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: August 7, 2017
Expires: December 31, 2020

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 7

IX. AQUATIC LIFE USE ATTAINMENT: NUMERICAL THRESHOLDS 9

X. REFERENCES.....14

XI. APPENDICES

A. MMI FINAL REPORT - RECALIBRATION OF THE COLORADO MACROINVERTEBRATE MULTI-METRIC INDEX FOR 2017

B. BENTHIC MACROINVERTEBRATE SAMPLING STANDARD OPERATING PROCEDURES

C. STANDARD OPERATING PROCEDURES FOR LABORATORY IDENTIFICATION AND ENUMERATION

D. METHODOLOGY FOR DETERMINING BIOLOGICAL CONDITION

E. THRESHOLD DEVELOPMENT - TECHNICAL UNDERPINNINGS

F. BIG RIVER ANALYSIS – DATA AND GRAPHS

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
CPW	Colorado Parks and Wildlife (formerly known as Colorado Division of Wildlife or CDOW)
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.
mi^2	Square mile
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).

Big Rivers are considered major rivers with a cumulative watershed area exceeding 2700 mi^2.

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.

Gray Zone means that the biological condition is neither impaired nor in attainment of the use based on Interval and Equivalence tests in which neither hypothesis is rejected. More information, in the form of two auxiliary metrics, are considered before reaching a conclusion about impairment for Class 1 waterbodies.

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 Nation's 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.

II. COLORADO'S AQUATIC LIFE USES

Colorado classifies water bodies for beneficial uses that, along with chemical and physical standards, assure their protection. Colorado has four classifications for waters: water supply,

---

[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.

recreation, agriculture and aquatic life. 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 River 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 their monitoring and assessment 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 CWA Section "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 the 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 – are defined briefly in the Definitions section, and then discussed in more detail in Section IV subsection A.

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 short life spans, limited migration patterns, 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 five metric categories: taxa richness, composition, pollution tolerance, functional feeding groups, and habit (mode of locomotion). 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]

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, habit (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.

[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.

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

[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 CPW in the Inventory and Status of South Platte River Native Fishes in Colorado, CPW, 1997).

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*).

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.

The 2012 Section 303(d) List contained 33 segments for aquatic life use impairment. Four segments were fully listed, meaning a pollutant was identified. Twenty-nine segments were provisionally listed, meaning a pollutant was not identified. 303(d) priority for provisionally listed segments ranged from low to high in some cases although most tended to be prioritized as "low". Twenty-two segments or portions of segments were placed on the Monitoring and Evaluation List. The division integrated sites within these segments into subsequent monitoring plans to be sampled.

The 2014 Section 303(d) List was not completed due to a re-prioritization of resources in the division.

The 2016 Section 303(d) List contained 81 segments for aquatic life use impairment. Sixteen segments were fully listed, meaning a pollutant was identified. Fifty segments were provisionally listed, meaning a pollutant was not identified. Fifteen segment or portions of segments were placed on the Monitoring and Evaluation List. The division integrated sites within these segments into subsequent monitoring plans to be sampled.

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 less than 2700 square miles and big rivers on the Western Slope of Colorado 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, 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, (i.e., 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 and subsequent iterations were 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.

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

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 and those big river watersheds listed in Section IX subsection B, 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 4, 2017, see Appendix A). The following documents describe the procedures for collecting data and determining MMI scores:

- Benthic Macroinvertebrate Sampling Standard Operating Procedures (Appendix B)
- Standard Operating Procedures for Laboratory Identification and Enumeration (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.

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	45	34
2	Mountains	48	40
3	Plains & Xeric	42	29

- Class 1 water bodies: An MMI score in the gray zone for a site on a Class 1 water requires examination of additional information – see Subsection 4 below.
- 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.

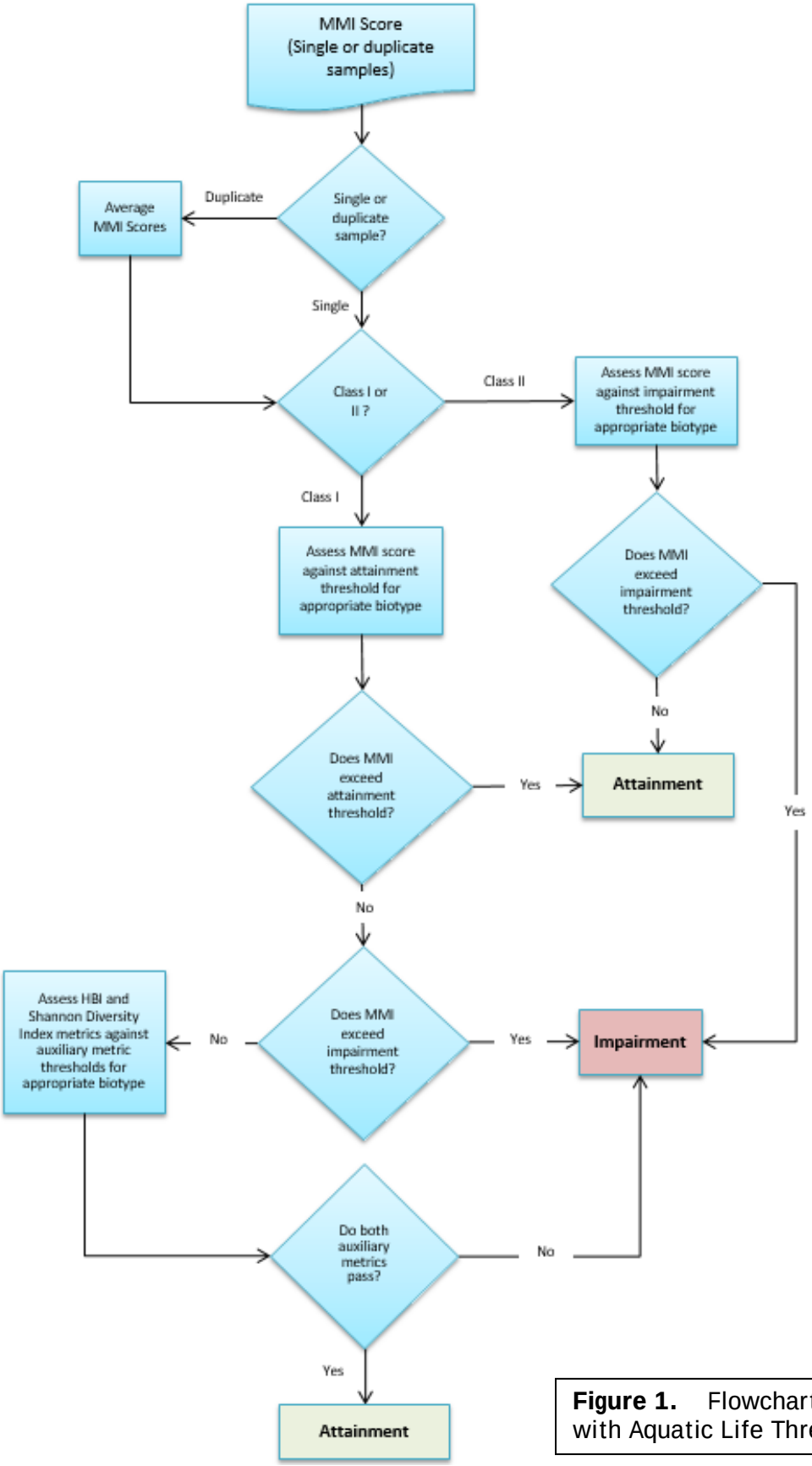


Figure 1. Flowchart: Comparison with Aquatic Life Thresholds

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	<5.8	>2.1
2	Mountains	<4.9	>3.2
3	Plains & Xeric	<7.6	>2.4

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. More than 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	>56	-22
2	Mountains	>62	-22
3	Plains & Xeric	>51	-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.

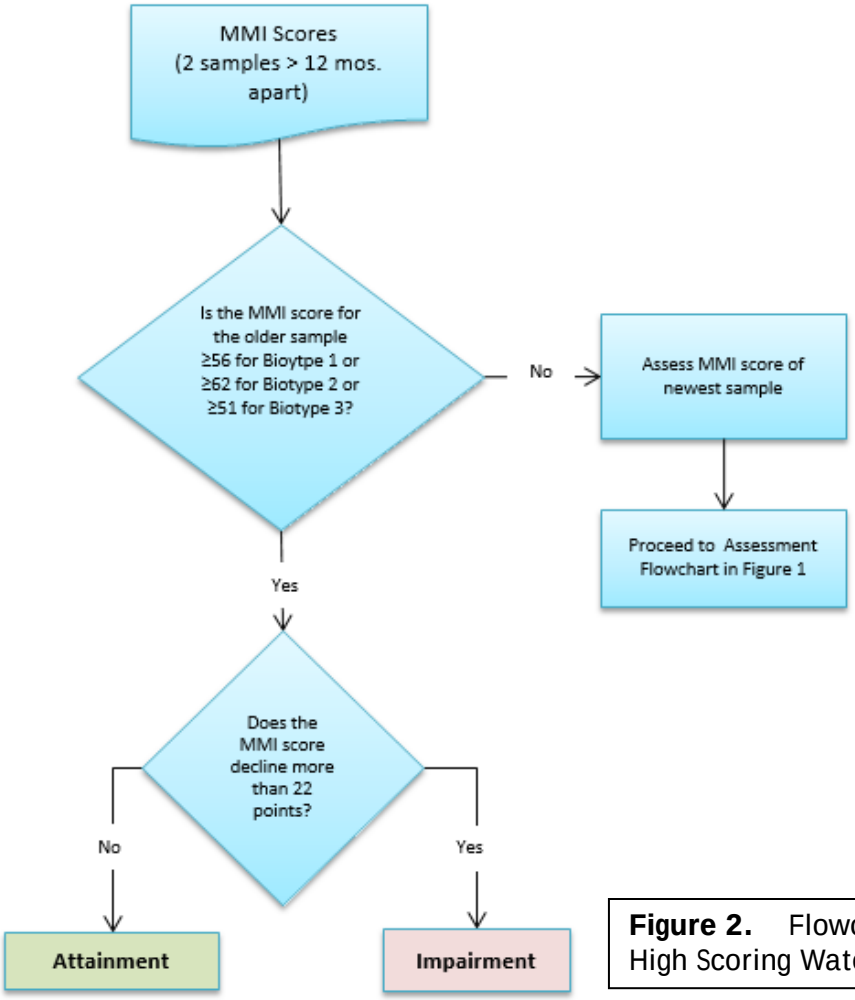


Figure 2. Flowchart: Protection of High Scoring Waters

B. Use Attainment Thresholds for Big Rivers:

The commission will apply the 2017 MMI-based Impairment Thresholds to watershed areas >2,700 mi^2 in the Yampa River, White River, Colorado River, Gunnison River, Dolores River, and San Juan River basins. Cluster analysis has demonstrated that the benthic macroinvertebrate composition, as defined by common genus taxonomy, in those bigger rivers are statistically comparable to the benthic macroinvertebrate compositions in smaller sized tributaries within the same basins. See Appendix F for data used in this analysis.

The commission will continue not to apply the 2017 MMI-based Impairment Threshold to watershed areas >2,700 mi^2 in the South Platte River, Arkansas River, including the Purgatoire River, and Rio Grande River basins.

C. Considerations of Representative Data

The 2016 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

2016 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 seasonal water level fluctuations that may dry/wet a portion of the targeted sampling habitat, 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. Therefore, it is important to ensure that samples are collected from habitat with similar depth/substrate/velocity conditions to minimize the potential effect of sampling bias. Furthermore, 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). It is also important to consider the potential change from perennial to seasonally intermittent/ephemeral flow conditions within the segment, especially during alternating dry/wet year type conditions.

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. The mechanism for establishing alternate methodologies for assessing the biological condition against the expected condition is best defined through the Regulation No. 93 Section 303(d) Listing Methodology, which is updated biennially.

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.

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

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

APPENDIX A

MMI FINAL REPORT

Recalibration of the Colorado Benthic Macroinvertebrate Multimetric Index 2017

Prepared for:

Colorado Department of Public Health and Environment
Water Quality Control Division – Environmental Data Unit

Prepared by:

Benjamin K. Jessup[1] and James B. Stribling[2]
Tetra Tech, Inc.
Center for Ecological Sciences

[1]73 Main Street, Suite 38
Montpelier, VT 05602
and
[2]10711 Red Run Boulevard, Suite 105
Owings Mills, MD 21117

Final Report
April 25, 2017

Abstract

The Colorado Department of Public Health and Environment (CDPHE) monitors streams throughout the state for assessment and protection of water resource quality. Their principal indicator is a multimetric index (MMI) which is based on direct benthic macroinvertebrate sample data collected from each waterbody. As a quantitative descriptor of stream conditions, the MMI and its components can be associated with known environmental gradients, both natural and those resulting from human disturbance. Prior index calibration in 2008 identified gradients related to natural conditions, and within them recognized three distinct sites classes as biological types, generally related to sites in the mountains (Biotype 2), plains (Biotype 3), and transitional regions (Biotype 1). Within each Biotype, variability was quantified and partitioned by adjusting metrics relative to factors such as summer temperature, elevation, and sample date. The purpose of this project is to recalibrate the Colorado MMI with sample data from 601 additional stream sites collected since 2008, and an expanded understanding of statewide stream characteristics. Defining responsiveness and robustness of metrics as those that, respectively, exhibit detectable change in the presence of stressors, and that represent diverse sample characteristics, we evaluated 104 metrics. Responsiveness of more than 150,000 index alternatives per Biotype were tested for their ability to detect differences between reference and stressor sites. Three separate indices, one for each Biotype, were selected, are comprised of eight metrics, and have a cumulative capacity for correctly distinguishing reference from stressed condition (discrimination efficiency [DE]) 97.1 percent of the time.

This page intentionally blank

Table of Contents

Abstract i
Table of Contents iii
List of Tables iv
List of Figures v
1 Introduction 1
Clean Water Act 1
2 Data Description 2
3 Analytical Methods 6
3.1 Site Classification 6
Biological Groupings 8
Discerning Environmental Factors 9
Site Classes 11
3.2 Metric Calculations 12
Metric Descriptions 12
Indicator taxa 15
Metric adjustment to environmental conditions 16
Metric Evaluation 17
3.3 Index Composition 19
Metric Scoring 19
Index Composition 20
Metric and Index Precision 21
4 Results 21
4.1 Site Classification 21
Ordination 22
Cluster Analysis 25
Metric and Environmental Variable Distributions 26
Discriminant Analysis 29
CART and Random Forest Analysis 29
Site Classification Conclusions 34
4.2 Metric Development 36
Indicator Taxa 36
Metrics Adjustments to Environmental Variables 37
Metric Performance 38
4.3 Index Performance 39
Biotype 1 39
Biotype 2 45
Biotype 3 51
4.4 Index Application 58
5 Conclusions and Recommendations 59
6 References Cited 60

Appendices

A Reference and Stressed Sites used in Analysis
B Metric Adjustments
C Metric Performance

List of Tables

Table 1. Reference and stressed site tally by ecoregion. 3
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. 10
Table 4. Metric descriptions. 13
Table 5. Variables tested for metric adjustments. 16
Table 6. Site Classification Rules. 22
Table 7. Random Forest importance of environmental variables in predicting reference PA clusters. 29
Table 8. Indicator taxa derived from an indicator species analysis on reference and stressed sites. 36
Table 9. Mid-elevation sensitive and tolerant families. 37
Table 10. Metrics included in the Biotype 1 MMI with scoring formulae. 39
Table 11. The best index alternatives considered for Biotype 1, including eight that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_61382. 42
Table 12. Precision statistics for the selected Biotype 1 index and its component metrics, based on both metric scores and values. 43
Table 13. Metrics included in the Biotype 2 MMI with scoring formulae. 45
Table 14. The best index alternatives considered for Biotype 2, including 8 that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_78149. 48
Table 15. Precision statistics for the selected Biotype 2 index and its component metrics, based on both metric scores and values. 49
Table 16. Metrics included in the Biotype 3 MMI with scoring formulae. 51
Table 17. The best index alternatives considered for Biotype 3, including 8 that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_70028. 54
Table 18. Precision statistics for the selected Biotype 3 index and its component metrics, based on both metric scores and values. 55
Table 19. MMI performance statistics. 58

List of Figures

Figure 1. Reference and stressed sites by Biotype (Group) overlaid on level 3 ecoregions in Colorado..... 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. 8

Figure 3. Example diagram of metric adjustment showing optimal expectations (black dashed line) for Beck's Biotic Index along the elevation gradient. The red line is the LOWESS regression line. 17

Figure 4. Illustration of metric discrimination efficiency (DE) between reference and stressed sites for EPT taxa in Biotype 1. The DE is between 50 and 75%, judging from the stressed metric value distribution relative to the reference 25th percentile. 18

Figure 5. Illustration of metric Z-score between reference and stressed sites for EPT taxa in Biotype 2. It appears that approximately 1.4 standard deviations of reference sites metric values fit into the gap between the reference and stressed mean values. 19

Figure 6. NMS ordination of taxa presence/absence data in 144 reference sites, showing the existing classification scheme with three Biotypes..... 23

Figure 7. NMS ordination of taxa relative abundance data in 144 reference sites, showing the existing classification scheme with three Biotypes..... 24

Figure 8. PCA ordination of taxa relative abundance data in 144 reference sites, showing the existing classification scheme with three Biotypes..... 24

Figure 9. Cluster analysis using taxa presence/absence data from 144 reference sites, showing color-coded existing Biotypes. 25

Figure 10. Cluster analysis using taxa presence/absence data from 144 reference sites, showing color-coded Bioregions (Mountains = red, Plains = blue, Xeric = black). 26

Figure 11. Distributions of environmental variables (catchment size, elevation, and summer air temperature) in the Biotypes. 27

Figure 12. Distributions of macroinvertebrate metrics (EPT taxa, predator + scraper taxa, and Hilsenhoff's Biotic Index [HBI]) in the Biotypes. 28

Figure 13. CART dendrogram showing distinctions for Biotypes 3 and 2 and a split of Biotype 1 sites by longitude..... 30

Figure 14. East/West divide for transitional sites (red dashed line) derived from CART analysis..... 30

Figure 15. NMS ordination of reference presence/absence data, showing Biotypes 2 and 3 and western and eastern sites in Biotype 1 (transitional)..... 31

Figure 16. Distributions of environmental variables (catchment size, elevation, and summer air temperature) in the CART groups, which include Biotypes 2 and 3 and western Biotype (10) and eastern Biotype 1 (11). 32

Figure 17. Distributions of macroinvertebrate metrics (EPT taxa, predator + scraper taxa, and Hilsenhoff's Biotic Index [HBI]) in the CART groups, which include Biotypes 2 and 3 and western Biotype (10) and eastern Biotype 1 (11)..... 33

Figure 18. Biotypes (Groups) for the macroinvertebrate index in Colorado. 35

Figure 19. Distribution of index scores (8_61382) in reference and stressed sites in Biotype 1, showing calibration and validation data sets. 40

Figure 20. Bi-plots of index metrics that were adjusted to summer temperature in Biotype 1, including EPT taxa, percent intolerant taxa, and clinger taxa. The plots show the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots) 41

Figure 21. Biotype 1 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions. 44

Figure 22. Distribution of index scores (8_78149) in reference and stressed sites in Biotype 2, showing calibration and validation data sets. 46

Figure 23. Bi-plot of index metric (EPT taxa) that was adjusted to Julian Day in Biotype 2, showing the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots). 46

Figure 24. Biotype 2 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions. 50

Figure 25. Distribution of index scores (8_70028) in reference and stressed sites in Biotype 3, showing calibration and validation data sets. 52

Figure 26. Sprawler taxa in relation to Julian Day in Biotype 3, showing the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots). 53

Figure 27. Biotype 3 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions. 56

Figure 28. Biotype 3 index values in relation to longitude, showing reference and stressed sites. 57

1 Introduction

Clean Water Act

The Clean Water Act of 1972 ([CWA] USGPO 1989) has as its primary objective the maintenance and restoration of "biological integrity" of the Nation's waters, which incorporates biological, physical, and chemical quality (§101[a]). This concept of biological integrity refers to environmental conditions that are undisturbed by human activities, and that have the capacity for supporting a natural assemblage of native organisms (Frey 1977, Karr et al. 1986, Karr 1991, Steedman 1994). As such, the most accurate, comprehensive, and robust primary indicator of undisturbed conditions is one that quantitatively describes the biota inhabiting them, with measured differences taken as response to degradation. The measurable definition of biological integrity is the *reference condition* (Barbour et al. 1995, 1999, Stoddard et al. 2006), which is characterized using data from minimally-disturbed reference sites within a region. In responding to the CWA, CDPHE has developed water quality standards for the protection of human and ecosystem health under the Colorado Water Quality Control Act ("CWQCA"), including establishing aquatic life uses (ALU) (Policy Statement 10-1 [CDPHE 2010a]). The Agency is committed to adhering to the CWA and the CWQCA, thus justifying continued implementation of Policy Statement 10-1 through refinement/recalibration of biological indicators, using them in routine stream monitoring, and fully developing their reference condition approach.

Why biological assessment?
The biological condition of minimally disturbed reference sites, as quantified and characterized by individual biological metrics and the MMI, becomes the standard to which those of other sites are compared. It is, thus, critical that reference sites are properly identified, biological data used to characterize those sites are of known and acceptable quality, natural variability is well described and accounted for, and that the metric selection process is structured to result in those that are most responsive to prevalent stressors. Following these principles provides a level of rigor and assurance that the resulting MMI is optimized for accurate comparisons and valid, defensible assessments that will strengthen most aspects of environmental management decision-making. In particular, CDPHE will have elevated the confidence with which objective judgments can be made on ALU attainment/non-attainment for individual waterbodies/streams.

Through the §303(d) and Total Maximum Daily Load (TMDL) framework outlined by the CWA (USGPO 1989), waterbodies are considered impaired and threatened if they fail to meet their designated uses, such as ALU. The CWA requires that waterbody impairment must be identified, and when identified, those impaired waterbodies must be improved to move them in the direction of attaining their designated use. The CDPHE is continuing use of the MMI as the biological criteria component of ALU.

The Colorado Department of Public Health and Environment established a statewide biological monitoring program for detecting and reporting stream conditions throughout the state. The Colorado MMI (Jessup 2009) has been applied to benthic macroinvertebrate sample data, allowing the biological assessment results to be used for the State water quality inventory (CWA §305[b] report) and list of impaired waters (CWA §303[d] report), most recently presented as the CDPHE *Integrated Water Quality Monitoring and Assessment Report 2016*[1] The objective of this project was to recalibrate the MMI using sample data from an additional 601 stream sites. The additional biological, chemical, and physical habitat data fill spatial and temporal gaps noted by Jessup (2009) during the prior MMI work, thus providing an improved understanding of variability in stream characteristics and ultimately greater robustness in MMI application and interpretation.

Index development process

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

1. Collect, compile, and subject to Quality Control (QC) analysis
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 the index as appropriate for biological assessment

These steps are presented as generally linear, although the process often is more iterative as analyses proceed and additional questions arise relevant to earlier steps. We received the dataset that had already been compiled and undergone QC analysis (Step 1) and for which assignment of reference status (Step 2) had been completed. The MMI development process focuses on optimizing sensitivity, accuracy, and precision of individual metrics and the final index. This will allow CDPHE to apply the index with confidence in the result. The purpose of this report is to present the analysis and discussion of the MMI recalibration for wadeable streams in Colorado, and to demonstrate applicability of the index to continued CDPHE implementation of Policy Statement 10-1 (CDPHE 2010a).

2 Data Description

The Colorado Department of Public Health and Environment provided a robust dataset in the Ecological Data Application System (EDAS; a Microsoft Access database) for use in this study. This database included benthic macroinvertebrate, physical habitat, and water chemistry data collected between 1994 and 2015. The dataset included data points obtained from other agencies. These additional datasets were of high quality and employed similar methods as those used by the CDPHE. The additional datasets included

[1] *https://www.colorado.gov/pacific/sites/default/files/WQ_2016-Integrated-Report_FINAL.pdf* (accessed March 23, 2017)

samples from 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). Since the last recalibration, which used data through 2008 (Jessup 2009), 601 new samples were collected. Of these new samples, all but 10 USFS samples were collected by CDPHE using standard field and laboratory protocols (CDPHE 2010b).

Although different programs do not use the exact same methods, the protocols used by each program were comparable and results from these samples have been compared in previous studies (Paul et al. 2005, Jessup 2009). All the samples were collected from 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, commonly genus and species. The principal difference among the programs was the subsample size. Most programs enumerated the whole sample, while CDPHE identified a fixed 300-count subsample. To standardize sample size, large samples were randomly subsampled to a fixed 300-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 and stressed sites. Sites that were not categorized (designated "Other") were generally not used in index calibrations, except to illustrate final index distributions. The reference and stressed sites were selected by CDPHE using designations assigned during the 2009 index recalibration as a starting point and then screening for intensities of land uses in the immediate and entire catchments of the sites, as described in Jessup (2009). In addition, a technical advisory committee reviewed and edited preliminary lists of reference and stressed sites. The screening process resulted in elimination of 6 previously designated reference sites and 27 previously designated stressed sites. New designations for 17 reference sites and 53 stressed sites were added, resulting in a total of 144 reference sites and 86 stressed sites (Appendix A). Samples were spatially distributed throughout the ecoregions in Colorado (Table 1, Figure 1). Sites were more numerous in ecoregions with greater land area within the state and with more wadeable streams. Only a single sample per site was selected for index calibration. Recent samples, collected since the 2009 index calibration, were substituted for previous samples in long-standing reference sites.

Table 1. Reference and stressed site tally by ecoregion.

Ecoregion (Code)	Reference	Stressed
Wyoming Basin (18)	2	1
Colorado Plateaus (20)	18	18
Southern Rockies (21)	97	43
Arizona/New Mexico Plateau (22)	1	2
Western High Plains (25)	5	18
Southwestern Tablelands (26)	21	4

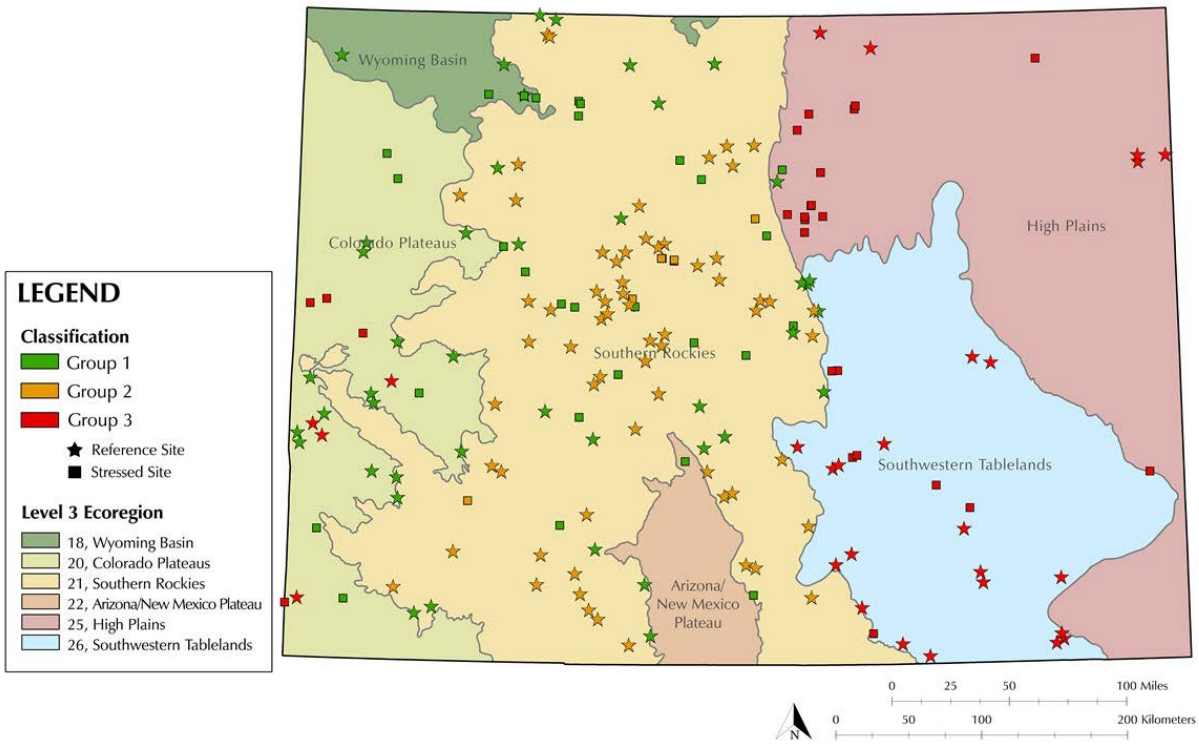


Figure 1. Reference and stressed sites by Biotype (Group) overlaid on level 3 ecoregions in Colorado.

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.

Table 2. Data treatments for consistency in analysis.

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

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 avoided because metrics are sensitive to 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, large samples were reduced through random re-sampling. In the analyzed data set, one stressed sample with 54 individuals in a mid-elevation site was used in index validation. Otherwise, samples were within the range of 150 – 360 individuals.

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.

3 Analytical Methods

The analytical methods are described in detail in the following sections and are summarized here. The multimetric index recalibration in Colorado was based on analysis of site and sample data provided by CDPHE. The least disturbed reference sites were identified by CDPHE. Site classes were established based on the premise that the existing Biotypes that were valid in the previous index calibration (2009) might still be valid for the re-calibration (2017). To resolve any remaining metric variability after classification, some metrics were adjusted to site-specific environmental conditions.

Metrics were calculated to describe multiple aspects of the macroinvertebrate samples. The ability of metrics to discriminate reference and stressed site conditions was measured using the Discrimination Efficiency (DE) and Z-score. Metrics were scored on a 0-100 scale so the standardized scores could be combined in an index for each site class. Indices were compiled using two methods: manual combination of the best performing metrics from each metric category and automated combination of all possible metric subsets. The indices were tested using the DE and Z-scores and a final index was selected from among the best performing alternatives based on performance, familiarity of metrics, diversity of metric responses, and performance of individual metrics.

The selected metrics in each site class were validated with independent sample data and precision of the indices was described. These results, along with the DE can be reported with assessment results to qualify index accuracy and precision. The index can be applied to new samples collected and processed using approved protocols from wadeable streams in Colorado.

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 important for accurate bioassessment.

In the site classification analysis, natural variability among biological community types is accounted for though explorations of sample characteristics in reference sites. The biological characteristics associated with the natural environmental setting are best recognized when they are not confounded by the effects of human disturbance. By accounting for such natural biological variability, an MMI can be specifically calibrated to the natural stream type and the responses to disturbance that might be unique to each type.

One way to account for the natural variability in biological metrics is to create distinct classes of streams that have similar biological expectations within each class. By examining the structure and composition of samples in least disturbed reference sites, a reasonable number of biological types can be recognized so that variability is reduced relative to the variability in all sites. Assigning sites into classes that reflect the biological types is accomplished by identifying the environmental conditions that determine the biological variation.

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 sites of a distinct site class are somewhat homogenous in those natural environmental factors that affect biological potential, such as soil types, topography, and climate (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 et al. 2000).

Ecoregions are commonly used as the primary classifying variable because they incorporate multiple environmental characteristics. The ecoregion concept is a trusted method for identifying general ecological categories of site characteristics (Omernik and Bailey 1997). Level 3 and 4 ecoregions have been used for classifying stream biota in several statewide assessment programs (Weigel 2003, Gerritsen et al. 2000, Van Sickle and Hughes 2000, Barbour et al. 1995, Bryce and Clarke 1996, Rohm et al. 1987, Paul et al. 2005, Stribling et al. 2000). Because ecoregions are such a prominent classification consideration they are mentioned in several steps of the index development process.

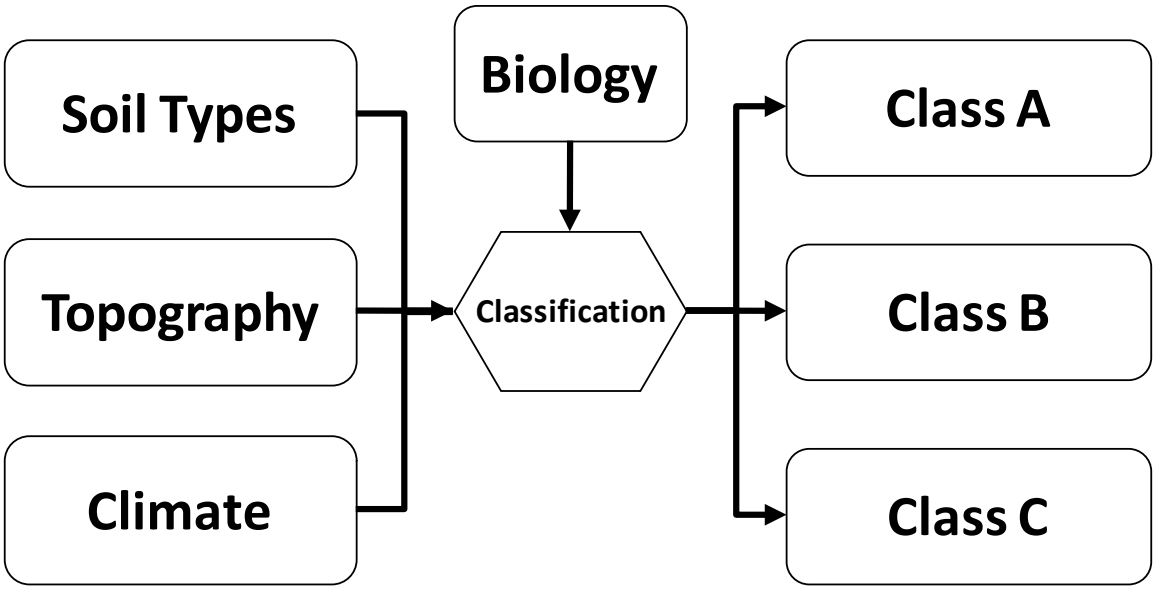


Figure 2. Typical inputs and results in a discrete classification process. Inputs include natural environmental variables and biological similarities among sites. Resulting site classes are discrete bins, where a site can be placed in only one bin.

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.

Both the NMS ordinations and the cluster analysis use the BC matrix as the basis for the diagram structures. In ordination, the arrangement of samples in the diagram places samples with similar taxonomic structure in close proximity and samples with few similar taxa at a greater distance from each other. Multiple ordination dimensions are compressed into the two or three dimensions that we can perceive and that explain large

amounts of taxonomic variability. Taxon presence does not take abundance into account, so that a single individual or 50 individuals would be counted equally in the presence/absence analyses.

Fidelity is the degree to which sites within Biotypes are distinguished by the major clusters. It is calculated as the percentage of sites within the three major clusters that are represented by the dominant Biotype of each cluster.

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[a]
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 Mean Summer Temperature Number of Frost Free Days Date of First Frost
Flow	Hydrologic Stability

a. 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.

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.

CART and Random Forest Analysis
Classification and Regression Trees (CART), also called Recursive Partitioning (RPART), were used to find the classification variables that were associated with differences in reference biological clusters. In CART analysis, the variables that explain differences between groups are identified one step at a time and the multiple splits in the data can be evaluated for agreement with the original groups. Random Forest analysis was used to identify the most important classification variables. Random Forest analysis is an iterative application of CART analysis that identifies those variables that are best at explaining differences among groups with repeated analyses. Analyses were conducted using R software.

Site Classes

Patterns in the environmental data distributions were observed in graphic displays of environmental variables and cluster assignments. The existing Biotypes were used as one

possible classification scheme. Those Biotypes were defined through mapping exercises displaying sites coded by potential groupings and overlain on landscape characteristics (such as ecoregion). 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, DFA, and CART analysis.

3.2 Metric Calculations

Metric Descriptions

Metrics were calculated as counts of taxa, relative percentages of individuals, or complex algorithms derived from the sample taxa lists. One hundred four (104) commonly applied, empirically proven, and theoretically responsive metrics were calculated for possible inclusion in a multimetric index (Table 4). The metrics were organized into five categories: richness, composition, pollution tolerance, functional feeding group, and habit (mode of locomotion). The calculations were made in the EDAS database. Large samples with more than 360 individuals were randomly subsampled to 300 organisms to control for high taxa richness that is associated with such unusually large samples.

Functional feeding group (FFG) traits were assigned to taxa by CDPHE and were recorded in EDAS. Each taxon could be associated with multiple FFGs. In the first generation of metric calculations, only the first (and presumably predominant) FFG was used in calculating richness and percent individuals. In this recalibration, we also calculated metrics based on any of the multiple FFG. These were termed ‘facultative’ (coded ‘_FAC’), assuming that a taxon with multiple feeding mechanisms might use all of the mechanisms depending on life stage. Both types of metrics were tested.

CDPHE provided information for calculation of the sediment (% fines) weighted average tolerance indicator value (TIV_{SED}) metric and for metrics related to indicator families in the mid-elevation (Biotype 1) sites. These metrics were calculated and tested in appropriate site groups.

Table 4. Metric descriptions.

Metric Code	Metric Description
Taxa Richness	**Counts of taxa, percentage of taxa**
TotalTax	Total number of taxa in the sample
InsectTax	Number of insect taxa in the sample
pt_noninsect	Non-insect taxa as a percentage of all taxa
EPTTax	Number of EPT taxa in the sample
pt_EPT	EPT taxa as a percentage of all taxa
EphemTax	Number of Ephemeroptera (E) taxa in the sample
PlecoTax	Number of Plecoptera (P) taxa in the sample
TrichTax	Number of Trichoptera (T) taxa in the sample
ColeoTax	Number of Coleoptera taxa in the sample
DipTax	Number of Diptera taxa in the sample
pt_Diptera	Diptera taxa as a percentage of all taxa
ChiroTax	Number of chironomid taxa in the sample
CrMolTax	Number of Crustacea or Mollusca taxa in the sample
OligoTax	Number of oligochaete taxa in the sample
TanytTax	Number of Tanytarsini (Chironomidae) taxa in the sample
Shan_2	Shannon-Weiner Index (base 2)
Evenness	Evenness
D_Mg	Margaleff's Index
D	Simpson's Index
PercDom01	% individuals in the dominant taxon of all individuals
Taxa Composition	**Percentages of taxa groups**
EPTPct	% EPT individuals of all individuals
pEPTnoB	% EPT excluding Baetidae of all individuals
EphemPct	% Ephemeroptera individuals of all individuals
PlecoPct	% Plecoptera individuals of all individuals
TrichPct	% Trichoptera individuals of all individuals
NonInPct	% Non-Insect individuals of all individuals
ColeoPct	% Coleoptera individuals of all individuals
PlCoPct	% Plecoptera + Coleoptera all individuals
OdonPct	% Odonata individuals of all individuals
DipPct	% Diptera individuals of all individuals
ChiroPct	% Chironomidae individuals of all individuals
CrCh2ChiPct	% *Cricotopus* + *Chironomus* of Chironomidae individuals
Orth2ChiPct	% Orthocladiinae of Chironomidae individuals
CCO2ChirPct	% *Cricotopus* + *Chironomus* + Orthocladiinae of Chironomidae
TanytPct	% Tanytarsini of all individuals
Tnyt2ChiPct	% Tanytarsini of Chironomidae individuals
OligoPct	% Oligochaeta individuals of all individuals
CrMolPct	% Crustacea and Mollusca individuals of all individuals
GastrPct	% Gastropoda individuals of all individuals

Metric Code	Metric Description
CorbPct	% Corbicula individuals of all individuals
IsoPct	% Isopoda individuals of all individuals
AmphPct	% Amphipoda individuals of all individuals
BivalPct	% Bivalvia individuals of all individuals
Habit	**Mode of locomotion or attachment**
BrrwrTax	Number of Burrower Taxa
ClmbrTax	Number of Climber Taxa
ClngrTax	Number of Clinger Taxa
SprwlTax	Number of Sprawler Taxa
SwmmrTax	Number of Swimmer Taxa
BrrwrPct	% Burrower individuals of all individuals
ClmbrPct	% Climber individuals of all individuals
ClngrPct	% Clinger individuals of all individuals
SprwlPct	% Sprawler individuals of all individuals
SwmmrPct	% Swimmer individuals of all individuals
Pollution Tolerance	**Sensitivity to stressors**
BeckBI	Beck's Index (taxa richness weighted by taxa sensitivity)
HBI	Hilsenhoff's Index (average tolerance value of individuals in the sample)
NCBI	North Carolina Biotic Index
IntolTax	Number of Intolerant Taxa
pt_Intol	Intolerant taxa as a percentage of all taxa
pt_Toler	Tolerant taxa as a percentage of all taxa
TolerTax	Number of Tolerant Taxa
InMolTax	Number of Intolerant Mollusca taxa
PercBaetEph	% Baetidae of Ephemeroptera individuals
PercIntol	% Intolerant individuals of all individuals
PercTol	% Tolerant individuals of all individuals
PercHydr2Tri	% Hydropsychidae of Trichoptera individuals
PercHydr2EPT	% Hydropsychidae of EPT individuals
TIV_{SED}	Sediment (% fines) weighted average tolerance indicator value
nt_DecrMtnTrn	Number of decreaser indicator taxa in Biotypes 1 and 2
pi_DecrMtnTrn	% decreaser indicator individuals in Biotypes 1 and 2
nt_IncrMtnTrn	Number of increaser indicator taxa in Biotypes 1 and 2
pi_IncrMtnTrn	% increaser indicator individuals in Biotypes 1 and 2
nt_DecrMidElev	Number of decreaser indicator taxa in Biotypes 1
pi_DecrMidElev	% decreaser indicator individuals in Biotypes 1
nt_IncrMidElev	Number of increaser indicator taxa in Biotypes 1
pi_IncrMidElev	% increaser indicator individuals in Biotypes 1
nt_DecrPlains	Number of decreaser indicator taxa in Biotypes 3
pi_DecrPlains	% decreaser indicator individuals in Biotypes 3
nt_IncrPlains	Number of increaser indicator taxa in Biotypes 3
pi_IncrPlains	% increaser indicator individuals in Biotypes 3

Metric Code	Metric Description
Functional Feeding Group	**Mechanism for obtaining food**
CllctPctFAC	% Facultative Collector individuals of all individuals
FiltrPctFAC	% Facultative Filterer individuals of all individuals
PredPctFAC	% Facultative Predator individuals of all individuals
ScrapPctFAC	% Facultative Scraper individuals of all individuals
ShredPctFAC	% Facultative Shredder individuals of all individuals
CllctTaxFAC	Number of Facultative Collector Taxa
FiltrTaxFAC	Number of Facultative Filterer Taxa
PredTaxFAC	Number of Facultative Predator Taxa
ScrapTaxFAC	Number of Facultative Scraper Taxa
ShredTaxFAC	Number of Facultative Shredder Taxa
GathFiltTaxFAC	Number of Facultative Gatherer or Filterer Taxa
PredScrapTaxFAC	Number of Facultative Predator or Scraper Taxa
PredShrTaxFAC	Number of Facultative Predator or Shredder Taxa
CllctPct	% Collector individuals of all individuals
FiltrPct	% Filterer individuals of all individuals
PredPct	% Predator individuals of all individuals
ScrapPct	% Scraper individuals of all individuals
ShredPct	% Shredder individuals of all individuals
CllctTax	Number of Collector Taxa
FiltrTax	Number of Filterer Taxa
PredTax	Number of Predator Taxa
ScrapTax	Number of Scraper Taxa
ShredTax	Number of Shredder Taxa
PredScrapTax	Number of Predator or Scraper Taxa
PredShrTax	Number of Predator or Shredder Taxa

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 site groups. 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 reference or stressed samples. By combining frequency of occurrence and relative abundance in the comparison of reference and stressed sites, an indicator value was calculated for each taxon and it was associated with an estimate of significance (p value). Indicator taxa were identified statewide and in the Biotypes.

Indicator families for mid-elevation sites were recommended by CDPHE based on professional judgment of their technical advisory committee. These included families that are sensitive or are tolerant of increasing stress. Taxa counts and percent individual

metrics were developed based on the families and applied in Biotype 1, the mid-elevation transitional sites.

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. Metrics were not adjusted based on strong correlation coefficients alone. Where high reference correlations existed (Spearman correlation coefficient > 0.40), correlations were also checked in all sites to confirm that the reference limitations were robust. Then, if the bi-plots confirmed that the environmental variable had a limiting effect on the metric, metrics were adjusted to the environmental variable through quantile regression.

The variables tested for metric adjustment included those that represented natural differences among site types (Table 5). Data were provided by CDPHE for most variables. Temperature was represented as average summer and winter air temperature from PRISM long-term averages[2] and as NorWeST[3] modeled stream temperature.

Table 5. Variables tested for metric adjustments.

Variable Code	Variable Description
JulianDay	Sampling date as the consecutive day of the year (0-364)
Order	Strahler stream order
Catchment Area	Catchment area in km^2
Lat_Dec	Latitude in decimal degrees
Long_Dec	Longitude in decimal degrees
NHDSLOPE	Stream slope from the National Hydrography Dataset
SUMMER	Average summer air temperature
ELEV_SITE	Elevation at the site in meters
WINTER	Average winter air temperature
LOG_XP_PT	Precipitation at the site (log transformed)
SQRT_TOPO	Topographic relief (square root transformed)
S2_02_11	Water temperature from the NorWest models

The 5th or 95th quantile regression line was used as the optimal value for the metric relative to the environmental factor. The metric score was calculated as a percentage of the optimal value. If a relationship appeared to have a plateau of metric values along the

[2] http://www.prism.oregonstate.edu/normals/

[3] https://www.fs.fed.us/rm/boise/AWAE/projects/NorWeST/ModeledStreamTemperatureScenarioMaps.shtml

environmental gradient, then the plateau value was calculated from the 95th quantile value at the point of the perceived plateau. In the range of the apparent plateau, scores used the plateau value as the optimal value in metric scoring.

The relationship between Beck's Biotic Index and elevation is presented as an example of metric adjustment during scoring. From the points in the bi-plot (Figure 3) and from correlation analysis, it was determined that an adjustment to the metric was warranted at elevations below 2500 meters. The plateau of metric values at 2500 meters was estimated by eye, based on the LOWESS regression line that levels off at that point and the coinciding peak in maximum metric values. The 95th quantile regression through reference sites within Biotypes was calculated in R software. This was set as the optimal value for metric scoring for elevations less than 2500 meters. Above 2500 meters, the plateau value was calculated from the 95th quantile line. Thus, scoring for Beck's Biotic Index would be scaled to the maximum index expectation, which was derived from a 95th quantile regression up to a saturation level.

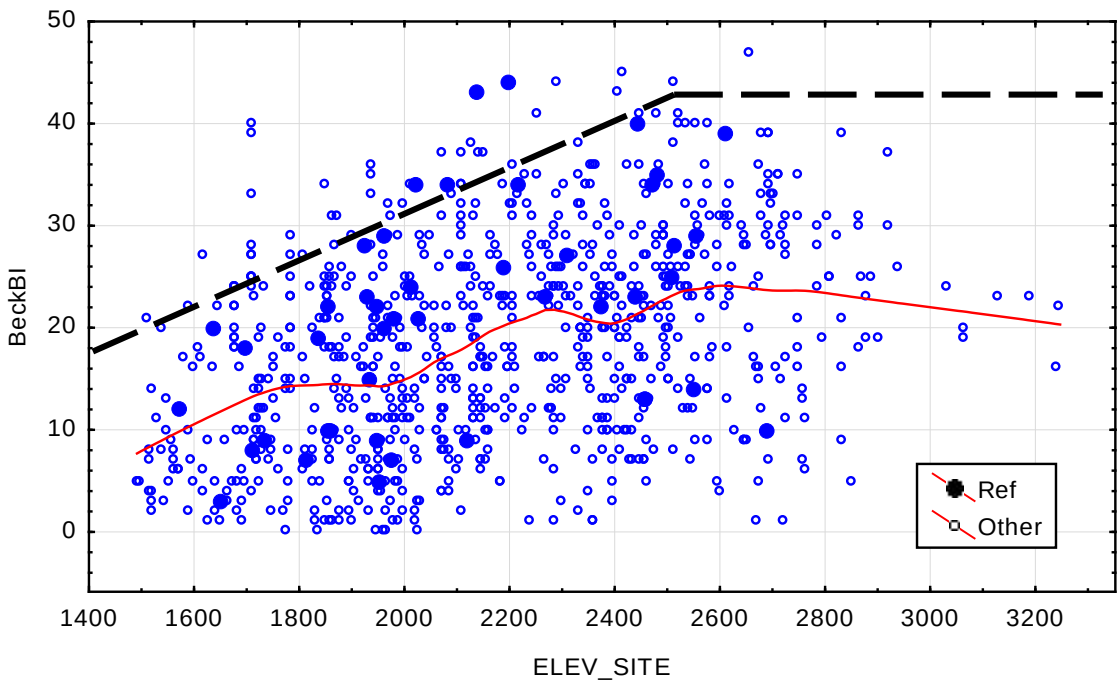


Figure 3. Example diagram of metric adjustment showing optimal expectations (black dashed line) for Beck's Biotic Index along the elevation gradient. The red line is the LOWESS regression line.

Metric Evaluation

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). 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 4). 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.

Metric discrimination between reference and stressed sites can also be quantified using the Z-score. 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 (Figure 5). The Z-score is similar to Cohen's D (Cohen, 1992) and gives a combined measure of index sensitivity and precision. There is no absolute Z-score value that indicates adequate metric performance, but among metrics or indices, higher Z-scores suggest better separation of reference and stressed values. Cohen proposed that Z-scores ≥ 0.80 indicated a "large" effect.

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.

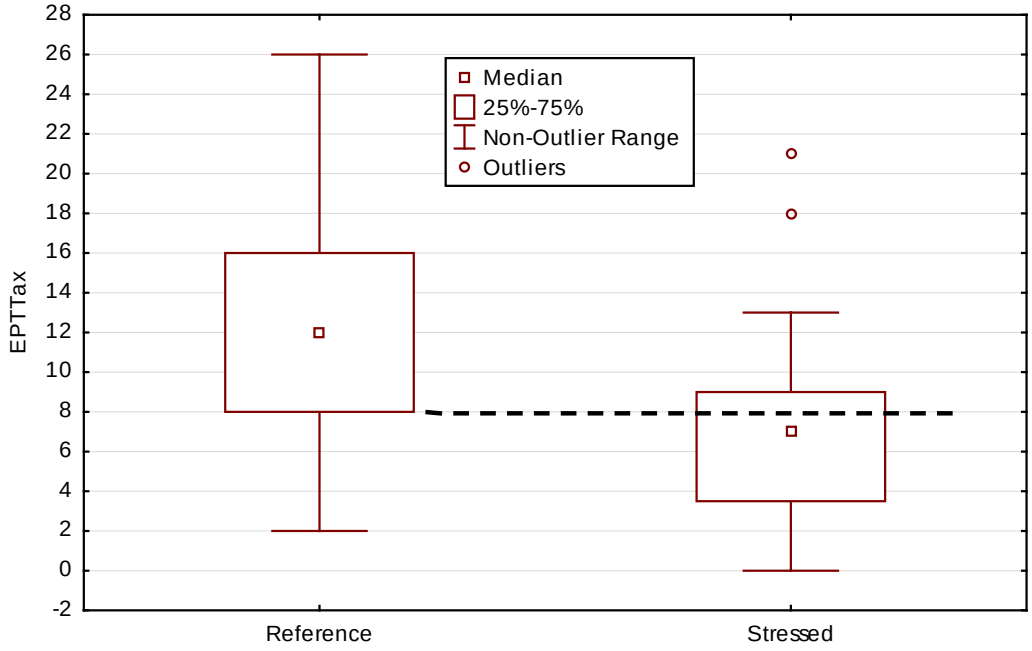


Figure 4. Illustration of metric discrimination efficiency (DE) between reference and stressed sites for EPT taxa in Biotype 1. The DE is between 50 and 75%, judging from the stressed metric value distribution relative to the reference 25th percentile.

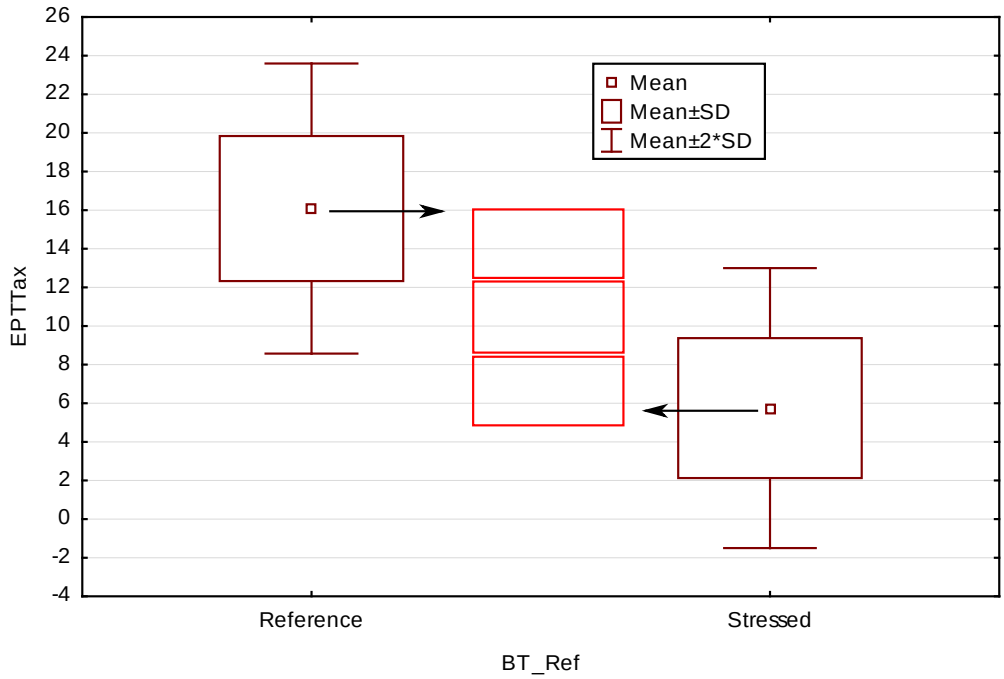


Figure 5. Illustration of metric Z-score between reference and stressed sites for EPT taxa in Biotype 2. It appears that approximately 2.9 standard deviations of reference sites metric values fit into the gap between the reference and stressed mean values.

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.

Metric Scoring

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. Scoring metrics on a continuous scale has proven to be an effective scoring method for multimetric index development (Blocksom, K.A. 2003). 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.

Index Composition

Alternative index compositions were formulated from the best performing metrics in each of the five metric categories. The metrics were combined by first scoring them and then averaging the scores. Each alternative index was then evaluated for DE and other measures of sensitivity and representativeness. Multiple index formulations were created and evaluated in two ways: manual metric substitutions and automatic all-subsets modeling.

In the manual substitution method, the best metrics were included in an initial index and subsequent index alternatives were formulated by removing, adding, or substituting metrics. The performance of each index formulation was used to determine whether the latest alternative was an improvement. New formulations built upon the best performing alternatives. In each Biotype, 20-40 index alternatives were considered.

In automatic all-subsets modeling, which was only applied in the macroinvertebrate assemblage, the best 19-21 candidate metrics in each Biotype were selected for inclusion in index trials. Metric selection was limited because computing time greatly increases with more than 21 metrics. Metric selections were based on the best performers in the manual index trials. An "all subsets" routine in R software was used to combine up to 8 metrics in multiple index trials and evaluate each index for performance using DE and Z-score to quantify sensitivity, number of metric categories, and redundancy of component metrics. DE and Z-score of the index alternatives were calculated as described for individual metrics.

As many metric categories as possible were represented in the index alternatives so that signals of various stressor-response relationships would be integrated into the index. While several metrics should be included to represent biological integrity, redundant metrics can bias an index to show responses specific to certain stressors or taxonomic responses, which might reduce sensitivity to a broader range of responses.

Models including correlated metrics (Spearman rho ≥ 0.85) were automatically excluded from consideration in the all-subsets routine. Redundancy was also considered from a conceptual perspective. For example, % EPT and % Ephemeroptera might not be highly correlated, but one measure is a component of the other and this is conceptually redundant. Index alternatives with conceptually redundant metrics were not selected.

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.

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.

Metric and Index Precision

Field sampling precision is calculated using metric and index values from the set of sample pairs (primary and duplicate) within Biotypes. Metric precision was quantified using replicate samples collected from adjacent reaches of the same site on the same day. Using ANOVA in Statistica software, the mean squared error was calculated for the replicates. The square root of the MSE is the RMSE, which is an estimate of the standard deviation within replicate sets. The coefficient of variability and 90% confidence intervals were calculated from the RMSE and mean values of the replicates (Zar 1999). Statistics were calculated for the selected indices and their component metrics and metric scores. The use of these statistics in characterizing bioassessment data quality is described by Flotemersch et al. (2006) and Stribling et al. (2008).

4 Results

4.1 Site Classification

The primary goal of site classification is to recognize natural variability in biological conditions before assessing effects of human disturbance. The site classes developed for existing macroinvertebrate MMI in Colorado describe how biological types (Biotypes) are essentially different between mountain and plains sites, and that a transitional zone has characteristics of both, but can be recognized as also supporting a distinct biological type (Jessup 2009). Because the existing Biotypes have been successfully applied and are well-understood in the context of stream macroinvertebrate assessments in Colorado, there was incentive to test the existing classification scheme for continued application. While continuity in site classification would provide a stable framework for assessments over time, alternative classification schemes were also explored to rule out the possibility of more robust Biotypes.

The site classification analysis proceeded to confirm the existing or alternative classification schemes by essentially repeating the original site classification analysis of Jessup (2009) using new data and new reference site designations. The new analysis addressed the questions: ‘Are the existing Biotypes still adequate for site classification?’ and ‘Should we lump, split, or redefine the classes?’ Existing Biotypes are defined as in Table 6.

Table 6. Site Classification Rules.

Biotype 3: If ecoregion = 25 or 26 (except sub-ecoregion 25l and 26i), then Biotype = 3 If elevation at the site < 1550, then Biotype = 3
Biotype 2: If sub-ecoregion (level 4) = 21a, 21b, 21e, or 21g, then Biotype = 2 If sub-ecoregion (level 4) = 21c and slope > 0.04, then Biotype = 2 (1 if shallower) If sub-ecoregion (level 4) = 21f and elevation >2500, then Biotype = 2 (1 if lower)
Biotype 1: If site is not predicted in classes 2 or 3, then Biotype = 1 (e.g., If sub-ecoregion (level 4) = 25l, 26i, 21d, 21h, 21i, or 21j, then Biotype = 1)
Subecoregions used in the rules Used in rules for Biotype 3. 25l = Front Range Fans 26i = Pine-Oak Woodlands Used in rules for Biotype 2. 21a = Alpine Zone 21b = Crystalline Subalpine Forests 21e = Sedimentary Subalpine Forests 21g = Volcanic Subalpine Forests 21c = Crystalline Mid-Elevation Forests (northern Front Range) 21f = Sedimentary Mid-Elevation Forests (primarily western slopes) Used in rules for Biotype 1: 21d = Foothill Shrublands 21h = Volcanic Mid-Elevation Forests 21i = Sagebrush Parks 21j = Grassland Parks

Ordination

The NMS ordination analysis primarily used presence/absence data from 144 reference sites. The resulting ordination diagram resulted in three dimensions and had a final stress of 16.2, within the acceptable range of < 20. The existing Biotypes were distinct, with no overlap of types 2 and 3, and type 1 transitioning between Biotypes 2 and 3, but with only a few sites overlapping in the ordination space (Figure 6). Because sites from different Biotypes are in distinct regions of the diagram and rarely overlap, we interpret that the three Biotypes have essentially distinct taxonomic structure. Julian day and longitude were the strongest sampling and environmental variables related to Axis 1. Samples collected earlier in the season in eastern Plains sites appear at the left of the diagram. Catchment size was strongly related to Axis 2. Smaller streams in the mountains appear in the lower portions of the diagram.

For this analytical data set, presence-absence data were more robust than relative abundance or metrics for identifying sample similarities. The ordination using taxa

relative abundance (RA) (Figure 7) was very similar to the PA ordination, suggesting that not only are the taxa different in the Biotypes, but the proportions of individuals for these taxa are also distinct. The metric PCA ordination of metrics separates the Plains (Biotype 3) from Biotypes 1 and 2 (Figure 8). There is considerable overlap in metric values for Biotypes 1 and 2 (Transitional and Mountains). This suggests that metric expectations might be similar for these two Biotypes.

- Ordination evidence suggests that existing site classes are robust with the new reference data.

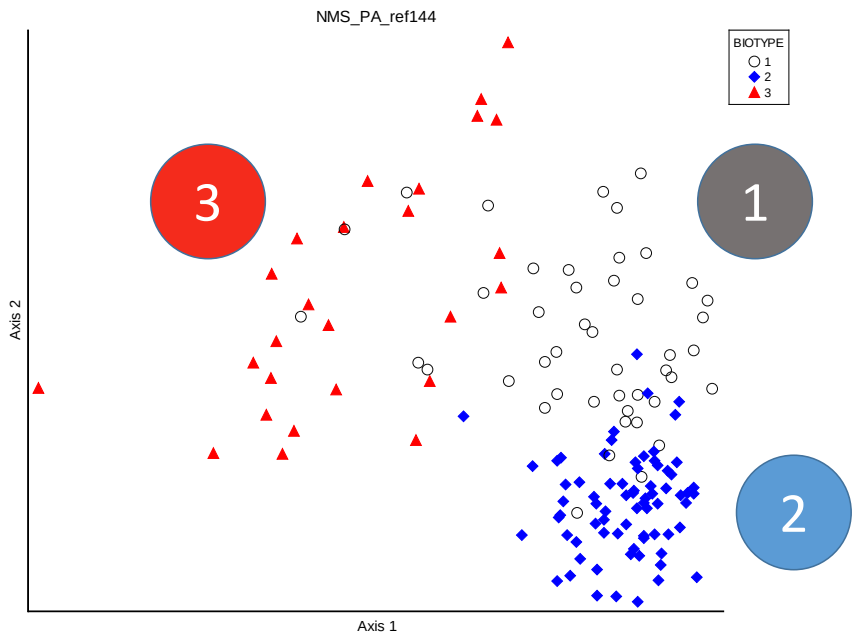


Figure 6. NMS ordination of taxa presence/absence data in 144 reference sites, showing the existing classification scheme with three Biotypes.

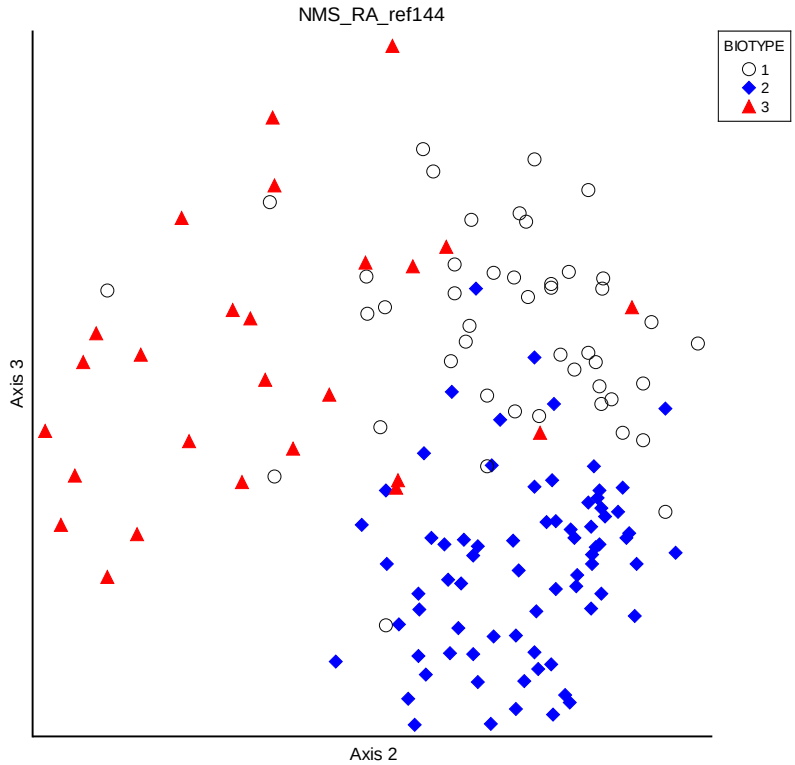


Figure 7. NMS ordination of taxa relative abundance data in 144 reference sites, showing the existing classification scheme with three Biotypes.

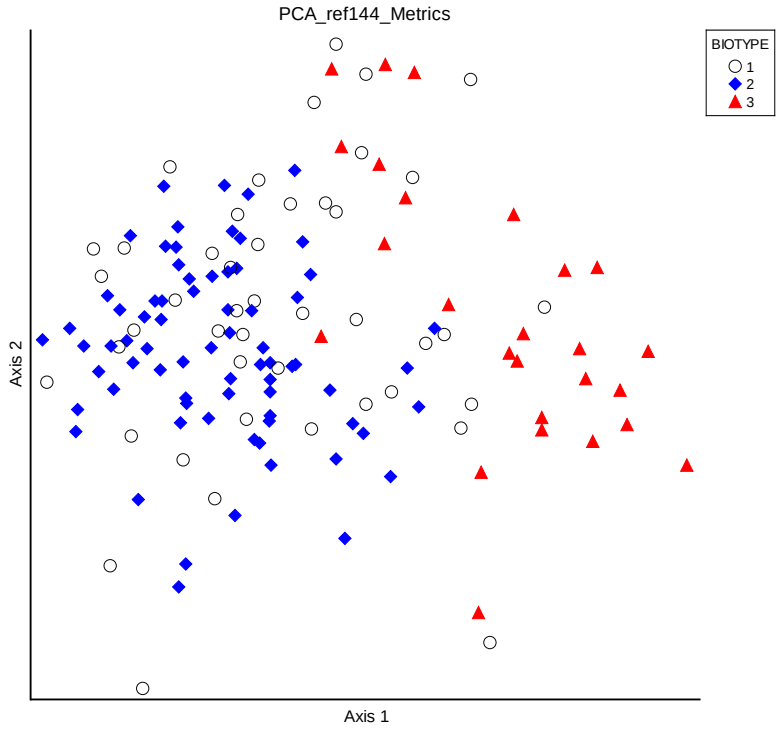


Figure 8. PCA ordination of taxa relative abundance data in 144 reference sites, showing the existing classification scheme with three Biotypes.

Cluster Analysis

The cluster analysis of reference presence-absence (PA) data with flexible beta clustering technique resulted in a cluster dendrogram with minimal chaining when beta was set at -0.3. In the cluster diagram, the major clusters are predominantly of a single Biotype (Figure 9). The fidelity to clusters was 87.5% with new data compared to 90.2% during the original classification. This degree of fidelity is high and suggests that the Biotypes account for major differences in taxonomic structure among reference sites. Bioregions (Xeric, Plains, and Mountains based on ecoregional groupings) were considered as an alternative classification scheme to Biotypes. Bioregions did not explain differences among clusters as well as Biotypes (Figure 10).

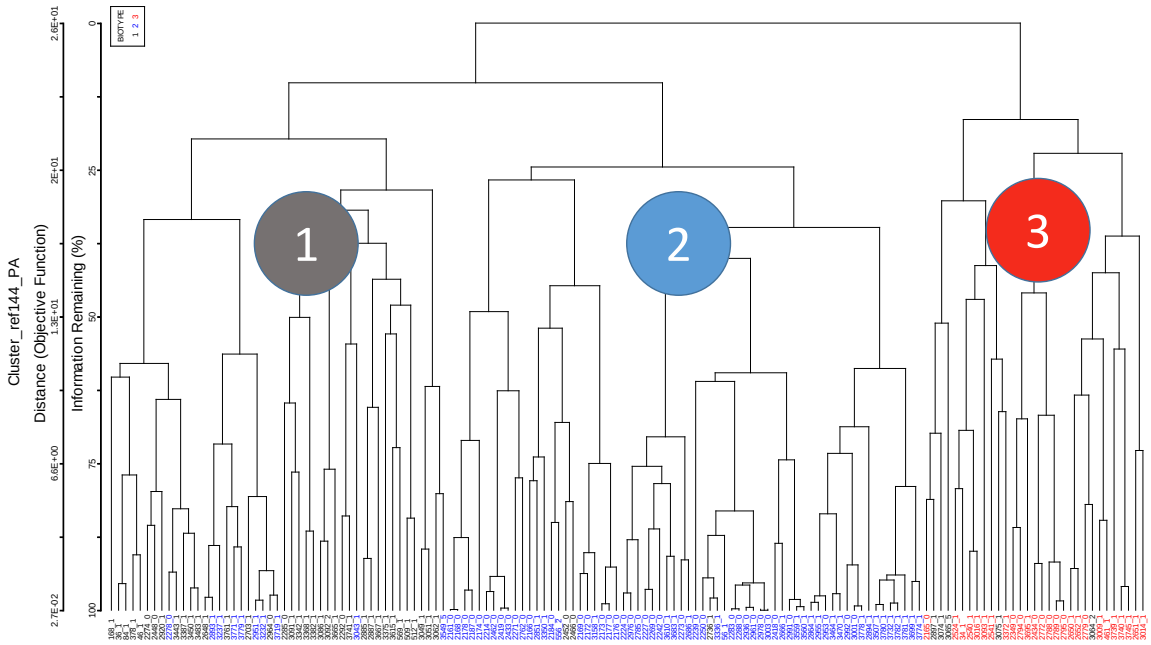


Figure 9. Cluster analysis using taxa presence/absence data from 144 reference sites, showing color-coded existing Biotypes.

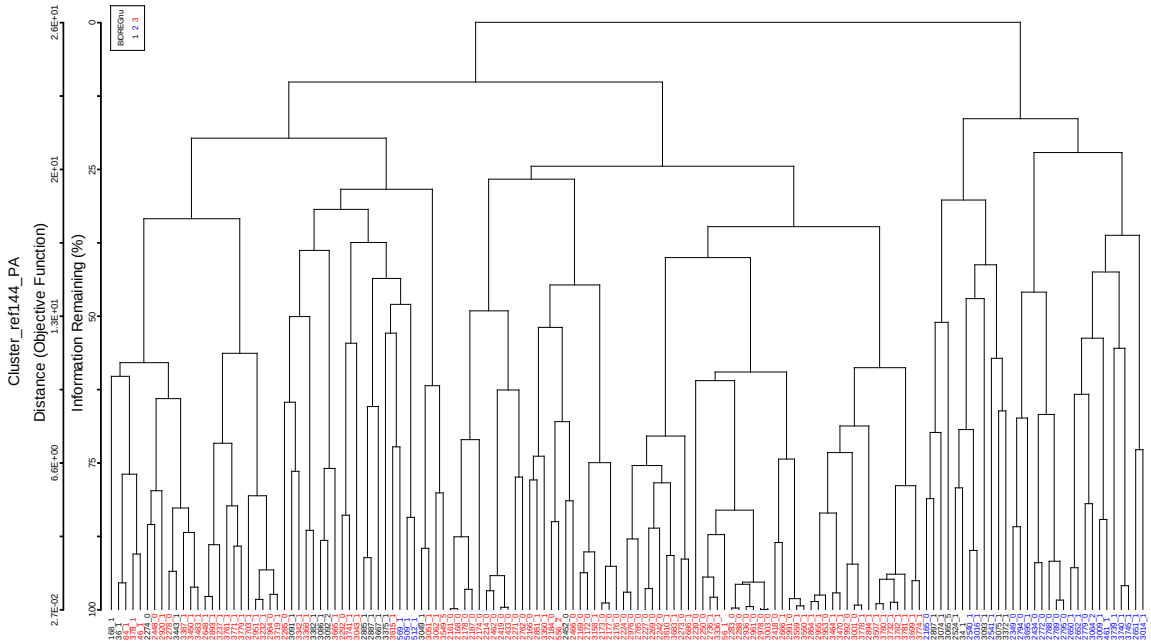


Figure 10. Cluster analysis using taxa presence/absence data from 144 reference sites, showing color-coded Bioregions (Mountains = red, Plains = blue, Xeric = black).

Metric and Environmental Variable Distributions

The distributions of environmental variables by Biotype show that the environmental characteristics of the Biotypes are distinctive (Figure 11). There are, in general, larger catchments and higher summer temperatures in Biotype 3 and higher elevations in Biotype 2. This figure includes only a subset of the environmental variables; those which showed the greatest distinctions among Biotypes and were not redundant with each other. Other variables, such as stream slope, winter temperature, and precipitation, also showed distinctions among Biotypes. In all cases, the transitional Biotype 1 values were intermediate to Biotypes 2 and 3.The plots of environmental variables include all sites, not just reference sites, because these variables are not expected to be influenced by human disturbance.

Plots of macroinvertebrate metrics among Biotypes are shown for reference sites only, because the metrics are expected to respond to stressors as well as natural settings. In general, more EPT, more predator and scraper taxa, and more sensitive individuals are in Biotype 2 samples compared to Biotypes 1 and 3 (Figure 12). Several metrics showed little distinction between Biotype 1 and 2 reference sites, while Biotype 3 metrics were often distinct.

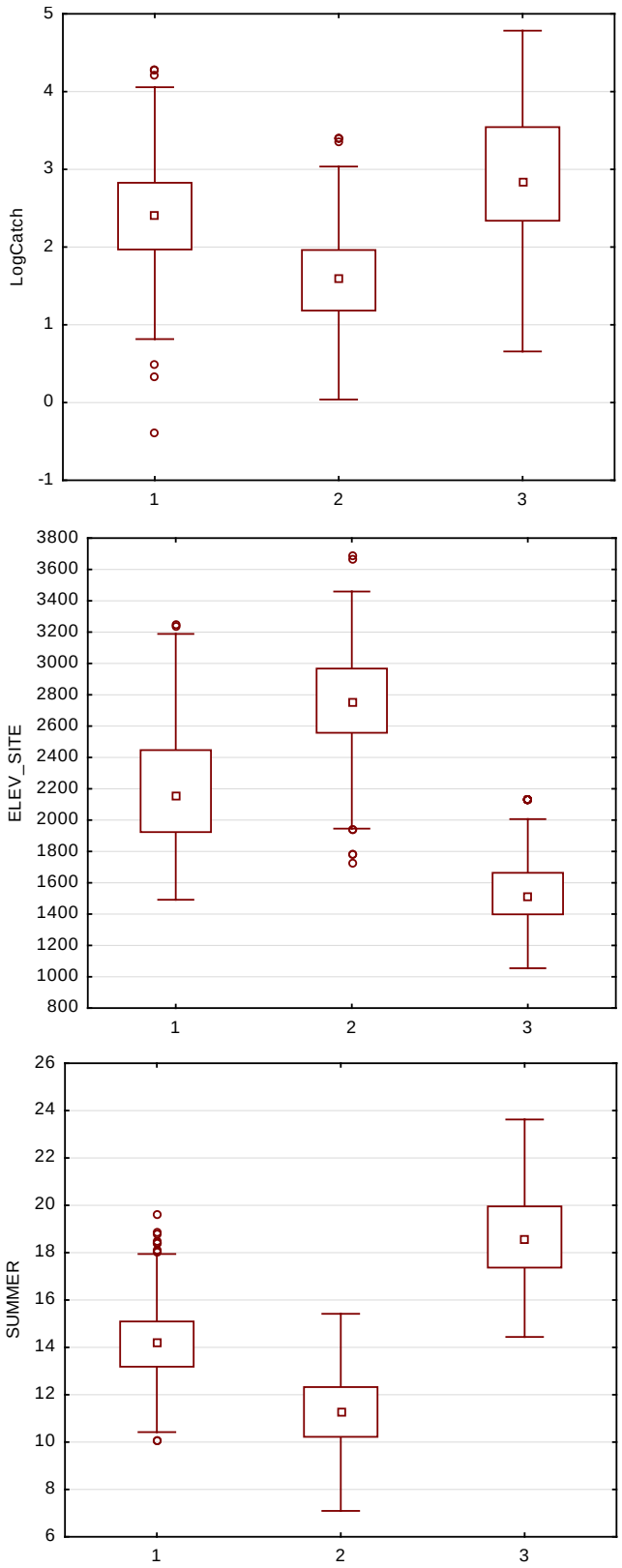


Figure 11. Distributions of environmental variables (catchment size, elevation, and summer air temperature) in the Biotypes.

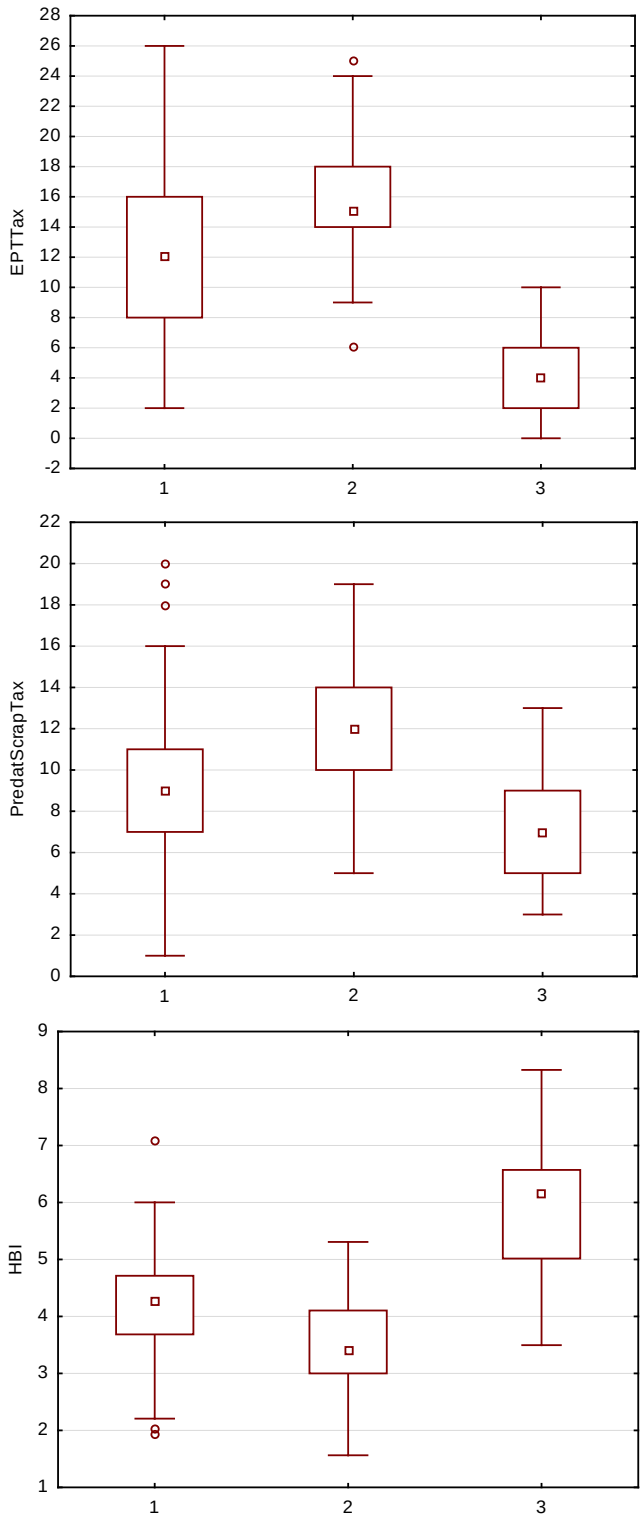


Figure 12. Distributions of macroinvertebrate metrics (EPT taxa, predator + scraper taxa, and Hilsenhoff's Biotic Index [HBI]) in the Biotypes.

Discriminant Analysis

A forward stepwise discriminant function analysis (DFA) was used to find the best predictors of the three most distinct clusters of the reference PA cluster analysis. When Biotypes were the only predictors, the classification accuracy was 87.5% among reference sites. When Biotypes were excluded, the classification accuracy improved by only one site. However, to achieve that improvement six variables were in the model and they were not all significant to the model ($p>0.05$). The variables entered the model in the following order (when Biotypes were not allowed): summer temperature, longitude, elevation, winter temperature, mean precipitation, and stream slope.

CART and Random Forest Analysis

A Random Forest analysis was used to identify the most important predictors of the reference PA clusters. Random Forest analysis runs multiple iterations of classification and regression tree (CART) analysis and identifies those variables that are important in the multiple iterations. CART is as effective at identifying categorical predictors as it is with continuous variables. Biotype and longitude were the most important predictors of the reference PA clusters (Table 7). The Random Forest analysis does not result in a predictive equation, but informs selection of variables that can be applied in a final CART analysis to generate the predictive results.

Table 7. Random Forest importance of environmental variables in predicting reference PA clusters.

Variable	Importance
Biotype	21.5
Longitude	11.2
Ecoregion	6.6
Catchment Size	4.9
Julian Day	3.3
Latitude	3.3
Elevation	2.8
BioRegion	2.6
Stream Order	1.5

A CART analysis was conducted with only the three most important variables: Biotype, longitude, and level 3 ecoregion. When the CART dendrogram was pruned to reduce error, only Biotype and longitude remained as predictors. The CART analysis accurately predicted clusters for the Biotypes 2 and 3 (mountains and plains). Biotype 1 (transitional) sites were split into 2 groups east and west of longitude -108.32 (Figure 13). Longitude -108.32 defines a strip of land in the west of the state that includes approximately one-eighth of the land area in Colorado (Figure 14). The east and west transitional sites are not distinct in ordination space, overlapping both Biotypes 1 and 2 (Figure 15).

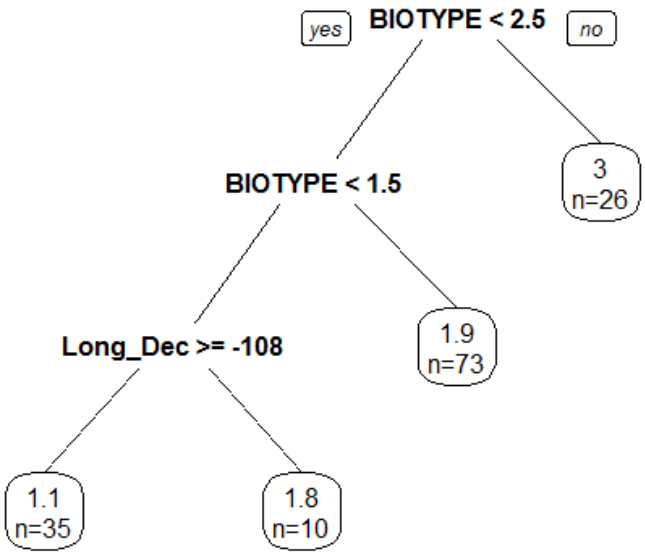


Figure 13. CART dendrogram showing distinctions for Biotypes 3 and 2 and a split of Biotype 1 sites by longitude.

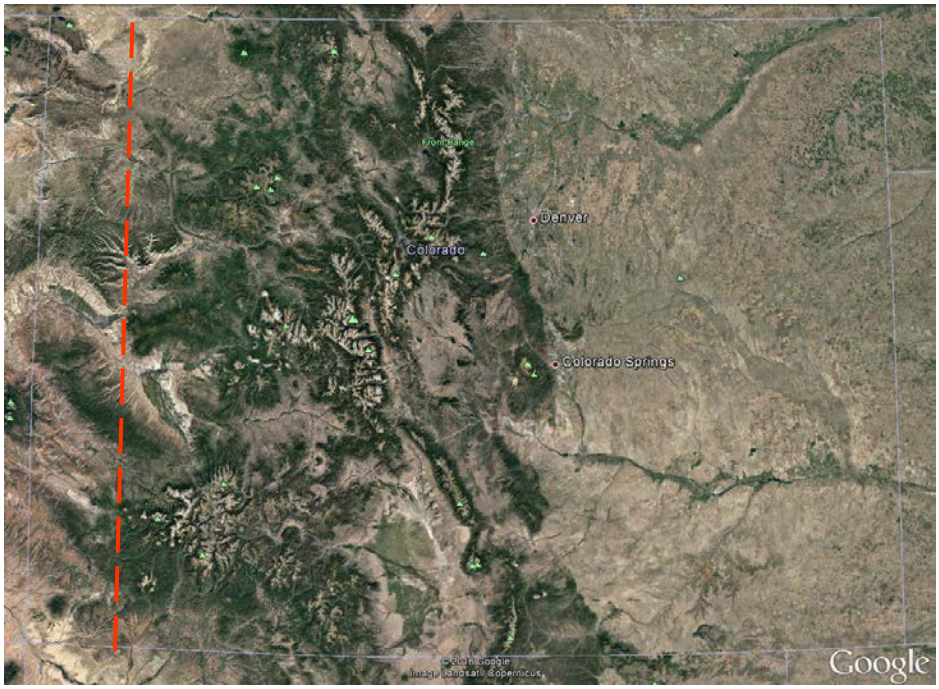


Figure 14. East/West divide for transitional sites (red dashed line) derived from CART analysis.

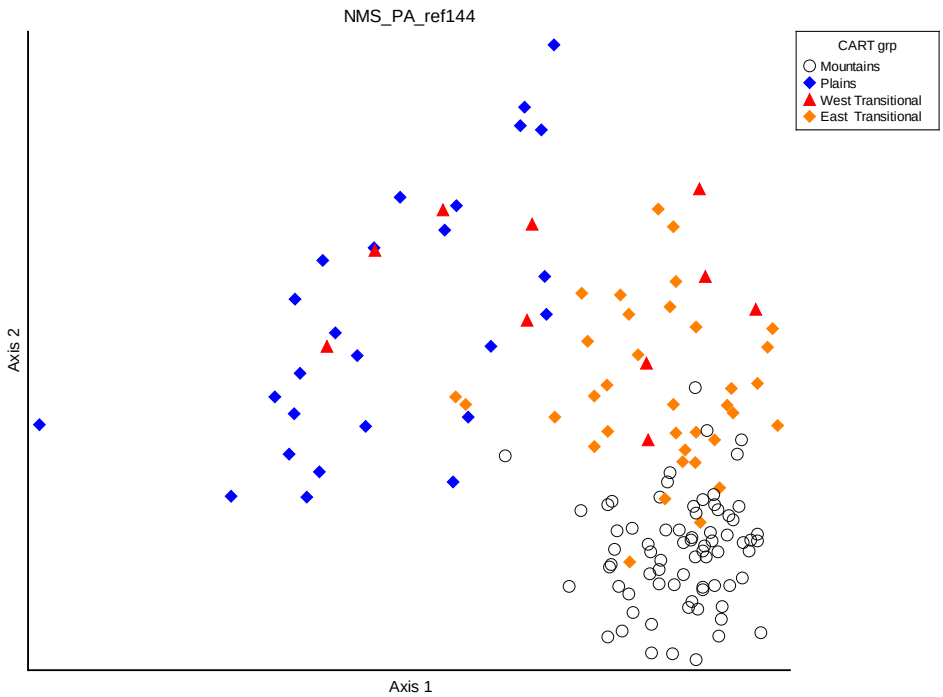


Figure 15. NMS ordination of reference presence/absence data, showing Biotypes 2 and 3 and western and eastern sites in Biotype 1 (transitional).

Distributions of environmental variables in all sites and reference metrics suggested that there were some differences in the western and eastern transitional sites. Western Transitional sites (code 10 in the figure) are distinct from Biotype 2 and are intermediate between Biotype 3 and eastern Transitional (code 11) in elevation and summer temperature (Figure 16). Western Transitional sites (code 10) have fewer EPT taxa than eastern Transitional sites (code 11) on average (Figure 17). Eastern and western transitional sites are fairly similar in the number of predator and scraper taxa and the HBI.

Although some differences were noted for the environmental variables and metrics in the eastern and western Biotype 1 sites, the site class was not split for index development. In the ordination diagram, it was unclear that the western sites had a distinct taxonomic structure. In addition, the sub-class had only 10 reference sites, which is minimal for the index development process. The differences in western sites were related to elevation and summer temperature. These environmental variables were explored in subsequent metric adjustment and scoring steps so that the remaining variability after site classification could be recognized for individual metrics.

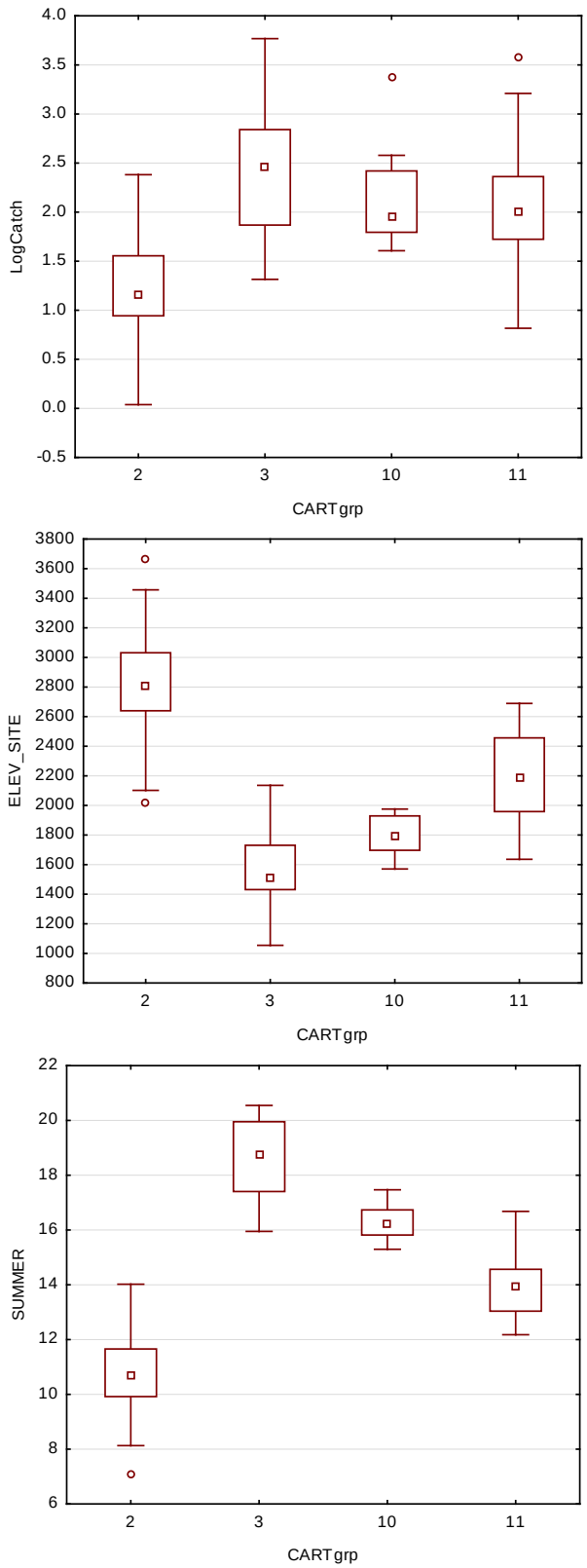


Figure 16. Distributions of environmental variables (catchment size, elevation, and summer air temperature) in the CART groups, which include Biotypes 2 and 3 and western Biotype (10) and eastern Biotype 1 (11).

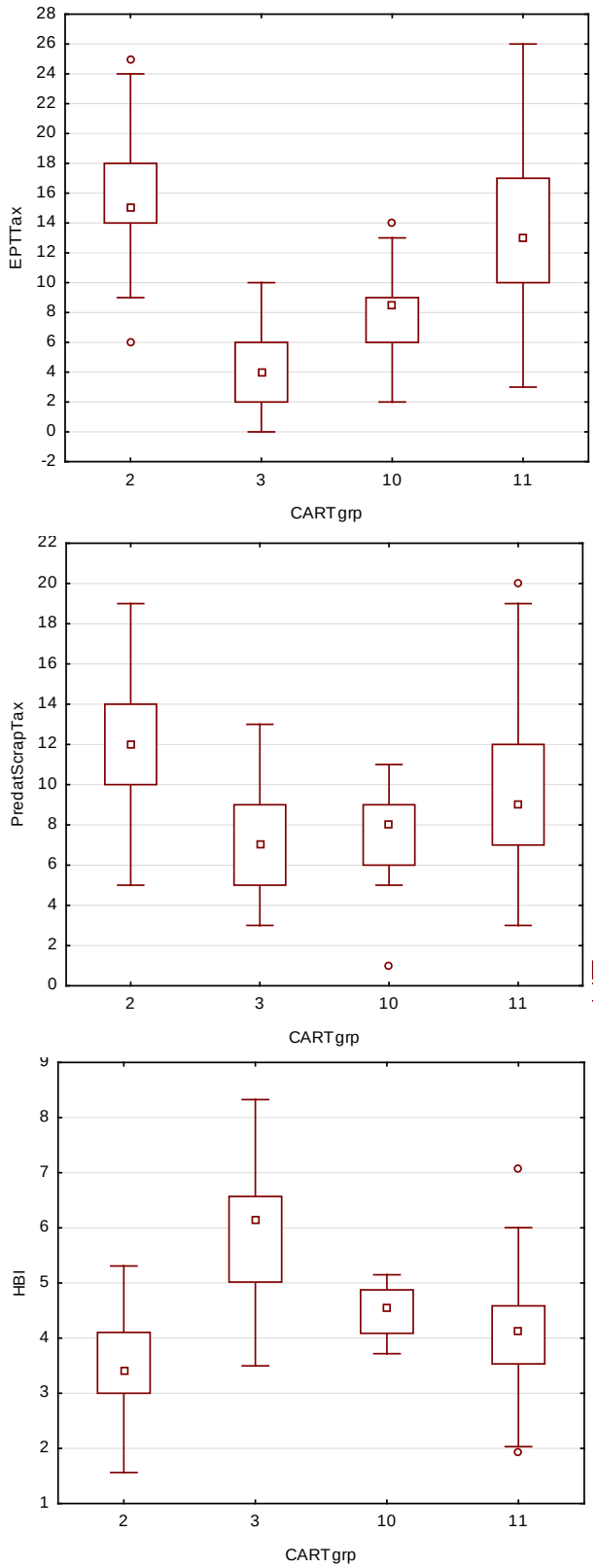


Figure 17. Distributions of macroinvertebrate metrics (EPT taxa, predator + scraper taxa, and Hilsenhoff's Biotic Index [HBI]) in the CART groups, which include Biotypes 2 and 3 and western Biotype (10) and eastern Biotype 1 (11).

Site Classification Conclusions

The existing Biotypes are strongly supported in the current analysis. Biotypes 2 and 3 are clearly distinct in all analyses. Transitional sites (Biotype 1) could be split by longitude into western and eastern regions, though this split was not strongly supported. Instead of splitting the class, metric expectations were scaled to the environmental variables to account for remaining variability in the existing site classes.

The three variables used in the classification rules (Table 6) include level 4 ecoregion, site elevation, and stream slope. These were decided upon because of corroborating results from ordination analysis, variable distribution evaluation, and DFA. Of the 144 sites used to derive the classification rules, 126 (87.5%) had Biotypes matching the biologically defined clusters. Class assignments for the re-calibration sites are listed in Appendix A.

The classes are conceptually described as:

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

The Biotypes are based on geographic location (ecoregion), slope, and elevation. Because stream slope is not well mapped across Colorado, a map of Biotypes was generated based on a Kriging technique that interpolates between individual sites with known Biotypes (Figure 18).

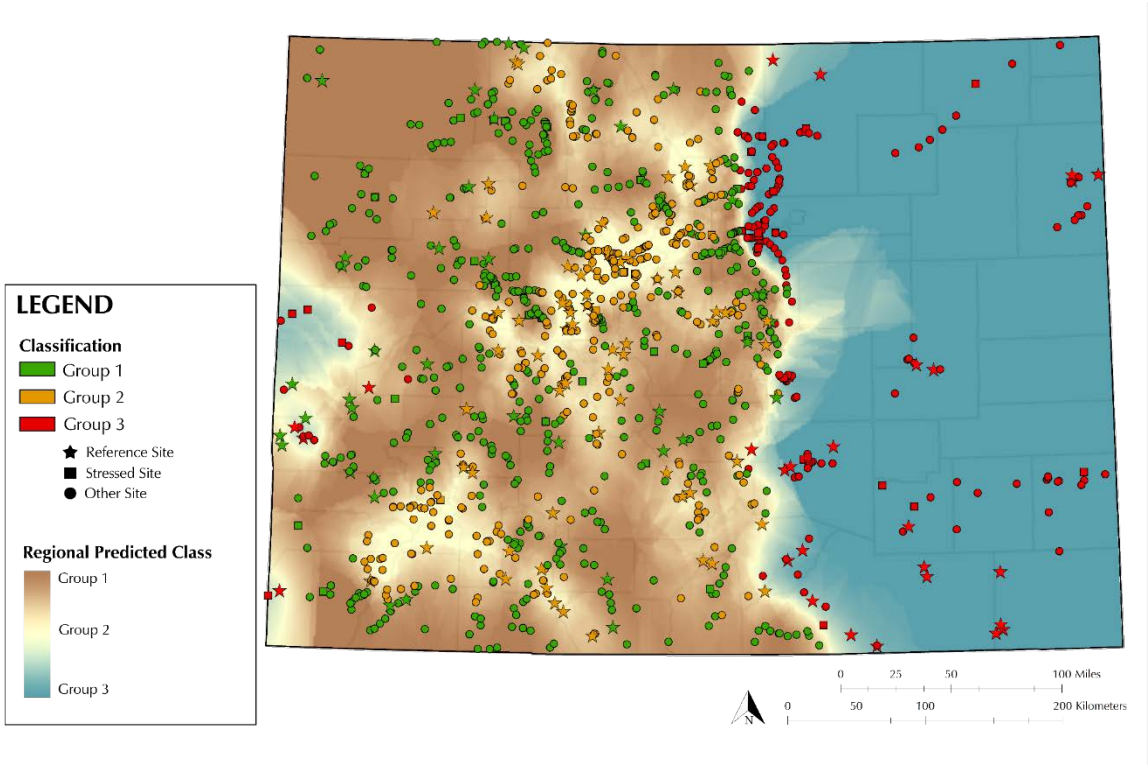


Figure 18. Biotypes (Groups) for the macroinvertebrate index in Colorado.

4.2 Metric Development

Indicator Taxa

Indicator taxa analysis revealed the taxa that were more common to either reference or stressed samples. Of the 289 taxa (at both family and genus level) that were tested in the analysis, 43 were identified as sensitive decreasers with increasing stress or as tolerant increasers. The taxa that were indicators in Biotypes 1 and 2 were similar. Indicator taxa in Biotype 3 were often uniquely indicative in that site class. Therefore, indicators were selected for two site types; mountain/transitional and plains.

In the mountain/transitional sites (Biotypes 1 and 2), a taxon was considered an indicator if the analysis showed that the difference between reference and stressed sites was significant at $p<0.05$ in one Biotype and $p< 0.10$ in the other. This resulted in 17 indicator taxa for the mountain/transitional sites (Table 8). In the plains, all 18 indicator taxa were significant at $p<0.05$ in Biotype 3 (Table 8). Families were selected over genera as indicators when the family was a significant indicator. There were 19 indicator families identified by CDPHE for the mid-elevation (Biotype 1) sites, including 11 sensitive families that decreased with increasing stress and 8 tolerant increaser families (Table 9).

Metrics were calculated to count taxa and percent of individuals for both sensitive and tolerant indicators. Metrics were tested in their respective Biotypes. Both sensitivity scales were used in Biotype 1, which had sensitive taxa derived from the indicator analysis for transitional sites and suggested by the technical advisory committee for mid-elevation sites.

Table 8. Indicator taxa derived from an indicator species analysis on reference and stressed sites.

Mountain/Transitional		**Plains**	
Sensitive Decreasers	**Tolerant Increasers**	**Sensitive Decreasers**	**Tolerant Increasers**
Ameletidae	*Erpobdellidae*	*Elmidae*	*Acentrella*
Chloroperlidae	*Naididae*	*Caenidae*	*Cricotopus/Orthocladius*
Dicranota	*Physidae*	*Ceratopogonidae*	*Cryptochironomus*
Elmidae	*Synorthocladius*	*Corynoneura*	*Erpobdellidae*
Ephemerellidae	*Tubificidae*	*Dytiscidae*	*Naididae*
Heptageniidae		*Leptophlebiidae*	*Saetheria*
Micrasema		*Micropsectra*	*Tubificidae*
Nemouridae		*Parametriocnemus*	
Perlidae		*Pentaneura*	
Perlodidae		*Tanytarsus*	
Rhyacophila		*Tvetenia*	
Stempellinella			

Table 9. Mid-elevation sensitive and tolerant families.

Sensitive Decreasers	Tolerant Increasers
Ameletidae	*Asellidae*
Blephariceridae	*Coenagrionidae*
Brachycentridae	*Corixidae*
Capniidae	*Enchytraeidae*
Glossosomatidae	*Lebertiidae*
Lepidostomatidae	*Naididae*
Nemouridae	*Physidae*
Perlodidae	*Tubificidae*
Philopotamidae	
Pteronarcyidae	
Rhyacophilidae	

Metrics Adjustments to Environmental Variables

Metric adjustments were applied to 13 metrics in Biotype 1, 13 in Biotype 2, and 9 in Biotype 3 (Appendix B). In Biotype 1 reference sites, several metrics were correlated with summer temperature, winter temperature, modeled temperature, elevation, and Julian day. The most consistent adjustment variable was summer temperature, which usually had stronger correlation coefficients than other variables, especially when considering correlations in all sites as well as in reference sites. All 13 adjusted metrics in Biotype 1 were made with summer temperature. The strongest relationships were with Trichoptera taxa, EPT taxa, clinger taxa, and intolerant taxa. Each of these metrics appeared to have a plateau at temperatures below 13.5 degrees C.

Metric adjustments were made when correlation coefficients were high in both reference sites and all sites and when the 95th quantile regression line appeared to be a good approximation of the limit evident in the bi-plots. Elevation measurements from maps might be easier to make than the summer temperature, which requires a GIS analysis. However, the effects of summer temperature are probably more direct than effects of elevation, and the two measures are correlated to each other. The NorWeST modeled temperature did not have correlation coefficients as strong as those with summer temperature.

In Biotype 2, Julian Day was the dominant adjustment factor, though three of the 13 adjusted metrics used summer temperature and one used elevation as adjustment factors. It appears the EPT and intolerant taxa increase over the sampling season while Diptera and collectors decrease. Coleoptera taxa and relative abundance increase with summer temperature, while percent Plecoptera decreases. Percent non-insects increase with elevation.

In Biotype 3, Julian Day was the most important adjustment factor, followed by summer temperature. Similar to Biotype 2, EPT taxa increase over the sampling season and Diptera taxa and relative abundance decreases. Sprawler taxa also decrease through the

season. Tolerant taxa increase with higher summer temperatures. Several relationships showed high correlation coefficients but were not adjusted due to inconsistencies with the quantile regressions. In some cases, the relationship appeared to be driven by outliers. Several metrics had high correlation coefficients with longitude. However, longitude is not a continuous variable in Biotype 3 because of the large gap in the mountainous middle of the state. Therefore, adjustment by regression is not appropriate for longitude.

Metric Performance

A total of 104 metrics were calculated using EDAS queries. These included some metrics that were experimental modifications of standard metrics because of some indications that became apparent while evaluating metric performance. For example, both percent Cricotopus + Chironomus of Chironomidae and percent Orthocladiinae of Chironomidae performed well on their own. A new metric combining the three midge taxa was calculated and tested to combine the signals into a more comprehensive metric of tolerant midges. The DE, Z-score, and CV of each metric was calculated within the site classes (Appendix C).

In Biotype 1 (transitional sites), metrics from each metric category had DEs greater than 50% and metrics in the richness and tolerance categories had DEs ranging to above 85%. The metrics with the highest DE's were Plecoptera taxa, percent intolerant, and percent mountain and transitional decreaser taxa (all adjusted for summer temperature).

Metrics from every category, except composition in Biotype 2 (Mountain sites), had DEs greater than 90%, including six metrics in the richness, tolerance, and habit categories with DEs of 100%. All metrics with high DEs decreased with increasing stress. Metric adjustments in Biotype 2 were with Julian day and summer temperature.

In Biotype 3 (Plains), richness, composition, and specialized tolerance metrics performed best, with tolerance values ranging up to 100%. Best performing metrics in each category included insect taxa, percent Coleoptera, sprawler taxa, predator taxa, and specialized tolerance metrics. Metrics were adjusted to Julian day when adjustments were needed.

4.3 Index Performance

Up to 275,000 index alternatives per Biotype were calculated and tested using reference and stressed sites in the iterative all-subsets routine. Metrics included in the all-subsets routine included 20-21 candidate metrics for each biotype. More than 21 metrics can overwhelm computational resources. The selected metrics included a variety of the best performers in each metric category. The index alternatives included all possible combinations of non-redundant metrics with a maximum of eight metrics per index.

The index alternatives that performed best were reconsidered and the following indices were recommended. The alternatives were evaluated for DE, *Z*-score, number of metric categories, and conceptual redundancy. With multiple index alternatives performing similarly, a final index was selected based on familiar metrics, individual metric DE, and a variety of expected metric responses. In each Biotype, comparisons were made to an alternative including the metrics used in the 2009 index version (Jessup 2009).

Biotype 1

In Biotype 1, 159,867 index formulations were tested using 20 candidate metrics in the all-subsets routine. Of these, 1,713 alternatives had DEs greater than 94%, included all 5 metric categories, and contained no conceptually redundant metrics. The alternative selected as the MMI for Biotype 1 (#8_61382**)** included eight metrics, had a DE of 94.6%, and *Z*-score of 2.34 (Table 10**,** Figure 19). Three of the metrics were adjusted to summer temperatures (Figure 20). All of the adjusted metrics had a plateau of values when temperatures were below 13.5 ^{o}C. The mid-elevation indicator taxa that increase in percentage of individuals with stress (pi_IncrMidElev) performed well and was included in the final index. The other mid-elevation indicator metrics did not perform as well.

Table 10. Metrics included in the Biotype 1 MMI with scoring formulae.

Metric	Scoring Formula
aEPTTax	SummerTemp≤13.5: 100*Metric / 24 SummerTemp>13.5: 100*Metric / (-3.73 * SummerTemp + 74.4)
NonInPct	100*(48.3-Metric) / 48.3
pEPTnoB	100*Metric / 72.9
ColeoPct	100*Metric / 38.4
apt_Intol	SummerTemp≤13.5: 100*Metric / 61 SummerTemp>13.5: 100*Metric / (-5.78 * SummerTemp + 139)
pi_IncrMidElev	100*(23.9-Metric) / 23.9
aClngrTax	SummerTemp≤13.5: 100*Metric / 20.8 SummerTemp>13.5: 100*Metric / (-2.23 * SummerTemp + 50.9)
PredShrTaxFAC	100*Metric / 14

Of the validation samples, 60% of reference index values were above the calibration 25th percentile and 80% of stressed were below it. While these statistics indicate index performance that is not as distinct as the calibration data, the box plots show that the distributions of reference values are similar though more variable in validation sites compared to calibration (Figure 19). The stressed site distribution is very similar for calibration and validation samples.

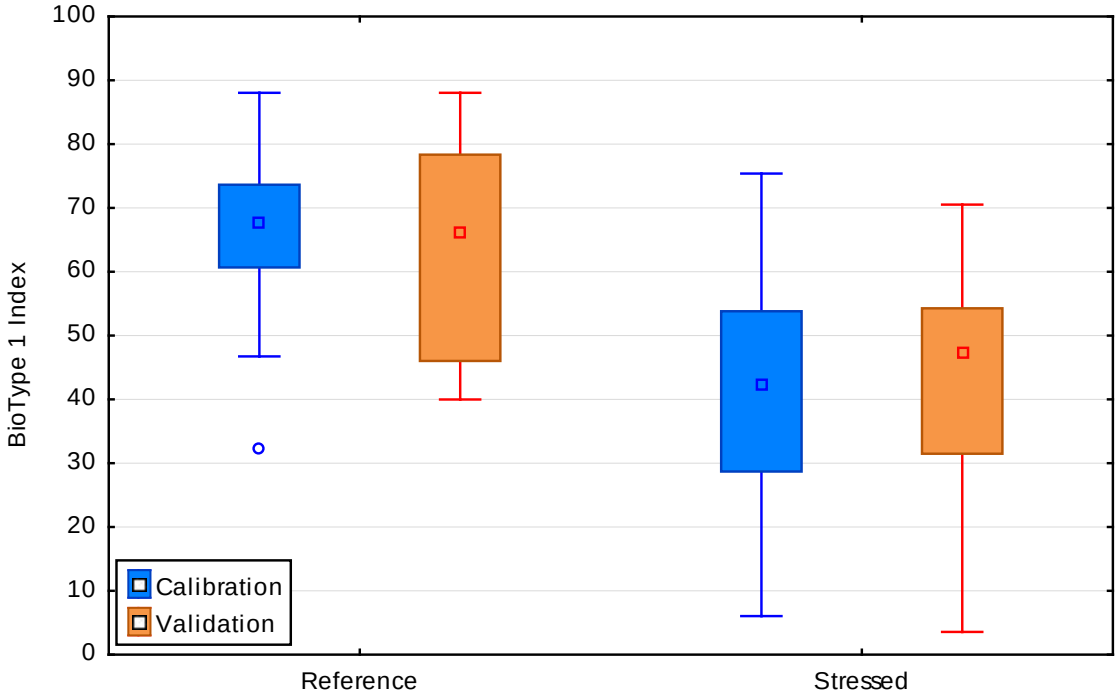


Figure 19. Distribution of index scores (8_61382) in reference and stressed sites in Biotype 1, showing calibration and validation data sets.

The eight best alternatives had 7-8 metrics, DEs of 94-97%, and *Z*-scores of 2.0 – 2.5 (Table 11). The alternative index with the highest *Z*-score was not selected because it had only seven metrics and the single richness metric was an increaser, which indicates taxa tolerance, but de-emphasizes sensitivity. The best manually composed index had 7 metrics, a DE of 97%, and a *Z*-score of 1.8. When the metrics from the 2009 version of the MMI for Biotype 1 were tested with the new dataset (new reference and stressed sites, new samples, etc.), the resulting DE (89.2) was lower than any of the recently derived alternatives.

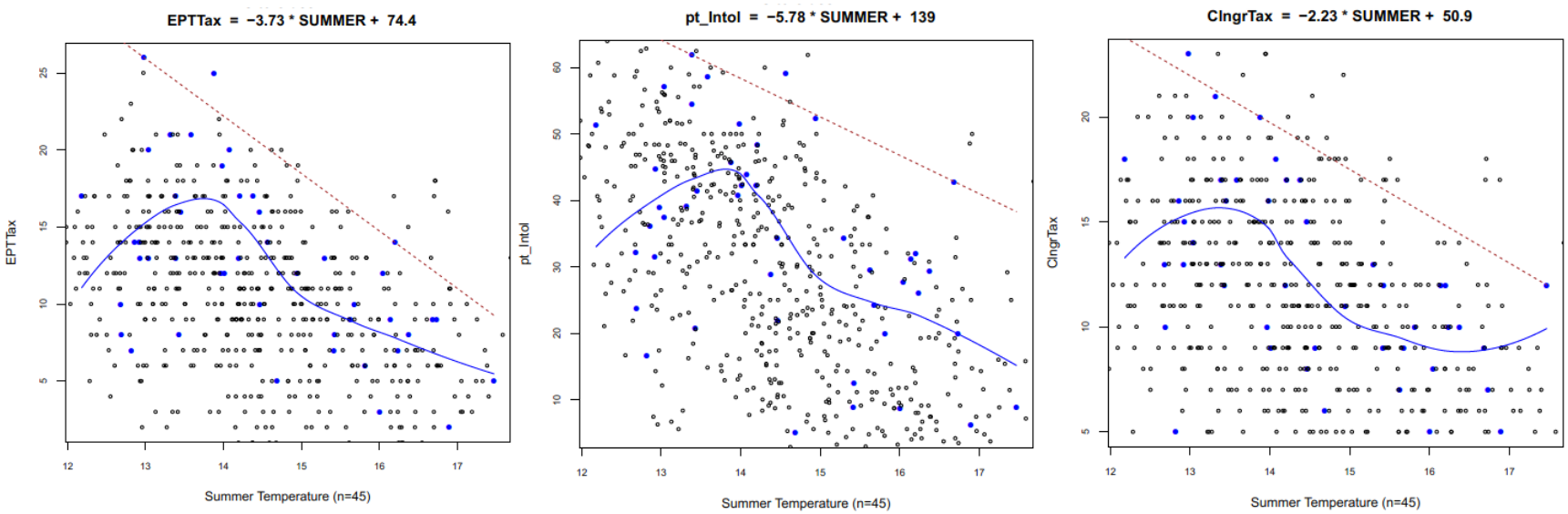


Figure 20. Bi-plots of index metrics that were adjusted to summer temperature in Biotype 1, including EPT taxa, percent intolerant taxa, and clinger taxa. The plots show the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots)

Table 11. The best index alternatives considered for Biotype 1, including eight that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_61382.

Metric \ ModelID	7_44411	8_67408	**8_61382**	8_67410	8_67939	7_31708	8_107957	7_38421	Best Manual	Existing Metrics
aInsectTax*		1		1	1		1		1	
aEPTTax*			**1**			1			1	
EPTax										1
pt_noninsect	1	1		1				1		1
NonInPct			**1**		1	1	1			
pEPTnoB	1	1	**1**	1	1	1	1	1		
PlecoPct									1	
ColeoPct	1	1	**1**	1	1	1	1	1	1	
ChiroPct										1
aIntolTax*								1		
apt_Intol*	1		**1**		1		1			
aBeckBI*										
aPercIntol*		1		1			1	1	1	
api_DecrMtnTrn*					1	1				1
pi_IncrMidElev	1	1	**1**	1						
aClngrTax*	1	1	**1**	1	1	1		1	1	1
ClngrPct							1			
PredShrTaxFAC	1	1	**1**		1	1	1			1
PredTaxFAC				1				1	1	
Ref.q25	57.9	55.2	**60.7**	54.4	53.7	52.1	54.9	47.8	41.9	57.8
Str.DE	94.6	97.3	**94.6**	97.3	94.6	94.6	97.3	97.3	97.3	89.2
Z-score	2.53	2.34	**2.34**	2.27	2.20	2.14	2.06	2.01	1.80	2.15

* = adjusted to summer temperature

Precision

In a precision analysis, nine replicate sets, each containing two samples collected at the same site on the same day, were used to estimate natural variability and sampling error of the index and component metrics. The 90% confidence interval (CI90) for the index was ±7.8 index points and the CV was 8.6% (Table 12). The pEPTnoB metric had a CV of 26.8% for both scores and the metric values, though other metrics were more variable by some measures.

Table 12. Precision statistics for the selected Biotype 1 index and its component metrics, based on both metric scores and values.

	Scores				Metrics			
	RMSE	mean	CV	CI90	RMSE	mean	CV	CI90
Index_Biotype1	4.7	55.0	8.6	7.8				
aEPTTax	7.2	54.1	13.3	11.9				
pEPTnoB	10.6	39.7	26.8	17.5	7.7	28.9	26.8	12.7
NonInPct	9.4	74.6	12.6	15.4	4.6	12.5	36.5	7.5
ColeoPct	6.3	30.2	20.8	10.3	2.4	11.6	20.8	4.0
aClngrTax	5.9	54.0	10.8	9.6				
apt_Intol	13.5	59.9	22.5	22.2				
PredShrTaxFAC	12.4	57.9	21.4	20.4	1.7	8.1	21.4	2.8
pi_IncrMidElev	3.0	69.9	4.3	4.9	4.6	10.2	45.0	7.6

Index response to stress

The index scores in Biotype 1 are highest in reference sites with low percentages of development in the catchments and when the reference scores do not indicate stress (Figure 21). There are several stressed index scores that are low despite low percentages of development. These are probably affected by other stressors. The sum of reference criterion scores is the number of stressors that were identified at a site during the reference determination process conducted by CDPHE. Two stressed site with low index values despite low sum of scores values, including Yellow Creek (StationID 11732A) and Noname Creek (CO050M). Sites with high MMI values and high sum of score values might have some unmeasured buffering capacity that reduces the effects of the measured stressors. Two sites have high index scores and sum of score values of 6, including the Slate River (StationID 151, stressed) and the North Platte River (NPlatte, reference).

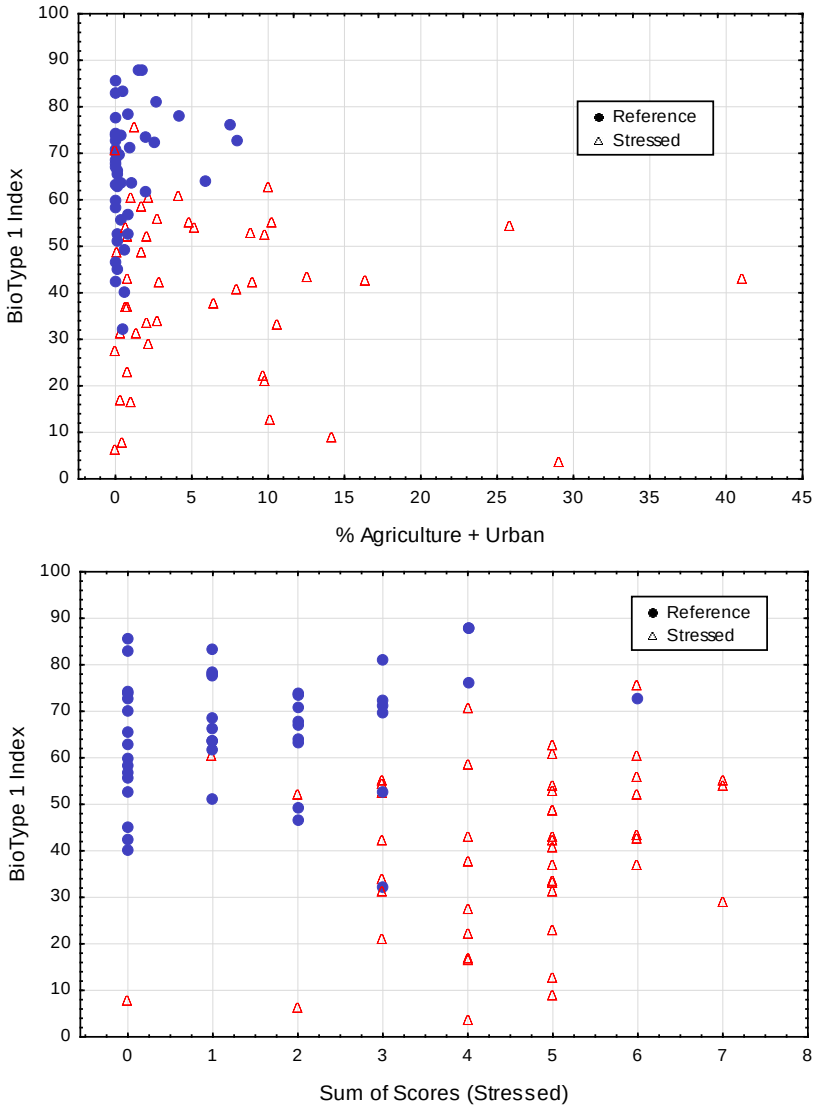


Figure 21. Biotype 1 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions.

Biotype 2

In Biotype 2, 158,591 index alternatives were tested using 20 metrics in the all-subsets routine. Of those, 5,383 alternatives had DEs of 100% and 7 or 8 metrics from all 5 metric categories after screening out metrics with conceptual redundancy. The index alternative selected as the MMI for Biotype 2 (#8_78149) included eight metrics and had a *Z*-score of 4.01 (Table 13**,** Figure 22). The EPT taxa are expected to be most diverse at the end of the sampling season (after day 250, September 7) (Figure 23).

Forty percent (40%) of the 15 reference validation sites were greater than the calibration reference 25th percentile, which is lower than the 75% of reference calibration sites that are above the 25th percentile (Figure 22). However, the range of validation reference sites is similar to the range of calibration reference sites (40.6-93.3 index points). In Biotype 2 there were no stressed sites reserved for validation.

The eight best alternatives had 7-8 metrics, DEs of 100%, and Z-scores of 3.5 – 4.5 (Table 14). The best manually composed index had 7 metrics, a DE of 100%, and a Z-score of 3.6. While some index alternatives had higher Z-scores than the selected index, the selected index had a metric combination that was well balanced among metric types. For example, the index alternative with the highest Z-score had two metrics from the indicator species analysis, which might overweight this metric type.

When the metrics from the existing MMI for Biotype 2 were tested with the new dataset (new reference and stressed sites, new samples, etc.), the resulting DE (100%) was as good as any of the derived alternatives. However, the Z-score (3.03) was somewhat lower than the selected alternative and the increased number of metrics in the new index (8 vs. 5) will decrease the likelihood that a single metric will overly influence the index results.

Table 13. Metrics included in the Biotype 2 MMI with scoring formulae.

Metric	Scoring Formulae
TotalTax	100* Metric / 42
aEPTTax	JulianDay≥250: 100*Metric / 24.5 JulianDay<250: 100*Metric / 0.0769 * JulianDay + 5.23)
pEPTnoB	100* Metric / 79.3
ClngrTax	100* Metric / 20
IntolTax	100* Metric / 21
pi_DecrMtnTrn	100* Metric / 72.3
PredTaxFAC	100* Metric / 13
ScrapPctFAC	100* Metric / 44.2

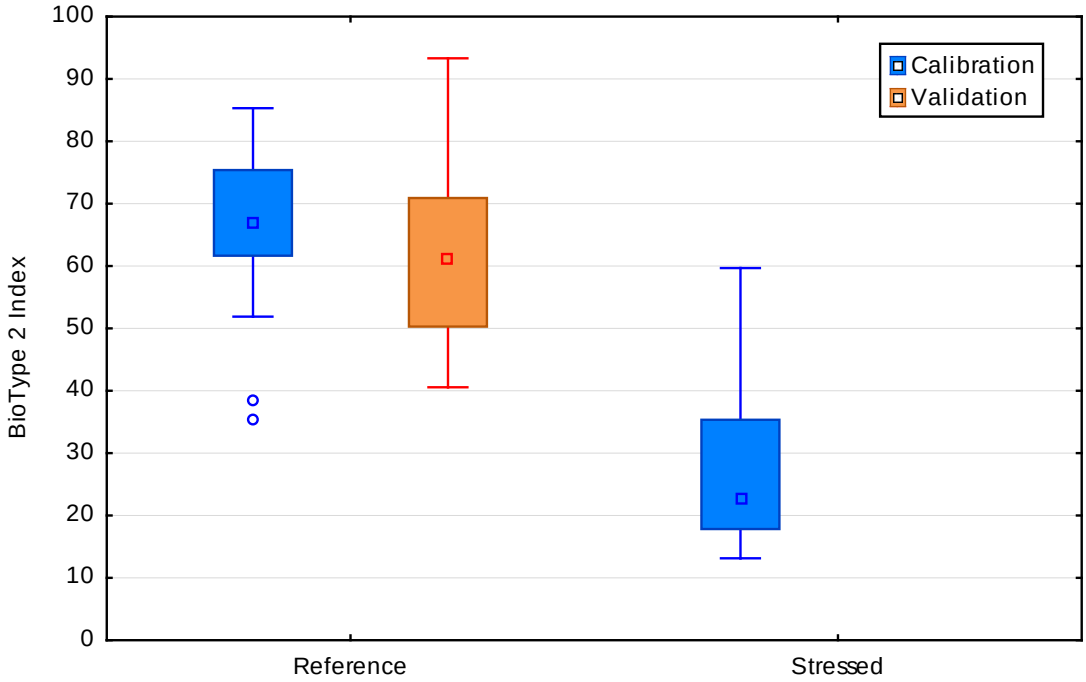


Figure 22. Distribution of index scores (8_78149) in reference and stressed sites in Biotype 2, showing calibration and validation data sets.

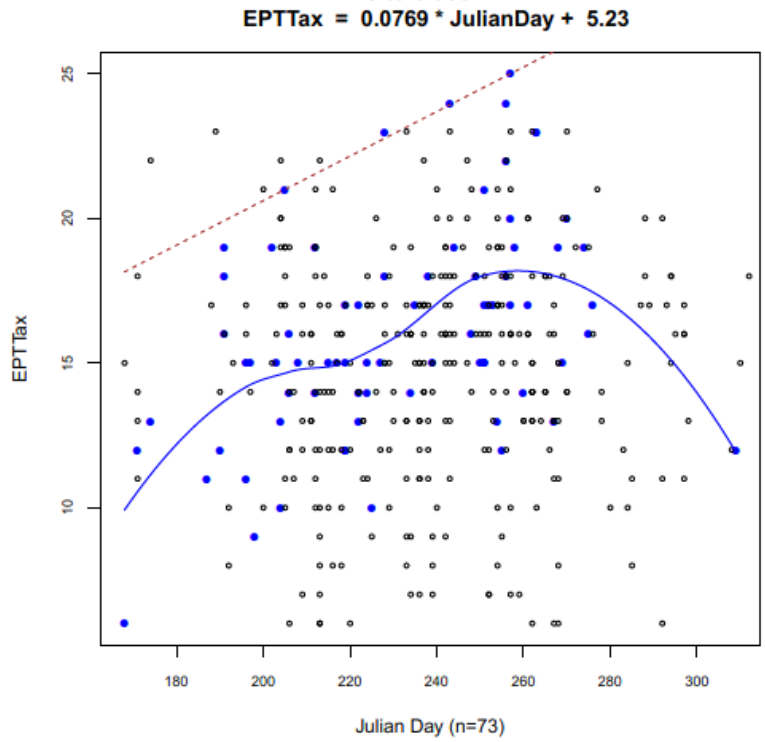


Figure 23. Bi-plot of index metric (EPT taxa) that was adjusted to Julian Day in Biotype 2, showing the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots).

Table 14. The best index alternatives considered for Biotype 2, including 8 that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_78149.

ModelID	8_97851	8_27551	7_6268	**8_78149**	7_12670	7_5140	7_6141	8_14927	Best Manual	Existing Metrics
TotalTax	1	1	1	**1**	1	1	1	1	1	1
aEPTTax*			1	**1**		1	1	1		
apt_EPT	1	1			1				1	
pEPTnoB	1	1	1	**1**		1	1	1		
EphemPct										1
PlecoPct									1	
aColeoPct*		1	1		1	1	1	1	1	
ClngrPct						1				
ClngrTax	1	1	1	**1**	1		1	1	1	1
IntolTax				**1**		1	1	1		
PercIntol					1					
BecksBI										1
nt_DecrMtnTrn	1	1	1						1	
pi_DecrMtnTrn	1			**1**				1		
PredScrapTaxFAC					1	1	1	1		
PredShrTaxFAC									1	1
PredTaxFAC	1	1		**1**						
ScrapPctFAC		1	1	**1**	1					
ScrapTaxFAC	1									
Ref.q25	66.8	58.2	53.4	**61.7**	56.7	58.0	58.2	58.5	57.4	64.3
Str.DE	100	100	100	**100**	100	100	100	100	100	100
Z-score	4.48	4.21	4.11	**4.01**	4.12	3.55	3.51	3.74	3.59	3.03

* = adjusted

Precision

In a precision analysis, six replicate sets, each containing 2 samples collected at the same site on the same day, were used to estimate natural variability and sampling error of the index and component metrics for Biotype 2. The 90% confidence interval (CI90) for the index was 10.4 index points and the CV was 16.0% (Table 15). The most variable metric was ScrapPctFAC, with a CV of 50.7% for both metric values and scores.

Table 15. Precision statistics for the selected Biotype 2 index and its component metrics, based on both metric scores and values.

	Scores				Metrics			
Index/Metric	RMSE	mean	CV	CI90	RMSE	mean	CV	CI90
Index_Biotype2	6.3	39.3	16.0	10.4				
TotalTax	5.5	60.3	9.1	9.0	2.3	25.3	9.1	3.8
aEPTTax	8.7	41.9	20.8	14.3				
pEPTnoB	7.3	22.1	32.9	11.9	5.8	17.5	32.9	9.5
ClngrTax	6.9	47.1	14.7	11.4	1.4	9.4	14.7	2.3
IntolTax	7.7	45.6	16.8	12.6	2.1	9.8	21.1	3.4
pi_DecrMtnTrn	8.7	28.9	30.2	14.4	6.3	20.9	30.2	10.4
PredTaxFAC	11.3	47.4	23.9	18.6	1.5	6.2	23.9	2.4
ScrapPctFAC	10.7	21.1	50.7	17.6	4.7	9.3	50.7	7.8

Index response to stress

The index scores in Biotype 2 are highest in reference sites with low percentages of development in the catchments and when the reference scores do not indicate stress (Figure 24). There are several stressed index scores that are low despite low percentages of development. The percentage of development in Biotype 2 sites was relatively low in general, with the highest percentages reaching only 3.5% in stressed sites. Other data in Biotype 2 reach as high as 94% developed land, but 90% of all samples have less than 9% developed land. In contrast, the range of % developed land in stressed sites of Biotypes 1 and 3 reach 41% and 72% respectively. Apparently, Biotype 2 index values respond when there are low levels of development in the stream catchment, though the stressors might be unrelated or in addition to land uses.

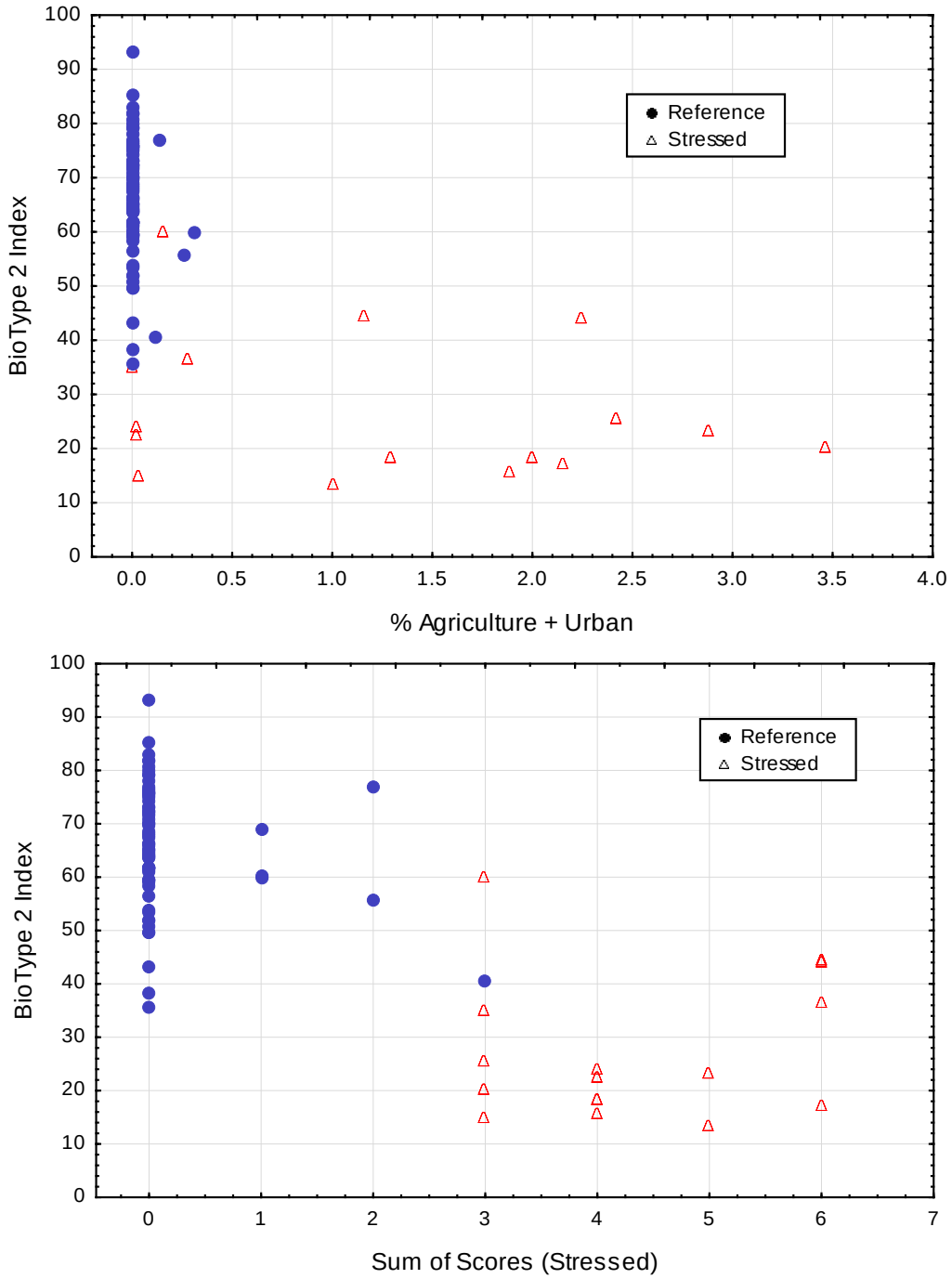


Figure 24. Biotype 2 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions.

Biotype 3

In Biotype 3, 273,806 index alternatives were calculated and tested. Of those, 30,273 alternatives had DEs of 100%, 5 metric categories, and non-redundant metrics. Therefore, other index evaluation factors were important in selecting the final MMI, such as the *Z*-score, individual metric DE, and selection of metrics with familiar response mechanisms. The alternative selected as the MMI for Biotype 3 (Table 16**,** Figure 25) included eight metrics, representing each of the five metric categories. Sprawler taxa decreased over the sampling season and adjustments were made when scoring this metric (Figure 26).

In the validation samples that were reserved and not used in index calibration, 62.5% of eight validation reference sites had index scores greater than the 25th percentile of calibration reference values. 100% of seven validation stressed sites had index scores below the 25th percentile of calibration reference values. The agreement between calibration and validation statistics suggest that the index is robust at distinguishing disturbance conditions.

Table 16. Metrics included in the Biotype 3 MMI with scoring formulae.

Metric	Scoring Formulae
TotalTax	100* Metric / 33.8
pt_noninsect	100*(45-Metric)/40
pEPTnoB	100* Metric / 54.6
aSprwlTax	100* Metric / (-0.02*JulianDay+13.9)
IntolTax	100* Metric / 4.75
pi_IncrPlains	100*(79.4-Metric)/79.4)
PredTaxFAC	100* Metric / 10
ScrapPctFAC	100* Metric / 24.1

The eight best alternative index compositions all had 8 metrics, DEs of 100%, and Z-scores of 2.6 – 5.5 (Table 17). The best manually composed index had 7 metrics, a DE of 100%, and a Z-score of 3.5. When the metrics from the existing MMI for Biotype 3 were tested with the new dataset (new reference and stressed sites, new samples, and new metric formulations), the resulting DE (100) was the same as the recently derived alternatives.

The selected index (#8_70028) includes two functional feeding group metrics, PredTaxFAC and ScrapPctFAC. These metrics decrease with increasing stress and suggest that better biological conditions include a complete trophic structure with both predators (higher consumers) and specialized primary consumers (scraping algae from hard substrates). The selected index does not include the CllctPctFAC metric, which depends on an increasing response in individuals that have a collector-gatherer feeding mechanism. These generalists take advantage of detritus as a food resource and indicate that more specialized feeding mechanisms and food resources are less common. While an alternative with this metric (#8_6699) has a high Z-score, the metric itself has a lower DE than the two other functional feeding group metrics.

The selected index has two metrics from the tolerance category, IntolTax and pi_IncrPlains. This metric combination was preferred over combinations with two

indicator taxa metrics, such as #8_6699 and others that included both nt_DecrPlains and pi_IncrPlains. The rationale for the preference was that a diversity in tolerance scales would lead to detection of more general response to stress.

Index alternative 8_67143, was also considered to be an excellent performer. It was not selected because it included the PredShrTaxFAC metric. While this metric was a good performer, the shredder taxa metric was not a good performer on its own and therefore the metric with only predators was preferred.

Model 8_24000 has the highest Z-score, but it also has several increaser metrics, which can increase the Z-score simply because the standard deviation of reference metric values is relatively small. It also has two non-insect metrics, which might overweight that aspect of the sample composition.

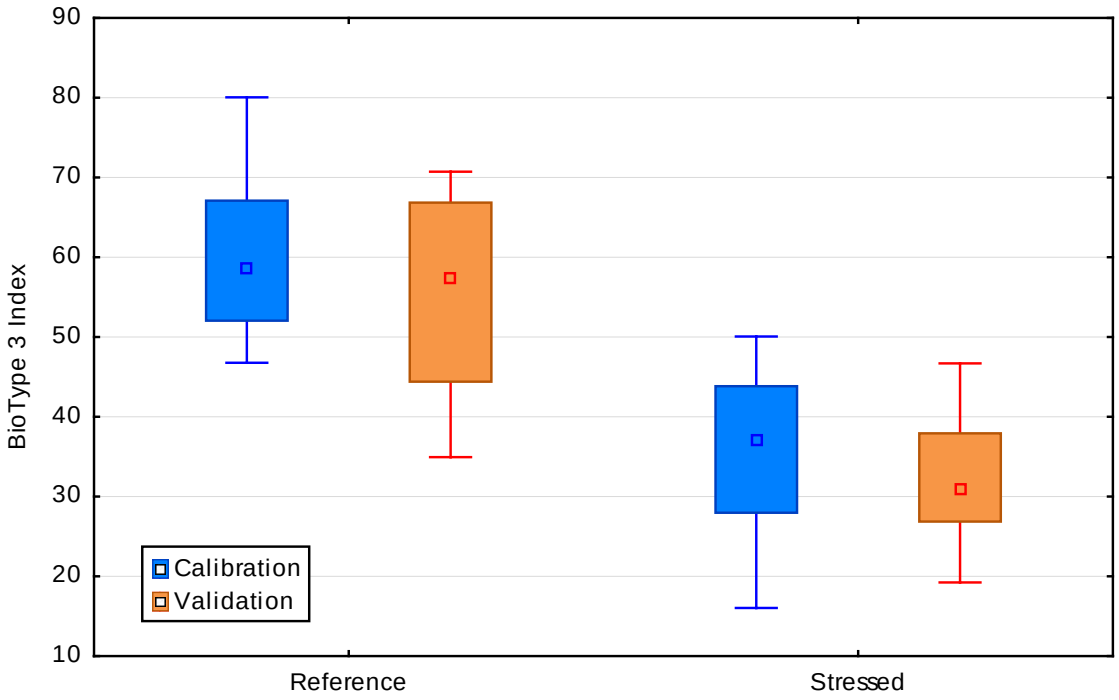


Figure 25. Distribution of index scores (8_70028) in reference and stressed sites in Biotype 3, showing calibration and validation data sets.

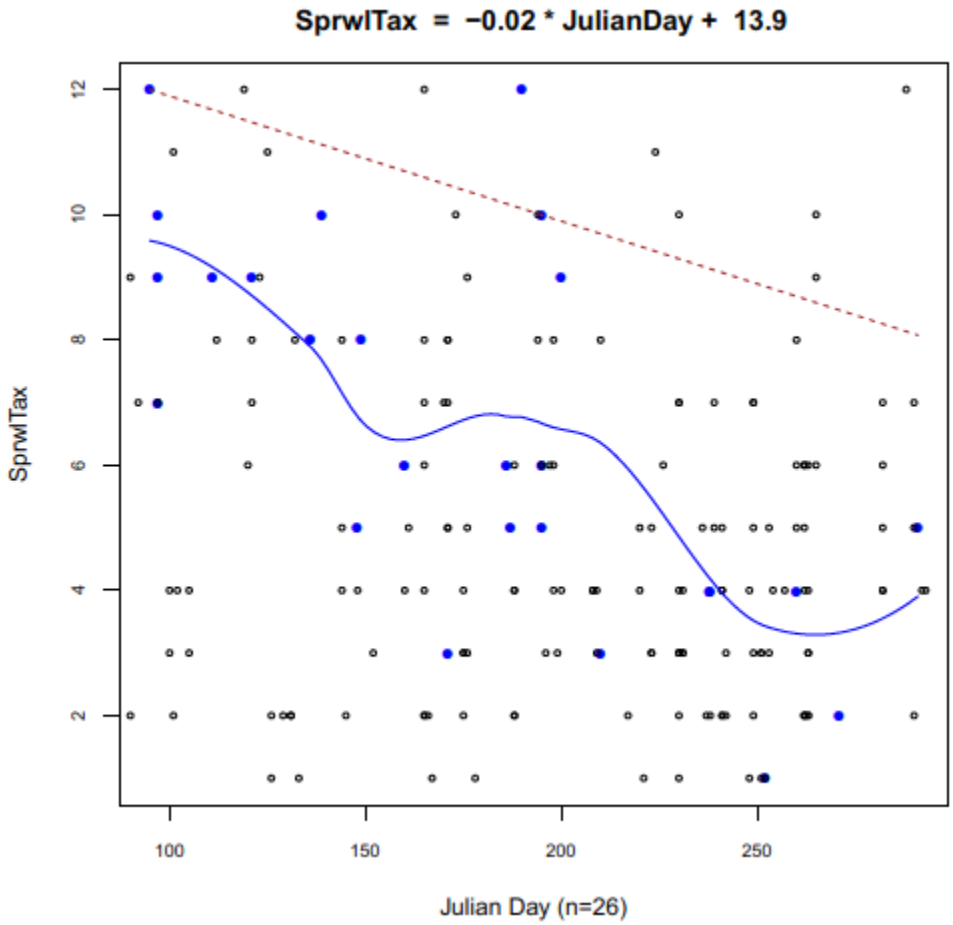


Figure 26. Sprawler taxa in relation to Julian Day in Biotype 3, showing the LOWESS regression curve (blue) and the 95th quantile regression line (dashed) with reference sites (blue dots).

When the metrics from the existing MMI for Biotype 3 were tested with the new dataset (new reference and stressed sites, new samples, etc.), the resulting DE (100%) was as good as any of the derived alternatives. However, the *Z*-score (2.44) was somewhat lower than the selected alternative and the increased number of metrics in the new index (8 vs. 6) will decrease the likelihood that a single metric will overly influence the index results. In addition, the new index included both insect taxa and percent non-insect taxa, which are not statistically redundant, but measure a similar aspect of the samples that might over-weight the index towards the non-insects.

Table 17. The best index alternatives considered for Biotype 3, including 8 that resulted from screening the all-subsets models, the best manually derived model, and the metrics from the existing MMI. The selected metric is #8_70028.

ModelID	8_24000	8_23042	8_6699	8_22242	8_12875	8_67143	8_48850	**8_70028**	Best Manual	Existing Metrics
TotalTax	1	1	1	1	1	1		**1**	1	
InsectTax							1			1
pt_noninsect	1		1					**1**	1	1
PercDom01							1			1
pEPTnoB		1	1	1	1	1	1	**1**		
NonInPct	1	1		1		1				
ColeoPct					1		1		1	
CCO2ChirPct	1	1	1	1	1					
aSprwlTax*	1	1	1	1	1	1	1	**1**		
SprwlPct									1	1
IntolTax						1		**1**		
pt_Intol	1						1		1	
nt_DecrPlains			1	1	1					
pi_DecrPlains		1								1
pi_IncrPlains	1	1	1	1	1	1	1	**1**	1	
CllctPctFAC			1							
PredScrapTaxFAC				1					1	
PredShrTaxFAC					1	1				1
PredTaxFAC								**1**		
ScrapPctFAC	1	1				1	1	**1**		
Ref.q25	64.1	60.8	64.0	66.3	60.5	53.8	49.7	**52.1**	57.5	57.9
Str.DE	100	100	100	100	100	100	100	**100**	100	100
Z-score	5.53	4.27	3.77	3.52	3.33	2.77	2.65	**2.64**	3.48	2.44

* = adjusted

Precision

In a precision analysis, seven replicate sets, each containing 2 or 3 samples collected at the same site on the same day, were used to estimate natural variability and sampling error of the index and component metrics. The 90% confidence interval (CI90) for the index was 11.8 index points and the CV was 15.3% (Table 18). The most variable metric was IntolTax, with a CV of 60.4%.

Table 18. Precision statistics for the selected Biotype 3 index and its component metrics, based on both metric scores and values.

	Scores				Metrics			
Index/Metric	RMSE	mean	CV	CI90	RMSE	mean	CV	CI90
Index_Biotype3	7.2	47.0	15.3	11.8				
TotalTax	6.8	67.2	10.1	11.1	2.3	22.7	10.1	3.8
pt_noninsect	11.8	48.8	24.2	19.5	4.7	25.5	18.6	7.8
pEPTnoB	5.2	20.5	25.1	8.5	4.4	11.6	37.8	7.2
aSprwlTax	19.5	58.1	33.5	32.0				
IntolTax	17.0	28.2	60.4	28.0	1.0	1.4	69.0	1.6
pi_IncrPlains	10.4	77.0	13.5	17.1	8.2	18.3	45.0	13.5
PredTaxFAC	13.3	52.4	25.4	21.9	1.9	5.4	34.9	3.1
ScrapPctFAC	9.0	24.3	37.0	14.8	4.7	6.8	69.5	7.7

Index response to stress

The index scores in Biotype 3 are highest in reference sites with low percentages of development in the catchments and when the reference scores do not indicate stress (Figure 27). There were two reference sites that had high index values despite relatively high % developed land in the catchments. These include the Purgatoire River (StationID EPA01-238) and Willow Creek (WCOP04-R001).

In the correlation analysis, there was indication that longitude was correlated to some reference metrics. Longitude is discontinuous in Biotype 3 and it was not used as a classification variable or for continuous metric adjustments. The index appears to have similar reference and stressed index value distributions in the east and west portions of the state, though the ranges are not as broad in the west (Figure 28).

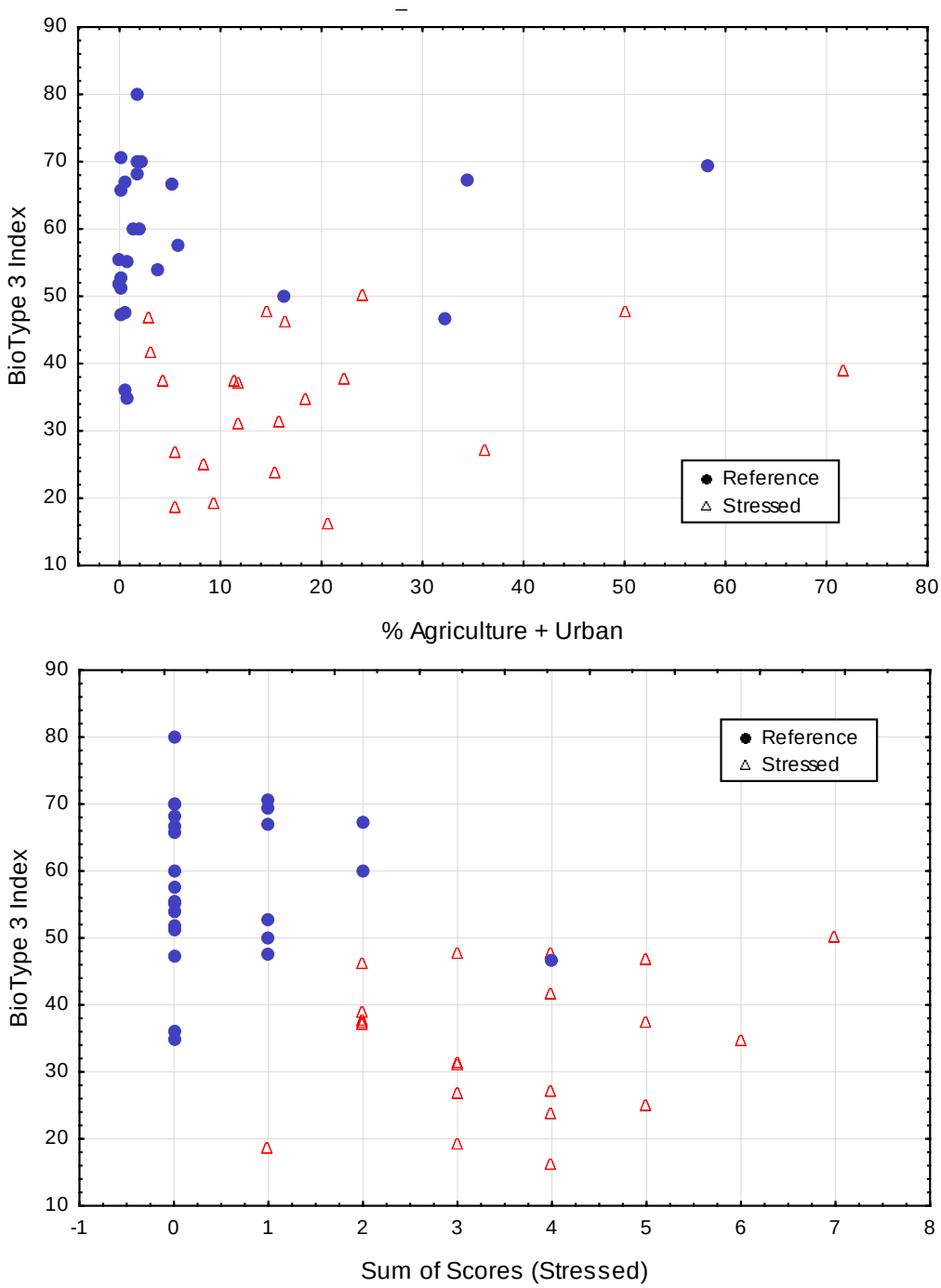


Figure 27. Biotype 3 index in response to the stressor gradients: % agriculture and urban land cover and the sum of reference criterion scores indicating stressed conditions.

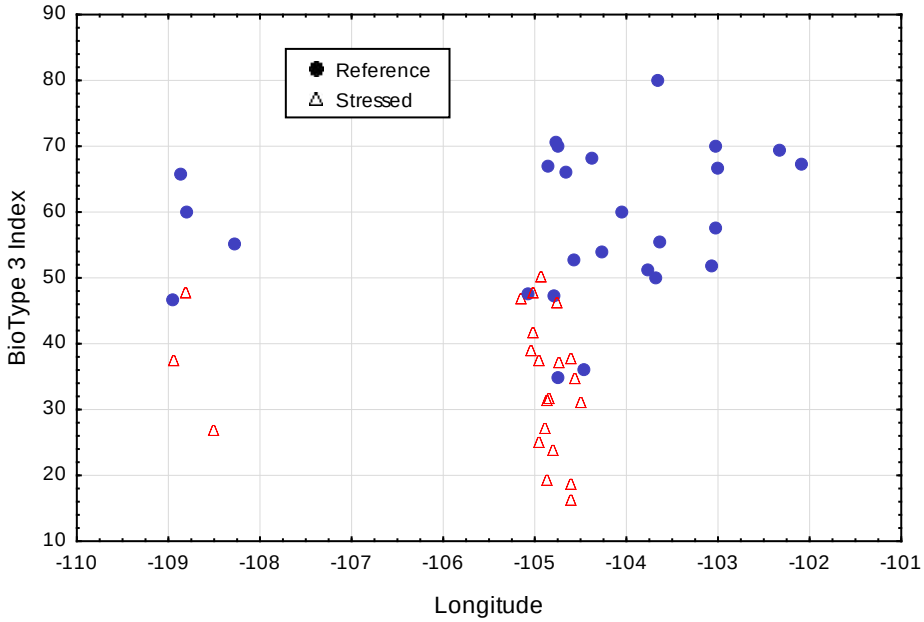


Figure 28. Biotype 3 index values in relation to longitude, showing reference and stressed sites.

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. Draft thresholds for impairment or other management actions will be determined by CDPHE and approved by Colorado's Water Quality Control Commission. The index should only be applied using samples from wadeable streams that were collected processed using CDPHE protocols and data manipulations as described in sections 2 and 3.2.

Index application should proceed as follows:

1) Enter the biological and environmental variable data into the analytical database.
2) Determine the site class of the site using the classification rules (Table 6).
3) Calculate index metrics and scores (Table 10, 13, or 16) and average as an MMI.
4) Compare the resulting MMI value to thresholds for the appropriate site class.
5) Include MMI performance statistics when reporting assessments (Table 19).

Summer temperature and Julian day are used in adjusting some metrics. Summer air temperature is determined as the 30 year average for the site based on PRISM data sets. Julian day is the sampling date expressed as the number of days after January 1 in any year. The Julian day for January 1 is 0 and for December 31 it is 364.

Table 19. MMI performance statistics.

Site Class	DE	CI_{90}
Biotype 1, Transitional	94.6	7.8
Biotype 2, Mountains	100	10.4
Biotype 3, Plains	100	11.8
All Biotypes	97.1	

5 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 both in an original analysis (Jessup 2009) and in the confirming analyses in this report. 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 12.5%.

The MMIs were selected from among several alternatives. All final MMIs included metrics from each of five metric categories and had a combined DE of 97.1% for all Biotypes. This high DE and low classification error indicate that assessments of Colorado stream benthic condition, made using the MMIs, will be accurate.

6 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.

Blocksom, K.A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. *Environmental Management*, *31*(5), 670-682.

Bryce, S.A., and S.E. Clarke. 1996. Landscape-level ecological regions: linking state-level ecoregion frameworks with stream habitat classifications. *Environmental Management*, 20(3):297-311.

CDPHE. 2010a. *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.

CDPHE. 2010b. Benthic Macroinvertebrate Sampling Protocols. Water Quality Control Division, Standard Operation Procedure WQCDSOP-001.

Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112(1):155.

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

Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006. Concepts and Approaches for the Bioassessment of Non-Wadeable Streams and Rivers. EPA/600/R-06/127. U.S. EPA, Office of Research and Development, National Exposure Research Laboratory, Cincinnati, OH.

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

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

Jessup, B.K. 2009. *Recalibration of the Macroinvertebrate Multi-Metric Index for Colorado.* Prepared for: Colorado Department of Public Health and Environment, Denver, CO. Prepared by: Tetra Tech, Inc., Center for Ecological Sciences, Owings Mills, Maryland. *For further information, contact CDPHE.*

Karr, J.R. 1991. Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications* 1(1): 66-84. Stable URL: http://www.jstor.org/stable/1941848.

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

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: Colorado Department of Public Health and Environment, Water Quality Control Division – Monitoring Unit, Denver, CO. Prepared by: Tetra Tech, Inc., Center for Ecological Sciences, Owings Mills, MD; and Utah State University, Logan, UT.

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* 20: 198-219.

Rohm, C. M., John W. Giese, and Charles C. Bennett. 1987. Evaluation of an aquatic ecoregion classification of streams in Arkansas. *Journal of Freshwater Ecology*, *4*(1):127-140.

Steedman, R.J. 1994. Ecosystem health as a management goal. *Journal of the North American Benthological Society* 13(4): 605-610.

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

Stribling, J.B., B.K. Jessup, and D.L. Feldman. 2008. Precision of benthic macroinvertebrate indicators of stream condition in Montana. *Journal of the North American Benthological Society* 27(1): 58-67. doi: 10.1899/07-037R.1.

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. Prepared by: Tetra Tech, Inc., Owings Mills, MD

USGPO. 1989. Federal Water Pollution Control Act (33 U. S. C. 1251 et seq.) as amended by a P. L. 92-500. In: *Compilation of selected water resources and water pollution control laws*. US Government Printing Office. Printed for use of the Committee on Public Works and Transportation. Washington, DC.

U.S. Environmental Protection Agency (USEPA). 2016. *A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems*. EPA-842-R-16-001. U.S. Environmental Protection Agency, Washington, DC.

Van Sickle, J., and R. M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society*, 19(3):370-384.

Weigel, B. M. 2003. Development of stream macroinvertebrate models that predict watershed and local stressors in Wisconsin. *Journal of the North American Benthological Society*, 22(1):123-142.

Zar, J. H. 1999. Biostatistical analysis. 4th edition. Prentice/Hall, Upper Saddle River, New Jersey.

Appendix A

Reference and Stressed Site Selection Process

Appendix A: Reference and Stressed Site Selection Process

1. Selection Process for 2010 Version of MMI

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.

<u>GIS Based Approach</u>

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)

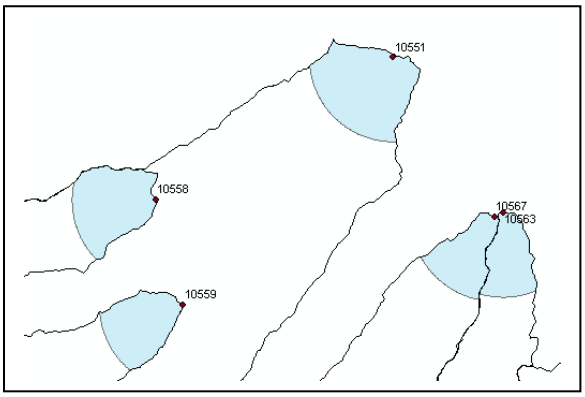
Sources of GIS Information

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.

2. Selection Process for 2017 Version of MMI

The 2017 reference and stressed site selection process began with the original 133 reference sites and 60 stressed sites used to develop MMI version 3. However, the division dropped two reference sites based on field observations made during subsequent visits to those sites, which indicated introduction of human disturbance. Additionally, during the 2018 Listing Methodology workgroup process, the Aquatic Life Technical Advisory Committee (TAC) subgroup agreed to remove four additional reference sites and one stressed site. These actions resulted in 127 reference sites and 59 stressed sites.

The division further refined the original suite of reference and stressed sites for 2017 by screening 372 additional candidate sites visited by the division, EPA, or USFS with samples collected between 2008 to 2014 and visually scrutinizing in more detail human disturbance at all sites in CO-EDAS (n=1285) via aerial reviews.

To remain consistent with the 2010 selection process, the division used the same GIS-based approach with some modifications to the sources of GIS information, as identified in Section 1. GIS calculations were restricted to the 372 additional candidate sites.

Modifications to Sources of GIS Information

Permitted point sources and Headgated diversions GIS layers were updated.

- Permitted point sources — WQCD GIS shapefile (updated spring 2015)
- Headgated diversions — CO Division of Water Resources (updated spring 2015)

Total Watershed Area vs. 5 km Radius

Total watershed areas were calculated for new candidate sites using EPA's BASINS 4.1 watershed delineation tool. The 5 km radius polygons were created in the same fashion as shown in Section 1.

GIS Calculations

Human disturbance values for the additional 372 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 comparison to existing thresholds or cutoffs.

The same thresholds from 2010 were used again in 2017. 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 23 additional candidate reference sites in the mountains bioregion, 8 in the plains bioregion and 5 in the xeric bioregion.

Site Duplication Removal

The division used the same "rule of thumb" approach to remove duplicate 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 the both sites were retained for further analysis. If the difference was <20% then the Division selected the site with the most recent sample year. The 2001-02 drought year exception did not apply to this selection process since the sites involved were collected from 2008-2014. However, the second exception was still used. 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 16 candidate reference sites in the mountains bioregion, 6 in the plains bioregion and 5 in the xeric bioregion.

Aerial Imagery Review

This phase of the reference screening process involved reviewing 27 aerial images on Google Earth. Division staff reviewed each site and a 3 km distance upstream from the site, scrutinizing for human activity. The 0-10 scoring scale from the Rapid Screening Disturbance Scoring instruction manual was not used during this selection process. Best professional judgment of division staff was used, based upon visual interpretation of aerial overviews, to remove sites from further consideration.

The outcome of this process yielded 14 reference sites in the mountains bioregion, 1 in the plains bioregion and 2 in the xeric bioregion. The final outcome of the GIS screening and aerial imagery review resulted in 17 new reference sites.

Stressed Site Screening Process

The stressed site screening process involved using the same techniques from the 2010 selection process. This process resulted in retaining 33 of the 59 remaining stressed sites and adding 53 additional stressed sites for a total of 86 stressed sites.

Appendix B

Reference and Stressed Sites Used in Analysis

Table B-1. Reference sites/samples used in index recalibration.

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
10262	Beaver Creek	38.552510	-107.053040	1	3483_1	8/1/2012
10350	Razor Creek	38.384550	-106.673283	1	3665_1	9/24/2014
10415	Leroux Creek	38.881110	-107.785830	1	3667_1	7/30/2014
10558	Nork Fork Escalante Creek	38.636890	-108.427220	1	3074_1	7/18/2007
10559	Middle Fork Escalante Creek	38.581180	-108.405880	1	3075_1	7/18/2007
10570	Kannah Creek	38.961210	-108.229650	1	3049_1	7/19/2007
10666	Billy Creek	38.294840	-107.701980	1	3051_1	7/16/2007
10805	San Miguel River	38.126000	-108.207660	1	36_1	10/19/1994
10834	Naturita Creek	38.159120	-108.403050	1	3086_1	9/11/2007
10905G	West Paradox Creek	38.382730	-108.996020	1	3382_1	7/19/2011
10906	La Sal Creek	38.320520	-108.976990	1	3091_1	9/12/2007
10917	North Fork Mesa Creek	38.503190	-108.790370	1	3065_5	9/10/2007
10922	Ute Creek	38.723700	-108.909700	1	3092_2	9/13/2007
11206	Carr Creek	39.566490	-108.501700	1	2885_1	9/13/2006
11208	Roan Creek	39.509620	-108.524820	1	2887_1	9/13/2006
11408	Little Snake River	40.997000	-107.150833	1	46_1	10/7/1998
11485	Vermillion Creek	40.719650	-108.751800	1	2897_1	8/15/2006
11535	Lost Creek	40.050600	-107.468700	1	3387_1	9/26/2011
12799C	Thompson Creek	39.3317833	-107.219083	1	3443_1	9/29/2011
12802	Yampa River	40.501330	-107.264330	1	168_1	8/28/2001
12832	Little Bear Creek	40.688833	-107.434500	1	2920_1	8/16/2006
12880	Service Creek	40.29453	-106.79903	1	3450_1	8/1/2011
12961	Illinois River	40.462400	-106.177000	1	84_1	9/23/1997
5588	South Boulder Creek	39.980500	-105.220833	1	569_1	6/24/2003
5771	Jackson Creek	39.345667	-104.981167	1	509_1	5/30/2003
5772A	Garber Creek	39.355240	-105.027180	1	3062_1	8/16/2007
5775	Cook Creek	39.181650	-104.896820	1	3064_2	8/16/2007
5777A	West Plum Creek	39.369167	-104.965500	1	512_1	5/30/2003
5971c1	Rule Creek	39.047694	-105.098333	1	2648_1	10/25/2001
7130	Hayden Creek	38.335580	-105.802200	1	2736_1	9/14/2005
7385	Little Fountain Creek	38.681670	-104.858060	1	3741_1	7/16/2015
8116	West Alder Creek	37.705000	-106.647800	1	2703_1	8/22/2005
8337	La Jara Creek	37.177500	-106.211930	1	3342_1	8/17/2010
8351a	Cat Creek	37.40495	-106.30519	1	3615_1	9/17/2013
8395	Rock Creek	37.490300	-106.258900	1	3761_1	8/25/2015
8530	Sand Creek	37.84293	-105.58335	1	3375_1	7/5/2011
CO072M	Junction Creek	37.333611	-107.909444	1	3368_1	6/20/2011
CO125M	Beaver Creek	37.999722	-108.193889	1	2274_0	8/21/1995
CO151M	Badger Creek	38.594167	-105.836389	1	2285_0	8/22/1995
CO162M	Fernleaf Gulch	38.406944	-105.638889	1	2292_0	8/25/1995
NPlatte	North Platte River	40.697280	-106.412530	1	378_1	11/4/1997
USFS Piney 2	Piney River	39.750400	-106.471300	1	2964_0	9/28/2005
WCOP99-0578	Middle Fork Little Snake River	40.971500	-107.019000	1	2448_0	8/14/2001
WCOP99-0595	East Rifle Creek	39.643050	-107.710400	1	2452_0	8/2/2001
WCOP99-0649	Noname Creek	39.581660	-107.288090	1	2466_0	7/23/2002
10742	Little Taylor Creek	37.59082	-108.21032	2	3350_1	6/21/2011

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
11462	Oliver Creek	40.877100	-107.087100	2	2893_1	10/3/2006
11463	Oliver Creek	40.864100	-107.067400	2	2894_1	10/4/2006
12182	Roaring Fork Arapaho Creek	40.130800	-105.767500	2	56_1	9/9/1997
12183A	Byers Creek	39.876523	-105.902246	2	3559_1	8/8/2013
12184A	East St. Louis Creek	39.887407	-105.877125	2	3560_1	8/8/2013
12759	South Fork Fryingpan River	39.237200	-106.590000	2	3507_1	9/24/2012
12771	Woody Creek	39.24511	-106.7521	2	3158_1	7/31/2008
12785A	Lost Man Creek	39.129200	-106.618000	2	2864_1	9/10/2006
12869B	Gilpin Creek	40.797	-106.695	2	3683_1	7/23/2014
5548	South St. Vrain Creek	40.078767	-105.577750	2	556_2	6/24/2003
5768B	Bear Creek	39.611358	-105.533367	2	3464_1	9/19/2011
5779B	East Plum Creek	39.184310	-104.930680	2	3043_1	8/14/2007
7120E	Grape Creek	37.92989	-105.456	2	3699_1	9/9/2014
7133	South Colony Creek	37.97547	-105.50822	2	3610_1	7/24/2013
7164	Browns Creek	38.669000	-106.161000	2	3732_1	9/16/2015
7170	Pine Creek	38.998833	-106.231767	2	2686_1	7/26/2005
7392	Severy Creek	38.897651	-104.98808	2	3336_1	9/21/2010
8715	La Manga Creek	37.116390	-106.377830	2	3771_1	8/24/2015
9294	Middle Fork Piedra River	37.51222	-107.14564	2	3719_1	8/27/2014
9716A	Box Canyon Creek	37.448888	-108.207777	2	2851_1	7/11/2006
CO038M	Middle Fork North Crestone	38.036111	-105.642500	2	2214_0	8/11/1994
CO03RS	Hope Creek	37.553056	-106.802222	2	2250_0	9/21/1995
CO04RS	Little Cimarron Creek	38.210000	-107.463611	2	2224_0	8/16/1994
CO051M	Lottis Creek	38.772500	-106.622500	2	2227_0	8/17/1994
CO067M	Silver Creek	37.428889	-106.758889	2	2239_0	8/1/1994
CO070M	Lime Creek	37.677000	-107.750900	2	2242_0	8/4/1994
CO113M	North St. Vrain Creek	40.204722	-105.406111	2	2269_0	9/14/1995
CO116M	Garner Creek	38.189722	-105.776389	2	2271_0	8/28/1995
CO122M	South Fork Saguache Creek	37.919167	-106.715278	2	2273_0	8/23/1995
CO133M	Trout Creek	37.666111	-107.070000	2	2278_0	8/17/1995
CO142M	North Anthracite Creek	38.981667	-107.191111	2	2283_0	9/13/1995
CO153M	Fourmile Creek	38.966111	-106.139722	2	3774_1	9/15/2015
CO155M	Middle Fork Brush Creek	38.954167	-106.858333	2	2288_0	9/6/1995
CO174M	Scott Gomer Creek	39.508056	-105.704722	2	3470_1	9/12/2011
EPA01-168	Fawn Creek	39.852500	-107.315000	2	2161_0	8/11/2001
EPA01-239	South Apache Creek	37.852200	-104.985000	2	2166_0	7/7/2001
EPA01-240	Wahatoya Creek	37.412300	-104.964000	2	3778_1	9/9/2015
EPA01-241	Huerfano River	37.613800	-105.472000	2	2168_0	7/10/2001
EPA01-242	El Rito Azul	37.274300	-106.619000	2	2169_0	7/11/2001
EPA01-246	Big Blue Creek	38.176000	-107.387000	2	2172_0	7/16/2001
EPA01-247	Agate Creek	38.451000	-106.341000	2	2173_0	7/17/2001
EPA01-248	Cottonwood Creek	38.056700	-105.579000	2	2174_0	7/18/2001
EPA01-249	Newlin Creek	38.266200	-105.189800	2	3779_1	7/16/2015
EPA01-250	Goose Creek	39.185600	-105.391000	2	2176_0	7/22/2001
EPA01-251	Wigwam Creek	39.247000	-105.358000	2	2177_0	7/25/2001
EPA01-252	Rock Creek	39.374600	-105.684000	2	2178_0	7/26/2001
EPA01-253	North Cottonwood Creek	38.870800	-106.266000	2	3780_1	9/15/2015

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
EPA01-259	Deeds Creek	39.297900	-106.659000	2	2184_0	8/8/2001
EPA01-260	Rock Creek	39.216900	-106.398000	2	3781_1	9/14/2015
EPA01-261	Lake Fork	39.285300	-106.449000	2	3782_1	9/14/2015
EPA01-262	Jefferson Fork	39.462600	-105.858000	2	2187_0	8/11/2001
USFS Avalanche	Avalanche Creek	39.231100	-107.199200	2	2936_0	9/7/2005
USFS Cataract	Cataract Creek	39.828500	-106.327200	2	2951_1	9/26/2006
USFS Cross 1	Cross Creek	39.541300	-106.433400	2	2992_0	9/15/2005
USFS Cross 4	Cross Creek	39.484100	-106.503900	2	2953_0	10/2/2006
USFS EFH Upper	East Fork Homestake Creek	39.355900	-106.454000	2	2955_1	9/18/2006
USFS Gore 1	Gore Creek	39.627700	-106.271000	2	3001_0	9/27/2004
USFS Meadow	Meadow Creek	39.595400	-106.123700	2	3003_0	9/11/2003
USFS N Tenmile	North Tenmile Creek	39.573000	-106.172200	2	2961_0	9/8/2005
USFS Ripple	Ripple Creek	40.074600	-107.302800	2	2976_0	8/12/2004
USFS Snowmass	Snowmass Creek	39.180100	-107.022000	2	2978_0	9/8/2004
USFS West Lake	West Lake Creek	39.539500	-106.616000	2	2991_0	9/27/2006
USFS-363463	Whalen Creek	40.60336	-106.61381	2	3232_1	9/1/2009
USFS-376014	Indian Creek	40.38417	-106.46071	2	3237_1	9/2/2009
USFSPIKE1	Pine Creek	39.240580	-105.282550	2	3549_5	7/23/2012
USFSPIKE4	Beaver Creek	39.028694	-104.944805	2	2666_1	11/6/2003
WCOP03-R009	East Fork Piedra River	37.481700	-107.097100	2	2785_0	8/8/2003
WCOP99-0502	Adams Fork Conejos River	37.329590	-106.689590	2	2418_0	8/5/2000
WCOP99-0503	Lost Man Creek	39.159490	-106.571760	2	2419_0	7/27/2000
WCOP99-0518	Ouzel Creek	40.199810	-105.625760	2	2433_0	7/10/2000
WCOP99-0633	Crystal River	38.723880	-106.672530	2	2762_0	6/18/2003
WCOP99-0634	Ute Creek	37.595500	-105.398900	2	2462_0	8/13/2002
10551	Escalante Creek	38.717800	-108.268610	3	2524_1	9/28/2004
10700	Dolores River	38.371100	-108.802000	3	34_1	10/19/1994
10980	Roc Creek	38.441790	-108.877370	3	3093_1	9/10/2007
5258	West Cherry Creek	39.17294	-104.74675	3	3695_1	9/18/2014
6753400	Lonetree Creek	40.898300	-104.868000	3	2349_0	7/20/1994
7226	Hardscrabble Creek	38.343430	-105.068240	3	3009_1	5/30/2007
7284	Rock Creek	38.207970	-104.793120	3	3739_1	7/15/2015
7285	Boggs Creek	38.229400	-104.746200	3	3740_1	7/15/2015
7512	Chico Creek	38.357700	-104.387300	3	3745_1	7/15/2015
7560	Trinchera Creek	37.044600	-104.051000	3	2540_1	8/26/2004
7571	San Francisco Creek	37.120617	-104.261350	3	2541_1	8/26/2004
7595	Del Agua Arroyo	37.347220	-104.574240	3	3014_1	5/29/2007
7630	Bear Creek	37.613800	-104.773800	3	3016_1	5/2/2007
7645	Cucharas River	37.679660	-104.652000	3	461_1	4/5/2000
7997a	East Carrizo Creek	37.135383	-103.015716	3	2650_1	4/7/2004
7997b	Cottonwood Creek	37.110850	-103.074400	3	2651_1	4/7/2004
7999a	East Carrizo Creek	37.168533	-103.034533	3	2652_1	4/7/2004
EPA01-238	Purgatoire River	37.558400	-103.652000	3	2165_0	7/6/2001
WCOP01-0777	Chacuaco Creek	37.494240	-103.631300	3	2772_0	4/22/2003
WCOP03-R003	Chief Creek	40.103333	-102.322500	3	2779_0	7/10/2003
WCOP04-R001	Willow Creek	40.800250	-104.459250	3	2788_0	7/29/2004
WCOP04-R003	Two Butte Creek	37.513370	-103.026670	3	2789_0	5/19/2004

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
WCOP04-R007	Yellow Jacket Creek	37.364200	-108.951100	3	3372_1	6/21/2011
WCOP04-R008	Rush Creek	40.099400	-102.097380	3	2794_0	6/9/2004
WCOP04-R009	Timpas Creek	37.827060	-103.772980	3	2795_0	7/6/2004
WCOP99-0563	South Rush Creek	38.886740	-103.682980	3	2434_0	5/17/2001

Table B-2. Stressed sites/samples used in index recalibration.

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
10115	Coal Creek	38.871200	-106.985790	1	3272_1	9/24/2010
10123	Woods Creek	38.896330	-106.967330	1	195_1	7/28/2000
10403	North Fork Gunnison River	38.791900	-107.726000	1	3181_1	9/2/2009
10847	Maverick Draw	38.192700	-108.378050	1	3378_1	7/21/2011
11140	Parachute Creek	39.569940	-108.109400	1	2879_1	9/12/2006
11345	Stinking Gulch	40.334700	-107.688800	1	210_1	8/23/2000
11601	Piceance Creek	39.964310	-108.268210	1	3388_1	9/27/2011
11732A	Yellow Creek	40.118080	-108.360890	1	2904_1	8/9/2006
12191	Fraser River	39.990278	-105.829722	1	3403_1	9/26/2011
12300A	Blue River	40.042480	-106.395730	1	3411_1	10/11/2011
12715	Castle Creek	39.195800	-106.832900	1	388_1	10/23/2001
12761B	Brush Creek	39.214000	-106.939000	1	2862_1	9/7/2006
12783	Roaring Fork River	39.210764	-106.839472	1	3440_1	10/20/2011
12783B	Roaring Fork River	39.217970	-106.854638	1	3572_1	8/22/2013
12805B	Yampa River	40.491510	-106.948010	1	3575_1	8/20/2013
12808P	Yampa River	40.287220	-106.828890	1	465_1	8/28/2001
12870	Fish Creek	40.466500	-106.824670	1	298_0	8/29/2001
12893	Burgess Creek	40.451660	-106.810160	1	436_1	8/29/2001
12940	Michigan River	40.745373	-106.296730	1	3580_1	8/19/2013
151	Slate River	38.852060	-106.953200	1	3508_1	7/30/2012
383103106594200	Gunnison River	38.517500	-106.995600	1	2300_0	8/6/1998
52	Eagle River	39.649900	-106.952400	1	3113_1	9/23/2008
5310	Cache La Poudre River	40.622333	-105.139833	1	3514_1	8/8/2012
5576	Boulder Creek	40.038120	-105.211600	1	3205_1	7/29/2009
5618	Clear Creek	39.758930	-105.217350	1	3211_1	9/22/2009
5762	Bear Creek	39.643150	-105.307200	1	3531_1	9/13/2012
5965	S.F. South Platte River	38.983488	-105.883436	1	2819_1	7/28/2006
5971a3	Trout Creek	39.088083	-105.096944	1	3321_1	8/2/2010
5988A2	Mosquito Creek	39.271330	-106.106670	1	87_1	9/10/1998
7252	Cripple Creek	38.711660	-105.183900	1	3228_1	8/5/2009
7587	Apache Canyon	37.148740	-104.921950	1	3546_1	8/21/2012
8105A	Willow Creek	37.822550	-106.910620	1	3707_1	8/19/2014
8300	Rio Grande River	37.078267	-105.756465	1	3752_5	9/16/2015
8375	Sangre de Cristo Creek	37.425000	-105.414722	1	3760_5	8/25/2015
8650	Kerber Creek	38.250000	-105.949830	1	3770_1	9/16/2015
9146200	Uncompahgre River	38.183900	-107.745900	1	2391_0	8/5/1998
9149480	Dry Creek	38.645800	-108.049000	1	2394_0	8/5/1998
9711	Mud Creek	37.295000	-108.366000	1	3357_1	7/7/2011

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
9713	Chicken Creek	37.332110	-108.325789	1	3358_1	7/7/2011
9717	Mancos River	37.304149	-108.357838	1	3722_1	8/6/2014
9871	McElmo Creek	37.327717	-108.701667	1	3349_1	7/13/2010
CO050M	Noname Creek	38.783056	-106.482222	1	2226_0	8/18/1994
SMR-3	San Miguel River	38.220200	-108.549200	1	2716_1	9/7/2005
UGRWCD-6	Gunnison River	38.517540	-106.994920	1	3634_1	8/16/2013
WCOP99-0568	La Plata River	37.134230	-108.163690	1	2438_0	6/26/2001
WCOP99-0622	Hartman Draw	37.365900	-108.593300	1	2456_0	6/19/2002
WCOP99-0638	White River	40.148870	-108.326480	1	2464_0	10/8/2002
10609	Uncompahgre River	37.989720	-107.647500	2	27_1	8/21/1996
10814	San Miguel River	37.948210	-107.877070	2	3555_1	8/23/2013
12303	Blue River	39.727200	-106.134700	2	3412_1	8/31/2011
12304b	Blue River	39.486500	-106.046283	2	3413_1	8/31/2011
12304D	Blue River	39.625200	-106.065650	2	3414_1	8/31/2011
5945	South Platte River	38.909933	-105.459367	2	3601_1	9/10/2013
12553	Gore Creek	39.620990	-106.426400	2	3152_1	9/23/2008
12555	Gore Creek	39.641670	-106.306670	2	3568_1	8/21/2013
12751G	Fryingpan River	39.365380	-106.826180	2	3494_1	9/25/2012
12756	Fryingpan River	39.362774	-106.829601	2	3499_1	9/25/2012
133	North Fork Clear Creek	39.747916	-105.399333	2	2575_1	5/27/2004
7187G	Lake Fork	39.251150	-106.375370	2	3737_1	9/14/2015
7227A	St. Charles River	37.985660	-105.049000	2	475_1	6/12/2002
7263	Cripple Creek	38.720300	-105.178000	2	3229_1	8/5/2009
9046530	French Gulch	39.493000	-106.044700	2	2371_0	8/4/1998
5971a5	Trout Creek	39.046476	-105.089960	2	2643_1	10/23/2003
22	South Platte River	40.412222	-104.562778	3	3550_5	9/11/2012
34	Clear Creek	39.827600	-104.952100	3	3300_1	8/30/2010
5164	South Platte River	39.831266	-104.948000	3	2579_1	7/7/2004
5208	Crow Creek	40.395320	-104.492800	3	3077_5	8/30/2007
5212a	Sand Creek	39.761100	-104.857600	3	2582_1	7/7/2004
5221	Big Dry Creek	40.069200	-104.832200	3	3308_1	9/20/2010
5306	Cache La Poudre River	40.422667	-104.598000	3	359_1	6/17/2002
5418	Big Thompson River	40.344100	-104.867500	3	2608_1	6/14/2004
5452	Little Thompson River	40.329100	-104.886200	3	2611_1	6/14/2004
5504	St. Vrain Creek	40.184590	-104.931000	3	3199_1	7/29/2009
5605	Clear Creek	39.774170	-105.141330	3	3312_1	8/30/2010
5791a	Lakewood Gulch	39.731333	-105.039333	3	557_1	6/21/2003
6713500	Cherry Creek	39.742500	-105.000000	3	2320_0	7/18/2002
6753990	Lonetree Creek	40.442500	-104.588800	3	2354_0	8/6/2002
7298B5	Salt Creek	38.227251	-104.598681	3	3703_1	8/19/2014
7309	Sand Creek	38.810060	-104.748030	3	3539_1	8/27/2012
7310	East Fork Sand Creek	38.822283	-104.726500	3	3271_1	6/15/2010
7316	Fountain Creek	38.797500	-104.786150	3	3332_1	8/31/2010
9153290	Reed Wash	39.211400	-108.803700	3	2398_0	11/12/1997
AR0018	Buffalo Creek	38.148700	-102.311000	3	3772_5	8/19/2015
WCOP99-0594	Colorado River	39.181700	-108.933100	3	2451_0	8/9/2001
WCOP99-0648	Gunnison River	39.004970	-108.506440	3	2465_0	8/26/2002

STATIONID	WaterbodyName	Lat_Dec	Long_Dec	BIOTYPE	SAMP_REP	COLLDATE
WCOP99-0671	South Platte River	39.663140	-105.003780	3	2766_0	5/13/2003

Appendix C

Correlations and Bi-Plots for Metric Adjustments

Table C-1. Spearman correlation coefficients for reference metrics and environmental variables in BioType 1, showing metric adjustment factors in the 'Adjust' column. Red font symbolizes significance of the correlation. Yellow highlights emphasize stronger correlation coefficients.

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
TotalTax	-0.01	-0.16	-0.16	0.31	0.25	-0.26	-0.37	0.34	-0.34	0.3	-0.07	-0.32	
InsectTax	0.12	-0.16	-0.15	0.27	0.2	-0.19	-0.46	0.43	-0.45	0.31	0	-0.39	SUMMER
pt_noninsect	-0.41	-0.13	-0.15	0.16	0.17	-0.08	0.15	-0.22	0.25	0.04	-0.06	0.13	
EPTTax	0.22	-0.15	0.01	0.18	0.1	0.06	-0.52	0.44	-0.48	0.29	0.13	-0.36	SUMMER
pt_EPT	0.3	-0.06	0.16	0.01	-0.06	0.2	-0.24	0.23	-0.23	0.13	0.07	-0.12	
EphemTax	-0.01	-0.27	-0.09	0.1	0.11	0.03	-0.37	0.3	-0.31	0.29	0.06	-0.23	
PlecoTax	0.3	-0.16	-0.04	0.18	0.15	0.07	-0.39	0.27	-0.27	0.31	0.21	-0.3	SUMMER
TrichTax	0.26	-0.01	0.09	0.25	0.05	0.1	-0.54	0.51	-0.59	0.21	0.04	-0.43	SUMMER
ColeoTax	0.12	-0.01	-0.09	0.15	-0.14	0.03	-0.27	0.24	-0.5	-0.01	0.17	-0.38	
DipTax	-0.19	-0.08	-0.23	0.14	0.27	-0.3	-0.23	0.18	-0.15	0.25	-0.1	-0.29	
pt_Diptera	-0.26	0.04	-0.23	-0.11	0.21	-0.21	0.05	-0.08	0.16	0.06	-0.08	-0.1	
ChiroTax	-0.23	-0.11	-0.11	0.08	0.27	-0.37	-0.2	0.2	-0.15	0.2	-0.18	-0.2	
CrMolTax	-0.41	-0.06	-0.01	0.29	0.14	-0.17	0.17	-0.22	0.26	0.17	-0.26	0.1	
OligoTax	-0.11	-0.28	-0.19	0	0.47	-0.15	-0.06	0.1	0.07	0.07	-0.09	0.07	
TanytTax	-0.09	-0.08	-0.01	0.16	0.07	-0.25	-0.19	0.21	-0.25	0.05	-0.05	-0.34	
Shan_2	0.08	-0.11	-0.12	0.41	0.24	-0.31	-0.22	0.3	-0.2	0.34	-0.16	-0.24	
Evenness	0.09	-0.12	-0.12	0.44	0.22	-0.32	-0.21	0.28	-0.19	0.38	-0.14	-0.23	
D_Mg	0.01	-0.2	-0.21	0.35	0.25	-0.21	-0.37	0.31	-0.28	0.32	-0.04	-0.34	
D	-0.07	0.07	0.07	-0.46	-0.22	0.33	0.14	-0.27	0.17	-0.36	0.21	0.21	
PercDom01	-0.08	0.13	0.14	-0.43	-0.24	0.28	0.17	-0.3	0.18	-0.42	0.19	0.28	
EPTPct	0.3	0.05	0.2	-0.24	-0.26	0.25	0.11	-0.01	0.08	-0.17	0.14	0.24	
EphemPct	-0.04	-0.21	-0.17	-0.28	-0.12	0.33	-0.02	0.01	0.08	-0.11	0.16	0.03	
PlecoPct	0.44	-0.23	-0.32	-0.06	0.2	-0.04	-0.22	0.3	-0.09	0.26	0.06	-0.21	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
TrichPct	0.39	0.27	0.4	0.04	-0.3	0.12	0.05	0.08	-0.15	-0.14	0.01	0.12	
NonInPct	-0.15	-0.1	-0.19	0.21	0.16	0.01	-0.09	0.04	-0.02	0.07	0.07	-0.14	
ColeoPct	-0.03	0.13	0.07	0.24	0.11	-0.22	-0.14	0.12	-0.27	0.09	-0.1	-0.29	
OdonPct	-0.01	0.15	0.08	-0.04	-0.25	-0.15	0.42	-0.25	0.2	-0.31	-0.2	0.5	SUMMER
DipPct	-0.43	-0.08	-0.2	0.05	0.28	-0.13	-0.1	0	-0.02	0.19	-0.12	-0.17	
ChiroPct	-0.24	-0.12	-0.11	0.19	0.22	-0.33	-0.17	0.13	-0.12	0.23	-0.22	-0.16	
CrCh2ChiPct	-0.1	-0.16	0.02	0.22	-0.03	0.1	-0.24	0.23	-0.19	-0.02	-0.19	-0.15	
Orth2ChiPct	-0.03	-0.11	-0.17	0.06	0.22	-0.16	-0.12	0.13	-0.18	0.13	0.04	-0.17	
TanytPct	-0.21	-0.15	-0.17	-0.06	0.13	-0.12	-0.33	0.36	-0.35	0.08	-0.11	-0.47	
Tnyt2ChiPct	-0.2	-0.17	-0.16	-0.12	0.07	0.03	-0.17	0.2	-0.13	0	-0.08	-0.35	
OligoPct	-0.03	-0.31	-0.26	-0.01	0.42	-0.06	-0.12	0.17	0.02	0.15	-0.03	-0.03	
CrMolPct	-0.39	-0.02	0.01	0.32	0.09	-0.13	0.12	-0.2	0.24	0.17	-0.16	0.05	
GastrPct	-0.16	0.14	0.14	0.3	-0.01	-0.32	0.19	-0.12	0.1	0.08	-0.25	0.16	
IsoPct	-0.25	-0.18	-0.09	0.06	0.22	-0.17	0.13	-0.17	0.23	0.14	-0.13	0.21	
AmphPct	-0.18	-0.16	-0.16	0.12	0.32	-0.25	0.14	-0.17	0.31	0.14	-0.25	0.18	
BivalPct	-0.33	0.03	0.13	0.01	0.01	0.14	0.09	-0.1	0.16	-0.02	-0.08	0.02	
CllctTax	-0.38	-0.27	-0.27	0.06	-0.07	0.12	0	-0.04	0.06	0.22	-0.13	-0.06	
FiltrTax	-0.13	0.12	0.19	0.05	-0.05	0.1	0.04	-0.08	0.1	-0.1	-0.06	-0.08	
PredTax	0.09	0.05	-0.09	0.29	0.08	-0.14	-0.18	0.06	-0.13	0.07	0.16	-0.07	
ScrapTax	0.03	-0.13	0.07	0.29	0.11	0	-0.28	0.3	-0.32	0.19	-0.14	-0.1	
ShredTax	0.08	-0.3	-0.35	0.13	0.19	0.08	-0.39	0.32	-0.15	0.44	0.06	-0.35	
CllctPct	-0.26	-0.18	-0.26	-0.08	-0.23	0.3	0.11	-0.02	0.13	0.08	-0.12	-0.02	
FiltrPct	-0.12	0.12	0.19	0.05	-0.04	0.1	0.04	-0.07	0.1	-0.1	-0.06	-0.08	
PredPct	0.09	0.21	0.1	0.35	0.18	-0.35	0.2	-0.1	0.11	-0.01	-0.3	0.24	
ScrapPct	0.06	0.09	0.19	0.18	0.16	-0.05	-0.32	0.38	-0.36	0.16	-0.17	-0.09	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
ShredPct	0.13	-0.21	-0.27	0.24	0.06	0.08	-0.32	0.1	0	0.47	0.26	-0.32	
BrrwrTax	-0.14	0.16	-0.18	0.15	0.3	-0.25	0.13	-0.09	0.29	-0.13	-0.22	0.22	
ClmbrTax	0.28	0.01	0.04	0.24	0.18	-0.27	-0.4	0.4	-0.51	0.38	-0.06	-0.48	
ClngrTax	0.15	-0.1	0.02	0.18	-0.02	0.12	-0.56	0.46	-0.52	0.15	0.21	-0.41	SUMMER
SprwlTax	-0.21	-0.13	-0.29	0	0.17	-0.21	-0.11	0.05	-0.05	0.28	0.02	-0.14	
SwmmrTax	-0.21	-0.39	-0.16	0.21	0.18	-0.19	0.03	-0.03	0.01	0.29	-0.3	0.02	
BrrwrPct	0.13	0.1	-0.25	-0.09	0.42	-0.19	-0.05	0.19	-0.01	-0.14	-0.21	-0.05	
ClmbrPct	0.23	-0.09	-0.24	-0.17	0.07	0.18	-0.48	0.35	-0.4	0.1	0.3	-0.52	
ClngrPct	0.24	0.12	0.37	0.08	-0.08	-0.09	-0.02	0.14	-0.18	0.03	-0.11	-0.01	
SprwlPct	-0.21	-0.04	-0.11	0.07	0.14	-0.18	0.11	-0.19	0.26	0.12	-0.01	0.14	
SwmmrPct	-0.29	-0.32	-0.18	0.02	-0.12	0.1	0.38	-0.28	0.45	0.1	-0.31	0.21	
BeckBI	0.29	-0.11	-0.05	0.23	0.16	0.03	-0.61	0.51	-0.55	0.31	0.17	-0.51	SUMMER
HBI	-0.4	-0.03	-0.19	-0.16	-0.1	0.06	0.29	-0.25	0.23	-0.14	-0.09	0.17	
NCBI	-0.44	0.01	0.04	-0.1	-0.14	-0.05	0.52	-0.44	0.43	-0.08	-0.23	0.48	SUMMER
IntolTax	0.28	-0.1	-0.02	0.24	0.16	0.06	-0.6	0.49	-0.53	0.29	0.14	-0.43	SUMMER
pt_Intol	0.36	-0.05	0.04	0.21	0.09	0.14	-0.45	0.36	-0.4	0.24	0.09	-0.32	SUMMER
pt_Toler	-0.27	-0.13	-0.02	0.12	0	-0.17	0.29	-0.32	0.34	-0.13	-0.14	0.37	
TolerTax	-0.25	-0.15	-0.06	0.29	0.11	-0.28	0.06	-0.14	0.1	0.06	-0.1	0.16	
PercBaetEph	-0.29	-0.02	-0.17	-0.41	-0.19	0.38	0.15	-0.27	0.2	-0.12	0.14	-0.03	
PercIntol	0.38	0.01	0.17	0.33	0.13	-0.07	-0.37	0.33	-0.33	0.2	0.01	-0.21	SUMMER
PercTol	-0.25	-0.08	-0.09	0.19	-0.01	-0.22	-0.02	0.05	-0.19	0.02	-0.23	-0.01	
PercHydr2Tri	0.2	0.16	0.2	-0.07	-0.33	0.04	0.37	-0.1	0.13	-0.27	-0.03	0.21	
PercHydr2EPT	0.29	0.38	0.42	0.08	-0.33	-0.07	0.26	-0.02	-0.04	-0.25	-0.12	0.19	
TIVsed	-0.1	-0.06	-0.13	-0.11	-0.02	-0.01	0.15	-0.02	0.13	-0.11	-0.01	0.05	
nt_DecrMtnTrn	0.18	-0.22	-0.18	0.13	0.26	0.34	-0.49	0.36	-0.24	0.26	0.26	-0.44	SUMMER
pi_DecrMtnTrn	0.21	-0.12	-0.06	0.33	0.36	-0.26	-0.38	0.39	-0.3	0.24	-0.12	-0.39	SUMMER

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
nt_IncrMtnTrn	-0.14	-0.11	0.05	0.29	0.13	-0.34	0.15	0	0.06	0.05	-0.34	0.22	
pi_IncrMtnTrn	-0.11	-0.1	0.02	0.26	0.03	-0.26	0.13	0	0.05	0.05	-0.23	0.19	
nt_DecrPlains	0.16	-0.25	-0.11	0.1	0.05	0.27	-0.54	0.41	-0.4	0.34	0.19	-0.43	
pi_DecrPlains	0.42	0.08	0.06	0.06	-0.07	0.17	-0.27	0.21	-0.12	0.2	0.23	-0.27	
nt_IncrPlains	-0.12	-0.07	-0.09	0.16	0.14	-0.3	0.11	-0.06	0.05	0.03	-0.1	0.27	
pi_IncrPlains	-0.05	-0.1	-0.19	0.11	0.02	-0.09	0.06	-0.03	0.05	-0.02	0.01	0.16	
CllctPctFAC	-0.26	-0.34	-0.41	-0.03	0.12	-0.02	-0.13	-0.01	-0.03	-0.03	0	-0.2	
FiltrPctFAC	0.06	0.15	0.27	-0.18	-0.16	0.23	0.2	-0.04	0.13	-0.02	-0.06	0.16	
PredPctFAC	0.23	0.11	0.01	0.28	0.24	-0.45	0.1	0.07	-0.04	0.02	-0.34	0.16	
ScrapPctFAC	-0.06	-0.03	-0.19	0.1	0.09	0.06	-0.22	0.13	-0.13	0.19	0.1	-0.33	
ShredPctFAC	-0.06	-0.23	-0.16	0.29	0.23	-0.1	-0.27	0.08	-0.06	0.36	0.04	-0.21	
CllctTaxFAC	-0.17	-0.27	-0.2	0.18	0.24	-0.26	-0.29	0.26	-0.27	0.34	-0.1	-0.33	
FiltrTaxFAC	-0.04	-0.03	0.13	0.13	-0.19	0.19	-0.18	0.15	-0.21	0.11	0.04	-0.08	
PredTaxFAC	0.14	0.06	-0.08	0.28	0.11	-0.17	-0.16	0.08	-0.13	0.04	0.11	-0.02	
ScrapTaxFAC	0.04	-0.09	0.17	0.22	0.07	-0.18	-0.29	0.36	-0.37	0.22	-0.14	-0.09	
ShredTaxFAC	0.03	-0.25	-0.31	0.21	0.2	-0.09	-0.4	0.37	-0.27	0.45	-0.08	-0.42	
GathFiltTaxFAC	-0.12	-0.26	-0.18	0.2	0.19	-0.27	-0.31	0.29	-0.32	0.35	-0.09	-0.33	
PredScrapTaxFAC	0.18	0.05	0.06	0.3	0.15	-0.25	-0.26	0.24	-0.26	0.13	0.01	-0.06	
PredShrTaxFAC	0.12	-0.09	-0.19	0.29	0.13	-0.1	-0.3	0.17	-0.2	0.19	0.16	-0.17	

Table C-2. Spearman correlation coefficients for reference metrics and environmental variables in Biotype 2, showing metric adjustment factors in the 'Adjust' column. Red font symbolizes significance of the correlation. Yellow highlights emphasize stronger correlation coefficients.

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
InsectTax	-0.07	-0.04	0.04	-0.14	-0.03	-0.05	0.16	-0.02	-0.25	-0.14	-0.11	0.1	
pt_noninsect	-0.13	-0.16	-0.2	-0.11	0.11	0.14	-0.03	0.19	0.01	0.13	-0.07	-0.03	
EPTTax	0.42	0.19	0.26	0.1	-0.15	-0.08	0.06	-0.16	-0.13	-0.22	-0.3	0.13	JulianDay
pt_EPT	0.63	0.28	0.28	0.25	-0.16	-0.09	-0.1	-0.17	0.08	-0.13	-0.23	-0.03	JulianDay
EphemTax	0.24	0.13	0.24	0	-0.17	-0.29	-0.2	0.07	-0.18	-0.06	-0.49	-0.03	
PlecoTax	0.55	0.07	0.04	0.14	-0.07	-0.05	-0.25	0.13	0	0.01	-0.23	-0.16	JulianDay
TrichTax	0.16	0.08	0.13	0.05	-0.02	0.11	0.4	-0.36	-0.05	-0.33	-0.04	0.33	
ColeoTax	0.07	-0.04	0.04	0.08	0.05	-0.09	0.57	-0.52	0.17	-0.37	-0.13	0.51	Summer
DipTax	-0.56	-0.16	-0.13	-0.27	0.1	0.05	0.02	0.19	-0.24	0.09	0.16	-0.04	JulianDay
pt_Diptera	-0.63	-0.14	-0.14	-0.24	0.1	0.08	-0.04	0.21	-0.17	0.15	0.3	-0.12	JulianDay
ChiroTax	-0.52	-0.14	-0.11	-0.17	0.08	0.08	-0.03	0.21	-0.14	0.23	0.19	-0.14	JulianDay
CrMolTax	-0.51	-0.1	-0.19	-0.04	0.23	0.13	0.16	-0.06	0.04	0.04	0.06	0.17	
OligoTax	0.3	0.02	0.02	-0.06	0.07	0.12	0.1	-0.03	0.07	-0.25	-0.16	0.13	
TanytTax	-0.07	0.08	0.11	-0.17	-0.01	0.13	0.03	-0.16	0.05	0.18	0.3	0.09	
Shan_2	-0.29	-0.11	-0.1	-0.15	-0.02	-0.02	-0.27	0.41	-0.34	0.18	0	-0.29	
Evenness	-0.29	-0.09	-0.08	-0.14	-0.03	-0.03	-0.29	0.43	-0.35	0.22	0.03	-0.31	
D_Mg	-0.26	-0.14	-0.08	-0.21	0.03	-0.04	0.12	0.11	-0.3	-0.01	-0.07	0.06	
D	0.27	0.1	0.07	0.1	-0.01	0.04	0.32	-0.41	0.28	-0.19	-0.03	0.33	
PercDom01	0.18	0.12	0.05	0.08	0.01	0.1	0.33	-0.38	0.2	-0.21	0.02	0.31	
EphemPct	0.14	0.22	0.32	-0.11	-0.26	-0.25	-0.01	-0.12	-0.06	0.03	-0.06	-0.07	
PlecoPct	0.33	-0.06	-0.22	0.11	0	0.13	-0.3	0.21	0.06	0.09	0.05	-0.27	Summer
TrichPct	0.23	0.08	0.09	0.09	-0.04	0.29	0.09	-0.19	-0.03	-0.12	0.12	-0.04	
NonInPct	-0.08	-0.29	-0.31	0.09	0.24	0.23	-0.33	0.5	-0.09	0.13	-0.11	-0.35	Elevation

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
ColeoPct	0.19	-0.03	0.04	0.12	0.14	-0.06	0.52	-0.3	0.03	-0.5	-0.13	0.39	Summer
OdonPct													
DipPct	-0.43	0	-0.05	-0.17	0.06	-0.08	-0.16	0.24	-0.07	0.23	0.07	0.03	JulianDay
ChiroPct	-0.36	-0.06	-0.11	-0.05	0.08	0.01	-0.24	0.3	-0.02	0.32	0.03	-0.1	
CrCh2ChiPct	-0.09	0.2	0.01	-0.17	-0.1	0	0.36	-0.19	-0.01	-0.12	0	0.29	
Orth2ChiPct	-0.06	-0.03	-0.07	0.22	0.01	0.04	0.08	-0.02	0.11	0.02	-0.04	-0.09	
TanytPct	-0.01	0.11	0.23	-0.13	0.05	-0.08	-0.17	0.06	0.06	0.21	0.04	0.08	
Tnyt2ChiPct	0.24	0.15	0.27	-0.11	0.02	-0.13	-0.08	-0.06	0.09	0.07	0.01	0.15	
OligoPct	0.39	-0.03	0.02	0.1	0.07	0.16	-0.03	0.1	0.11	-0.21	-0.14	-0.08	
CrMolPct	-0.5	-0.1	-0.18	-0.03	0.22	0.12	0.13	-0.02	0.02	0.04	0.05	0.14	
GastrPct	-0.12	-0.18	-0.09	0.04	0.17	0.04	0.2	-0.2	0.2	-0.18	-0.01	0.2	
IsoPct	-0.17	0.07	-0.15	-0.17	-0.2	0.15	0.16	-0.12	0.04	0.06	0.08	0.14	
AmphPct													
BivalPct	-0.38	-0.11	-0.09	0.08	0.34	0.08	0.13	-0.13	0.27	0.06	0	0.17	
CIlctTax	-0.53	-0.24	-0.29	-0.27	0.09	0.03	0.08	0.14	-0.22	0.1	0.1	0.16	JulianDay
FiltrTax	-0.1	-0.04	-0.1	0.08	0.14	0.13	-0.01	0	0.15	0.2	0.02	-0.03	
PredTax	0.02	0.07	0.2	-0.06	-0.26	-0.14	-0.13	0.08	-0.21	0.11	-0.19	-0.16	
ScrapTax	-0.08	-0.04	-0.05	0.06	-0.16	-0.03	-0.05	-0.07	-0.05	0.08	-0.01	-0.16	
ShredTax	0.26	-0.07	-0.04	0.06	0.15	0.09	0.15	-0.02	0.06	-0.23	-0.07	0.08	
CIlctPct	-0.58	-0.25	-0.26	-0.23	0.1	-0.06	0.13	0.1	-0.2	0.09	0.1	0.14	
FiltrPct	-0.1	-0.04	-0.1	0.08	0.14	0.13	0	0	0.15	0.2	0.02	-0.03	
PredPct	0.03	0.15	0.15	-0.02	-0.36	0.01	-0.29	0.05	-0.09	0.19	0.1	-0.26	
ScrapPct	-0.11	0.06	0.01	0	-0.36	-0.23	-0.18	0.05	-0.12	0.26	0.12	-0.17	
ShredPct	0.3	-0.09	-0.25	0.01	0.11	0.23	-0.18	0.22	-0.02	0.08	0.01	-0.22	
BrrwrTax	-0.23	-0.17	-0.16	-0.15	-0.09	-0.16	0.12	0.12	-0.35	-0.04	0.02	-0.02	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
ClmbrTax	0.1	0.06	0.19	0.01	0.07	0.13	0.11	-0.18	0.09	-0.16	-0.1	0.22	
ClngrTax	0.13	0.17	0.27	0.04	-0.14	-0.1	0.27	-0.23	-0.23	-0.33	-0.16	0.12	
SprwlTax	-0.07	-0.2	-0.16	-0.12	0.11	0.18	-0.13	0.21	-0.04	0.07	0.04	-0.13	
SwmmrTax	-0.17	-0.15	-0.2	-0.04	0.14	-0.04	0.16	0.03	-0.03	-0.05	-0.06	0.24	
BrrwrPct	-0.29	-0.22	-0.29	-0.16	-0.12	-0.04	0	0.17	-0.27	0.23	0.05	-0.07	
ClmbrPct	0.11	0.16	0.26	-0.04	0	0	0	-0.12	0.07	-0.1	-0.18	0.06	
ClngrPct	0.26	0.06	0.08	0.16	-0.11	0.07	0.05	-0.19	-0.02	-0.22	-0.03	-0.05	
SprwlPct	-0.05	-0.04	-0.18	-0.03	0.09	0.01	-0.37	0.39	-0.08	0.17	0	-0.22	
SwmmrPct	-0.27	-0.17	-0.16	0.03	0.13	-0.12	0.16	-0.01	0	0.04	-0.02	0.24	
BeckBI	0.43	0.06	0.19	0.12	-0.03	-0.06	0.18	-0.18	-0.11	-0.4	-0.29	0.19	
HBI	-0.41	-0.02	-0.05	-0.18	0.11	0	0.13	0.08	-0.07	0	0.07	0.19	
NCBI	-0.09	0.07	0.09	0.05	0.14	0.11	-0.12	0.13	0.1	-0.07	-0.08	-0.11	
IntolTax	0.37	0.05	0.19	0.11	0.01	-0.06	0.05	-0.07	-0.11	-0.32	-0.3	0.07	
pt_Intol	0.59	0.15	0.23	0.25	-0.01	-0.07	-0.12	-0.09	0.1	-0.23	-0.23	-0.05	JulianDay
pt_Toler	-0.36	0.08	0.02	-0.27	0.15	0.13	0.06	0.1	-0.03	0.1	0.32	0.06	
TolerTax	-0.41	0.06	0.04	-0.32	0.12	0.12	0.06	0.12	-0.11	0.09	0.22	0.08	
PercBaetEph	-0.12	0.15	0.07	-0.21	0.08	0.05	0.19	-0.13	-0.04	-0.01	0.08	0.13	
PercIntol	0.4	-0.01	0	0.15	-0.08	0.1	-0.31	0.12	0.07	0.1	-0.09	-0.35	
PercTol	-0.2	-0.09	-0.1	-0.01	0.11	0.17	-0.19	0.31	-0.02	0.18	-0.04	-0.09	
PercHydr2Tri	0.3	0.07	0.24	0.03	-0.06	-0.21	0.12	-0.17	0.02	-0.14	-0.16	0.31	
PercHydr2EPT	0.28	0.07	0.22	0.02	-0.05	-0.15	0.16	-0.24	0.04	-0.16	-0.13	0.34	
TIVsed	-0.14	0	-0.1	0.05	0.26	0.01	0.26	-0.15	0.08	-0.2	0.09	0.32	
nt_DecrMtnTrn	0.11	-0.04	0.08	0.09	0.1	-0.12	0.12	0.03	-0.13	-0.1	-0.21	0.02	
pi_DecrMtnTrn	0.18	0.07	0.07	0.02	-0.12	-0.05	0.02	-0.12	-0.05	-0.08	-0.01	0	
nt_IncrMtnTrn	-0.03	-0.21	-0.12	-0.22	0.25	0.19	-0.01	0.06	-0.03	-0.12	-0.05	0.08	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
pi_IncrMtnTrn	-0.02	-0.22	-0.12	-0.22	0.24	0.2	-0.01	0.06	-0.04	-0.14	-0.05	0.08	
CllctPctFAC	-0.03	-0.13	-0.05	0.1	0.3	0.11	0.25	-0.06	0.02	-0.28	-0.23	0.18	
FiltrPctFAC	-0.09	0.21	0.33	-0.09	0.08	-0.22	0.19	-0.19	-0.11	-0.25	-0.01	0.33	
PredPctFAC	0.01	0.15	0.14	0.01	-0.35	0.03	-0.31	0.08	-0.12	0.21	0.09	-0.29	
ScrapPctFAC	-0.23	-0.11	-0.11	0.15	-0.05	0.02	-0.04	-0.06	0.11	0.08	0.22	-0.13	
ShredPctFAC	0.26	-0.14	-0.32	0.18	0.09	0.14	-0.13	0.22	0.01	0.09	-0.08	-0.21	
CllctTaxFAC	-0.33	-0.24	-0.19	-0.18	0.19	0.03	0.18	0.07	-0.09	-0.05	-0.02	0.2	
FiltrTaxFAC	-0.08	0.23	0.33	-0.09	0.1	-0.18	0.43	-0.32	-0.09	-0.33	-0.02	0.47	
PredTaxFAC	0.01	0.11	0.25	-0.04	-0.25	-0.16	-0.08	0.03	-0.18	0.07	-0.17	-0.13	
ScrapTaxFAC	-0.31	-0.24	-0.31	-0.05	0.06	0.15	-0.02	0.09	-0.23	-0.02	0.02	-0.08	
ShredTaxFAC	0.2	-0.05	-0.05	0.17	0.12	0.07	0.23	-0.05	0.03	-0.2	-0.17	0.06	
GathFiltTaxFAC	-0.31	-0.12	-0.06	-0.22	0.19	-0.03	0.27	-0.02	-0.13	-0.14	0	0.33	
PredScrapTaxFAC	-0.1	-0.04	0.04	-0.04	-0.2	-0.05	-0.09	0.08	-0.26	0.09	-0.15	-0.2	
PredShrTaxFAC	0.15	0.04	0.17	-0.05	-0.16	-0.05	-0.05	0.05	-0.16	0.01	-0.16	-0.1	

Table C-3. Spearman correlation coefficients for reference metrics and environmental variables in Biotype 3, showing metric adjustment factors in the 'Adjust' column. Red font symbolizes significance of the correlation. Yellow highlights emphasize stronger correlation coefficients.

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
InsectTax	-0.36	0.04	0.31	-0.11	0.54	-0.17	0.46	-0.27	0.39	0.28	-0.31	0.11	
pt_noninsect	-0.05	-0.39	-0.36	0.27	0.34	0.01	0.07	-0.12	0.34	0.22	-0.04	-0.01	
EPTTax	0.31	-0.09	0.25	0	-0.12	0.11	0.05	-0.06	-0.23	-0.01	0.15	-0.06	
pt_EPT	0.47	-0.03	0.18	0	-0.41	0.16	-0.15	0.06	-0.42	-0.22	0.3	-0.11	Julian Day
EphemTax	0.16	-0.06	0.15	0.03	0.08	-0.14	0.15	0.04	-0.05	0.1	-0.22	-0.11	
PlecoTax	0.17	-0.4	-0.1	-0.14	-0.24	0.25	-0.35	0.27	-0.37	0.11	0.09	-0.18	
TrichTax	0.25	0	0.27	-0.04	-0.13	0.25	0.07	-0.13	-0.13	-0.11	0.29	0.05	
ColeoTax	0.04	-0.37	-0.37	-0.29	0.24	-0.05	0.04	-0.15	0.11	-0.1	-0.06	0.11	
DipTax	-0.66	0.16	0.28	0.03	0.59	-0.03	0.43	-0.29	0.5	0.44	-0.29	0.06	Julian Day
pt_Diptera	-0.66	0.28	0.12	-0.08	0.27	0.07	0.18	-0.04	0.29	0.32	-0.21	-0.04	Julian Day
ChiroTax	-0.67	0.11	0.22	0	0.65	-0.06	0.5	-0.39	0.57	0.42	-0.24	0.08	Julian Day
CrMolTax	-0.01	-0.43	-0.22	0.41	0.4	0.03	0.27	-0.22	0.39	0.25	-0.03	0.17	
OligoTax	-0.11	-0.16	-0.27	-0.33	-0.12	0.2	-0.33	0.34	0.13	0.02	-0.07	-0.23	
TanytTax	-0.53	0.12	0.2	-0.05	0.53	0.06	0.43	-0.29	0.54	0.36	-0.23	0.15	
Shan_2	-0.06	-0.23	0.12	0.17	0.56	-0.03	0.39	-0.2	0.33	0.47	-0.26	0.14	
Evenness	-0.03	-0.2	0.15	0.22	0.55	-0.05	0.41	-0.23	0.33	0.45	-0.25	0.16	
D_Mg	-0.38	-0.04	0.29	0.06	0.66	-0.25	0.59	-0.42	0.52	0.37	-0.31	0.15	
D	-0.12	0.31	0	-0.16	-0.47	-0.05	-0.31	0.18	-0.25	-0.4	0.09	-0.11	
PercDom01	-0.09	0.31	0.05	-0.11	-0.44	-0.03	-0.31	0.11	-0.26	-0.35	0.11	-0.1	
EPTPct	0.4	0.09	0.27	-0.12	-0.25	0.03	0.08	-0.09	-0.17	-0.38	0.05	-0.09	
EphemPct	0.34	0.21	0.27	-0.1	-0.1	-0.19	0.2	-0.03	-0.04	-0.3	-0.25	-0.14	
PlecoPct	0.16	-0.39	-0.1	-0.13	-0.25	0.25	-0.35	0.27	-0.37	0.11	0.1	-0.18	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
TrichPct	0.32	-0.04	0.25	0.07	-0.24	0.29	-0.08	-0.13	-0.24	-0.17	0.37	0	
NonInPct	-0.15	-0.45	-0.34	0.26	0.43	-0.01	0.16	-0.13	0.37	0.39	0.02	0.07	
ColeoPct	0.46	-0.21	-0.15	-0.07	-0.18	0.01	-0.24	-0.1	-0.46	-0.19	0.18	0.14	
OdonPct	0.15	0.03	0.25	-0.08	0	-0.31	0.13	0.07	0.01	-0.19	-0.31	0.24	
DipPct	-0.66	0.05	-0.17	-0.17	0.27	0.13	-0.01	0.14	0.31	0.44	-0.2	-0.02	Julian Day
ChiroPct	-0.59	-0.04	0	-0.07	0.51	0.11	0.27	-0.14	0.49	0.51	-0.12	0.06	Julian Day
CrCh2ChiPct	0.2	0.42	0.5	0.35	-0.03	-0.42	0.24	-0.08	-0.08	-0.25	-0.2	0.12	
Orth2ChiPct	0.03	-0.01	-0.12	-0.21	-0.17	0.13	-0.4	0.25	-0.35	0.02	-0.16	-0.27	
TanytPct	-0.46	-0.01	0.15	0.2	0.64	-0.02	0.57	-0.46	0.61	0.4	-0.13	0.15	
Tnyt2ChiPct	-0.36	0.04	0.15	0.22	0.49	0.01	0.59	-0.44	0.57	0.2	-0.08	0.14	
OligoPct	-0.14	-0.23	-0.33	-0.41	-0.03	0.27	-0.36	0.35	0.13	0.14	-0.06	-0.21	
CrMolPct	-0.07	-0.38	-0.25	0.35	0.42	0.13	0.29	-0.28	0.49	0.23	0.07	0.21	
GastrPct	-0.35	-0.24	-0.13	0.13	0.64	-0.01	0.58	-0.49	0.61	0.17	0.06	0.21	
CorbPct												0.16	
IsoPct	0.28	-0.1	-0.17	0.23	-0.12	-0.07	-0.33	0.33	0.01	0.25	-0.04	-0.04	
AmphPct	0.07	-0.32	-0.2	0.46	0.27	0.11	0.2	-0.29	0.37	0.24	0.04	0.14	
BivalPct	0.16	-0.46	-0.36	0.26	0.17	-0.05	-0.08	0.13	0.08	0.4	-0.26	0.16	
CllctTax	-0.12	-0.17	0.02	0.44	0.44	-0.17	0.42	-0.47	0.54	0.05	-0.13	-0.08	
FiltrTax	0.35	-0.22	-0.06	0.18	0.06	-0.23	0.06	-0.04	-0.2	0.09	-0.04	0.19	
PredTax	-0.02	-0.02	0.01	-0.08	0.4	-0.3	0.26	-0.27	0.28	0.2	-0.16	0.23	
ScrapTax	-0.37	0.02	0.24	0.14	0.48	-0.05	0.61	-0.41	0.39	0.07	-0.06	0.13	
ShredTax	-0.02	0.1	0.41	0.23	0.16	0.04	0.26	-0.08	0.2	0.05	-0.17	0.14	
CllctPct	-0.14	-0.06	0.21	0.39	0.46	-0.23	0.51	-0.53	0.5	-0.01	-0.2	0.13	
FiltrPct	0.35	-0.18	-0.04	0.21	0.04	-0.24	0.06	-0.07	-0.2	0.09	0	0.19	
PredPct	-0.07	-0.22	-0.25	-0.26	0.28	-0.05	0.02	-0.14	0.14	0.22	0.12	0.15	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
ScrapPct	-0.33	-0.08	0.1	0.15	0.57	-0.03	0.68	-0.59	0.58	0.03	0.15	0.16	
ShredPct	-0.03	0.2	0.37	0.27	0	0.09	0.11	0.02	0.09	0.05	-0.1	0.15	
BrrwrTax	-0.37	-0.09	0.24	0.12	0.62	-0.34	0.62	-0.47	0.48	0.27	-0.21	0.21	
ClmbrTax	-0.14	-0.07	0.15	0.16	0.38	-0.29	0.45	-0.2	0.3	0.15	-0.27	0.13	
ClngrTax	0.21	0.09	0.24	-0.06	-0.25	0.09	-0.21	0.2	-0.33	-0.02	-0.11	-0.03	
SprwlTax	-0.67	0.09	0.13	-0.07	0.49	0.23	0.34	-0.21	0.55	0.43	-0.09	-0.01	Julian Day
SwmmrTax	0.09	-0.16	-0.01	-0.05	0.1	0.01	0.19	-0.06	0.25	-0.03	-0.09	0.09	
BrrwrPct	-0.32	-0.18	0.01	0.19	0.42	-0.04	0.22	-0.27	0.31	0.41	-0.05	0.18	
ClmbrPct	-0.26	-0.18	0.09	-0.16	0.47	-0.1	0.54	-0.26	0.33	0.15	-0.09	0	
ClngrPct	0.21	0.24	0.1	0.02	-0.48	0.1	-0.48	0.25	-0.45	-0.25	-0.02	-0.11	
SprwlPct	-0.46	0.1	0.07	-0.03	0.28	0.14	0.21	-0.07	0.45	0.32	-0.23	0.04	
SwmmrPct	0.19	-0.08	-0.04	-0.21	-0.07	0.08	0	0.05	0.04	-0.18	-0.17	-0.03	
BeckBI	0.15	-0.04	0.31	0.04	-0.12	-0.02	0.07	-0.1	-0.15	0.02	0.16	-0.07	
HBI	-0.46	-0.14	-0.23	0	0.51	-0.08	0.24	-0.14	0.59	0.38	-0.11	0.01	
NCBI	-0.13	0.23	0.3	-0.04	0.19	-0.23	0.25	-0.36	0.23	-0.33	-0.15	0.18	
IntolTax	0.02	-0.11	0.31	0.12	-0.06	-0.01	0.1	-0.15	-0.06	0.08	0.13	0.07	
pt_Intol	0.14	-0.02	0.33	0.13	-0.23	0.01	0	-0.1	-0.2	-0.09	0.22	0.04	
pt_Toler	-0.36	-0.21	-0.18	0.03	0.64	-0.23	0.51	-0.37	0.64	0.27	-0.13	0.2	Summer
TolerTax	-0.39	-0.27	-0.04	0.12	0.77	-0.21	0.6	-0.45	0.7	0.43	-0.19	0.24	Summer
PercBaetEph	0.31	0.02	-0.02	0.01	-0.52	-0.06	-0.5	0.42	-0.47	-0.19	-0.16	-0.27	
PercIntol	0.36	0.03	0.48	0.25	-0.2	0.03	0.01	-0.2	-0.3	-0.08	0.3	0.07	
PercTol	-0.37	-0.12	0	0.14	0.56	-0.17	0.38	-0.23	0.6	0.44	-0.11	0.06	
PercHydr2Tri	0.33	0.09	0.04	-0.01	-0.52	0.14	-0.42	0.35	-0.49	-0.23	0.11	-0.17	
PercHydr2EPT	0.31	0.02	0.16	0.05	-0.43	0.23	-0.29	0.1	-0.41	-0.23	0.31	-0.03	
TIVsed	0.8	0.32	0.4	0.2	0.4	-0.4	-0.4	-1	-1	0.6	1	0.54	

Metric	Julian Day	Order	Catch Area	Lat	Long	NHD Slope	Sum Temp	Elev Site	Wint Temp	Log XP_PT	SQRT TOPO	S2 02_11	Adjust
nt_DecrPlains	-0.62	0.03	0.08	-0.1	0.57	0.03	0.43	-0.26	0.62	0.39	-0.23	0.24	
pi_DecrPlains	0.14	0.1	0.37	0.05	0.22	-0.17	0.4	-0.53	0.25	-0.22	0.21	0.54	
nt_IncrPlains	0.1	-0.16	-0.12	-0.1	-0.11	0.05	-0.28	0.25	0.02	0.2	-0.11	-0.16	
pi_IncrPlains	-0.07	-0.19	-0.27	-0.25	-0.06	0.29	-0.29	0.26	0.06	0.25	-0.06	-0.2	
CllctPctFAC	0.21	-0.07	0.13	0.06	0.32	-0.41	0.42	-0.2	0.26	0.02	-0.22	0.31	
FiltrPctFAC	0	0.18	0.1	-0.12	-0.17	0.13	-0.15	-0.13	-0.22	-0.33	0.11	0.11	
PredPctFAC	-0.04	-0.18	-0.29	-0.25	0.21	-0.01	0	-0.09	0.18	0.18	0.13	0.12	
ScrapPctFAC	-0.2	-0.3	-0.26	-0.09	0.14	0.42	-0.09	-0.12	0.03	0.23	0.39	0.16	
ShredPctFAC	-0.06	0.16	0.31	0.06	0.06	0.11	-0.04	0.02	-0.18	0.03	-0.04	0.14	
CllctTaxFAC	-0.48	-0.18	0.01	-0.06	0.6	-0.03	0.34	-0.21	0.46	0.45	-0.28	0.3	
FiltrTaxFAC	0.2	-0.01	0.05	0.04	-0.21	0.3	-0.23	0.2	-0.17	0.11	0.04	-0.04	
PredTaxFAC	0.01	-0.11	-0.04	-0.07	0.41	-0.21	0.23	-0.17	0.29	0.26	-0.19	0.33	
ScrapTaxFAC	-0.49	-0.15	0.04	-0.05	0.33	0.34	0.15	-0.16	0.13	0.38	0.11	-0.01	
ShredTaxFAC	-0.11	0.24	0.49	0.28	0.08	-0.07	0.07	0.03	-0.09	0.06	-0.33	0.02	
GathFiltTaxFAC	-0.46	-0.22	-0.04	-0.07	0.54	0.09	0.25	-0.14	0.42	0.52	-0.28	0.26	
PredScrapTaxFAC	-0.27	-0.23	-0.1	-0.18	0.57	-0.05	0.31	-0.27	0.41	0.4	-0.1	0.33	
PredShrTaxFAC	-0.03	-0.01	0.16	-0.02	0.46	-0.24	0.36	-0.28	0.34	0.21	-0.23	0.4	

Appendix D

Macroinvertebrate Metric Performance

All statistics are based on calibration data only, except for the 5th and 95th percentiles of all data.

Header	Description
Adjust	Adjustment factor (STemp = Summer temperature, JulianD = Julian day)
MeanRef	Average metric value in reference sites
RefSD	Standard deviation of reference metric values
RefCV	Reference metric coefficient of variation
Ref25th	Reference 25th percentile
Ref75th	Reference 75th percentile
StrsDE25	Discrimination efficiency for decreasing trends (cells shaded blue emphasize high DE)
StrsDE75	Discrimination efficiency for increasing trends (cells shaded blue emphasize high DE)
Trend	Metric trend with increasing stress
Z-score	Z-score
95th BT_	95th percentile of all metric values in the Biotype
5th BT_	5th percentile of all metric values in the Biotype

Table D-1. Biotype1 Metric Performance Statistics.

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT1	5th BT1
Richness metrics												
TotalTax		28.9	8.5	29.2	21.5	33.5	43.2	13.5	Decr	0.7	40	15
InsectTax		26	7.3	27.9	20.5	31.5	70.3	5.4	Decr	1	35	12
aInsectTax	STemp	65	16.3	25.2	51.8	75.5	56.8	10.8	Decr	0.9		
pt_noninsect		9.6	6.4	67	4.8	13.1	5.4	73	Incr	-1.6	30	0
EPTTax		12.1	5.2	42.7	8.5	15	64.9	2.7	Decr	1.1	19	3
aEPTTax	STemp	61.4	19	31	52.2	75	78.4	13.5	Decr	1.1		
pt_EPT		42	13.8	32.8	32.8	51	59.5	5.4	Decr	1.1	60	13.3
EphemTax		5.1	2.7	51.7	3	7	40.5	5.4	Decr	0.8	9	1
PlecoTax		2.9	1.9	63.8	2	4.5	78.4	0	Decr	1.1	6	0
aPlecoTax	STemp	48.3	28.8	59.6	31.5	68.6	86.5	0	Decr	1.2		
TrichTax		4	1.9	46.8	3	5	56.8	8.1	Decr	0.7	8	1
aTrichTax	STemp	54.3	21.9	40.3	46	69	62.2	16.2	Decr	0.7		
ColeoTax		2.8	1.5	54.6	2	3	43.2	8.1	Decr	0.7	5	0
DipTax		10.4	3.9	37.5	7	13.5	27	16.2		0.1	18	5
pt_Diptera		35.6	8.3	23.2	29.2	42	24.3	48.6	Incr	-0.8	60	20
ChiroTax		6.6	3.6	53.9	4	9	10.8	32.4	Incr	-0.3	15	3
CrMolTax		0.6	0.9	154.7	0	1	0	32.4	Incr	-0.7	3	0
OligoTax		0.9	1.1	121.8	0	1	0	45.9	Incr	-0.6	4	0
TanytTax		1.1	0.9	76.9	0.5	2	37.8	10.8	Decr	0.1	3	0
Shan_2		3.4	0.8	23.3	2.9	3.9	48.6	8.1	Decr	0.7	4.3	1.9
Evenness		0.4	0.1	23.1	0.4	0.5	48.6	8.1	Decr	0.6	0.5	0.2
D_Mg		5.1	1.5	29.1	3.9	5.9	51.4	13.5	Decr	0.7	7	2.6
D		0.2	0.1	75.8	0.1	0.2	10.8	37.8	Incr	-0.5	0.4	0.1
PercDom01		33.6	16.4	49	22.3	39.3	10.8	45.9	Incr	-0.5	63.3	15.3
Composition metrics												
EPTPct		56.8	25	44	39.3	75.5	54.1	10.8	Decr	0.9	86.1	4.5

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT1	5th BT1
pEPTnoB		37.5	24.2	64.5	19.3	50.6	56.8	8.1	Decr	0.8	72.9	1
EphemPct		32.6	21.8	67	16.2	40.7	54.1	16.2	Decr	0.5	66.9	1
PlecoPct		6.6	8.3	126.8	1.4	8.5	78.4	0	Decr	0.7	15.6	0
TrichPct		17.6	18.5	105.1	5.5	25.7	56.8	16.2	Decr	0.3	49.5	0.3
NonInPct		4.9	9.5	194.6	0.6	3.4	5.4	67.6	Incr	-1.8	48.3	0
ColeoPct		16.4	14.8	90.5	6.4	21.2	75.7	5.4	Decr	0.7	38.4	0
PlCoPct		22.9	16.3	71	11.6	33	81.1	5.4	Decr	0.9		
OdonPct		0.4	1.7	393.3	0	0	0	27		-0.6	1.5	0
DipPct		21	15.9	75.5	9.4	28	16.2	62.2	Incr	-0.9	72.7	4.7
ChiroPct		12.3	15.1	122.6	2.8	16.4	10.8	54.1	Incr	-0.7	56.2	1.7
CrCh2ChiPct		12.2	23.6	194.2	0	8.8	0	16.2		0.3	47.3	0
Orth2ChiPct		10.7	16.3	152.3	0	16.2	0	32.4	Incr	-0.3	62.1	0
TanytPct		1.7	2.6	149.6	0.1	2.3	37.8	24.3	Decr	-1	14.4	0
Tnyt2ChiPct		16.5	19.7	119.3	0.6	23.6	40.5	16.2	Decr	0.2	59.4	0
OligoPct		1.8	7.8	423.5	0	0.7	0	62.2	Incr	-0.9	24	0
CrMolPct		0.9	2.1	237.8	0	0.3	0	54.1	Incr	-3.3	13	0
GastrPct		0.3	1.1	381.5	0	0	0	45.9	Incr	-3.1	3.5	0
CorbPct		0	0		0	0	0	0			0	0
IsoPct		0	0.2	591.6	0	0	0	10.8		-0.3	0.9	0
AmphPct		0	0.1	295.8	0	0	0	18.9		-27.6	2.7	0
BivalPct		0.1	0.5	363.9	0	0	0	5.4		0.1	0.9	0
Habit metrics												
BrrwrTax		1.9	1.3	68	1	2.5	24.3	16.2		0.4	4	0
ClmbrTax		2	1.1	56.9	1	3	2.7	5.4		0	4	1
ClngrTax		12.3	4.2	33.9	9	15.5	67.6	2.7	Decr	1.3	19	3
aClngrTax	STemp	67.4	18.4	27.2	52.2	80.5	75.7	0	Decr	1.5		
SprwlTax		6	2.6	43	4	8	40.5	16.2	Decr	0.2	11	2
SwmmrTax		1.6	1	64.2	1	2	37.8	16.2	Decr	0.4	3	0

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT1	5th BT1
BrrwrPct		2.9	4.5	154.3	0.8	2.7	48.6	24.3	Decr	0	9.1	0
ClmbrPct		12.6	14.4	114.7	1.6	14.5	18.9	35.1		-0.2	53.2	0.3
ClngrPct		48.7	19	39	37.7	60.6	54.1	13.5	Decr	0.8	79.2	6.9
SprwlPct		12.5	18.6	148.2	3.2	13.6	24.3	37.8	Incr	-0.1	37	1
SwmmrPct		8.8	13.2	149.3	0.8	10	54.1	13.5	Decr	0.4	21.2	0
Tolerance metrics												
BeckBI		20.7	10.4	50.2	12.5	27	62.2	2.7	Decr	1	35	3
aBeckBI	STemp	62.7	24.2	38.6	44.1	81.1	73	2.7	Decr	1.2		
HBI		4.1	1	25.2	3.7	4.7	0	83.8	Incr	-1.6	6.9	2.4
NCBI		2.8	1.1	39.8	2.1	3.3	8.1	56.8	Incr	-1.2	6	1.5
IntolTax		10.1	5.4	53.1	5.5	13	67.6	2.7	Decr	1.1	19	1
aIntolTax	STemp	60.1	24.9	41.4	48.2	79.5	78.4	5.4	Decr	1.3		
pt_Intol		34.1	15.2	44.5	24	43.8	73	5.4	Decr	1.2	57.1	4.2
apt_Intol	STemp	62.1	24.6	39.7	48.7	75.7	78.4	5.4	Decr	1.3		
pt_Toler		14	8	57.1	9.5	17.8	10.8	73	Incr	-1.6	38.5	4
TolerTax		3.9	2.2	55.7	2	5	2.7	43.2	Incr	-0.9	11	1
InMolTax		0	0		0	0	0	0			0	0
PercBaetEph		63.4	30.2	47.6	38.1	85.4	24.3	54.1	Incr	-0.1	100	0
PercIntol		24.9	18.4	74.1	13	34.7	86.5	2.7	Decr	1	64.7	0.3
aPercIntol	STemp	42.3	28.6	67.6	21.8	57.9	89.2	2.7	Decr	1.1		
PercTol		8.9	13	146	2	8.8	2.7	59.5	Incr	-1.6	54.6	0.7
PercHydr2Tri		52.3	33.8	64.8	21.9	78.9	35.1	35.1		0.1	100	0
PercHydr2EPT		19.1	19.7	103.4	3.5	36.4	37.8	18.9	Decr	0.1	67.1	0
TIVsed		5.4	1	18	4.7	5.9	4.8	38.1	Incr	-0.8	7.5	4
nt_DecrMtnTrn		6.2	3	48.2	3.5	8.5	75.7	2.7	Decr	1.3	9	0
ant_DecrMtnTrn	STemp	61.1	24.5	40.1	46.2	81.2	78.4	2.7	Decr	1.5		
pi_DecrMtnTrn		30.4	18.1	59.6	16.5	45.6	86.5	2.7	Decr	1.2	55.6	0
api_DecrMtnTrn	STemp	50.6	29.1	57.5	31.7	75.3	91.9	2.7	Decr	1.3		

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT1	5th BT1
nt_IncrMtnTrn		0.5	0.7	128	0	1	0	40.5	Incr	-1.3	3	0
pi_IncrMtnTrn		1.7	7.8	448.1	0	0.5	0	64.9	Incr	-1	23.9	0
nt_DecrMidElev		3.3	1.9	57.4	2	5	59.5	2.7	Decr	1.1	6	0
ant_DecrMidElev	STemp	54.6	28.9	52.9	39.3	78.4	73	2.7	Decr	1.1		
pi_DecrMidElev		8.9	9.6	108.8	2.6	12.1	73	8.1	Decr	0.6	33.4	0
api_DecrMidElev	STemp	32.9	32.6	99.1	7.7	50.6	70.3	8.1	Decr	0.7		
nt_IncrMidElev		1.1	1	90.4	0	1.5	0	64.9	Incr	-0.9	4	0
pi_IncrMidElev		2.3	8.2	355.5	0	1.2	0	73	Incr	-1.4	32.6	0
Feeding Group metrics												
CllctPctFAC		58.8	20.3	34.5	45.1	77.2	29.7	29.7		0	87	20.2
FiltrPctFAC		17.8	17.7	99.2	5.6	27	32.4	35.1		-0.3	56	0.7
PredPctFAC		11.3	7	62.3	5.2	15.4	56.8	10.8	Decr	0.5	31.5	1.7
ScrapPctFAC		21.8	13.1	60.2	11.3	29.9	62.2	18.9	Decr	0.5	44.3	1.2
ShredPctFAC		8	7.2	90.5	2.3	11.3	32.4	27		-0.3	35.3	0.6
CllctTaxFAC		13.1	4.9	37.6	9	17.5	21.6	13.5		0.2	21	6
FiltrTaxFAC		2.8	1.1	38.2	2	4	21.6	5.4		0.3	5	1
PredTaxFAC		7.7	3	38.8	6	9	64.9	5.4	Decr	0.9	12	2
ScrapTaxFAC		4.5	1.9	42	3	5.5	40.5	8.1	Decr	0.7	8	1
ShredTaxFAC		3.6	1.6	44.9	2	5	18.9	0		0.7	6	1
GathFiltTaxFAC		15.8	5.1	32.1	12	20	35.1	16.2		0.3	24	8
PredScrapTaxFAC		12.1	4.2	34.3	10	14	64.9	5.4	Decr	1	18	4
PredShrTaxFAC		9.8	3.2	32.8	8	12	62.2	0	Decr	1.2	14	3
CllctPct		38.2	20.4	53.4	22.4	47.8	18.9	32.4		-0.3	77.1	10.4
FiltrPct		17.6	17.7	100.5	5.1	25.8	40.5	32.4	Decr	-0.2	54.9	0.4
PredPct		10.4	6.8	65.8	5.1	13	56.8	13.5	Decr	0.5	29.8	1.4
ScrapPct		5	6.8	134.3	0.3	6.9	35.1	10.8	Decr	0	22	0
ShredPct		4.9	5.3	109.2	1.1	6.5	51.4	5.4	Decr	0.4	23.2	0
CllctTax		0.8	0.8	99.6	0	1	0	18.9		0.1	2	0

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT1	5th BT1
FiltrTax		2.8	1.1	38.2	2	4	21.6	5.4		0.3	5	1
PredTax		7.3	2.9	40.3	6	9	67.6	2.7	Decr	0.9	11	2
ScrapTax		1.8	1.5	82.2	1	3	27	5.4		0.5	4	0
ShredTax		2.5	1.2	48	2	3	64.9	0	Decr	1	5	0
PredScrapTax		9	3.8	42.2	7	10	62.2	5.4	Decr	0.9	14	3
PredShrTax		9.8	3.2	32.8	8	12	62.2	0	Decr	1.2	14	3

Table D-2. Biotype 2 Metric Performance Statistics.

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT2	5th BT2
Richness metrics												
TotalTax		33.6	6.3	18.8	29.3	37.8	93.8	0	Decr	2.2	42	16
InsectTax		30.4	5.7	18.9	26	34	100	0	Decr	2.5	38	13.6
pt_noninsect		9.3	5.6	60.4	5.6	12.4	6.3	35.1	Incr	-1.6	24.6	0
EPTTax		16.4	3.7	22.5	14.3	18.8	100	0	Decr	2.9	21	4
aEPTTax	JulianD	72	14.7	20.5	62.2	80	100	0	Decr	3.2		
pt_EPT		49.6	10.3	20.8	41.6	57	81.3	2.7	Decr	2.1	67.1	20
apt_EPT	JulianD	77.2	13.5	17.5	68.8	86.9	93.8	2.7	Decr	2.5		
EphemTax		6.7	1.6	23.9	6	7.8	81.3	0	Decr	2.2	10	2
PlecoTax		5	1.5	30.7	4	6	87.5	0	Decr	2.6	7	0
aPlecoTax	JulianD	69.1	19.6	28.3	55.9	85.7	87.5	0	Decr	2.9		
TrichTax		4.7	2.1	44.1	3	6	87.5	0	Decr	1.5	8	1
ColeoTax		1.6	1.3	84.4	1	2	43.8	5.4	Decr	0.5	4	0
aColeoTax	STemp	45.1	26.8	59.4	27.6	61	75	8.1	Decr	0.8		
DipTax		12.3	4.4	35.6	10	14	50	0	Decr	0.7	19	5
aDipTax	JulianD	63.6	19.1	30	51.4	74.8	43.8	2.7	Decr	0.6		
pt_Diptera		36.4	9.7	26.7	29.8	42.9	6.3	29.7		-1.2	56.3	20
apt_Diptera	JulianD	57.9	22.3	38.6	44.1	74.2	68.8	2.7	Decr	1.3		
ChiroTax		8.7	4	46.1	6	11	18.8	0		0.3	15	2
aChiroTax	JulianD	58.6	22.4	38.2	42.4	72.7	18.8	2.7		0.1		
CrMolTax		0.2	0.4	264.5	0	0	0	10.8		-0.4	1	0
OligoTax		0.7	0.9	122.6	0	1	0	24.3		-0.8	3	0
TanytTax		1.3	0.9	67.6	1	2	56.3	0	Decr	1	3	0
Shan_2		4	0.4	10.5	3.7	4.2	93.8	0	Decr	3	4.5	2.2
Evenness		0.5	0.1	10.6	0.5	0.5	93.8	0	Decr	3	0.5	0.3
D_Mg		6	1.3	20.8	5.2	6.8	100	0	Decr	2.2	7.5	2.7
D		0.1	0	42.4	0.1	0.1	6.3	40.5	Incr	-3.2	0.4	0.1

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT2	5th BT2
PercDom01		24.4	8.6	35.2	18.2	28.4	6.3	27		-1.8	59.5	14.1
Composition metrics												
EPTPct		59.7	20.5	34.3	40	74.9	62.5	5.4	Decr	1.2	92.5	11
pEPTnoB		43.5	18.1	41.6	27.5	57	75	5.4	Decr	1.4	79.3	2.8
EphemPct		36.6	16.1	44	23.9	47.5	68.8	10.8	Decr	0.7	68.6	4.4
PlecoPct		12.9	6.9	53.7	7.9	17.4	81.3	5.4	Decr	1.2	37.8	0
TrichPct		10.3	9.7	94.1	3.5	13.5	68.8	10.8	Decr	0.4	28.9	0.3
NonInPct		5.6	5.7	101.6	1.8	7.5	31.3	16.2		-2.2	27.7	0
ColeoPct		7.9	11.7	147.8	0.9	11.1	75	0	Decr	0.6	31.8	0
aColeoPct	STemp	19	25.4	133.1	3.1	28.4	75	0	Decr	0.6		
OdonPct		0	0		0	0	0	2.7			0	0
DipPct		26.8	18.4	68.7	12.4	38.1	18.8	24.3		-1	74.4	3.7
ChiroPct		21.9	17.6	80.3	7.9	34	18.8	21.6		-0.9	62.8	0.9
CrCh2ChiPct		4.1	12.7	311.4	0	0	0	2.7		0.3	23.9	0
Orth2ChiPct		6.5	10.9	167.2	0	8.2	0	27		-1.3	56.3	0
TanytPct		5.1	8.1	158.6	0.6	6	62.5	2.7	Decr	0.5	20.8	0
Tnyt2ChiPct		22.2	21.8	97.9	3.5	38.4	81.3	0	Decr	0.8	66.7	0
OligoPct		1.6	3.1	190.4	0	1.4	0	18.9		-2.1	15.7	0
CrMolPct		0.3	1.6	473.4	0	0	0	10.8		-3.6	2.7	0
GastrPct		0	0		0	0	0	5.4			0	0
CorbPct		0	0		0	0	0	0			0	0
IsoPct		0	0	761.6	0	0	0	2.7		-144.2	0	0
AmphPct		0	0		0	0	0	0			0	0
BivalPct		0	0.1	534.1	0	0	0	2.7		-3.1	0.4	0
Habit metrics												
BrrwrTax		1.6	1.3	80.1	1	2	25	5.4		0.5	3	0
ClmbrTax		2.3	1.2	51.9	2	3	43.8	0	Decr	0.5	4	1
ClngrTax		15	3.6	23.9	13	17	100	0	Decr	2.8	20	4

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT2	5th BT2
SprwlTax		7.1	2.5	35.9	5	8	31.3	2.7		0.5	12	2.6
SwmmrTax		1.3	1.2	91.8	0	2	0	0		0.6	3	0
BrrwrPct		1.8	2.2	123.2	0.3	2.5	43.8	10.8	Decr	-0.7	6.7	0
ClmbrPct		14.6	12.6	86.6	3.5	23.1	18.8	13.5		-0.4	52	0.6
ClngrPct		49	14.8	30.2	39.8	59.4	87.5	2.7	Decr	2.2	75	8.2
SprwlPct		14	10.5	74.7	7	18.2	18.8	24.3		-1	41.8	1.8
SwmmrPct		3.9	6.5	167	0	4.8	0	0		0.5	14.6	0
Tolerance metrics												
BeckBI		29.8	7	23.5	24.3	33.8	100	0	Decr	2.7	39	8
HBI		3.5	0.8	23.7	3	4.1	6.3	35.1	Incr	-1.8	5.7	1.8
NCBI		2.5	1.4	54.9	1.7	2.9	12.5	27		-0.7	5.8	0.9
IntolTax		15.9	3.8	23.9	13.3	18	100	0	Decr	2.7	21	4
pt_Intol		47.9	10.2	21.2	40.7	54.8	75	2.7	Decr	1.9	66.7	19.1
apt_Intol	JulianD	75.9	13.5	17.8	66.8	85.2	87.5	2.7	Decr	2.4		
pt_Toler		11.8	4.8	40.8	8.4	14.7	6.3	24.3		-1.7	26.5	3.5
TolerTax		4	2	50.1	3	5.8	18.8	5.4		0.1	7.4	1
InMolTax		0	0		0	0	0	0			0	0
PercBaetEph		45.4	21.1	46.5	32.6	59.1	6.3	29.7		-1.5	100	5.6
PercIntol		43.4	17.4	40.1	28	55.6	81.3	2.7	Decr	1.3	84.4	4.9
PercTol		10.7	8.6	80.3	4.1	15.8	25	18.9		-1.1	38.6	0.3
PercHydr2Tri		13.3	21.3	159.9	0	15.1	0	13.5		-0.6	80	0
PercHydr2EPT		2	3.4	168	0	1.9	0	13.5		-0.9	14.7	0
TIVsed		5	0.9	18.3	4.4	5.8	12.5	21.6		-1	7.1	3.3
nt_DecrMtnTrn		9.7	2.2	22.8	8	11	100	0	Decr	3.2	12	2
pi_DecrMtnTrn		44.7	13.7	30.6	35.9	54.4	93.8	2.7	Decr	2.6	72.3	2.9
nt_IncrMtnTrn		0.1	0.3	252.2	0	0	0	29.7		-2.3	2	0
pi_IncrMtnTrn		0.3	1.4	437.2	0	0	0	29.7		-4.7	5	0
Feeding Group metrics												

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT2	5th BT2
CllctPctFAC		53.1	17.2	32.4	42.4	67.7	18.8	24.3		-0.9	85.8	19.4
FiltrPctFAC		4.9	4.9	100.2	1.5	6.5	37.5	13.5	Decr	-0.8	33.7	0
PredPctFAC		17.9	8.2	46.1	12.1	24.6	68.8	2.7	Decr	0.8	34.8	2.9
ScrapPctFAC		20.6	13.1	63.5	10.4	31.3	81.3	2.7	Decr	0.8	44.2	2.6
ShredPctFAC		11.8	8	67.7	6.7	15.5	25	27		-0.8	37.6	1.1
CllctTaxFAC		13.7	4.3	31.4	11	16	43.8	0	Decr	0.7	20	6
aCllctTaxFAC	JulianD	66.3	18.4	27.8	53.5	79.9	50	5.4	Decr	0.5		
FiltrTaxFAC		2.1	1.1	52.6	1	3	31.3	0		0.9	4	0
PredTaxFAC		9.8	2.6	26.4	8	12	93.8	0	Decr	2	13	3
ScrapTaxFAC		5.7	1.5	26.2	5	6.8	87.5	0	Decr	1.9	8	2
ShredTaxFAC		4.8	2	41.5	3	6	50	2.7	Decr	0.9	7	1
GathFiltTaxFAC		15.7	4.7	29.7	12	19	31.3	0		0.8	23	7.6
PredScrapTaxFAC		15.5	3	19.1	13	18	93.8	0	Decr	2.7	19	5.6
PredShrTaxFAC		12.9	2.9	22.2	11	15	93.8	0	Decr	2.4	17	4
CllctPct		40.2	16.3	40.4	25.4	54.3	12.5	18.9		-0.6	69.4	12
FiltrPct		4.6	4.8	104.8	1.4	5.8	37.5	13.5	Decr	-0.8	33	0
PredPct		16.9	8	47.3	11.1	22	68.8	13.5	Decr	0.7	33	2.5
ScrapPct		12.5	11.4	91.1	4	14.2	93.8	2.7	Decr	0.9	35.7	0
ShredPct		8.2	5.2	64.1	4.9	10.8	75	8.1	Decr	0.5	27.1	0
CllctTax		0.8	0.8	105.4	0	1	0	8.1		0	2	0
FiltrTax		2.1	1.1	52.6	1	3	31.3	0		0.9	4	0
PredTax		9.6	2.5	26.4	8	11.8	93.8	0	Decr	2.1	12.5	3
ScrapTax		2.9	1.1	38.6	2	4	75	0	Decr	1.9	5	0
ShredTax		3.3	1.7	51.1	2	4	50	2.7	Decr	1	5	0
PredScrapTax		12.5	3	24.3	10	14.8	93.8	0	Decr	2.4	16	4
PredShrTax		12.9	2.9	22.2	11	15	93.8	0	Decr	2.4	17	4

Table D-3. Biotype 3 Metric Performance Statistics.

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT3	5th BT3
Richness metrics												
TotalTax		26.3	6.4	24.4	21.3	30	86.7	0	Decr	1.4	33.8	10.3
InsectTax		22.1	6	27	17.3	24	93.3	0	Decr	1.6	28	7
pt_noninsect		15.9	8.8	55.4	8.8	20.8	6.7	66.7	Incr	-1.4	45.7	5.6
EPTTax		4.4	2.9	65.1	2	6	20	20		0.2	9	0
pt_EPT		17	12	70.8	9.6	23.3	20	46.7	Incr	-0.4	45.5	0
apt_EPT	JulianD	54.5	28	51.4	31.2	76.6	20	33.3		-0.2		
EphemTax		2.8	2.1	73.9	1.3	4.5	33.3	20		0.1	6	0
PlecoTax		0	0		0	0	0	0			1	0
TrichTax		1.6	1.6	100.2	0	2	0	20		0.2	3	0
ColeoTax		2.3	1.7	72	1	3	80	0	Decr	1.3	3	0
DipTax		13.9	5.1	36.9	11.3	15	86.7	0	Decr	1.2	18	4
aDipTax	JulianD	81.4	15.4	18.9	68.3	93.3	73.3	0	Decr	1.7		
pt_Diptera		52.6	13.3	25.3	44.4	63.8	40	6.7	Decr	0.6	68.2	25.3
apt_Diptera	JulianD	30.9	29.3	94.8	0.5	47.2	6.7	53.3	Incr	-0.6		
ChiroTax		10.6	4.2	40.2	9	11.8	86.7	6.7	Decr	1.1	15	2
aChiroTax	JulianD	67.3	23.9	35.5	56.4	86.1	80	0	Decr	1.1		
CrMolTax		1.9	1.6	81.7	1	3	33.3	20		0.2	5	0
OligoTax		0.9	1.1	127.3	0	1.8	0	46.7	Incr	-0.6	5	0
TanytTax		1.4	0.8	54.3	1	2	66.7	0	Decr	1.4	3	0
Shan_2		3.3	0.5	14.8	3	3.6	80	6.7	Decr	1.5	3.9	1.4
Evenness		0.4	0.1	14.6	0.4	0.4	80	6.7	Decr	1.5	0.5	0.2
D_Mg		4.6	1.2	26.7	3.8	5.2	86.7	0	Decr	1.4	6	1.7
D		0.2	0.1	40.5	0.1	0.2	6.7	46.7	Incr	-1.7	0.6	0.1
PercDom01		33	11.9	36	24.7	40	13.3	46.7	Incr	-1.1	78	18.2
Composition metrics												
EPTPct		26.1	27.6	105.8	5.5	36.4	33.3	20		0	80.8	0

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT3	5th BT3
pEPTnoB		15.9	19.3	121.1	1.5	23.6	46.7	13.3	Decr	0.4	54.6	0
EphemPct		19.9	25	125.2	2.1	25.9	40	26.7	Decr	0	66.9	0
PlecoPct		0	0		0	0	0	0			1	0
TrichPct		6.2	9.4	152	0	7.7	0	20		0.2	27.9	0
NonInPct		18.4	18.3	99.7	1.2	28.8	6.7	53.3	Incr	-1.1	78.6	0.7
ColeoPct		5.4	9	167.3	1.4	6.7	100	0	Decr	0.6	9.4	0
OdonPct		0.9	1.6	184.3	0	0.6	0	26.7		0.2	5.6	0
DipPct		48.1	25.4	52.8	28.3	65.7	46.7	6.7	Decr	0.5	91.2	5.6
aDipPct	JulianD	43.5	32.3	74.3	18.1	67.4	6.7	46.7	Incr	-0.5		
ChiroPct		31.9	20.7	64.9	17.6	41.3	53.3	13.3	Decr	0.3	86.4	2.1
aChiroPct	JulianD	61.1	28.7	47	48.2	80.1	53.3	20	Decr	0.5		
piChirNoTany		25.9	20.3	78.2	9.7	33.3	33.3	26.7		0.2		
CrCh2ChiPct		12.9	23.5	182	0	9.2	0	60	Incr	-0.7	77.7	0
Orth2ChiPct		2.2	4	181	0	3.4	0	46.7	Incr	-4	75.8	0
CCO2ChirPct		15.1	22.7	150.7	0.9	12.9	0	86.7	Incr	-1.4	80	0
TanytPct		6	6.9	114.5	1.1	12	86.7	6.7	Decr	0.5	18.7	0
Tnyt2ChiPct		21.2	21.9	103.4	4	43.5	73.3	0	Decr	0.8	58.8	0
OligoPct		1.2	2	162.3	0	1.3	0	60	Incr	-8.8	50.3	0
CrMolPct		9.5	13.7	144.3	0.3	10.5	33.3	13.3		0.2	49.7	0
GastrPct		4.9	7.4	150.3	0	6.3	0	6.7		0.2	12.3	0
CorbPct		0	0		0	0	0	6.7			0	0
IsoPct		0	0		0	0	0	13.3			4.5	0
AmphPct		3.8	7.5	197.1	0	2.6	0	13.3		0.1	20.5	0
BivalPct		0.3	0.8	258.1	0	0	0	20		0.3	2.2	0
Habit metrics												
BrrwrTax		3.4	2.4	68.4	2	4.8	46.7	0	Decr	0.7	6	0
ClmbrTax		1.7	1.1	62.4	1	2	33.3	0		0.9	4	0
ClngrTax		5.1	3.3	64.7	2	8	6.7	0		0.5	9	1

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT3	5th BT3
SprwlTax		7.4	2.6	35.5	5.3	9	86.7	0	Decr	1.6	11	1
aSprwlTax	JulianD	69.2	20.7	29.9	54.1	84	86.7	0	Decr	1.7		
SwmmrTax		2.2	1.8	81	1	3	20	6.7		0.4	4.8	0
BrrwrPct		6.6	12.3	187.2	1.1	4.7	33.3	33.3		0.1	45	0
ClmbrPct		10	12.3	123.3	1.1	15.9	60	6.7	Decr	0.5	28.9	0
ClngrPct		29.5	21.2	72	11.8	40.3	20	6.7		0.3	67	0.5
SprwlPct		23.3	14	59.9	9.4	34.4	73.3	6.7	Decr	1	45	0.7
SwmmrPct		7.1	12.2	171.7	1	4.9	33.3	53.3	Incr	-0.6	36.5	0
Tolerance metrics												
BeckBI		3.9	3.6	91.4	1.3	6	20	6.7		0.2	9	0
HBI		5.9	1	16.2	5.6	6.6	26.7	40	Incr	-0.5	8	4.4
NCBI		4.8	1.6	34.5	3.4	6.2	20	33.3		-0.2	7.9	2
IntolTax		1.6	1.3	86	0.3	2	60	0	Decr	0.8	4.8	0
pt_Intol		5.8	5	87.2	0.8	8.8	60	0	Decr	0.6	22.5	0
pt_Toler		33.7	12.7	37.8	20.8	43.6	20	46.7	Incr	-0.3	58.2	6.9
apt_Toler	JulianD	56.3	33.4	59.4	45.2	78.5	46.7	26.7	Decr	0.4		
TolerTax		8.9	4	44.8	5	12	33.3	6.7		0.6	14	1
aTolerTax	JulianD	53.7	28.2	52.5	33.3	70.9	13.3	46.7	Incr	-0.5		
InMolTax		0	0		0	0	0	0			0	0
PercBaetEph		31.6	39.5	124.9	0	67.8	0	53.3	Incr	-0.6	100	0
PercIntol		5.4	11.3	210.2	0.1	3.6	60	6.7	Decr	0.4	19.2	0
PercTol		33.6	23	68.5	13.7	52.7	33.3	40	Incr	-0.3	88.3	1.2
PercHydr2Tri		44.3	47.7	107.6	0	92.8	0	26.7		0.1	100	0
PercHydr2EPT		19.2	28.6	149.1	0	35.9	0	6.7		0.4	73.6	0
TIVsed		5.1	0.3	6.8	5	5.2	0	100	Incr	-8.3	7.7	4.2
nt_DecrPlains		4.7	1.4	29.8	5	5.8	100	0	Decr	2.8	6	0
pi_DecrPlains		22.1	12.9	58.4	12.8	33.8	100	0	Decr	1.7	39.3	0
nt_IncrPlains		1.1	0.8	74.9	1	1.8	6.7	80	Incr	-2.3	5	0

Metric	Adjust	Mean Ref	Ref SD	Ref CV	Ref 25th	Ref 75th	Strs DE25	Strs DE75	Trend	Z-score	95th BT3	5th BT3
pi_IncrPlains		1.8	1.8	101.5	0.3	3.3	6.7	86.7	Incr	-16	79.4	0
Feeding Group metrics												
CllctPctFAC		54.1	14.6	26.9	44.7	63.9	13.3	53.3	Incr	-0.7	95.8	22.5
FiltrPctFAC		24.2	14.5	59.8	14.8	33.7	60	20	Decr	0.7	52.6	0.5
PredPctFAC		11	8.1	73.5	4.6	14.7	66.7	13.3	Decr	0.6	31.3	0.3
ScrapPctFAC		10.3	8.7	84.5	4.5	13.4	53.3	20	Decr	0.1	24.1	0
ShredPctFAC		8	12.9	162.5	1.7	7.3	0	66.7	Incr	-0.5	61.7	0.3
CllctTaxFAC		13.2	4.7	35.4	11	14	73.3	0	Decr	0.9	19	5
FiltrTaxFAC		2.2	1.4	64.7	1	3	13.3	6.7		0.3	4	0
PredTaxFAC		6.6	2.6	39.2	5.3	8.5	93.3	0	Decr	1.3	10	1
ScrapTaxFAC		2.3	1.4	60.6	1	3	20	6.7		0.4	4	0
ShredTaxFAC		2.5	1.5	58.6	1.3	3.8	13.3	13.3		0	5	1
GathFiltTaxFAC		15.3	4.9	32.1	13	16.8	73.3	0	Decr	1	21	6
PredScrapTaxFAC		8.9	2.5	27.5	7.3	11	93.3	0	Decr	1.6	13	1
PredShrTaxFAC		7.7	2.8	36.1	6	9.8	93.3	0	Decr	1.4	11	1
CllctPct		43.6	14.6	33.3	35.1	51.6	26.7	53.3	Incr	-0.6	86.7	7.7
FiltrPct		19.3	16.2	84	5.6	31.2	53.3	13.3	Decr	0.5	46.2	0
PredPct		9.9	7.4	74	4.5	13.9	73.3	6.7	Decr	0.7	30.2	0.3
ScrapPct		5.5	7.2	130.6	0.5	7.3	40	13.3	Decr	-0.1	15.1	0
ShredPct		6.2	12.7	207	0	5.9	0	40	Incr	-0.2	34	0
CllctTax		1.2	1.1	86.7	0	2	0	0		0.4	3	0
FiltrTax		2.2	1.4	64.7	1	3	13.3	6.7		0.3	4	0
PredTax		6.1	2.4	39.8	5	7.5	86.7	0	Decr	1.4	9	1
ScrapTax		1.2	0.9	71.9	1	1.8	26.7	20		0.3	2.8	0
ShredTax		1.6	1.4	85.5	0	2	0	6.7		0.3	3	0
PredScrapTax		7.3	2.3	31.5	6	8.8	93.3	0	Decr	1.6	11	1
PredShrTax		7.7	2.8	36.1	6	9.8	93.3	0	Decr	1.4	11	1

APPENDIX B

Benthic Macroinvertebrate Sampling Standard Operating Procedure

Water Quality Control Division • Environmental Data Unit

1.0 Introduction

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. They 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 multi-metric 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.

2.0 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 un-wadeable large rivers or intermittent type streams are beyond the scope of this procedure.

3.0 Index Period

The index period is the period of time that samples shall be collected to minimize seasonal variation. The standard index period utilized by the Water Quality Control Division ("division") in Biotypes[1] 1 and 2 shall be late June to November 30. For Biotype 3, the index period shall be May 1 to November 30.

These periods are congruent with the central tendency of sample dates of macroinvertebrate replicates used to regionally calibrate the multi-metric indices.

4.0 Sampling Frequency

[1] Biotype – biological similar regions in Colorado as established during multi-metric index development.

This protocol recommends that benthic macroinvertebrate samples be collected once per year per site and within the standardized index periods provided in Section 3.0.

5.0 Site Selection

The study reach length shall be one of the following: 1) 40 times the average wetted width of the wadeable waterbody or 2) long enough to encompass multiple riffles (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). A glide generally refers to a calm stretch of shallow water flowing smoothly.

Although riffle areas with hard-bottomed substrates are generally the most diverse and productive habitat type in mountainous 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.

5.1 Hard-Bottomed Streams

A hard-bottomed stream is one where the stream substrate is dominated by particles gravel size or larger. Riffle 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 6.0.

5.2 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 multi-habitat approach described in Section 6.0.

6.0 Sample Collection Information

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

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

- Semi-quantitative sample collection of soft-bottomed streams that focuses collecting macroinvertebrates from non-riffle 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 multi-metric 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 multi-habitat 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, multi-habitat(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 sample will need to be collected.

6.1 Equipment and Supplies

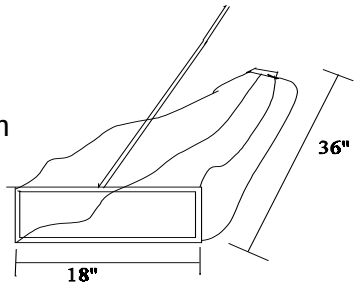
- Kick-net (see specifications in Section 6.1.1)
- Sieve dolphin bucket (504 µm mesh)
- 1-liter wide-mouth sample jars with screw tight lids
- 95% ethanol stored in sealed and labeled polyurethane carboys or bottles
- 1-liter rinse bottle
- Fine-tip forceps
- Number 30 (600 µm) **or** 35 (500 µm) standard sieve
- 5-gallon bucket with handle
- Dissecting tray
- Pre-printed, rectangular "*Rite in the Rain*" labels

- Standard #2 pencil(s)
- Waterproof stop watch
- Tape measurers (100 ft or more)
- Galvanized metal stakes or nails
- 48 quart or larger ice chests or sealable Rubbermaid totes
- Safety glasses
- Rubber or nitrile gloves
- Hip or chest waders with wading boots
- Applicable field sheets

6.1.1 Kick Net Specifications

The kick net is comprised of the following components:

- 18” x 8” rectangular frame
- 500 to 600 µm mesh nylon bag with canvas reinforced bottom and shroud reinforced opening (≈ 36” long)
- 1 or 2 piece long handle (≈ 70” long)
- Sieve dolphin bucket (504 µm mesh)
- Sieve dolphin bucket adaptor



7.0 Sample Collection

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

7.1 Riffle Habitat Method

- Ensure that the sampling net and sieve bucket are clean prior to usage.

- Select the dominant riffle habitat within the study reach according to Section 5.0.

- 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.

- 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.

- 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 and not exceed 1 minute. Approximately 0.25 meters should be disturbed for every 15 seconds.

- 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.

- 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.

- Release back into the stream any fish, amphibians, reptiles or crayfish/rusty crayfish caught in the net.

- If excessive or large debris items are present refer to Section 7.4.1.

- The kick-net should be rinsed clean by backwashing with site water before collecting additional samples.

- At this point refer to Section 7.4 for sample processing.

7.2 Multi-Habitat Method

- Ensure that the sampling net and sieve bucket are clean prior to usage.

- 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 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 6.0.

  Woody Debris or Snag

  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.

- 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.

- Release back into the stream any fish, amphibians, reptiles or crayfish/rusty crayfish caught in the net.

- If excessive or large debris items are present refer to Section 7.4.1.

- The kick-net should be rinsed clean by backwashing with site water before collecting additional samples.

- At this point refer to Section 7.4 for sample processing.

7.3 Field Duplicates

One out of ten (10%) sample events shall include a duplicate field sample to ensure quality control (QC). For example, when a biological survey 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 7.1 or 7.2.

7.4 Sample Processing (On-site)

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

7.4.1 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 half full of sample material. Consequently, splitting the sample into two or more sample jars is acceptable. See Section 7.4.4 for labeling split samples.

7.4.2 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 7.4.1.

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 a 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 7.4.4 for labeling split samples.

7.4.3 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 then filling the jar with 95% ethanol (ETOH) so the ETOH is 1" above the detritus/specimen matrix. 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.

7.4.4 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. Pull back a corner of the packaging tape prior to affixing the label so the tape/label can be easily removed later once the taxonomist returns the 1-liter jars.

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
- Check if kicknet was used
- Check 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.

7.4.5 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 beds.

8.0 Invasive Species

Precautions must be taken to avoid collecting invasive species from areas where the Colorado Parks and Wildlife ("CPW") 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) or sampling equipment, as instructed by CPW.

9.0 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. 57p.

10.0 Document Version

5.0_111716

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**).

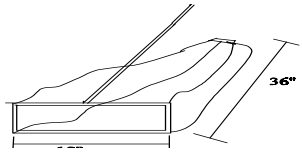


Figure 1. Modified Kick Net

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 2016. 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., K. W. Cummins, and M.B. Berg (eds.) 2008. An Introduction to the Aquatic Insects of North America. Fourth edition. Kendall Hunt Publishing Company, Dubuque, Iowa.

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

Additional literature sources may include:

- 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 all of the River Invertebrate Prediction and Classification System ("RIVPACS") Observed over Expected ("O/E") predictor variables to the Predictor Variable Import Sheet, with the exception of the "Summer" temperature variable, since they are not relevant to the generation of MMI scores. The "Summer" (temperature) is now used to adjust three metrics in the Biotype 1 MMI. The Predictor Variable Import Sheet will upload correctly without the remaining 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 will take upwards to 5 minutes. When the calculations are finished, the Excel spreadsheet will open and MMI scores can be viewed immediately in Column AT.

F. Manually Generating Index Scores

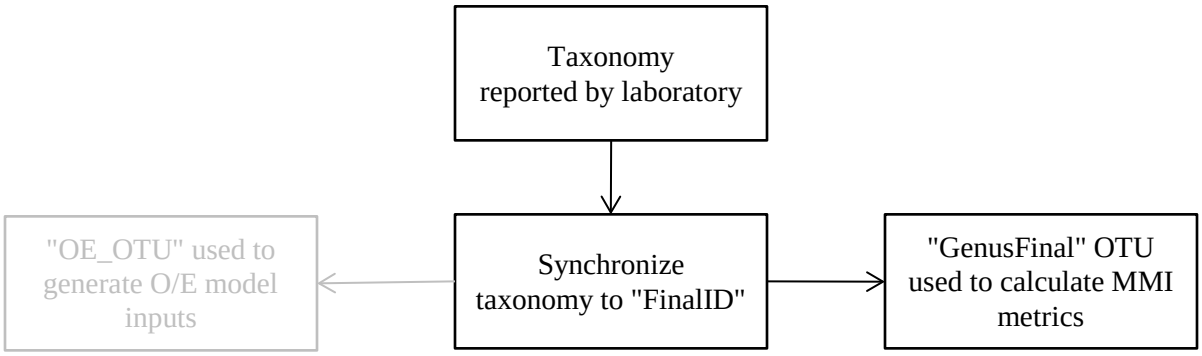
Manually deriving MMI scores, independent of EDAS, is best handled through formulaic calculations in an Excel spreadsheet. However, there are many steps that must occur between the receipt of laboratory reported taxonomic data and setting up the MMI calculation spreadsheet. These steps are organized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

Taxonomic Resolution

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 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

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. Some taxon have multiple FFG's. These are termed "facultative", assuming that a taxon with multiple feeding mechanisms might use all of its mechanisms depending on its life stage.

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 39-50 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 each MMI. For instance, if MMI #1 is comprised of eight metrics, then the scores for those eight metrics are summed and divided by eight. 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 2017

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 United States Environmental Protection Agency (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 Tiered Aquatic Life Use ("TALU") condition gradient, but it would be hard to know this without having first developed

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

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

[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

of a rationale is puzzling in 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%).

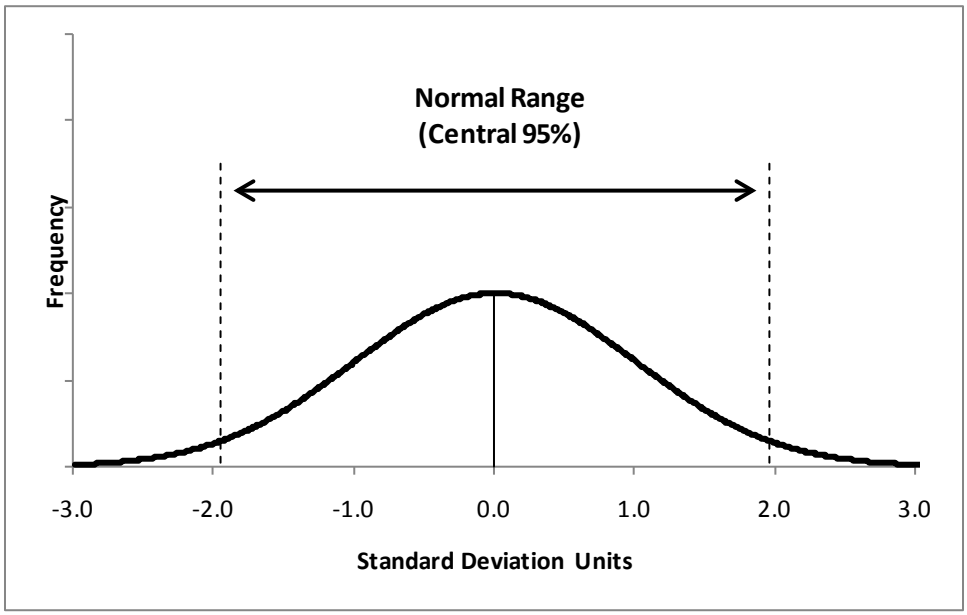


Figure 1. Graphical representation of the "normal range" based on the central 95% of the distribution (bold line represents the normal frequency distribution). The dashed vertical lines signify the boundaries of the central 95% at ±1.96σ.

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

[4] Kilgour et al. 1998

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 tests, 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

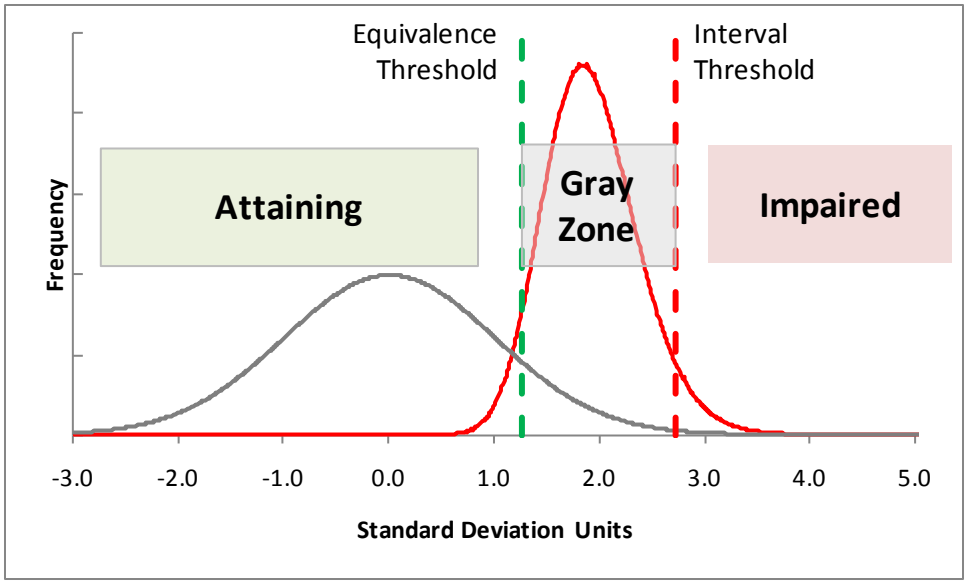


Figure 2. Graphical representation of thresholds derived from interval and equivalence tests. The distribution from which the normal range is derived is bell-shaped curve centered on 0.0 standard deviation units. The non-central F distribution is shown as the taller curve centered on the Gray Zone. Equivalence 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 144 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 2 (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.

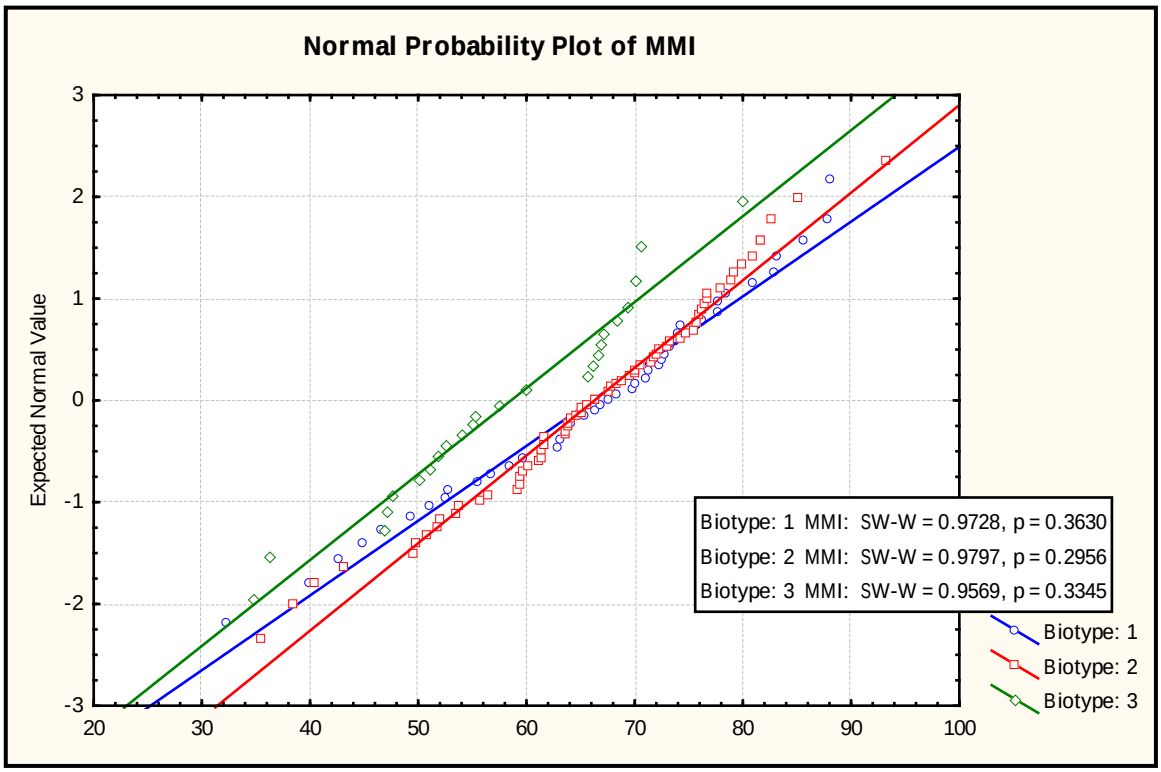


Figure 3. Probability plots of MMI scores at reference sites in each of the three Biotypes. Biotype membership was determined by cluster analysis. One outlier was removed from Biotype 2, as explained in the text.

Biological Thresholds Options

Two options were revisited in 2017; 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. The two 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	45	66.05	13.11	48.9	38.6
1	2	45	66.05	13.11	45.2	33.7
2	1	72	66.25	11.29	50.8	43.8
2	2	72	66.25	11.29	47.5	39.8
3	1	26	58.53	11.15	44.9	33.1
3	2	26	58.53	11.15	41.9	28.7

Table 1. Summary of group characteristics and biological 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. In subsequent sections of this document, all attention is focused on the Option 2 criteria.

Evaluation in the "Gray Zone"

As explained previously, using interval and equivalence tests to derive biological 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 classification 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 sensitivity 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 Index ("SDI") and the Hilsenhoff Biotic Index ("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 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 SDI, which is calculated in EDAS for every sample. However, it should be noted that the new MMI 4.0 bioassessment tool capitalizes on both species-level and genus-level taxonomy. Accordingly, whereas previously, for example, three species under a single genus were counted as only "one genus", now the three species are counted separately resulting in two additional taxa. The practical outcome from this change associates to either zero gain in taxa or an increase in taxa depending on whether the overall taxonomic resolution of a given sample was to species or genus-level. No taxa were lost from a sample.

For each Biotype, diversity values were assembled for all reference sites and the distributions were examined (Figure 4). After removing one outlier from the lower end of the distributions for Biotype 2, 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.

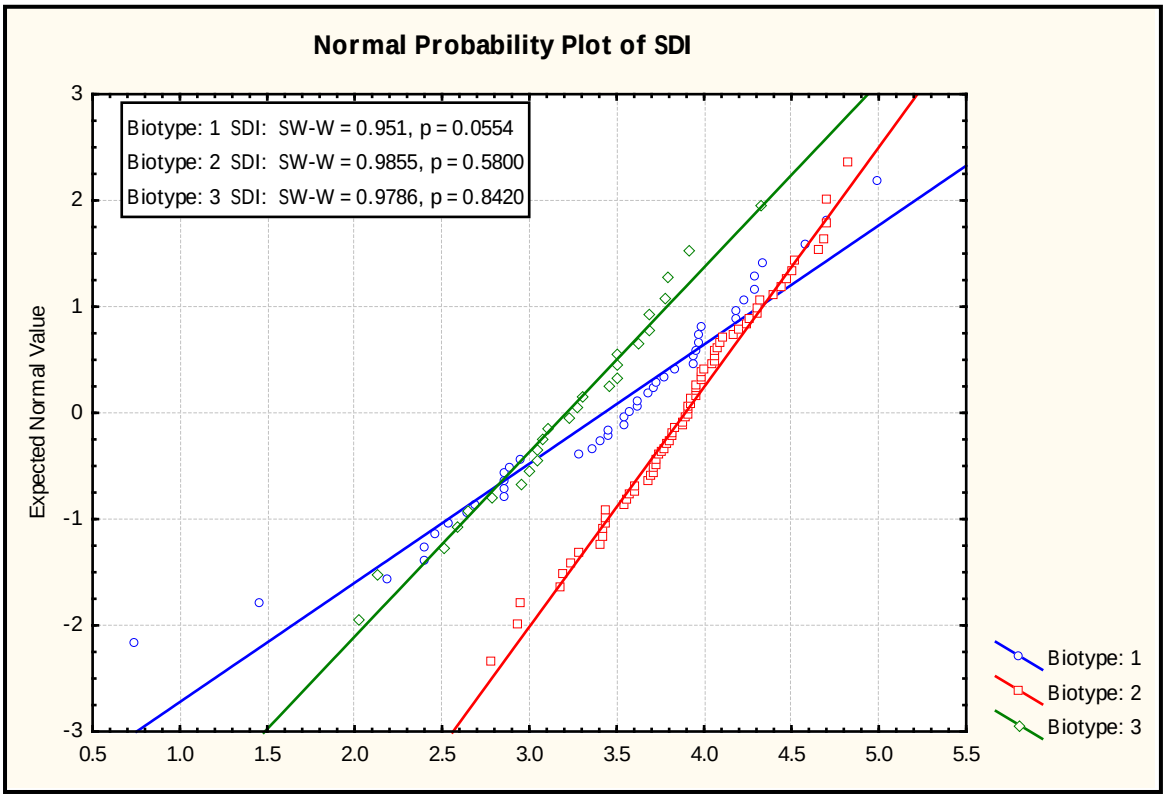


Figure 4. Probability plots for Shannon Diversity Index values at reference sites in each biotype.

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 3.2 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	SDI	HBI
1	45.2	33.7	2.1	5.8
2	47.5	39.8	3.2	4.9
3	41.9	28.7	2.4	7.6

Table 2. Boundaries of the gray zone and thresholds of equivalence for two auxiliary metrics – SDI 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 SDI, the HBI is not part of the MMI and the values are routine output from EDAS. Distributions of the HBI were normal for reference sites in each of the Biotypes and there were no outliers (Figure 5). Variances were not significantly different among the Biotypes, but means were. Accordingly, equivalence thresholds are developed separately for each Biotype (Table 3).

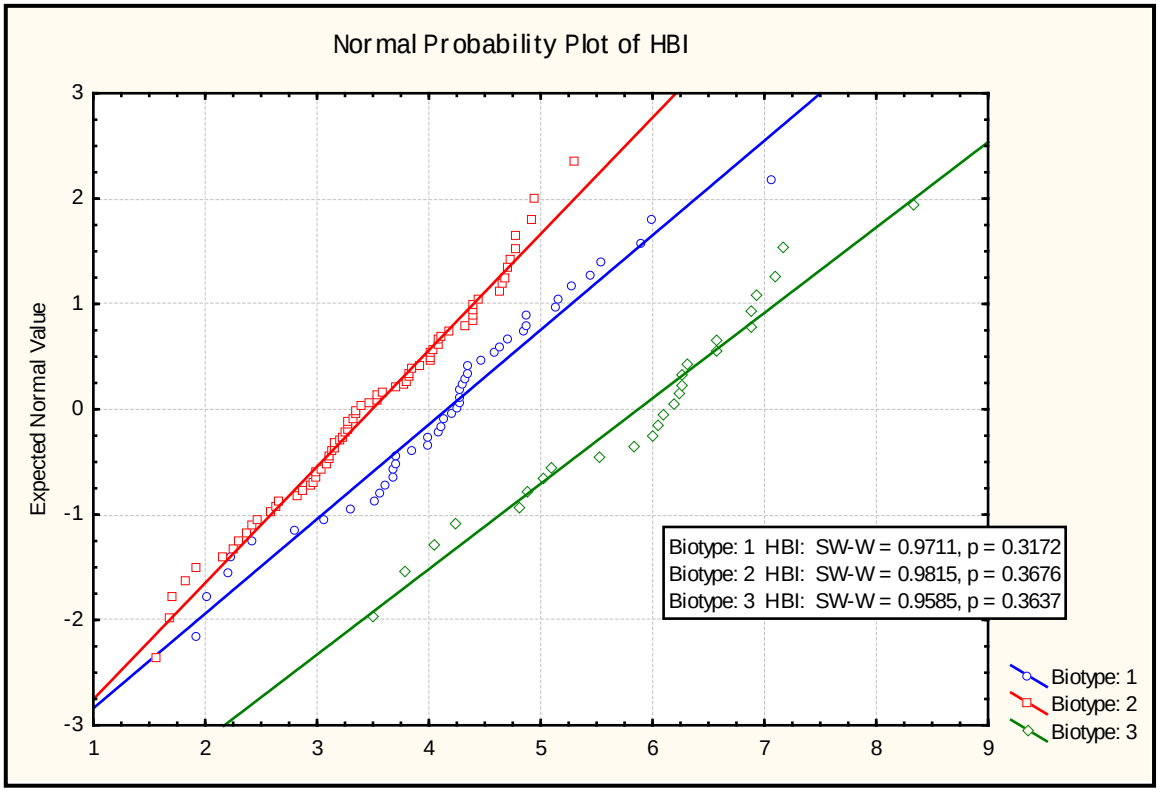


Figure 5. Probability plots for HBI values at reference sites in each Biotype.

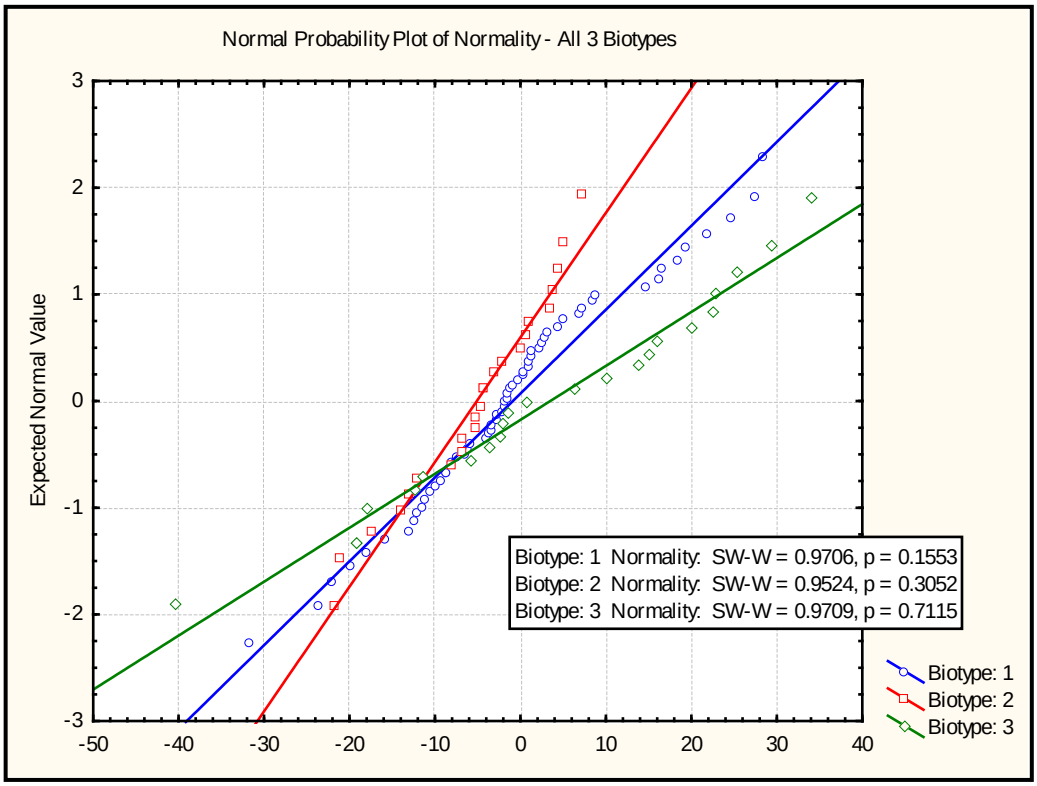
In contrast to the SDI, low scores of the HBI indicate better conditions. Consequently, using the same example presented in the preceding section, the HBI would have to be *less than* 4.9 for the site to be considered in attainment. Moreover, both auxiliary metrics would have to show attainment before 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 4.9 *or* the SDI was not greater than 3.2 – 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 80 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) and any pairings that included a first sample collected in 2001 with a second sample collected in 2003. 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 for all Biotypes were normal with no outliers identified with Rosner's sequential outlier test. 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=-0.90; s.d.=13.35; N=107). The d-MMI scores showed considerable variability. The "normal operating range" was 27.1 to -27.1[8]. For a change to be considered significant at the 0.05 level, the MMI would have to increase or decrease by 27.1 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.



[8] The normal operating range is centered on the mean of the number of cases (n=109) to reflect the assumption that no change in biological condition would yield no change in MMI. .

Figure 6. Probability plot of d-MMI scores.

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 -21.9 and -29.4 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 21.9 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 56 for Biotype 1, 62 for Biotype 2, or 51 for Biotype 3, before a site would qualify for consideration as high-scoring.

APPENDIX F

Big River Analysis – Data and Graphs

Materials presented in this appendix are the data underpinnings used to support application of MMIs and biological thresholds to bigger rivers on the Western Slope of Colorado. Statistical cluster analysis concludes that these big rivers and their tributaries share similar taxonomic composition when comparing common genera between the two types of waterbodies. Provided in the tables below are the mainstem and tributary sites as well as common genera used in the aforementioned statistical analysis. For additional details see Policy Statement Section IX subsection B.

Table F1. Yampa and White River Basins Data and Cluster Analysis Graph

*Due to low number of sites in the White River basin, the mainstem and tributary sites from both the Yampa and White River basins were combined to improve statistical rigor.

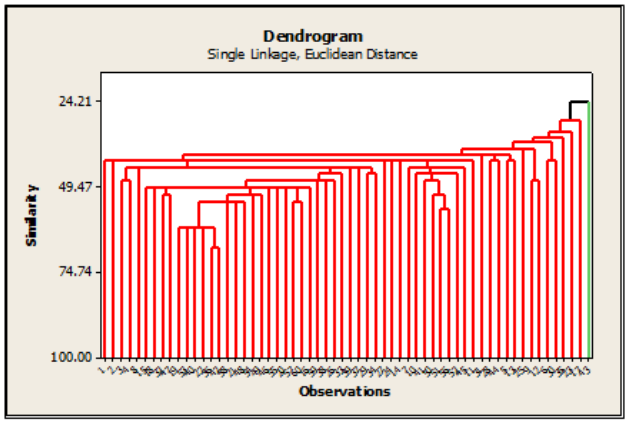
Yampa River Sites

STATIONID	MAINSTEM	STATIONID	TRIBUTARY
38	Yampa River	153	Oak Creek
12802	Yampa River	11325	Good Spring Creek
12803	Yampa River	11341	Wilson Creek
12804	Yampa River	11345	Stinking Gulch
12805	Yampa River	11368	Temple Gulch
12807	Yampa River	11387	Fortification Creek
12808	Yampa River	12823	Morapos Creek
12809	Yampa River	12832	Little Bear Creek
12811	Yampa River	12840	Elkhead Creek
12814	Yampa River	12843	Elkhead Creek
12815	Yampa River	12846	Elkhead Creek
12805B	Yampa River	12854	Fish Creek
12806d	Yampa River	12860	Elk River
12806F	Yampa River	12863	Mad Creek
FW08CO037	Yampa River	12870	Fish Creek
FW08CO066	Yampa River	12876	Trout Creek
WCOP99-0601	Yampa River	12891	Oak Creek
		12892	Oak Creek
		12893	Burgess Creek
		12894	Walton Creek
		12895	Wolf Creek
		11343c	Milk Creek
		12851B	Sage Creek
		12898A	Bear River
		RW-702	Good Spring Creek
		WCOP99-0512	Elkhead Creek
		WCOP99-0565	Elkhead Creek

White River Sites

STATIONID	MAINSTEM	STATIONID	TRIBUTARY
FW08CO050	White River	11531	Ute Creek
RW-529	White River	11533	Marvine Creek
RW-531	White River	11535	Lost Creek
RW-754	White River	11720	West Douglas Creek
WCOP99-0638	White River	11722	Cathedral Creek
		11726	East Douglas Creek
		11770	South Fork White River
		FW08CO003	East Douglas Creek
		FW08CO021	Douglas Creek
		FW08CO085	East Douglas Creek
		WCOP99-0516	East Douglas Creek

YAMPA AND WHITE BASIN - COMMON GENUS			
Acentrella	Diamesa	Lebertia	Physa
Acerpenna	Dicranota	Lepidostoma	Polypedilum
Ambrysus	Diphetor hageni	Lopescladius	Protoptila
Antocha	Drunella	Microcylloepus	Pteronarcella
Arctopsyche	Dubiraphia	Micropsectra	Rheocricotopus
Atherix	Epeorus	Microtendipes	Rheotanytarsus
Atractides	Ephemera	Nais	Rhithrogena
Baetis	Ephemerella	Nanocladius	Rhyacophila
Brachycentrus	Eukiefferiella	Neotrichia	Simulium
Cardiocladius	Fallceon quilleri	Oecetis	Skwala
Centroptilum	Glossosoma	Oligochaeta	Sperchon
Cheumatopsyche	Helicopsyche	Ophiogomphus	Stempellinella
Choroterpes	Hemerodromia	Optioservus	Sweltsa
Cinygmula	Heptagenia	Orconectes	Tanytarsus
Claassenia sabulosa	Hexatoma	Parakiefferiella	Thienemanniella
Cladotanytarsus	Hyalella	Paralauterborniella	Thienemannimyia
Corynoneura	Hydropsyche	Paraleptophlebia	Torrenticola
Crangonyx	Hydroptila	Parametriocnemus	Tricorythodes
Cricotopus	Hygrobates	Pentaneura	Tvetenia
Cryptochironomus	Isogenoides	Pericoma	Zaitzevia
Culoptila			



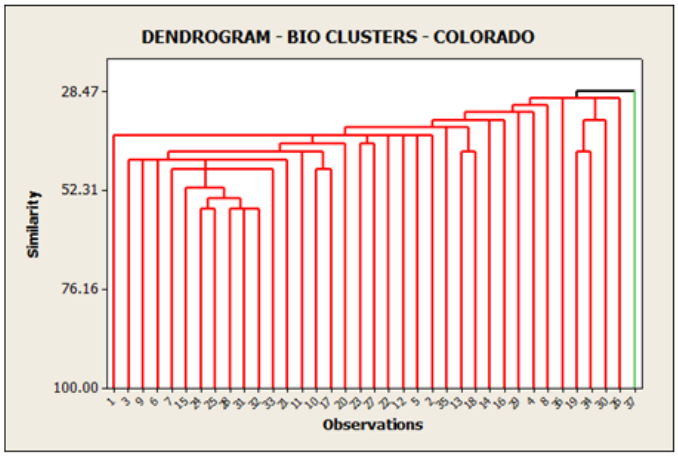

```
Final Partition
Number of clusters: 2


                                          Average   Maximum
                                   Within  distance  distance
             Number of  cluster sum      from      from
          observations   of squares  centroid  centroid
Cluster1            59      720.610   3.47097   4.73607
Cluster2             1        0.000   0.00000   0.00000
```

Table F2. Colorado River Basin Data and Cluster Analysis Graph

STATIONID	MAINSTEM	STATIONID	TRIBUTARY
9070500	Colorado River	11140	Parachute Creek
9095500	Colorado River	11175	Middle Rifle Creek
9163500	Colorado River	11187	East Divide Creek
FW08CO046	Colorado River	11193	Garfield Creek
FW08CO062	Colorado River	11206	Carr Creek
RW-550	Colorado River	11207	Dry Fork
RW-553	Colorado River	11208	Roan Creek
WCOP99-0571	Colorado River	11217	Clear Creek
WCOP99-0594	Colorado River	11250	Buzzard Creek
		12115	Sweetwater Creek
		12116	Sheephorn Creek
		12118	Big Alkali Creek
		12144	Troublesome Creek
		12146	Muddy Creek
		9095300	Dry Fork Roan Creek
		9153290	Reed Wash
		11190A	Main Elk Creek
		EPA01-166	Buzzard Creek
		EPA01-169	Deep Creek
		FW08CO006	East Rifle Creek
		RW-331	Grizzly Creek
		RW-350	West Mammoth Creek
		RW-351	Middle Mammoth Creek
		RW-548	Rifle Creek
		RW-731	Noname Creek
		WCOP04-R011	Roan Creek
		WCOP99-0595	East Rifle Creek
		WCOP99-0649	Noname Creek

COLORADO BASIN - COMMON GENUS			
Acari	Cryptochironomus	Lepidostoma	Polypedilum
Acentrella	Diamesa	Microcylloepus	Protzia
Agabus	Dicranota	Micropsectra	Pteronarcella
Arctopsyche	Diphetor	Microtendipes	Rheocricotopus
Argia	Drunella	Nematoda	Rheotanytarsus
Atherix	Epeorus	Neoplasta	Rhithrogena
Baetis	Ephemerella	Ochrotrichia	Rhyacophila
Brachycentrus	Eukiefferiella	Oecetis	Simulium
Caloparyphus	Helichus	Optioservus	Sperchon
Cardiocladius	Hemerodromia	Orthocladius	Sweltsa
Chelifera	Heptagenia	Pagastia	Thienemanniella
Cheumatopsyche	Heterlimnius	Parakiefferiella	Thienemannimyia
Cinygmula	Hexatoma	Paraleptophlebia	Tipula
Claassenia	Hydropsyche	Parametriocnemus	Tricorythodes
Cleptelmis	Hydroptila	Pentaneura	Tvetenia
Corynoneura	Hygrobates	Pericoma	Zaitzevia
Cricotopus	Lebertia	Phaenopsectra	Zapada



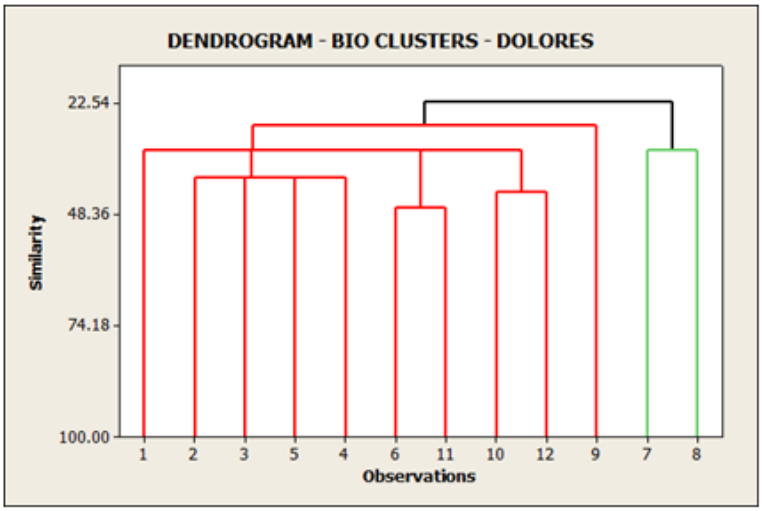

```
Final Partition
Number of clusters: 2


                          Within   Average   Maximum
                         cluster  distance  distance
             Number of    sum of      from      from
          observations   squares  centroid  centroid
Cluster1            36     425.5   3.41374   4.08418
Cluster2             1       0.0   0.00000   0.00000
```

Table F3. Dolores River Basin Data and Cluster Analysis Graph

STATIONID	MAINSTEM	STATIONID	TRIBUTARY
61	Dolores River	10906	La Sal Creek
10700	Dolores River	10916	Mesa Creek
RW-290	Dolores River	10917	North Fork Mesa Creek
RW-725	Dolores River	10922	Ute Creek
RW-742	Dolores River	10980	Roc Creek
		10905G	West Paradox Creek
		RW-750	West Creek

DOLORES BASIN - COMMON GENUS		
Ablabesmyia	Hemerodromia	Petrophila
Acentrella	Heptagenia	Choroterpes
Argia	Hesperoperla pacifica	Polypedilum
Atrichopogon	Hexatoma	Rhithrogena
Baetis	Hydropsyche	Simulium
Callibaetis	Hydroptila	Smicridea
Cordulegaster	Isoperla	Sperchon
Corydalus	Cryptochironomus	Stilobezzia
Cheumatopsyche	Microcylloepus	Taenionema
Cricotopus	Micropsectra	Lebertia
Ephemerella	Optioservus	Thienemannimyia
Eukiefferiella	Orthocladius	Tipula
Fallceon quilleri	Parakiefferiella	Tricorythodes
Helichus	Parametriocnemus	



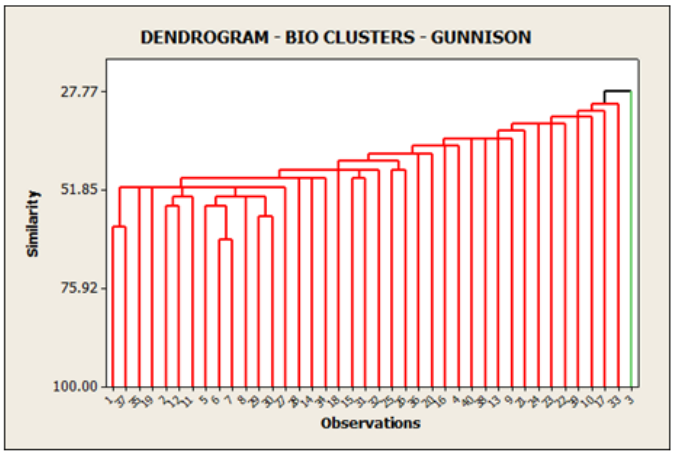

```
Final Partition
Number of clusters: 2


                          Within   Average   Maximum
                         cluster  distance  distance
              Number of   sum of      from      from
           observations  squares  centroid  centroid
Cluster1             10     69.4   2.62430   3.00998
Cluster2              2      5.5   1.65831   1.65831
```

Table F4. Gunnison River Basin Data and Cluster Analysis Graph

STATIONID	WATERBODY	STATIONID	WATERBODY
9128000	Gunnison River	10125	Smith Fork
9152500	Gunnison River	10212	Smith Fork
383103106594200	Gunnison River	10215	Smith Fork
FW08CO025	Gunnison River	10227	Curecanti Creek
RW-261	Gunnison River	10230	Cimarron River
RW-747	Gunnison River	10232	Pine Creek
RW-753	Gunnison River	10235	Willow Creek
RW-7898	Gunnison River	10240	Lake Fork Gunnison River
UGRWCD-1	Gunnison River	10249	Soap Creek
UGRWCD-6	Gunnison River	10250	Soap Creek
WCOP99-0624	Gunnison River	10262	Beaver Creek
WCOP99-0648	Gunnison River	10270	Ohio Creek
		10280	Cebolla Creek
		10282	Cebolla Creek
		10551	Escalante Creek
		10558	North Fork Escalante Creek
		10559	Middle Fork Escalante Creek
		10567	Potter Creek
		10570	Kannah Creek
		10575	Kannah Creek
		10586	Milk Creek
		10589	Oak Creek
		10271A	South Beaver Creek
		CO054M	Powderhorn Creek
		CO139M	Antelope Creek
		FW08CO023	Crystal Creek
		RW-332	Surface Creek
		UGRWCD-2	Lake Fork Gunnison River

GUNNISON BASIN - COMMON GENUS			
Acentrella	Diphetor hageni	Nais	Pteronarcella
Agapetus	Drunella grandis	Nematoda	Rheocricotopus
Ameletus	Epeorus	Neoplasta	Rheotanytarsus
Antocha	Ephemerella	Oecetis	Rhithrogena
Arctopsyche	Eukiefferiella	Optioservus	Rhyacophila
Atherix	Glossosoma	Orthocladius	Simulium
Attenella	Helicopsyche	Pagastia	Skwala americana
Baetis	Hemerodromia	Parakiefferiella	Sperchon
Brachycentrus	Hesperoperla pacifica	Paraleptophlebia	Stempellinella
Cardiocladius	Hexatoma	Parametriocnemus	Sweltsa
Chelifera	Hydropsyche	Pentaneura	Tanytarsus
Cheumatopsyche	Hydroptila	Pericoma	Thienemanniella
Chironomus	Hygrobates	Phaenopsectra	Thienemannimyia
Claassenia sabulosa	Lebertia	Physa	Torrenticola
Corynoneura	Lepidostoma	Pisidium	Tricorythodes
Cricotopus	Microcylloepus	Polypedilum	Tvetenia
Cryptochironomus	Micropsectra	Protzia	Zaitzevia
Dasyhelea	Microtendipes	Psychomyia flavida	




```
Final Partition
Number of clusters: 2


                                   Average   Maximum
                          Within  distance  distance
          Number of  cluster sum      from      from
       observations  of squares  centroid  centroid
Cluster1        39     481.385   3.48733   4.36071
Cluster2         1       0.000   0.00000   0.00000
```