Repeat until the upload results in an empty "TTImport" table.

EDAS is programmed to allow transfer of data from the two aforementioned import sheets into a series of relational database tables. Station metadata and predictor variables are stored in the "Stations" table in EDAS. EDAS automatically predicts a "biotype" for each new station based upon a sequence of classification rules using the three predictor variables. The predicted "biotype" is stored within the "Stations" table and becomes the prompt for which MMI will be used to generate an index score.

E. Automatically Generating Index Scores

EDAS allows users to automatically generate MMI index scores.

From EDAS' home page, click on the "Multimetric and Predictive Models" box to open the "Model Analysis" form. On the Model Analysis form any number of stations can be selected upon which to run the MMI tool. If new macroinvertebrate samples were recently uploaded into EDAS, it is necessary to click the "Verify 300 organism subsample" box. This runs a Visual Basic macro that subsamples any sample count greater than 360 individuals down to a 300 count. Samples with counts less than 360 individuals remain as reported by the laboratory. The 300 count subsample is stored in a table and cannot be sub-sampled again. All metric calculations and MMI scores are based upon this one and only subsample.

Once sub-sampling is completed, click the "Calculate MMIs" box to open the "Multimetric Index Model (MMI) Calculation and Reporting" form. The user is prompted to select the output of the MMI calculations. Selecting "View MMI results report" opens a reporting form within EDAS. Selecting "Export MMI results as 'Excel'" opens a save file window and allows the user to save the output spreadsheet to a file location of their choice.

Depending on the number of stations selected on the Model Analysis form, the MMI calculation can take anywhere between 30 seconds and 3 minutes. When the calculations are finished, the Excel spreadsheet will open and MMI scores can be viewed immediately.

F. Manually Generating Index Scores

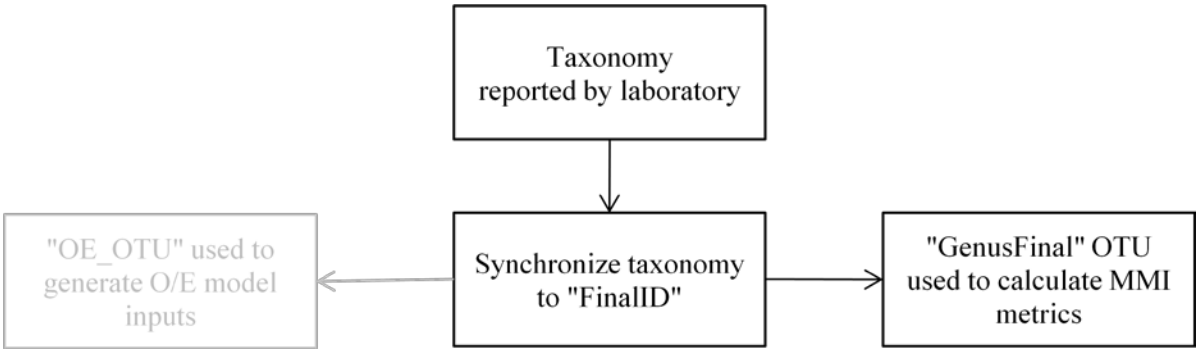
Manually deriving MMI scores, independent of EDAS, is best handled through formulaic calculations in an Excel spreadsheet. However, there are many steps that must occur between the receipt of laboratory reported taxonomic data and setting up the MMI calculation spreadsheet. These steps are organized in sequence and explained below:

<u>Taxonomic Resolution</u>

Biological assessment tools require that a consistent taxonomy be applied across all samples because ambiguous taxa amongst samples can be confusing. This issue was first encountered during the first version of the O/E model completed in 2005. To counteract this problem, consistent taxonomy was identified and referred to as operational taxonomic units or OTU's (Hawkins, 2009). A comparable taxonomic structure was developed for the MMI. The MMI equivalent to the O/E OTU was given the name of "GenusFinal". Thus, the MMI utilizes high-resolution taxonomy that includes identifying midges to genus and other taxa treatments that improve the indices' sensitivity to stress.

Accordingly, the Division strongly encourages that any identified and enumerated data first be synchronized with the consistent taxonomy provided in the "FinalID" column in the Benthics Master Taxa table in EDAS. Each taxon with a "FinalID" will have an equivalent OTU designation associated with each bioassessment tool. Figure 2 illustrates this intention.

Figure 2. Organized Taxonomic Unit Assignments for each Bioassessment Tool



Sample Standardization

Because the MMI is based on estimates of the number of taxa present in a sample, MMI scores can be affected by any modification to sampling technique, such as increasing or decreasing sampling area or timed effort. For instance, increasing sampling area or time generally results in increased individual counts, which in turn increases richness. Conversely, decreasing sampling area or time may result in low individual counts, which in turn under-estimates richness. For this reason, samples shall be standardized on area and count.

The Division recommends that raw sample counts greater than 360 individuals be randomly resampled by computer to a fixed 300 count. A free executable program (Matrify.exe) is available for download at Utah State University's The Western Center for Monitoring and Assessment of Freshwater Ecosystems. Additionally, a Visual Basic macro can be scripted to perform this function. Raw sample counts less than 360 individuals need not be resampled and can be used in further analysis. However, raw sample counts less than 150 individuals shall be flagged as not conforming to the data treatments ensconced in the MMI development.

Metric Prerequisites

Taxonomic hierarchy, regional tolerance values, functional feeding designations and habit/behavior assignments are requirements for computing accurate metrics. The best method to compute metrics would be to set up another Excel spreadsheet. Include columns for taxonomic identifications (harmonized beforehand), individual counts (subsampled beforehand) and columns for the following information:

Taxonomic Hierarchy

Several richness and composition metrics require information pertaining to an organism's taxonomic hierarchy. This will become evident when metrics are calculated further in this process. For instance, in Biotype 2 one of the composition metrics is % Ephemeroptera. For those organisms with an OTU designation of sub-family (rare), tribe or genus, the taxonomic hierarchy is vital to those affected richness and composition

metrics. The taxonomic hierarchy shall include phylum, class, order, family, tribe, genus and species.

Regional Tolerance Values

Several tolerance metrics are components of all three MMI's. It will be necessary to research information pertaining to each organism's regional tolerance value. Tolerance values do not need to be specifically from the Southern Rocky Mountain region. It is acceptable to use tolerance values from other regions in the United States as long as they are cited from valid sources. Valid tolerance values can be found in the literature sources cited under Subsection C.

Functional Feeding Group Designations

The predator/shredder taxa metric is a component of all three MMI's. Therefore, it will be necessary to research information pertaining to each organism's FFG. Functional feeding group designations can be found in the literature sources cited under Subsection C. Functional feeding designations include parasite, predator, omnivore, scraper, gatherer/collector, filter/collector, shredder, and piercer.

Habit/Behavior Assignments

Several habit metrics are components of all three MMI's. It will be necessary to research information pertaining to each organism's habit preferences. Habit designations can be found in the literature sources cited under Subsection C. Habit and behavior designations include clinger, climber, sprawler, burrower, swimmer, diver, and skater.

Metric Calculations

Computing individual metrics will involve setting up Excel spreadsheet formulas or equations whereby the breadth of taxonomic information is transformed into measureable statistics. Guidance on building these formulae will not be provided within this document.

MMI Calculation

The same Excel spreadsheet can be used to compute the MMI scores. Computing the MMI score involves applying a specific scoring formula to each metric. Scoring formulas for all metrics inclusive to each MMI are provided on pages 20-22 of the MMI report located in Appendix A of WQCC Policy Statement 2010-1. This results in an individual metric score. The sum of these individual metric scores is divided by the number of metrics that comprise that particular MMI. For instance, if MMI #1 is comprised of metrics A, B, C, D and E, then the scores for metrics A, B, C, D, and E are summed and divided by five. The output is the MMI score. The range of MMI scores is 0-100.

Of note, individual metric scores greater than 100 or less than zero will occur. It is necessary to modify those occurrences by adjusting scores greater than 100 down to 100 and scores less than

zero up to zero. This will have to be done manually within the spreadsheet cell. If this data treatment is ignored, there is a chance that an MMI score can exceed 100 or fall below zero.

APPENDIX E

Threshold Development – Technical Underpinnings

Water Quality Control Division
August 2010

The Multi-metric Index (MMI) assesses biological condition on a scale of 0 to 100. The numeric characterization is analogous to measuring the concentration of a pollutant, but the number alone does not define what constitutes attainment (or impairment) of the use. Establishing biological thresholds is largely a statistical endeavor, but it must be preceded by a policy decision that defines use support on the basis of site characteristics. Once the basis for use support is established, appropriate statistical methods can be applied to determine the thresholds defining biological condition consistent with use support. Finally, there are implementation issues that the Division anticipates regarding differences between Class 1 and Class 2 waters, as well as a concern about protection of high-scoring waters.

Defining Use Support

The definition of use support is central to the derivation of biological thresholds because it controls membership in the group of sites that are in attainment. Once group membership is established, the statistical properties of biological condition (MMI) can be determined for sites in the group. Sample size, central tendency, and variance will shape the development of thresholds.

It is common practice to tie the definition of use support to the concept of reference conditions. In EPA guidance documents, for example, "use support" is equated solely with the technical definition of "reference conditions" (i.e., the basis for tool development). Moreover, some percentage of low-scoring reference sites is excluded. However, when use support is equated with reference conditions, the range of policy options can be adjusted only by the relatively trivial method of varying the percentage of reference sites excluded.

Defining reference conditions was an integral part of bioassessment tool development because it established one end of the spectrum for biological condition. Because selections are made before the tool is developed, there is a presumption that selection criteria will yield sites with good biological condition. Generally, this means selecting sites where anthropogenic stressors are minimal or absent. There is little doubt that minimally-disturbed[1] reference sites can serve the *technical* purpose of bioassessment tool development because they clearly represent one end of the spectrum for biological condition.

It can also be argued that stringent selection criteria, which minimize potential stressors, yield only those sites most easily associated with use support. Higher stressor levels may also be consistent with use support, as is clearly acknowledged in the TALU condition gradient, but it

[1] Minimally-disturbed in Colorado is likely to include atmospheric deposition of some pollutants (e.g., nitrate and mercury of anthropogenic origin).

would be hard to know this without having first developed the tool. Once the tool is available, it becomes possible to assess biological condition independent of labels.

The point to be emphasized is that, at least for Colorado, the Division believes reference sites represent a subset of all sites at which the aquatic life use is supported. Whether that subset constitutes 50% or 80% or 95% of sites supporting the use is guided to a large extent by policy that the Commission will set. The concept of reference condition, which was essential for bioassessment tool development, is not indispensable for reaching conclusions about use support.

Biological Thresholds Derivation

Development of biological thresholds falls largely in the policy realm, albeit with a strong reliance on statistical tools. The goal is a characterization of biological conditions demonstrating support of the aquatic life use. Because policy options are both enabled and constrained by this characterization, the definition of use support represents a critical exercise of policy prerogative.

The Division believes the Commission should be able to consider a range of policy options in setting biological thresholds. These options, which are described below, may include biological conditions that are more expansive (or more restrictive) than those circumscribed by the technical definition of reference conditions. The key elements for option development include the basis for membership in the group representing use support and the characteristics of the MMI distribution for that group.

The statistical approach recommended by EPA involves the application of interval and equivalence tests for the purpose of locating attainment and impairment thresholds. The method is described in Kilgour et al.[2] It begins with a definition of the normal range, which establishes the acceptable range of biological condition within the group of sites representing use support.

Normal Range

According to Kilgour et al., the normal operating range is "typically defined as the range of values enclosing 95% of the population…, regardless of the discipline." Previous EPA guidance bears superficial resemblance to the normal range concept in that a low percentile of the reference set is often used to set the impairment threshold. Beginning with Ohio's work on biological thresholds, the 25th percentile of reference was set as the threshold for impairment. However, EPA guidance frames the choice somewhat differently by stating that the "actual percentile chosen … is arbitrary and represents the amount of uncertainty that a monitoring program can tolerate."[3] EPA's arbitrariness in selecting a percentile is evident in the range of values that has been applied in other states: 2.5% to 25%. No rationale is available to explain the preference for one percentile over another. Moreover, the absence of a rationale is puzzling in

[2] Kilgour, BW, KM Somers, and DE Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience 5(4): 542-550.
[3] Barbour, MT et al. 1996. Biological Criteria – Technical Guidance for Streams and Small Rivers, Revised Edition. See http://www.epa.gov/bioindicators/pdf/EPA-822-B-96-001BiologicalCriteria-TechnicalGuidanceforStreamsandSmallRivers-revisededition1996.pdf

view of the primary goal, which is to determine "whether an individual site is a member of the least-impaired reference population."[4]

Given the level of effort that Colorado has invested in selecting reference sites, application of the normal operating range (central 95% of the distribution of MMI scores; Figure 1) seems appropriate. Nevertheless, in deference to concerns raised by EPA Region 8, an option also is presented with a narrower operating range (central 90%).

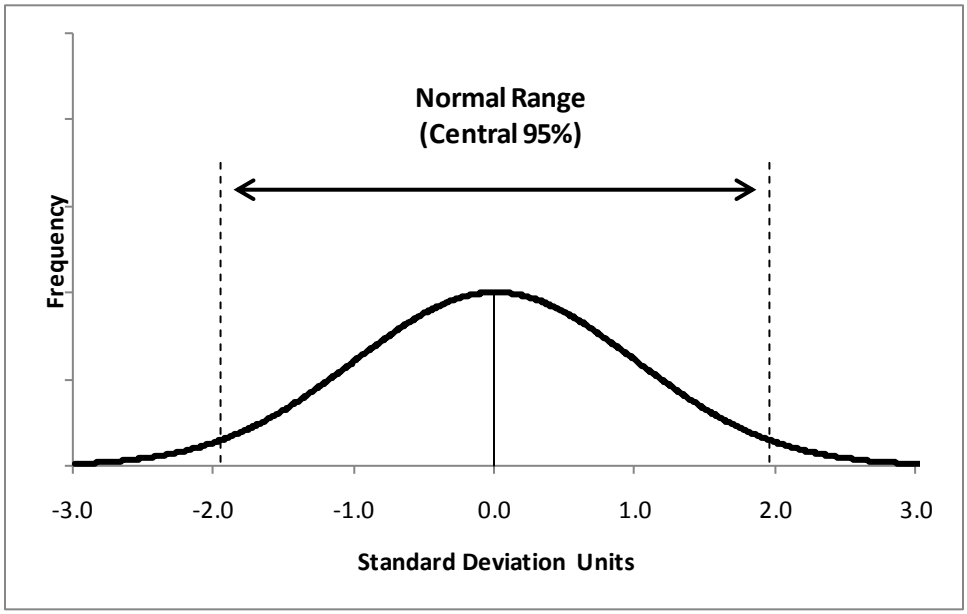


Figure 1. Graphical representation of the "normal range" based on the central 95% of the distribution (bold line represents the normal frequency distribution). The dashed vertical lines signify the boundaries of the central 95% at ±1.96σ .

Statistical Basis for Thresholds – Interval and Equivalence Tests

The Division is proposing to use MMI scores as the numeric biological thresholds that determine whether biological condition at a particular site is impaired or in attainment of the use. Establishing those criteria is a statistical exercise that incorporates policy considerations related to the definition of use support, as well as the risk of being wrong about concluding that a site is impaired or in attainment. The proper framing of the statistical question is, of course, central to the conclusions that can be drawn.

The traditional statistical approach tests the hypothesis that the site in question is a member of the group of sites with MMI scores representing use support (e.g., reference conditions). A test of this nature is straightforward and provides a clear-cut conclusion regarding the hypothesis (i.e., reject or not). The alpha value for the test defines the risk of mistakenly concluding that a reference-quality site is impaired (usually, α=0.05, which means 1 chance in 20 of false rejection). However, the results would not reveal whether an ecologically-important difference is present, or even if one is detectable. Moreover, the power of the test may be low, meaning

[4] Kilgour et al. 1998

that there would be a relatively high risk of mistakenly concluding that an impaired site was in attainment.

An alternative statistical approach is possible when the ecologically-important difference (also the critical effect size) is defined in advance. Although it might be difficult to decide in advance that a difference of 10 MMI would be considered ecologically-important, it would be both clear and simple to define the difference in terms of normal range (see above). When the normal range is defined in terms of standard deviation units, the concept is readily transferable among the biotypes and can be adjusted easily if group membership is augmented with additional sites (see below).

The alternative statistical approach, which is recommended by EPA, sets thresholds based on interval and equivalence tests. For these test, the observed difference between the trial MMI and the mean of the use support group is compared to the "ecologically important effect size"[5]. The interval test evaluates the hypothesis that the difference between the trial MMI and the group mean is smaller than the effect size; if it is rejected (i.e., the difference is significantly larger than the effect size), the trial site would be considered impaired. The equivalence test evaluates the hypothesis that the difference between the trial MMI and the mean is larger than the effect size; if it is rejected (i.e., the difference is significantly smaller than the effect size), the trial site would be considered in attainment.

Advantages of this statistical approach can be better appreciated by stepping through possible outcomes. For example, if the interval test is rejected, the outcome is simple: the site is impaired. Failure to reject the interval test, on the other hand, indicates no significant impairment, but that outcome does not necessarily mean that the site is in reference condition. The status of the site can be clarified by the equivalence test, rejection of which indicates that the site is probably in reference condition. If neither hypothesis is rejected, the trial value falls in a gray zone that could be interpreted as *possibly impaired*.[6]

[5] Kilgour, BW, KM Somers, and DE Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience 5(4): 542-550. The Division has followed this paper in setting the effect size to correspond to the normal operating range (1.96 standard deviation units on either side of the mean encompasses the central 95% of the distribution) and in adopting a 5% level of significance for hypothesis testing.

[6] Bowman, MF and KM Somers. 2006. Evaluating a novel Test Site Analysis (TSA) bioassessment approach. Journal of the North American Benthological Society 25(3): 712-727. A useful graphical representation of the hypotheses appears in this paper, which also outlines implementation of the computational procedures for Excel spreadsheets

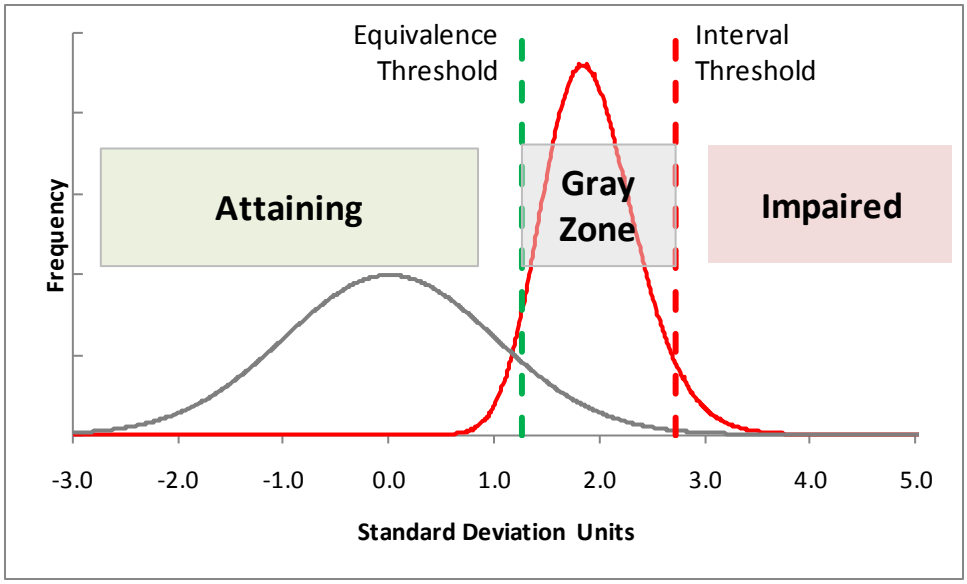


Figure 2. Graphical representation of thresholds derived from interval and equivalence tests. The distribution from which the normal range is derived is bell-shaped curve centered on 0.0 std deviation units. The non-central F distribution is shown as the taller curve centered on the Gray Zone. Equivalence and interval thresholds, which are shown as vertical dashed lines, divide assessments into three categories – attaining, gray zone, and impaired – as explained in the text.

Group Membership

Within each of the biotypes, some sites have been placed in the reference category, and some have been placed in a stressed category. The reference and stressed categories were important for tool development, as mentioned previously. However, a somewhat different perspective is appropriate for biological thresholds development. Biological thresholds focus on use support, which encompasses, but may not be restricted to, reference conditions.

Consistent with EPA guidance, reference conditions are used to define use support for one set of options developed by the Division. A total of 133 reference sites were identified through a rigorous screening process that included evaluation of human disturbance using GIS coverages, to produce the candidate reference list, and evaluation of aerial photos, to produce the final list. Reference sites were assigned to biotypes based on community composition (using cluster analysis as described in MMI development). For sites other than reference, biotype membership was based on predictive relationships. For each of the three biotypes, biological condition scores (MMI) were tested for normality and screened for outliers. All were normally distributed, and a single outlier[7] was excluded from biotype 1 (Figure 3).

[7] The Division has applied Rosner's sequential procedure to test for outliers because it handles the problem of masking when outliers may be close together. See: Gilbert, RO. 1987. Statistical Methods for Environmental Pollution Monitoring. Wiley, New York. 320p.

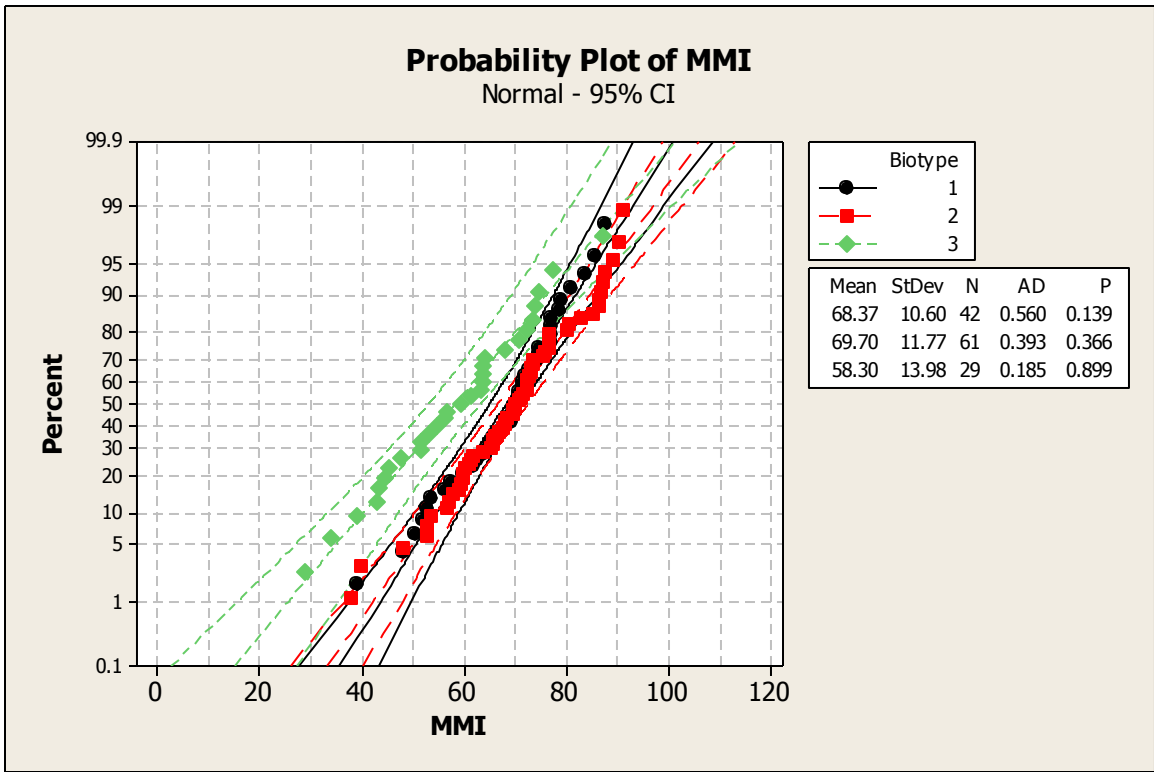


Figure 3. Probability plots of MMI scores at reference sites in each of the three biotypes. Biotype membership was determined by cluster analysis. One outlier was removed from biotype 1, as explained in the text.

The Division also has defined a group that expands membership beyond the reference sites. The added sites have human disturbance scores comparable to those measured at reference sites. Scores for human disturbance factors for all sites were derived from GIS coverages. Within each biotype, the range of scores recorded for reference sites was used to screen all non-reference sites. Any non-reference site for which scores for all disturbance factors were within the ranges observed for reference sites was added to the expanded set. The selection criteria are shown in Table 1.

Disturbance Factor	Units	Biotype 1	Biotype 2	Biotype 3
Irrigated Agriculture	% of WSA	<15.2	0	<36
Dryland Agriculture	% of WSA	<6.3	0	<49
Urban	% of WSA	<0.44	0	<0.31
Permitted Point Sources	#/km^2	0-1	0	0
Diversions	#/km^2	0-10	0-1	0-5
Road Density	Mi/mi^2	<3.6	<1.2	<2.8
Abandoned Mines	#/km^2	<0.6	0	<0.014
Oil & Gas	#/km^2	0-4; NA	--	0-1; NA
CAFO	#/km^2	0; NA	--	--
Sites added to group	#			

Table 1. Threshold values of human disturbance factors used to define membership in the Expanded set. WSA= watershed area.

The premise for the expanded set is that it defines a group with relatively low human disturbance scores, all of which are comparable to scores at reference sites. Consequently, sites in the expanded group are just as likely to support the use as the reference sites. Distributions of MMI scores for reference, expanded, and all sites are shown for each biotype (Figure 4-Figure 6).

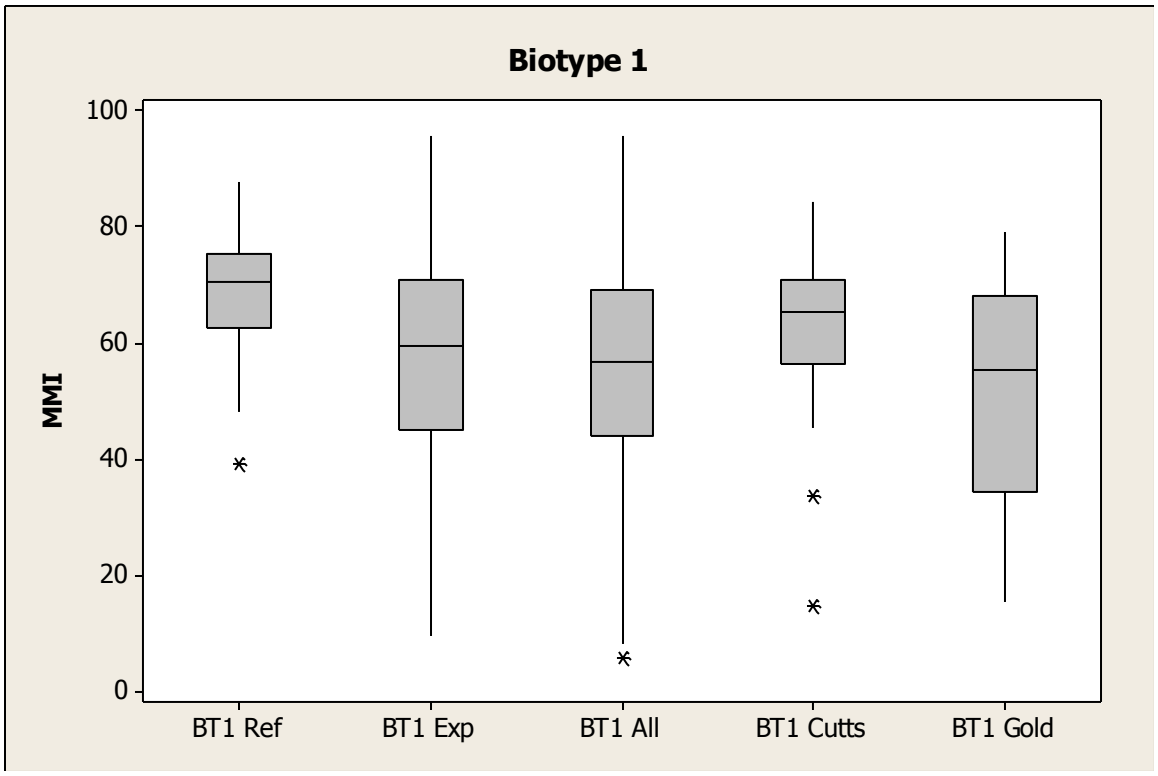


Figure 4. Boxplots of MMI scores for sites in biotype 1 aggregated by group membership (Reference, Expanded, All) or associated with Designated Cutthroat Trout Habitat or Gold Medal fisheries.

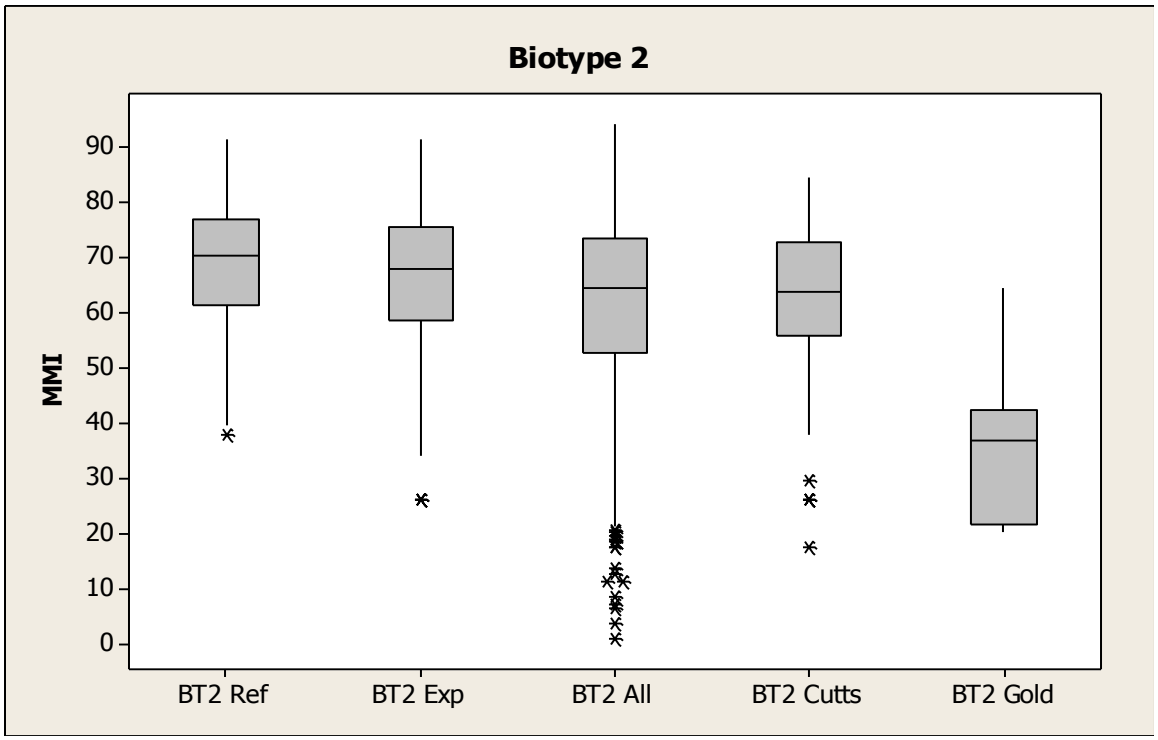


Figure 5. Boxplots of MMI scores for sites in biotype 2 aggregated by group membership (Reference, Expanded, All) or associated with Designated Cutthroat Trout Habitat or Gold Medal fisheries.

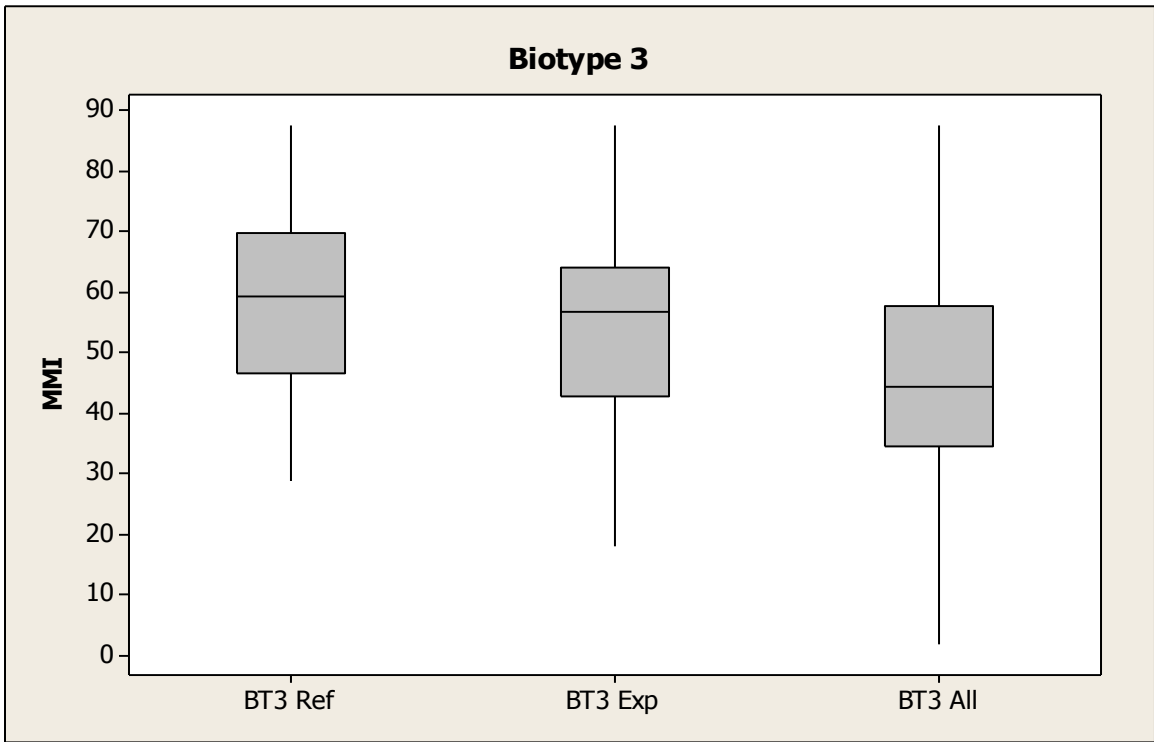


Figure 6. Boxplots of MMI scores for sites in biotype 3 aggregated by group membership (Reference, Expanded, All).

Biological Thresholds Options

Three options are developed; they differ in terms of group membership for use support and definition of normal range. For each option, thresholds are determined and the potential implications for streams statewide are considered. Option 1 defines use support based on a narrow operating range (central 90%) of MMI values at reference sites. Option 2 defines use support based on a normal operating range (central 95%) of MMI values at reference sites. Option 3 defines use support based on a normal operating range (central 95%) of MMI values at sites in the expanded set. Group membership for option 3 in biotype 1 was modified to reduce the number of sites by making the human disturbance criteria slightly more restrictive than values recorded for the group of reference sites. The three options define a range of policy options for characterizing use support (Table 2). In addition, it is clear that large increases in sample size can diminish the width of the gray zone.

Biotype	Option	N of Sites	Average	Standard Deviation	Attainment Threshold	Impairment Threshold
1	1	42	68.37	10.60	54.6	45.9
1	2	42	68.37	10.60	51.6	42.0
1	3	174	59.28	18.28	27.1	19.1
2	1	61	69.70	11.77	53.8	45.9
2	2	61	69.70	11.77	50.4	41.6
2	3	108	67.11	12.81	45.2	38.1
3	1	29	58.30	13.98	41.0	27.0
3	2	29	58.30	13.98	37.1	21.6
3	3	57	54.64	14.26	31.4	20.4

Table 2. Summary of group characteristics and biological thresholds thresholds for each option and each biotype.

The Division recommends adoption of criteria derived from Option 2. From a regulatory perspective, it is based on a methodology that most closely matches EPA guidance; from a practical perspective, it yields thresholds consistent with other evidence of use support (see next section). In subsequent sections of this document, all attention is focused on the Option 2 criteria.

Perspectives on Use Attainment

One of the challenges inherent in developing biological thresholds is that there are few direct measures of use support. Inferring use support from biological condition at reference sites is a reasonable starting point, but it may result in a more exclusive set of sites than is necessary. Too much exclusivity in group membership is likely to underestimate variance, which may lead to an unrealistic basis for developing thresholds.

Although there is no *a priori* basis for deciding what constitutes use support for the macroinvertebrate community, use support may be inferred from data available for some fish communities. For example, the Division of Wildlife applies special designations to stream segments that support native trout populations, naturally-reproducing trout populations, and gold

medal fisheries. Self-sustaining wild trout populations, especially for the native trout species, require good quality habitat and an adequate food base (i.e., bugs). The gold medal fisheries, which are highly prized by anglers, represent "the highest quality cold water habitats"[8] that DOW seeks to protect. The Division believes that these special fisheries designations are direct evidence for support of the aquatic life use.

For many of the segments with special fisheries designations, biological condition also can be determined from existing macroinvertebrate data. The set of MMI values for Gold Medal fisheries in Biotype 1, for example, can be compared with other use support information from other sources (e.g., the set of reference sites). MMI values in streams with special fishery designations are generally lower than those for reference sites in the same biotype (Figure 4-Figure 5). The differences are not large, but it would be hard to argue that the aquatic life use is not being supported in these special fisheries where biological condition is apparently below that of reference.

Evaluation in the "Gray Zone"

As explained previously, using interval and equivalence tests to derive biological thresholds thresholds creates a gray zone within which neither hypothesis is rejected. Although the statistical evaluation does not lead to a clear-cut answer, MMI values are low enough that more information should be considered before reaching a conclusion about impairment. A key piece of information is the aquatic life class for the segment in question.

Within each of the two major aquatic life use classifications – cold water and warm water aquatic life – streams may be designated as class 1 or class 2 based on expectations for the abundance and diversity of the biota. Class 1 streams support "a wide variety of cold water biota, including sensitive species" whereas class 2 streams do not. The presumption is that a "substantial impairment of the abundance and diversity of species" occurs as the result of "physical habitat, water flows or levels, or uncorrectable water quality conditions."

When the MMI score falls within the gray zone and it is from a class 2 stream, the stream is considered to be in support of the use because expectations are diminished. However, the same score in a class 1 stream would elicit a different response requiring examination of additional data. The Division proposes evaluation of two auxiliary metrics that were not included in the MMI. These two metrics – Shannon diversity and the HBI – were selected because they provide information on the two characteristics – diversity and sensitivity – that separate class 1 and class 2 streams.

Evaluation of the auxiliary metrics is based on the equivalence test, rather than the interval test or both, in order to require a convincing demonstration that, in spite of the low MMI, the site is not in bad shape. If the null hypothesis of the equivalence test is rejected, it means that the auxiliary metric score is equivalent to those of the reference distribution. In addition, because the regulation addresses diversity *and* sensitive species, a site will be considered impaired when

[8] Colorado Wildlife Commission Policy on Wild and Gold Medal Trout Management; September 18, 1992, revised Jun 12, 2008.

the MMI is in the gray zone and either auxiliary metric fails to show equivalence to the reference set.

An example is given for evaluation of attainment for each of the auxiliary metrics. The example is derived from one of the three biological thresholds options described above, but the logic is transferable to any of the options.

Shannon Diversity Index

The diversity of taxa observed at each site is captured with the Shannon index, which is calculated in EDAS for every sample. For each biotype, diversity values were assembled for all reference sites and the distributions were examined (Figure 7). After removing one outlier from the lower end of the distributions for biotypes 2 and 3, the three distributions were found to be normal and the variances were not significantly different. Means were significantly different, however, meaning that equivalence thresholds must be established separately for each biotype.

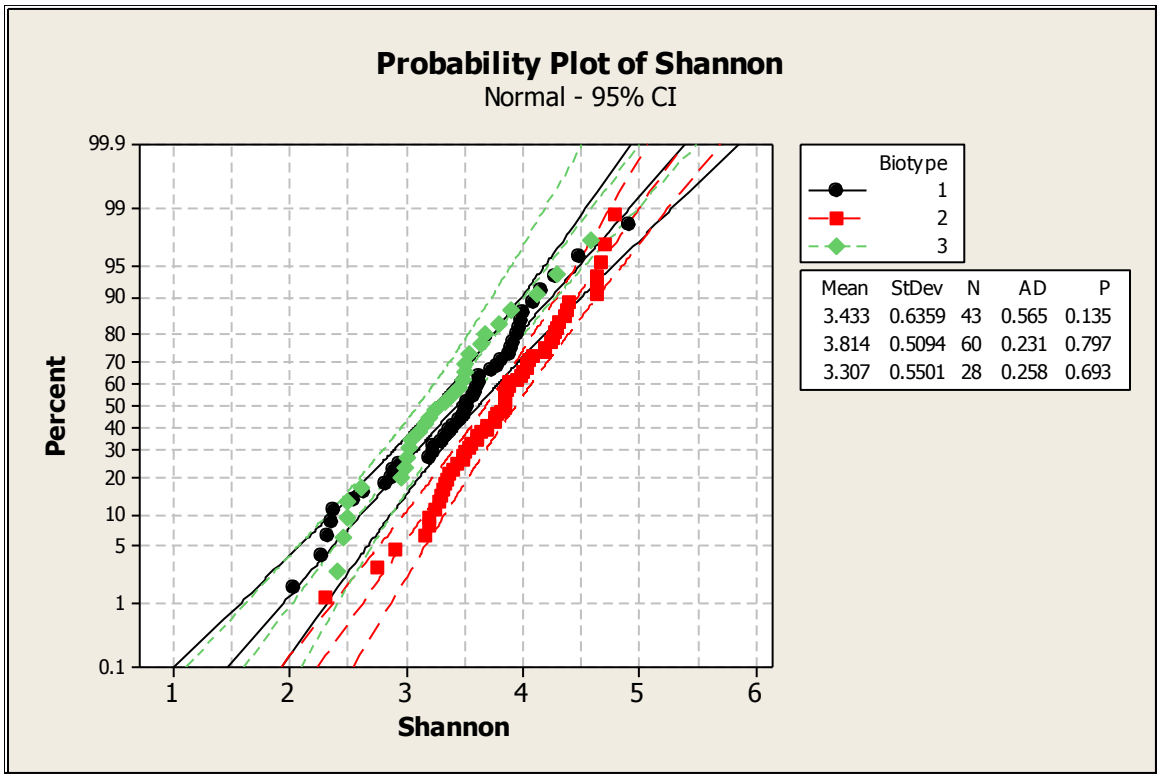


Figure 7. Probability plots for Shannon diversity values at reference sites in each biotype.

Using option 2 (as described above), the dimensions of the gray zone and the equivalence threshold are presented for each biotype in Table 3. If a site in biotype 2 had an MMI score of 45, for example, it would fall in the gray zone and be considered possibly impaired. A diversity score 2.8 would lead immediately to a conclusion of impairment. A diversity score greater than 3.0 would be a necessary, but not a sufficient, condition for considering the site in attainment. A decision regarding attainment could not be reached, however, until the HBI also had been evaluated.

Biotype	Attainment	Impairment	Shannon	HBI
1	51.6	42.0	2.4	5.4
2	50.4	41.6	3.0	5.1
3	37.1	21.6	2.5	7.7

Table 3. Boundaries of the gray zone and thresholds of equivalence for two auxiliary metrics – Shannon diversity and HBI.

Hilsenhoff Biotic Index

The Hilsenhoff Biotic Index (HBI) is a widely used indicator of organic pollution. High values of the index indicate a predominance of tolerant organisms (i.e., the sensitive species have been lost). Like the Shannon index, the HBI is not part of the MMI and the values are routine output from EDAS. Distributions of the HBI are normal for reference sites in each of the biotypes and there are no outliers (Figure 8). Variances were not significantly different among the biotypes, but means were. Accordingly, equivalence thresholds are developed separately for each biotype (Table 3).

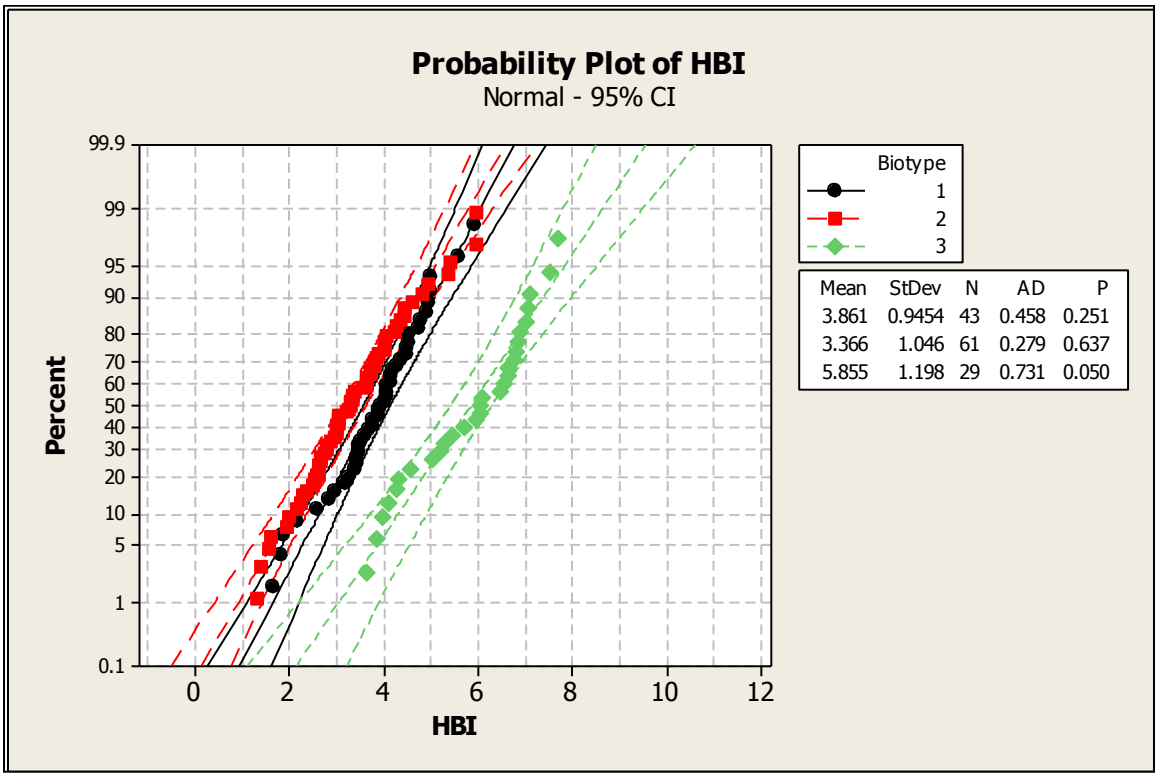


Figure 8. Probability plots for HBI values at reference sites in each biotype.

In contrast to the diversity index, low scores of the HBI indicate better conditions. Consequently, using the same example presented in the preceding section, the HBI would have to be *less than* 5.1 for the site to be considered in attainment. Moreover, both auxiliary metrics would have to show attainment before the site, which had an MMI in the gray zone, would be

considered in attainment. If either auxiliary metric failed to show attainment – if the HBI was at least 5.1 *or* the Shannon index was not greater than 3.0 – the site would be considered impaired.

Protection of High-Scoring Sites

The MMI thresholds that have been presented as options for biological thresholds are appropriate for detecting impairment on the basis of individual samples. However, because some sites have much higher MMI scores, it would seem to leave open the possibility of undetected degradation at high-scoring sites. For example, a site with an MMI value of 90 in biotype 2 would have to experience a drop of more than 40 points before it would be declared impaired. It would be prudent to take action sooner if significant degradation was occurring.

In order to detect a significant decline in biological condition, there must first be an understanding of what constitutes a normal range or year-to-year variability in MMI scores at sites where there is no known water quality or habitat trend. Fortunately, there are a number of sites where biological condition has been examined more than once over a period of years. The change in MMI (d-MMI) was calculated for all available pairs, after excluding any data from the drought year (2002). Because the typical time window for assessments is usually five years, and because a short interval diminishes the likelihood that the change will be influenced by a trend, the data set was restricted to pairs spanning five or fewer years. Statistical comparison of d-MMI values for different intervals (1, 2, 3, 4, or 5 years) showed no significant difference; therefore, all pairs were combined for analysis. For sites with more than two samples, no sample was used more than once, and preference was given to shorter intervals (but at least a year).

Distributions were examined first for each biotype. Distributions were normal, and two outliers were removed from biotype 1. After outliers were removed, variances were found to be homogeneous and there was no significant difference among the biotypes. Accordingly, d-MMI scores for all three biotypes were combined (mean=-0.20; s.d.=12.93; N=123). The d-MMI scores show considerable variability. The "normal operating range" is 25.3 to -25.3[9]. For a change to be considered significant at the 0.05 level, the MMI would have to increase or decrease by 25.3 points. The inter-year variability may seem relatively large, but the standard deviation is similar in magnitude to those observed among reference sites in each of the three biotypes.

[9] The normal operating range is centered on mean of zero to reflect the assumption that no change in biological condition would yield no change in MMI. The observed mean of -0.20 MMI units is indistinguishable from zero.

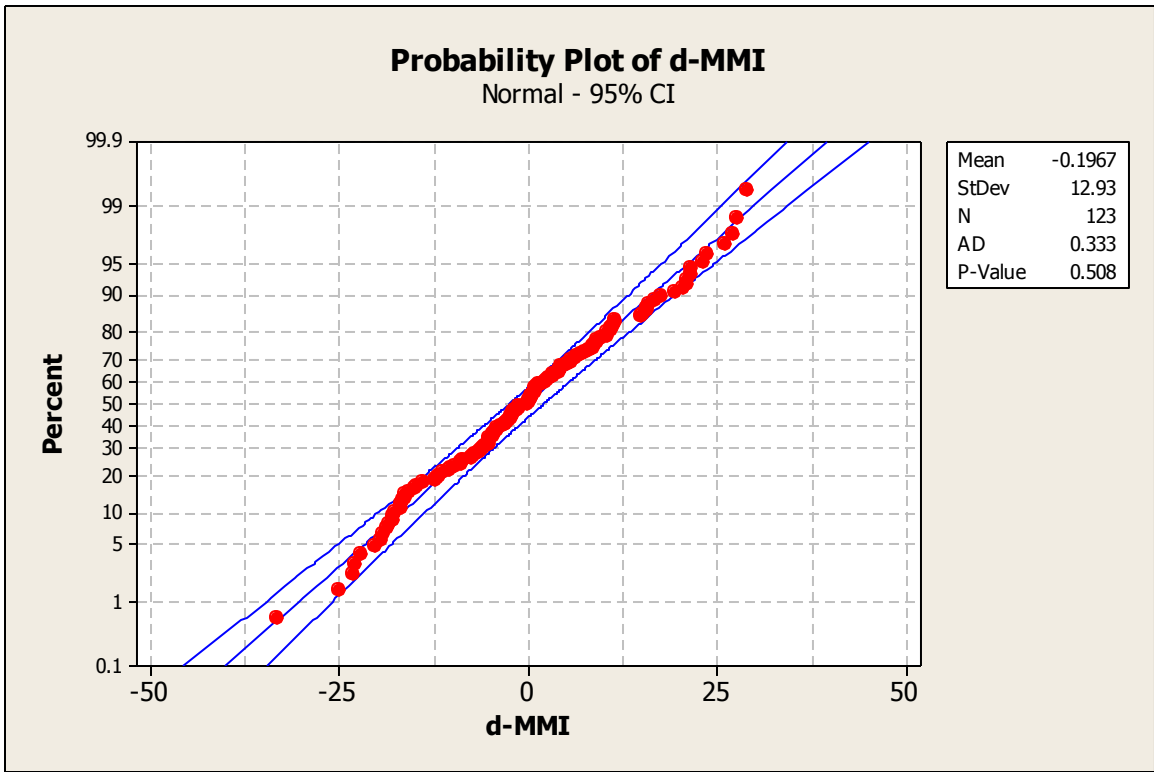


Figure 9. Probability plot of d-MMI scores after removing two outliers as described in the text.

The threshold for decline in MMI can be derived from the same interval and equivalence tests applied previously to set thresholds for each of the biotypes. Using the normal operating range (central 95% of the d-MMI scores) and a 5% level of significance for the interval and equivalence tests, the thresholds for attainment and impairment would be set at -22.3 and -29.0 MMI units, respectively. Given the required magnitude of the change and the high initial level for biological condition, the Division proposes a binary decision system based solely on the equivalence test. If the MMI declines by at least 22.3 units, it should register a serious concern about biological condition, and the site should be considered impaired.

Application of this test proposed for protection of high-scoring sites use requires having at least two MMI scores. Although the test value was derived from samples separated in time by one to five years, there is no need to apply similar restrictions for the regulatory decisions. As a practical matter, the initial MMI must be at least 22 MMI units higher than the impairment threshold for the appropriate biotype (see Table 2). Thus, the initial MMI would have to exceed 64 for biotypes 1 and 2, or 44 for biotype 3, before a site would qualify for consideration as high-scoring.