The West Virginia GLIMPSS

Genus Level Index of Most Probable Stream Status

A Benthic Macroinvertebrate Index of Biotic Integrity for West Virginia's Wadeable Streams

The West Virginia GLIMPSS

Genus Level Index of Most Probable Stream Status

A Benthic Macroinvertebrate Index of Biotic Integrity for West Virginia's Wadeable Streams

Prepared by:

Gregory J. Pond

U.S. EPA Region III
Environmental Assessment and Innovation Division
Office of Monitoring and Assessment
Wheeling, WV

Jeffrey E. Bailey
and
Benjamin Lowman
and
Michael J. Whitman

West Virginia Department of Environmental Protection
Division of Water and Waste Management
Watershed Assessment Branch
Charleston, WV

Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgements

The monitoring staff from WVDEP's Watershed Assessment Branch was involved in the collection of benthic, chemical, and habitat data used in the development of the GLIMPSS. Extended thanks go to Patrick Campbell and John Wirts for programmatic support. U.S. EPA personnel involved in field support, taxonomy, and review included Margaret Passmore, Frank Borsuk, Louis Reynolds, Kelly Krock, and Carole Rose.

Early draft input by Dr. Ben Stout, Dr. J. Todd Petty, and Karen Blocksom.

WVDEP personnel involved in internal review included Nicholas Murray, Janice Smithson Ryan Pack, Mathias Hickman, and Sarah Kuhn.

External technical reviews were completed by:

Dr. Mindy Yeager Armstead-Senior Scientist, Potesta and Associates, Inc.
Karen Blocksom-Biostatistician, USEPA-Office of Research and Development
Dr. Jeroen Gerritsen-Aquatic Ecologist, Center of Ecological Sciences-Tetra Tech, Inc.
Dr. Nathaniel P. Hitt, USGS Leetown Science Center-Aquatic Ecology Branch
Dr. Tom Jones-Associate Professor of Integrated Sciences and Technology-Marshall University
Ed J. Kirk-Director, Biological Division-REI Consultants, Inc.
Dr. George T. Merovich Jr.-Research Assistant Professor, Wildlife and Fisheries Resources-West Virginia University
Dr. J. Todd Petty-Associate Professor, Forestry and Natural Resources-West Virginia University

Suggested Citation:

Pond, G.J., J.E. Bailey, B. Lowman, and M. J. Whitman. 2011. The West Virginia GLIMPSS (genus-level index of most probable stream status): a benthic macroinvertebrate index of biotic integrity for West Virginia's wadeable streams. West Virginia Department of Environmental Protection, Division of Water and Waste Management, Watershed Assessment Branch, Charleston, WV.
DOI: 10.13140/RG.2.1.4536.3682

On the cover: Otter Creek (WVMC-60-F) near its mouth in Otter Creek Wilderness, Monongahela National Forest, WV. Photo by Michael Whitman.

On the back: Hickman Creek (WVO-77-O-1), Marshall County, WV. Photo by Natalie Mancuso.

Executive Summary

The Watershed Assessment Branch (WAB) of the West Virginia Department of Environmental Protection (WVDEP) is charged with monitoring the State's waters in order to address the rules outlined in the Clean Water Act for assessing aquatic life uses. This report documents the development of a Genus-Level Index of Most Probable Stream Status (GLIMPSS) for benthic macroinvertebrates that can distinguish between reference (least-disturbed) and environmentally stressed benthic communities inhabiting wadeable streams with riffle-run habitat that are not ephemeral in West Virginia. The index covers most of the streams in WV; however, an index for true limestone streams (which represent only a small fraction of eastern WV streams) will be developed in a subsequent analysis. Macroinvertebrates and their role as bioindicators are paramount in stream assessments. As such, the WAB, in conjunction with USEPA Region III, has sought to improve existing methods of stream health evaluation in the State through benthic multi-metric index (MMI) development. Therefore, a benthic macroinvertebrate-based stream health index was created, incorporating genus-level taxonomic information and representing regional and seasonal potential.

The development of reference conditions is a key component of stream bioassessments. The regional reference approach is based on the range of conditions found in a population of sites or streams with similar physical characteristics and minimal human impact. Currently, the WAB uses Level III ecoregions as a framework to establish reference conditions that are used to interpret regional differences in benthic macroinvertebrate communities. Within this framework, several combinations of regions and seasons were evaluated to optimize benthic macroinvertebrate classification and MMI performance.

An EPA recommended procedure for testing the sensitivity of indicator metrics is to compare the range of values among all reference sites to a population of sites known to be stressed by chemical or physical factors. Overall, a total of 3737 sample sites were filtered from the database, which excludes larger rivers, winter samples, duplicate samples, and re-visits. These data represented 391 reference (REF) and 962 stressed (STRESS) sample sites. This dataset was randomly divided into calibration (CAL; 70% of all sites) and validation (VAL; remaining 30% of all sites). The resulting numbers of sites are shown in the table below:

	REF	Non-REF	STRESS
Calibration	273	1669	674
Validation	118	715	288
Totals	391	2384	962

Separate GLIMPSS indices were developed for different strata that included seasonal, regional, and stream size groupings. Within each stratum, two separate GLIMPSS were developed differing only in the treatment of a somewhat difficult group of dipterans known as Chironomidae (non-biting midges). The rationale for developing two indices was to provide an index that could be used for datasets that either included or lacked genus-level chironomid identifications, could distinguish REF from STRESS, was responsive, and lacked redundancy. These two indices were strikingly similar in their overall performance; details on the GLIMPSS (Chironomidae Family, or CF), are reported in Appendix G.

Classification of benthic communities was carried out with Non-metric multidimensional scaling and mean similarity analysis. We found that a combination of ecoregions and season produced the best separation of communities and were used to develop region and season-specific GLIMPSS. Regions include Mountains (ecoregion 67 and 69) and Plateau (ecoregion 70), and seasons include Spring, Summer, and Winter. The Fall season was not included since WAB sampled very few sites in this season. Within each of these strata, we tested 41 metrics that spanned a wide scope of ecological attributes (richness and composition, tolerance, feeding, habit, dominance/diversity). Metrics were tested for discrimination efficiency (sensitivity), correlation to stressors combined by PCA (response), correlation to other metrics (redundancy), range and variability. GLIMPSS for Spring/Winter and Summer Mountain strata each include 10 metrics, Plateau Spring uses 8 metrics, and Plateau Summer uses 9 metrics. A summer GLIMPSS for larger mountain streams and rivers (>60 sq. mi.) uses 7 metrics. Within each stratum, best standard values (BSVs) and worst standard values (WSVs) (*i.e.*, as ceilings and floors using 95th or 5th percentiles) were calculated from the dataset and used to score individual metrics. The GLIMPSS score was calculated as the average score of all metrics in the stratum. Site scoring examples are provided in Appendix D.

Overall discrimination efficiency of the CAL GLIMPSS was ~80%, while classification efficiency of the VAL dataset was ~90%. Both CAL and VAL scores responded similarly to the PCA stressor gradient indicating that GLIMPSS performance was highly repeatable with independent data. Metric BSVs and WSVs were then calculated from the full dataset and final GLIMPSS scores were re-calculated with these final SVs. This was done for both GLIMPSS and modified GLIMPSS (CF).

Criteria used to assess individuals sites are based on the calculated 5th percentile of the reference distribution within each stratum. Since metric scoring and the actual reference distributions differ across strata, it is impossible to directly compare GLIMPSS scores (0-100) between the strata (season and region). However, sample scores collected in different seasons or regions can be compared by calculating a “percent of threshold” value for each sample. Examples of this simple procedure are provided in Appendix D.

The GLIMPSS is a powerful yet practical tool for evaluating stream conditions and aquatic life uses. Improvements over the family-level WVSCI were noted as benthic assessments using genus-level taxonomy provided WVDEP with distinct seasonal and geographical classification strata to help refine aquatic life uses and ecological expectations across the State. Refinement of the WAB’s methodology to bioassess wadeable streams includes benefits that will apply to a broad spectrum of management programs including the following:

- characterizing the existence and extent of point and nonpoint source stressors;
- targeting and prioritizing watersheds for remedial or preventive programs;
- evaluating the effectiveness of nonpoint source best management programs; and
- assessing ecosystems for use attainability.

1.0 Introduction

The Federal Water Pollution Control Act, also known as the Clean Water Act (CWA) of 1972, is a comprehensive ruling aimed at restoring and maintaining the chemical, physical, and biological integrity of the nation's waters. The Watershed Assessment Branch (WAB) of the West Virginia Department of Environmental Protection (WVDEP) is charged with monitoring the State's waters in order to address the rules outlined in the CWA. To accomplish this, the WAB collaborated with the United States Environmental Protection Agency (USEPA) Region III to develop a multi-metric index for biological assessments called the West Virginia Stream Condition Index (WVSCI) (Gerritsen et al. 2000a). The WVSCI summarizes family level identifications of benthic macroinvertebrate assemblages to "bioassess" the condition of wadeable streams. This index includes six biological metrics that represent elements of the structure and composition of benthic macroinvertebrate communities. Since its publication in 2000, the WAB has used the WVSCI to characterize patterns of stream degradation and measure biological impairment of "Designated Aquatic Life Uses" under the CWA. Furthermore, the WVSCI is the quantitative measure that the WAB uses to interpret the narrative water quality standard which states that "…no significant adverse impact to the chemical, physical, hydrologic, or biological components of aquatic ecosystems shall be allowed" (WV Code of State Rules, 47CSR2-3.2.i).

Although the family-level WVSCI was designed using sound ecological and statistical principles and has typically met the needs of the WAB, the availability of genus-level benthic macroinvertebrate data has led to the development of a more accurate tool for measuring biological impairment in wadeable streams. While the scientific debate over the cost-effectiveness of finer taxonomic resolution continues, it is widely accepted that genus or species-level data more accurately represent the "aquatic life" that the CWA intends to protect. Moreover, most research acknowledges that surrogate family-level data can detect obvious impacts to streams (Lenat and Resh 2001), but genus and species-level information can detect subtle effects as well (Waite et al. 2004, Arscott et al. 2006). This report documents the development of a statewide Genus-Level Index of Most Probable Stream Status (GLIMPSS) for benthic macroinvertebrates that can distinguish between reference (least-disturbed) and environmentally stressed benthic communities within West Virginia streams. The index covers most of WV; however, an index for true limestone streams (which represent only a small fraction of eastern WV streams) will be developed in a subsequent analysis.

As with the WVSCI, there are several benefits expected from using the GLIMPSS to bioassess wadeable streams. These benefits will apply to a broad spectrum of management programs including the following:

- characterizing the existence and extent of point and nonpoint source stressors;
- targeting and prioritizing watersheds for remedial or preventive programs;
- evaluating the effectiveness of nonpoint source best management programs; and
- assessing ecosystems for use attainability.

2.0 Benthic Macroinvertebrates as Indicators

For decades, benthic macroinvertebrates have served as long-term indicators of biological health in flowing waters (Carter et al. 1996). The CWA specifically defines laws for states and tribes to protect aquatic life through establishing designated uses and water quality criteria. Aquatic insects are a functionally irreplaceable component of stream ecosystems (Rosenberg and Resh 1993). However, limitations in the use of benthic macroinvertebrates as a measure of ecological health have resulted from a lack of taxonomic resolution (*i.e.*, the level to which the study organisms were identified) (Waite et al. 2004), limited knowledge of many organisms' life histories (Rosenberg and Resh 1993), and unique seasonal changes experienced by aquatic insect populations (Merritt and Cummins 1996). Additionally, topographical extremes, geological complexities, and attributes of the surrounding landscape influence a region's macrobenthos, and make predictive studies difficult (Johnson et al. 2004; Townsend et al. 2003; Vinson and Hawkins 2003). Yet, regardless of the challenges presented by their study, macroinvertebrates and their role as bioindicators are paramount in stream assessments. As such, the WAB, in conjunction with USEPA Region III, has sought to improve existing methods of stream health evaluation in the State through benthic multi-metric index (MMI) development. Therefore, a benthic macroinvertebrate-based stream health index, incorporating detailed taxonomic information and representing ecoregion, seasonal potential, and stressor response was initiated—the Genus-Level Index of Most Probable Stream Status (GLIMPSS).

West Virginia's diverse geologic and topographic features have facilitated colonization by a specialized flora and fauna within its streams. The diversity that exists among macrobenthos, particularly insects, not only lends itself to the research of species attributes but is also well-suited to biogeographical studies. Physiography, a term applied to broadly-synthesized environmental variables, has been used, for example, to demarcate conditions between highland, montane, and lowland/plateau regions, like those found across West Virginia (Jezerinac et al. 1995). Biologists have long recognized that selective pressures leading to speciation and subsequent diversity are introduced by such physiographic conditions, among other variables (Beauchard et al. 2003). In preparation for the refinement of analytical tools involved in the biological monitoring of streams, the WAB has assembled nearly 5,000 benthic macroinvertebrate samples since 1999, which have been taxonomically classified to the genus level. Concurrent with the biological collections, assessments of a stream's physical condition, including abiotic factors and water chemistry analyses, were also performed.

Biological diversity increases with habitat complexity, thus an equivalent account of taxonomic information is warranted in biotic integrity analyses of dynamic environs (Stanford and Ward 1983; Vinson and Hawkins 2003). At coarser levels of identification, many organisms are similar in function and tolerance; however, at more detailed taxonomic levels, important differences are often revealed (Doledec et al. 1999; Lenat and Resh 2001; Bady et al. 2005). Improvements in biological integrity indices often result from more concise classification of the study organisms into lower taxonomic units and use of the additional taxon-specific information (Resh and McElravy 1993; Thompson and Townsend 2000; Lenat and Resh 2001). In fact, an improvement in the ability to identify impaired communities (*e.g.*, those based on family-level identifications) was a primary impetus for this research. Statistical measures applied to the genus-level dataset were expected to be more predictive of biological integrity, both through the distinction of reference communities from assemblages incurring stress, as well as through detection of more subtle community responses to environmental features (Waite et al. 2004).

The partitioning of macroinvertebrate communities driven by ecoregional distributions (Gerritsen et al. 2000b, Beauchard et al. 2003), or the biogeography of the organisms, known by biologists to occur within the statewide benthic fauna was a central theme in development of this improved index. Differentiation of communities collected within distinct ecoregional boundaries (*e.g.*, Woods et al. 1996) has often served as a means to partition biological assemblages with like attributes and potential. The use of more taxonomically resolute metrics (*i.e.*, genus-level) was expected to make comparative evaluations between similarly structured assemblages more meaningful.

Common Macroinvertebrate Taxa in WV Streams

Ecoregional groupings also served as an opportunity for incorporation of seasonal emergence information, a most unique (and analytically difficult) feature of the aquatic macroinvertebrate life cycle. As flying insects, many benthic macroinvertebrates are capable of dispersion into distant environs; however, they are limited in colonization potential by the instream conditions to which offspring are subjected (Merritt and Cummins 1996). Yet, through the life history requirement of vacating the water at a certain time, macroinvertebrate collections often reflect phenological emergence patterns (Dobrin and Giberson 2003; Merritt and Cummins 1996; Stark et al. 1998). For example, winter stoneflies (Taeniopterygidae) may be absent from benthic collections made during late spring or summer due to the timing of their emergence, or hatch (Stark et al. 1998). A potential remedy to account for seasonal variations in community structure is through multiple collections representing distinct seasons, replicated at specific stations (Norris and Georges, 1993). Once sufficiently documented, seasonal differences in macroinvertebrate community structure can not only be accounted for in index calculations, but also provides a means for more accurate assessments of collections made at various times throughout the year. The ability to evaluate benthic communities that are representative of distinct seasons was an inherent goal of this process, and was accomplished through a study of seasonal macroinvertebrate community structure conducted by the WAB.

A final goal of improved index development involved diagnostic abilities in assemblages subsidized by stressors (Perrin and Richardson 1997). For example, a select group of organisms with distinct functional characters, capable of proliferation in degraded conditions, may be indicative of stress via shared biological traits and may not be accounted for in more traditional indices (Rueda et al. 2002). Since taxonomy is not always reflective of functional behaviors, entomological knowledge of species/group attributes was integrated into the development of the GLIMPSS. In comparison to benthic indices excluding such considerations, organisms capable of subsidizing resources may also be appropriately accounted for through quantitative measures. Equally important, tolerant and ubiquitous genera may be segregated from related taxa. It is for these reasons that the WAB has pursued genus-level taxonomy for stream assessments.

3.0 Geographic Setting

3.1 General Physiography

The study region examined for the GLIMPSS included the entire area within the bounds of West Virginia. West Virginia is the third most forested state in the United States, with forests covering about 78.0% of the State's 24,282 square land miles (Childs 2005). The mean elevation is 1,500 feet, higher than any other state east of the Mississippi River. West Virginia's highest point, Spruce Knob, reaches 4,862 feet (1482 m) above sea level (Stephenson 1993). Major lowlands lie along the larger river basins, especially the Potomac, Ohio, and Kanawha. A point on the Potomac River near Harpers Ferry has the lowest elevation in West Virginia (240 feet above sea level). Such variety in elevation contrasts the steep, rugged streams of the high mountains in the eastern counties to the gentler, meandering streams draining the lowlands in the western portions of the State (Figure 1).

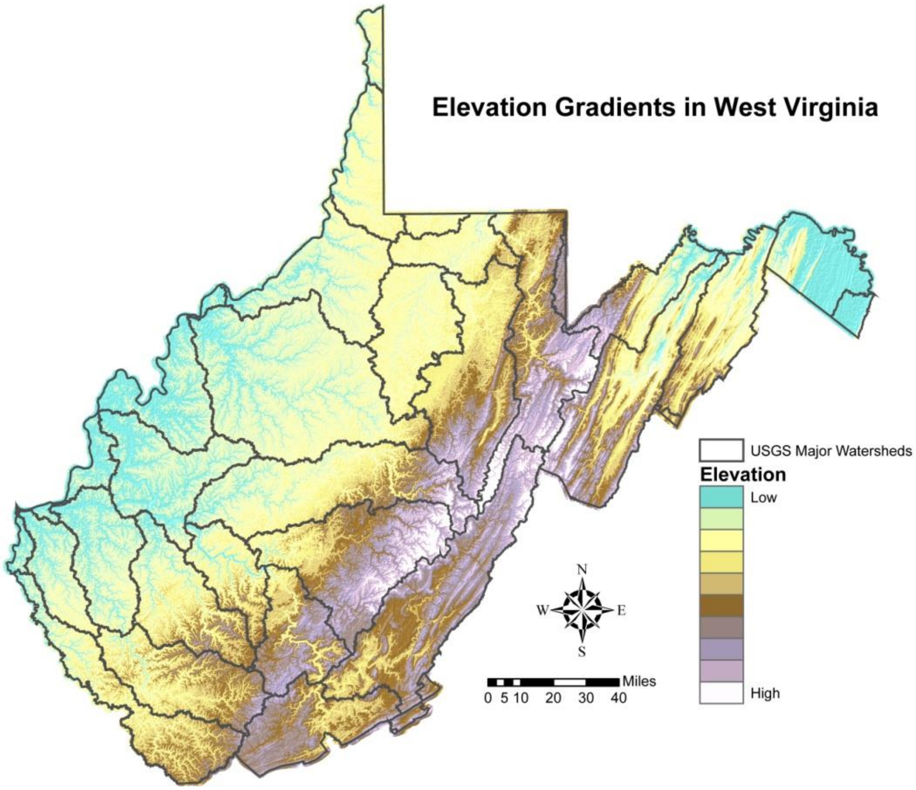


Figure 1. Elevation gradients and major watersheds found across West Virginia.

At the 1:24,000 NHD (National Hydrography Dataset) mapping scale, West Virginia has about 55,000 miles of rivers and streams, most of which are wadeable. The WAB uses the U.S. Geological Survey (USGS) scheme of hydrologic units to divide the State's streams into 32 major watersheds (Figure 2). These watershed units include entire stream basins bounded by natural hydrologic divides, clusters of small tributaries that drain directly into larger mainstem streams, West Virginia parts of interstate basins, and divisions of large watersheds into smaller units.

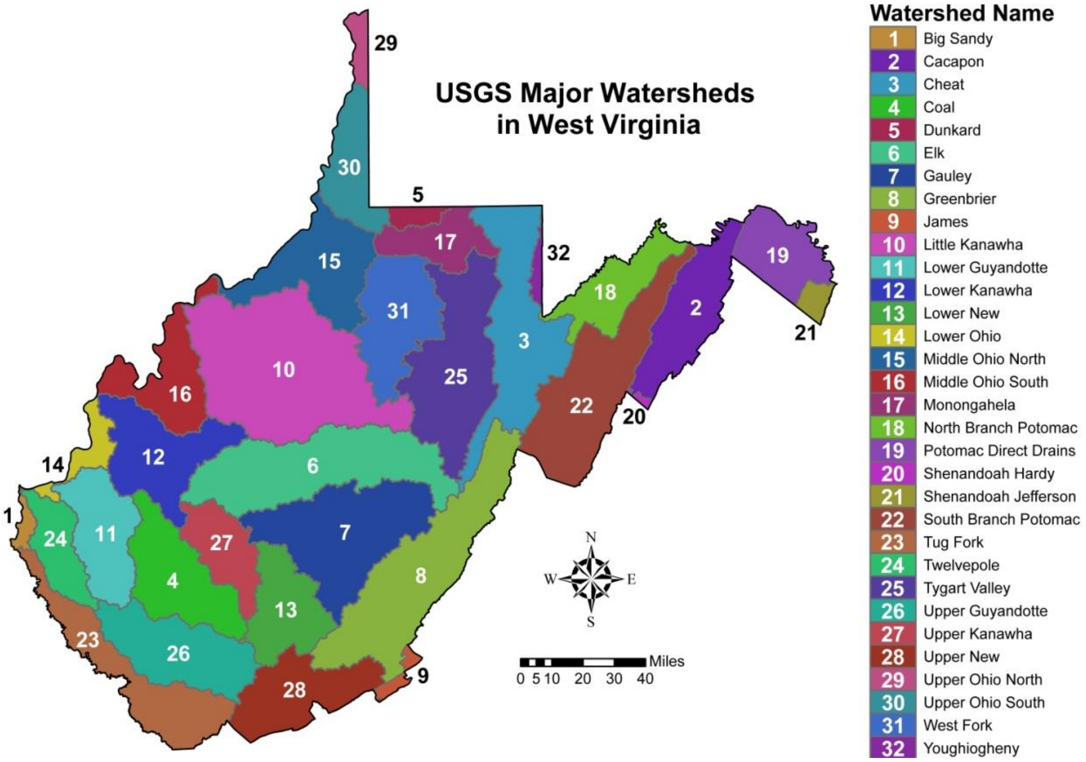


Figure 2. Major watersheds of West Virginia

3.2 Geologic Formations

Most of western and central West Virginia, which makes up about three-fourths of the State, is composed of cyclic sequences of relatively flat-lying Mississippian- and Pennsylvania-age shales, sandstones, limestones, and coals, punctuated by relatively gentle anticlines and synclines. Stream drainage is dendritic in nature, and it is the long-term erosion by the streams and rivers that gives the area its relief, not the folding or faulting of the geological strata. Most of this area is rugged, with steep hillsides and narrow river valleys; however, this ruggedness moderates in the southeastern part of the State where the limestones of the Greenbrier Group become the dominant bedrock. Here, there are wide valleys and flatter highlands, and farming dominates the land use (Dasher 2001).

In eastern West Virginia the rocks are folded into a series of tight anticlines and synclines. The valleys are wider and much more linear than on the plateau, with larger mountains in between, and relief of 1,500 feet is not uncommon. Drainages are trellis in nature, and the streams and rivers flow predominantly to the north-northeast. This area is a mixture of older Ordovician and Cambrian limestones, dolomites, and shales, all of which have been severely folded and faulted. Some of this area is a part of the lower Shenandoah Valley, and contains the most complex geology within West Virginia. The lower Shenandoah Valley and Opequon Creek region is comprised mostly of karst (Dasher 2001).

The extreme eastern edge of West Virginia's Eastern Panhandle is comprised of over-thrusted Cambrian metamorphic rock. This geologic formation is very narrow in West Virginia, and is only found east of the Shenandoah River (Dasher 2001).

4.0 Stream Classifications

In order to maximize the performance of the GLIMPSS, aspects of natural variability that are known to influence benthic macroinvertebrate communities were evaluated. This included an examination of the variability associated with regions (*i.e.*, ecoregions), seasons (*e.g.*, spring, summer, winter), and stream size (*e.g.*, watershed area, stream width).

4.1 Classification by Region

A convenient and ecologically relevant method to account for natural environmental variability on a large spatial scale is to delineate using a regional classification scheme like ecoregions (Omernik 1987). Ecoregions are based on the premise that ecological regions can be identified through the analysis of the patterns and the composition of biotic and abiotic factors that affect or reflect differences in ecosystem quality and integrity. These factors include geology, physiography, hydrology, vegetation, climate, soils, land use, and wildlife.

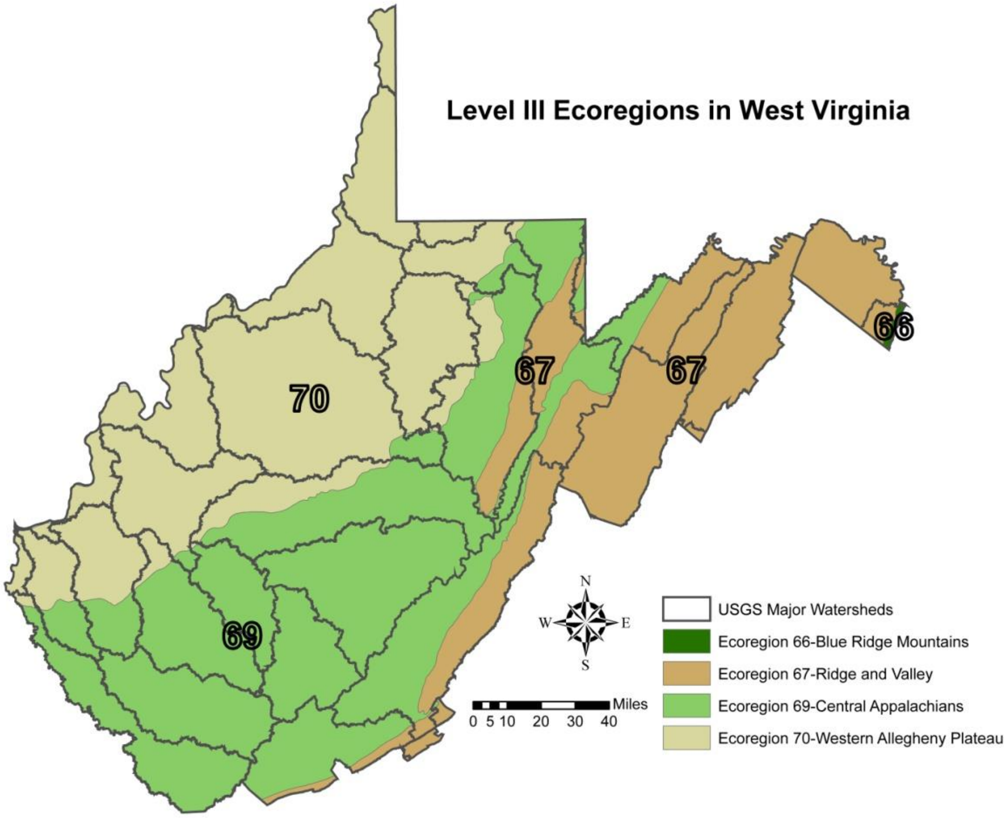


Figure 3. Level III ecoregions delineated in West Virginia.

Level III ecoregions of the United States were originally defined by Omernik (1987) and later modified (U.S. EPA 2000). West Virginia includes portions of four Level III ecoregions including the Blue Ridge Mountains (66), Ridge and Valley (67), Central Appalachians (69), and Western Allegheny Plateau (70) (Figure 3).

The Blue Ridge Mountains (66) ecoregion extends from southern Pennsylvania to northern Georgia, varying from narrow ridges to hilly plateaus to more massive mountainous areas, with high peaks reaching over 2000 meters. This ecoregion is insignificant in West Virginia, as it covers only a small area in the eastern panhandle of Jefferson County primarily to the east of the Shenandoah River. It is characterized by forested slopes, high-gradient, cool, clear streams, and rugged terrain. It is one of the most floristically diverse ecoregions, and includes Appalachian oak forests, northern hardwoods, and, at the highest elevations, Southeastern spruce-fir forests. Shrub, grass, and heath balds, hemlock, cove hardwoods, and oak-pine communities are also significant. Due to its very limited extent in WV, this ecoregion was analyzed with ecoregion 67 due to its geographic proximity.

The Ridge and Valley (67) ecoregion is a northeast-southwest trending, relatively low-lying, but diverse ecoregion sandwiched between generally higher, more rugged mountainous regions with greater forest cover. As a result of extreme folding and faulting events, the region's roughly parallel ridges and valleys have a variety of widths, heights, and geologic materials, including limestone, dolomite, shale, siltstone, sandstone, chert, mudstone, and marble. Springs and caves are relatively numerous. Present-day forests cover the majority of the region. The ecoregion has a diversity of aquatic habitats and species of fish. Agriculture is common in the valleys of this ecoregion in West Virginia**.**

The Central Appalachians (69) ecoregion, stretching from central Pennsylvania to northern Tennessee, is primarily a high, dissected, rugged plateau composed of sandstone, shale, conglomerate, and coal. The rugged terrain, cool climate, and infertile soils limit agriculture, resulting in a mostly forested land cover. The high hills and low mountains are covered by a mixed mesophytic forest with areas of Appalachian oak and northern hardwood forest.

The hilly and wooded terrain of the Western Allegheny Plateau (70) ecoregion was not muted by glaciation and is less rugged and not as forested as ecoregion 69 to the east and south. Extensive mixed mesophytic forests and mixed oak forests originally grew in the Western Allegheny Plateau. Today, most of its rounded hills remain in forest; dairy, livestock, and general farms as well as residential developments are concentrated in the valleys. Horizontally-bedded sedimentary rock underlying the region has been mined for bituminous coal. Wadeable streams typically have lower gradients here than streams in ecoregions 67 and 69.

During the development of the family level WVSCI, classification by ecoregions was not supported by the analysis of the benthic assemblages from cobble substrate in wadeable streams in West Virginia (Gerritsen et al. 2000). However, it was stated that using ecoregions to stratify or partition the aquatic community might be valid if the level of taxonomy changes, *i.e.*, identifying to the genus level of taxonomy.

4.2 Classification by Season

Seasonal variability in benthic macroinvertebrate communities is well known and can be attributed to life cycles and emergence of resident species, the availability of food and resources, and changes in habitat and the local environment. Several studies have related the timing of life cycle events and invertebrate size at emergence with variables such as temperature, photoperiod, food resources, and discharge (Sweeney 1984).

Greens Run (WVMC-16) in the Winter and Summer.

During the development and application of the family level WVSCI, the WAB recognized the potential variability associated with seasons and benthic macroinvertebrate assemblages. Therefore, analyses were performed to evaluate the influence of sample date. It was determined that the broad collection timeframe (mid-April to early-October) of the WAB benthic data introduced variability, but there was no clear differentiation of sampling periods discernable for the family level data used in the WVSCI (Gerritsen et al. 2000). However, further analysis suggested that a narrower sampling window of late spring to early summer would improve the precision of the index assessments by reducing variability. Using seasons to stratify or partition the benthic macroinvertebrate community for GLIMPSS development is valid, particularly when considering the change in taxonomy from family (WVSCI) to genus level (GLIMPSS).

4.3 Classification by Size

West Virginia is a headwater State and most of its streams are small (< 7 square miles drainage area) and wadeable. In the Strahler (1957) ordering system, they would be designated as 1^{st}, 2^{nd}, and some 3^{rd} order streams. These streams are an extremely important water resource in West Virginia as it has been estimated that around 90% of the stream miles fit into these small orders. An important function of headwater streams is the maintenance of the ecological health of larger streams and rivers by controlling sediment deposition, nutrient enrichment, and by lessening the impacts of flooding (Meyer and Wallace 2001; OH EPA 2002; Alexander et al. 2007). Headwater streams also provide habitat for specialized wildlife and may supply clean water for human consumption.

In the undisturbed condition, most of these small streams are riffle/run dominated with cobble and coarse gravel/boulder substrates. They are typically narrow, shallow, cool, and heavily shaded. In general, they are low in nutrients and dissolved ion concentrations. They are dependent on allochthonous organic material such leaves, sticks, and large woody debris for energy which is provided by the dense forests that make up the riparian areas. By contrast, larger streams might have more open canopies, have greater diel flux of temperature, contain higher dissolved ions and nutrients, exhibit different habitat conditions, and offer different food types to invertebrate consumers.

Example of a small 1st order stream and a larger 5th order stream.

In general, small streams represent the majority of the biological assessment efforts conducted by the WAB. However, the WAB routinely conducts assessments on larger streams with watershed areas exceeding 500 square miles (sq. mi.). Because it is well known that aquatic communities change longitudinally from headwaters to large rivers, the WAB felt it was important to consider stream size during GLIMPSS testing. A major tenet describing this phenomenon, called the River Continuum Concept, states that relatively predictable gradients of biological and chemical processes correspond to a stream’s physical attributes, and that the structure and function of biological communities changes in a downstream direction (Vannote et al. 1980). Phenomena related to the observed longitudinal changes in natural streams should be considered when developing biological monitoring and assessment tools.

5.0 Reference Conditions

Reference conditions represent the characteristics of stream reaches that are least disturbed by human activities and are used to define benchmarks for chemical, biological, and habitat conditions for a region. The development of reference conditions is a key component of stream bioassessments. In West Virginia streams, historical data were not collected prior to human disturbances and activities. Therefore, a logical method of determining the health of streams is to compare them to reference conditions, accomplished by using a regional reference approach (Hughes 1995). The regional reference approach is based on the range of conditions found in a population of sites or streams with similar physical characteristics and minimal human impact. Currently, the WAB uses Level III ecoregions (Omernik 1987) as a framework to establish reference conditions that are used to interpret regional differences in benthic macroinvertebrate communities.

Typical reference sites: Back Fork/Elk River (WVKE-111) and Bee Run/Gandy Creek (WVMC-60-T-5).

In 1998, the WAB developed methodologies, including a list of reference site selection criteria, to select reference quality assessment sites (*i.e.*, minimally- or least-disturbed) and ultimately establish reference conditions for wadeable streams (Table 1). The WAB uses a combination of quantitative physical and chemical attributes and narrative criteria to identify reference quality streams. Additionally, candidate reference sites are selected by examining historic data (if available), and by consulting with regional professionals of various agencies and entities that have knowledge of their local streams. To be classified as reference, a site must meet all of the listed conditions. However, in areas where high quality reference sites are scarce, a site could be listed as reference even if it failed one or more of the criteria (especially if they are highly correlated). For example, a Rapid Bioassessment Protocol (RBP) habitat parameter that is one or two points below the criterion may still qualify as a reference site. Table 1 lists criteria used to screen reference sites.

Establishing reference sites throughout some regions of West Virginia can be difficult. For example, few relatively undisturbed streams exist in the Western Allegheny Plateau ecoregion of the State. Conversely, the Ridge and Valley and parts of the Central Appalachians have many relatively undisturbed streams (*i.e.*, minimally disturbed) located mostly in the mountains of the Monongahela National Forest. Therefore, the term “least disturbed” might describe more accurately the reference conditions in the Western Allegheny Plateau. Best professional judgment by experienced personnel is an important part of the initial and final selection of reference sites.

Table 1. Summary of West Virginia DEP reference site selection criteria for West Virginia[1].

	Parameter and Criterion	Explanation
1	D.O. > 5.0 mg/l	Taken from "WV Water Quality Standards" (47CSR2).
2	pH between 6.0 and 9.0 S.U.	Taken from "WV Water Quality Standards" (47CSR2).
3	Conductivity < 500 µmhos/cm	Criterion for conductivity was established from analysis of DEP data. A value > 500 may indicate the presence of dissolved ions exceeding the background levels for the area. A conductivity reading can be used as a means of flagging a site for further investigation before it can be considered a reference site.
4	Fecal coliform bacteria < 800 colonies/100 ml	Fecal coliform bacteria data is used as a means of flagging a site for further investigation before it can be considered a reference site.
5	Epifaunal substrate/ available fish cover ≥ 11	Lowest score possible for sub-optimal rating - USEPA-RBP habitat score - 0 to 20 point scale. [2]
6	Channel alteration ≥ 11	Lowest score possible for sub-optimal rating - USEPA-RBP habitat score - 0 to 20 point scale. [2]
7	Sediment deposition ≥ 11	Lowest score possible for sub-optimal rating - USEPA-RBP habitat score - 0 to 20 point scale. [2]
8	Bank vegetative protection (right bank ≥ 6 & left bank ≥ 6)	Lowest score possible for marginal rating - US EPA-RBP habitat score - 0 to 10 point scale for each bank. [2]
9	Undisturbed riparian vegetative zone width (right bank ≥ 6 & left bank ≥ 6)	Lowest score possible for marginal rating - US EPA-RBP habitat score - 0 to 10 point score for each bank. [2]
10	Total habitat score ≥ 130	Mid suboptimal score - U.S. EPA-RBP habitat score - 0 to 200 point scale. [2]
11	No known point source discharges upstream of assessment site (*i.e.*, NPDES)	GIS coverages provide easy access to locations of many permitted point sources. Field reconnaissance is also performed to ensure that point sources do not exist above the site.
12	Evaluation of anthropogenic activities and disturbances at the assessment site	Visual inspection is performed within the stream assessment area. Best professional judgment is employed to make reference site inclusions based on the number and type of disturbance(s). GIS coverages are also used to validate the reference sites.
13	No obvious sources of NPS (Non-Point Source) pollution near assessment site	Obvious sources of NPS are documented within the assessment area. If sources of NPS are documented for areas above the assessment site, they are also considered. Best professional judgment is employed to make reference site inclusions based on the type and intensity of the NPS.
14	No known violations of state water quality criteria	Because of their toxicity, metals are the primary consideration when evaluating data for violations. If there is a violation of a water quality criterion as set forth in 47CSR2, the site is eliminated from reference site consideration.

[1] As provided in "WVDEP Watershed Branch 2010 Standard Operating Procedures (WVDEP 2010). [2]EPA Rapid Bioassessment Protocols (RBP) Habitat Assessment scoring from Barbour et al. (1999).

6.0 Field and Laboratory Methods

Streams in West Virginia are predominantly high gradient with coarse substrate materials such as boulder, cobble, and gravel. These physical conditions are responsible for the typical riffle/run habitats commonly found in most areas of the State. Consequently, the data used to develop the GLIMPSS is based on benthic macroinvertebrate samples collected from riffle/run habitats in wadeable streams only. The WAB database currently contains over 4,700 samples with genus level identifications and associated water quality and habitat information.

The WVDEP WAB employs a method consistent with USEPA's protocols for conducting biological assessments of streams and rivers (Barbour et al. 1999). Field and laboratory protocols are briefly described below. A detailed description of the WAB's stream assessment procedure can be found in WVDEP Watershed Assessment Branch's Standard Operating Procedures (WVDEP 2010).

Prior to collecting benthic macroinvertebrates, a 100-meter assessment reach was established. Water quality samples were collected and habitat assessments were performed within the confines of the reach. Field water quality parameters collected at each site included dissolved oxygen (D.O.), pH, conductivity, and temperature. Additional water chemistry sampling (*e.g.*, nutrients, metals) was conducted at many of the sites used in this study.

Habitat quality was evaluated with the U.S. EPA Rapid Bioassessment Protocol (RBP) Habitat Assessment procedure following Barbour et al. (1999). This procedure evaluates important habitat components such as epifaunal substrate quantity and quality, embeddedness, velocity/depth regimes, sediment deposition, riffle frequency, channel flow status, channel alteration, stream bank stability, bank vegetative protection, and riparian zone width.

Benthic macroinvertebrates were collected in riffle/run habitats using a 0.5 meter wide rectangular frame kick net with 500 µm mesh openings. Four 0.25 m^2 kick samples were collected at each site and composited into one sample that represented approximately 1 square meter of stream bottom substrate. Larger stones within each 0.25 m^2 quadrat were gently scrubbed with a small brush to dislodge clinging organisms into the net. Quadrats were then kicked for approximately 20 s to an approximate depth of 10 cm. The samples were preserved in 95% ethanol and returned to the WAB's biology laboratory for sorting and identification.

Sorting involved placing the entire benthic sample into a gridded rectangular sieve and removing a random 200-organism (±20%) subsample. A fixed 200-count sample offers a compromise of laboratory cost-benefit and an equally sensitive ability to detect differences in reference versus degraded stream conditions compared to larger subsamples or full picks. Subsampling also eliminates the need for rarefaction of samples in order to standardize richness estimates across multiple streams or treatments. Studies on fixed-count sampling (Barbour and Gerritsen 1996; Vinson and Hawkins 1996; Sovell and Vondracek 1999) often showed asymptote curves of richness that reflect that 200 organisms provides a decent estimate of richness compared to larger subsamples. Moreover, Sovell and Vondracek (1999) found that abundance-based metrics did not differ significantly between 100, 150, 200, 250, and 300 fixed counts.

The organisms were identified to the genus level of taxonomy or lowest level possible. Taxonomic QA/QC (Quality Assurance/Quality Control) was performed on 5% of the identified samples. Some highly degraded sites (*e.g.*, acid mine drainage) did not yield the target number of organisms after picking the entire sample. For index development purposes only, we excluded these sites when <100 organisms were collected, but GLIMPSS scores will be applied in these highly degraded sites for assessment purposes.

7.0 Index Development

7.1 Data Set

All data are stored in WVDEP's WAB database, a relational database that includes biological, chemical, habitat, and land use information. Nearly 4600 genus-level macroinvertebrate samples spanning 1999-2009 were used in various portions of this analysis (see Figure 4). Because the index is intended to be applicable to all wadeable streams with riffle-run habitat that are not ephemeral, data were first filtered from the WAB database for all methodologically comparable samples (*e.g.*, benthos collection, processing, and identification, habitat scoring, in situ p-chem collection, and fecal coliform analysis). True limestone streams, which are known to have distinctive communities from non-limestone streams in the region (Botts 2009), were excluded from the dataset and are not discussed further; these stream types will be analyzed in future index development efforts. In addition, many samples containing too few individuals (<100) were also omitted from the dataset. While most of these low abundance sites indicate severe chemical or habitat impairment, they might also include samples influenced by drought or spate conditions. Therefore, benthic samples used in this analysis contained between 100-240 individuals.

Although WVDEP has sampled benthic communities in all calendar months, very few samples were collected between November and February. This late-fall to late-winter period was not used in index development but is explored in a separate section of this report (see Figure 4 and Section 8.7 *Winter Index Period Development*). Additionally, larger rivers were initially omitted from GLIMPSS development. Preliminary analyses by EPA and WVDEP biologists found that summer collections at sites with a catchment area of >60 sq. mi. (generally >20 m stream width) were different in terms of the plecopteran fauna (both abundance and richness) and genus-level Hilsenhoff Biotic Index (HBI). Therefore, by excluding these larger sites in the initial development phase, the GLIMPSS dataset ultimately included >90% of the sampled streams throughout West Virginia. These larger stream and river sites were examined separately with respect to the final GLIMPSS (see Figure 4 and Appendix E).

Same day duplicate samples, and any additional samples collected from the same site within a 5 year period were omitted from the development dataset as a means to control for pseudoreplication. Duplicates and annual re-visits were analyzed separately (see Index Precision Sections 8.9.1 and 8.9.2).

The final main data set (n=3737) used in this analysis represented 2354 uniquely named or coded streams and 3411 unique stations (see Table 2). Those streams that had more than one station located on them averaged a distance of 2.4 miles between stations. The data set also consisted of approximately 33.3% probabilistically selected sites and 66.7% target sites sampled from March to Early October at stream sites with catchment areas of <60 sq. mi.

Biological samples were not always accompanied by a full suite of environmental data. 1617 of the 3737 samples were subsequently used to generate a stressor gradient as they possessed a symmetric dataset of habitat and water quality parameters (see Section 7.3.2 *Response to Stress*).

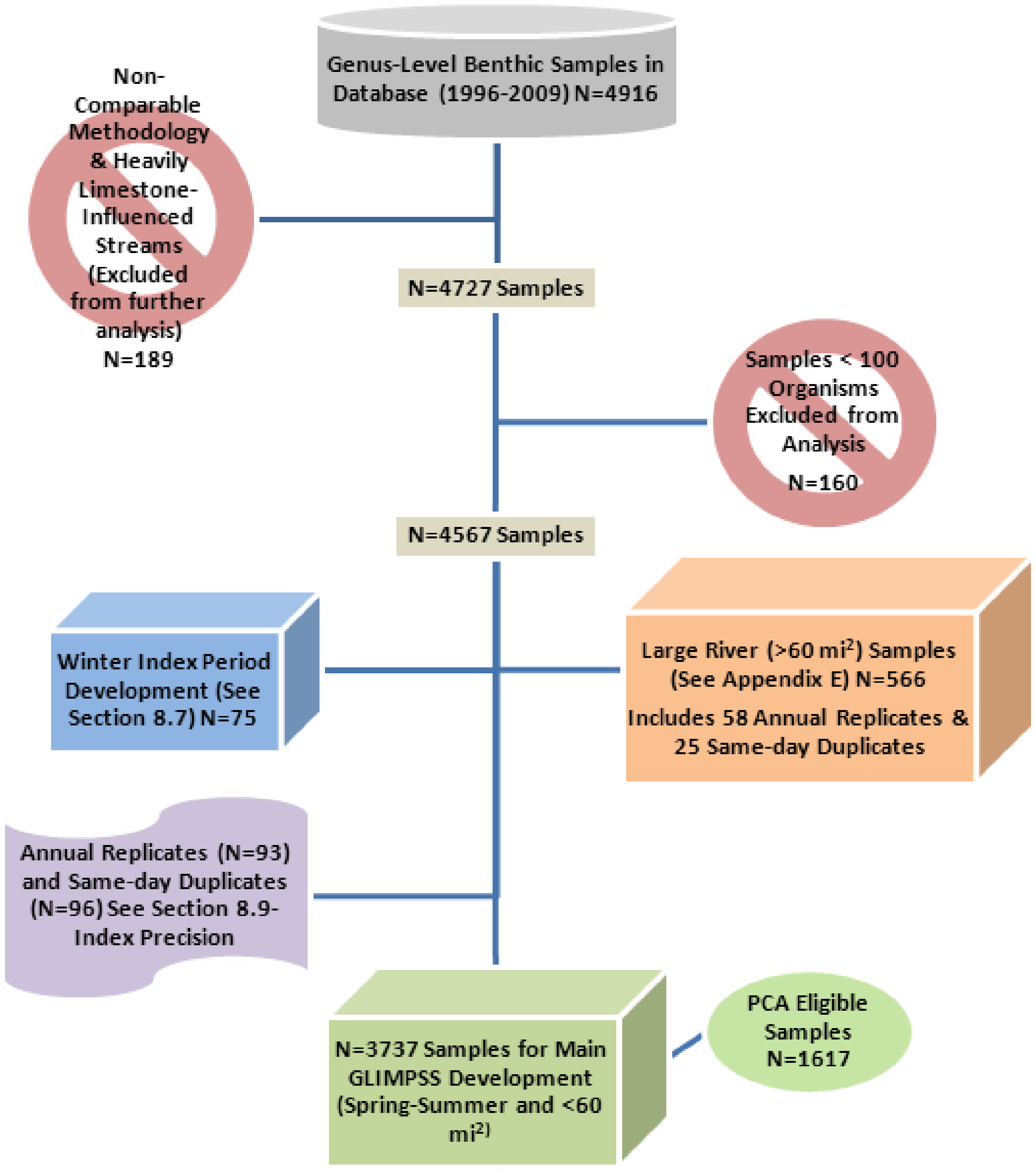


Figure 4. Conceptual diagram of the filtering and partitioning of the data set used in index development.

Table 2. Breakdown of Main Data Set Contents

Unique Streams (Names and Codes) Sampled	**2354**	Number of Streams Sampled only 1 time	**1596**
		Number of Streams Sampled >1 time (*Average Distance between Stations on a Unique Stream = 2.4 miles*)	**758**
Unique Stations Sampled	**3411**	Number of Stations Sampled only 1 time	**3102**
		Number of Stations Sampled >1 time (*>5 years apart*)	**309**
Unique Stations x Season	**3551**	Number of Stations Sampled in 1 Season Only	**3365**
		Number of Stations Sampled in 2 Seasons	**186**
Total Number of Samples	**3737**		

As described in Section 5.0, the development of a biological assessment multi-metric index is predicated upon the use of the reference condition approach (Barbour et al. 1999). An EPA recommended procedure for testing the sensitivity of indicator metrics is to compare the range of values among all reference sites to a population of sites known to be stressed by chemical or physical factors.

Following EPA guidance (Barbour et al. 1999) the data were divided into reference (REF), stressed (STRESS) and other, or non-reference (Non-REF) populations, and further divided into calibration (CAL) and validation (VAL) sets for index development and testing (see below). First, a database query was used to tabulate all sites qualifying as REF and STRESS sites. Sites that met REF criteria stated in Section 5.0 (Table 1) are currently specifically designated in the WAB database. STRESS sites were considered to be "abiotically" stressed (physically, chemically) if they met any of the database-filtered conditions shown in Table 3. Note that these values are similar to the original WVSCI STRESS site criteria (Gerritsen et al. 2000a) and cover a broad range of potential stressor response variables across West Virginia. Non-REF sites included all other sites that were not classified as either REF or STRESS and were used in combination with REF and STRESS sites for purposes of calculating stressor correlations (Section 8.2), metric standard values (Section 8.4), and in precision analyses (Section 8.9).

Table 3. Criteria for assigning stressed sample sites (STRESS); only an exceedance of any one of the criteria is required.

STRESS Site Classification Criteria	
pH	<4 or >9 S.U.
D.O.	<4mg/l
Fecal Coliform	>5000 col./100 ml
Specific Conductance	>1000 uS/cm
Epifaunal Substrate Score	<7 and Total Hab Score <120
Channel Alteration Score	<7 and Total Hab Score <120
Sediment Deposition Score	<7 and Total Hab Score <120
Total Bank Vegetation Score	<7 and Total Hab Score <120
Total Riparian Zone Score	<4 and Total Hab Score <120

Both pH and dissolved oxygen (D.O.) criteria represent the lower or upper physiological thresholds of many West Virginia taxa and are either above or below water quality standards for aquatic life uses (ALUs). Fecal coliform counts above this criterion indicate obvious organic enrichment from humans or domestic animals. Specific conductance incorporates many potentially harmful chemicals such as chloride and sulfate, and serves as an overall indicator of nutrient enrichment and urbanization (Dow and Zampella 2000, Paul and Meyer 2001, Black et al. 2004, Kratzner et al. 2006) and resource extraction effects (Rikard and Kunkle 1990, Pond 2004). The RBP habitat score and the selected indicator metrics elucidate non-chemical factors such as excess sediment, channelization, and riparian zone degradation (Barbour et al. 1999). Habitat metrics values below these criteria indicate low-marginal, to poor quality.

Overall, a total of 3737 sample sites were initially filtered from the database (excludes larger rivers, winter samples, duplicate samples, and re-visits). These data represented 391 REF and 962 STRESS sample sites. This dataset was randomly divided into CAL (70% of all sites) and VAL (remaining 30% of all sites) using a random number function in Excel (Microsoft Corp.). This resulted in an index development dataset (CAL) of 2,616 distinct samples (273 REF, 674 STRESS, and 1,669 Non-REF) and a validation dataset (VAL) consisting of 1,121 samples (118 REF, 288 STRESS, and 715 Non-REF). Thus, the initial 947 REF and STRESS samples were used to develop and calibrate a multi-metric index capable of distinguishing stream status across the State. The remaining 1,121 validation samples were used to independently verify the performance of the GLIMPSS. Figure 5 depicts the distribution of CAL and VAL sites across the State.

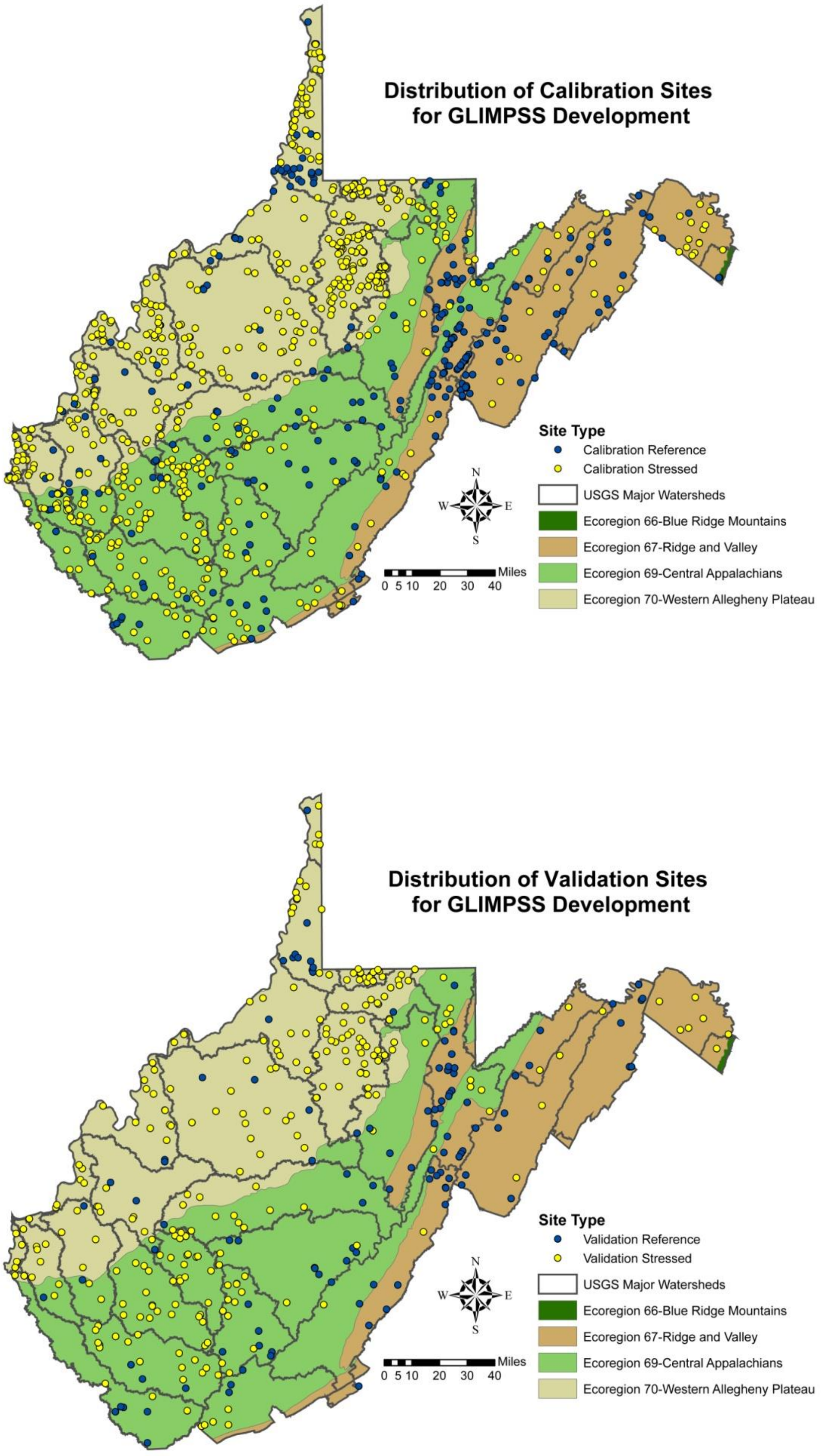


Figure 5. Site locations for CAL REF and STRESS (top), and VAL REF and STRESS (bottom).

7.2 Community Classification

Because benthic communities often differ across seasons or regions, we evaluated several combinations of seasonal and geographic strata in an effort to explain natural variability in macroinvertebrate assemblages found in least-disturbed West Virginia streams. Seasonal cutoffs given below were mostly biologically-based (*i.e.*, they relate to known life history phenomena of resident taxa) and also corresponded to those established for macroinvertebrates in neighboring Kentucky (Pond et al. 2003). Geographical stratification followed Level III ecoregions after Woods et al. (2000). The effect of stream size (*e.g.*, catchment area or stream width) on community structure and final GLIMPSS scores was also investigated. Final GLIMPSS scores were analyzed in relation to stream size and are presented in 8.6 *GLIMPSS Relation to Stream Size* section.

Multivariate ordination using non-metric multidimensional scaling (NMDS) (Ludwig and Reynolds 1988) and mean similarity analysis (MEANSIM) (Van Sickle 1997) were used to help select the best classification scheme. The following combinations of strata were evaluated:

- Level III Ecoregion (67, 69, 70)
- Season (Spring [March-through May], Summer [June-early October])
- Bioregion (combined Level III mountain ecoregions [67/69], Plateau [70])
- Level III Ecoregion x Season
- Bioregion x Season

For comparisons, a recommended Bray-Curtis similarity matrix of reference site communities (Hawkins and Norris 2000) was used and MEANSIM calculated the average within-class similarity ($\overline{W}$) to the average between-class similarity ($\overline{B}$) among strata combinations. The classification strength (CS) is simply $\overline{W}$ - $\overline{B}$. We used the multi-response permutation procedure (MRPP) (PC-ORD^{TM}, Version 6, MjM Software^{TM}, Gleneden Beach, OR) with rank-transformed distances to calculate the between-class similarity; these values were then input into MRPPCONV (John VanSickle, USEPA, http://www.epa.gov/wed/pages/models/dendro/meansim6.htm).
MRPP uses $\overline{W}$ (relative to its expected value and variance under the null hypothesis) as its test statistic, rather than the $\overline{W}/\overline{B}$ and ($\overline{W}$ - $\overline{B}$) used by VanSickle's RNDTST6. However, all three statistics are nearly equivalent and will give nearly identical P-values in practice (John VanSickle, USEPA, http://www.epa.gov/wed/pages/models/dendro/meansim6.htm).

NMDS of REF sites was run with PC-ORD^{TM}. NMDS is a distance-based ordination technique that maximizes rank-order correlation between the distance measure and distance in ordination space. Through many iterations, sample points are moved to minimize "stress", a measure of lack of correspondence between the original distance matrix and the final ordination. When plotted, the distance or spread of sites represents the similarity (or dissimilarity) in community composition. Although NMDS is non-parametric, and thus distribution-free, genus-level invertebrate abundances were log (x+1) transformed to reduce any effect of skewed abundance distributions on the outcome of the ordination. NMDS was run using the Bray-Curtis coefficient and we excluded infrequently occurring taxa found at less than 2.5% of all REF sites (using 158 taxa out of 322 total taxa). By excluding these infrequent taxa, multivariate analyses are more robust and patterns are more evident

(McCune and Grace 2002). These omitted taxa were added back into the calculation for all metrics and the final GLIMPSS. NMDS was run using the Bray-Curtis coefficient with PC-ORD^{TM} set on the "autopilot" slow and thorough mode (six dimensions with 250 real runs and 250 randomized runs), a setting recommended by the software's authors.

7.3 Metric Selection

Metrics combine ecological attributes of macroinvertebrate populations in order to summarize community level data. There are potentially hundreds of metrics that can be calculated from macroinvertebrate assemblage data (Resh and Jackson 1993, Barbour et al. 1999, Karr and Chu 1999). We chose to evaluate forty-one (41) biological metrics that were calculated from queries built in the WAB database (Table 4, also see Appendix A). These metrics spanned a wide scope of ecological attributes that included recommended measures of richness, composition, dominance, tolerance, trophic or functional feeding groups, and habit (Barbour et al. 1999) and included some of the metrics used in the original WVSCI (Gerritsen et al. 2000a). We excluded proportional richness metrics (*e.g.*, proportion of plecopteran genera to total genera, etc.) used by other investigators (Whittier et al. 2007, Blocksom and Johnson 2009) after checking that these metrics did not improve sensitivity compared to the original richness metric. Because of the fixed-count subsample (a means to standardize richness expectations), actual richness metrics were preferred over proportional richness metrics. Many of the metrics we tested have been applied successfully by other state, tribal and federal assessment programs both nationally and regionally (Barbour et al. 1999). In relation to West Virginia, similar metrics have been used in the Mid-Atlantic Highlands (MAH) (Klemm et al. 2003), Ohio (Deshon 1995), Tennessee (Arnwine and Denton 2001, Kerans and Karr 1994), Virginia (Burton and Gerritsen 2003), Kentucky (Pond et al. 2003), Maryland (Southerland et al. 2007), Pennsylvania (Chalfant 2007) and the Potomac River Basin (Astin 2006).

Because WVDEP samples riffle/run communities, it was deemed that certain metrics which specifically evaluate, for example, the odonate, coleopteran, hemipteran or crustacean taxon groups, would not be responsive since (1) many taxa in these groups reside in non-riffle habitats, (2) often have insufficient ranges of metric values, or (3) exhibit unclear trends of water quality responsiveness (WVDEP, unpub. data). For example, odonate richness insufficiently ranged from 0-4 genera at REF sites, with many zero occurrences. Genus-level taxonomy allowed for refinement of certain common order-level metrics (*e.g.*, %EPT) by excluding known facultative or tolerant genera (*e.g.*, *Baetis*, *Cheumatopsyche*) that often become hyperdominant in samples under various levels of stress.

Metrics were compared in each stratum for their ability to distinguish between REF and STRESS sites (discrimination efficiency, or DE), redundancy (correlation between two similar metrics), range, variability, and their response to human disturbance (see below).

Table 4. List of the 41 metrics analyzed for index development.

Richness (Generic)
- Total
- Intolerant (TV <3)
- Intolerant (TV <4)
- EPT
- Mayfly
- Stonefly
- Caddisfly

Functional/Habit Richness (Generic)
- Scrapers
- Shredders
- Collectors
- Filterers
- Predators
- Clingers

Composition (Order, Family, Tribe)
- % Mayflies
- % Mayflies (No Baetis)
- % Stoneflies
- % Caddisflies
- % EPT
- % mEPT (No *Cheumatopsyche*)
- % mEPT2 (No *Cheum* + *Baetis*)
- % Hydropsychidae
- % Chironomidae
- % Annelida
- % Chironomini
- % Orthocladiinae
- % Tanytarsini
- % Chironomidae + Annelida
- % Non-Insects

Trophic/Feeding Group Composition
- % Scrapers
- % Shredders
- % Collectors
- % Filterers
- % Predators

Habit Composition
- % Clingers
- % Sprawlers
- % Swimmers
- % Climbers
- % Burrowers

Tolerance/Dominance
- HBI
- % Tolerant (TV >6)
- % 5 Dominant Genera

7.3.1 Discrimination Efficiency

Discrimination efficiency (DE) was used to evaluate metric sensitivity (Gerritsen et al. 2000a). Percent DE was calculated as the number of STRESS site metric values that fell below the reference set 25th percentile (or >75th percentile for negative response metrics) divided by the total number of STRESS sites, and multiplied by 100. Most metrics that had DE values of less than 65% were automatically dropped from further analysis unless it was deemed that the metric offered additional ecological information desired for the index (*e.g.*, habit, trophic, tolerance) and the metrics passed other selection criteria (see below). Those metrics that had the highest DE were considered for inclusion in the multi-metric index.

7.3.2 Response to Stress

To evaluate metric and index response to stress, a human disturbance gradient was constructed from the linear combinations of abiotic factors at 1617 sites (approximately 70.3% of which were probabilistically selected) using principal components analysis (PCA) (Klemm et al. 2003). PCA requires a symmetric dataset, so only sites that had all physical and chemical measurements were included. Mathematically, PCA consisted of an eigenanalysis of a correlation matrix calculated on the original measurement data. PCA is often used over multiple regression when input variables are highly correlated. Graphically, it can be described as a rotation of a cloud of data points in multidimensional space so that the longest axis (the axis accounting for the greatest variance) is the 1st PCA axis, the second longest axis perpendicular to the first is the 2nd PCA axis, and so forth (MVSP, Kovach Computing, London). Thus, the 1st axis is often used to define the strongest gradient of abiotic variation from sites ranging from least-disturbed to most-disturbed in the dataset (Klemm et al. 2003). Spearman correlation was then used to evaluate individual metric response along the disturbance gradient (PCA 1 scores). Physicochemical and habitat parameters used in the PCA were checked for skewness (a rule of thumb in this case, where we were not explicitly testing hypotheses, thus, multivariate normality of the variables was not required, and reducing skewness was the main goal of the transformations (McCune and Grace 2002)). The variables included same-time measures of: pH, and log transformed temperature, D.O., fecal coliform, conductivity, sulfate, chloride, total phosphorus, nitrite-nitrate, total suspended solids, total aluminum, total iron, total manganese, and 7 of the 10 RBP habitat metrics (channel flow status, velocity regime, and frequency of riffles excluded) . We selected metrics with significance ($p<0.05$) and Pearson correlation ($|r|>=0.25$) to the disturbance gradient.

7.3.3 Redundancy

Metric pairs that are highly correlated often provide similar information and thus one metric would be considered unnecessary. While there is no consensus on “hard cutoffs” for detecting metric redundancy, several workers have chosen r-values in excess of 0.75 (Maxted et al. 2000, Blocksom and Johnson 2009), 0.85 (Butcher et al. 2003, Gerritsen et al. 2000a), and 0.90 (Barbour et al. 1996) to screen for metric redundancy. For GLIMPSS development, we chose Pearson r-values of 0.75 as the cutoff, but metric pairs approaching or slightly exceeding this value were further examined using scatterplots to see if nonlinear relationships were apparent or if there was sufficient dispersion (*i.e.*, scatter) of the paired metric data points (Barbour et al. 1999). In this case, inclusion of both metrics could be beneficial to the multi-metric index.

7.3.4 Range and Variability

Metrics with insufficient range within the reference dataset are not acceptable for inclusion in the index because they often cannot detect deviations of new sites from reference conditions. Richness metrics with a range of 5 or more taxa and abundance metrics with a range >10% were considered acceptable (Blocksom and Johnson 2009). Variability was visualized using boxplots of the REF site metrics within individual strata. If considerable spread of reference interquartile (25th and 75th) ranges were observed (*i.e.*, if an interquartile range was greater than the range between zero and the lower quartile), these metrics were deemed to have excessive variability and thus be grounds for rejection. This was referred to as “scope of impairment” (SOI) (modified after Klemm et al. 2003, Blocksom and Johnson 2009) and expressed as a ratio (or interquartile coefficient). Metrics with an interquartile coefficient of >1 were rejected.

7.4 Index and Metric Scoring

The most sensitive, responsive, and non-redundant metrics were aggregated for each stratum so that indicators could best contribute to the final GLIMPSS. These metrics included “positive” (*i.e.*, increase with improving water quality) and “negative” (decrease with improving quality) scoring metrics. For scoring purposes, metrics were first normalized by calculating the 95th percentile (or 5th percentile for negative responding metrics) of each metric based on all sites within each stratum (only samples containing between 100 and 240 individuals were used). We calculated both ceiling and floor values (95th and 5th percentile) for each metric following Blocksom (2003), Blocksom and

Johnson (2009), and Whittier et al. (2007). This use of best standard values (BSVs) and worst standard values (WSVs) results in unit less and equally-weighted scoring of metrics that not only excludes outlier metric values, but also overcomes the problem of normalization so that metrics using counts, proportions, and logarithmic functions can be compared uniformly when applied to the aggregate index (Gerritsen et al. 2000b). Here, metrics were scored by standardizing the metric value by the BSV and WSV on a continual scale of 0−100 %, and then averaged to produce the final index score (after Gerritsen et al. 2000). Blocksom (2003) showed that this continual scoring method performed better than categorical scoring from a MAH dataset, and neighboring states' (VA, KY, PA) multi-metric indices are based on this scoring procedure. If a metric scored over 100 (*e.g.*, a value above the 95th percentile) then it was corrected to the maximum score of 100. Alternatively, if that metric scored below the 5th percentile, it received a score of zero. See Appendix D for examples of scoring formulae.

7.5 Index Performance

7.5.1 Discrimination and Classification Efficiency

The calibration GLIMPSS was evaluated for DE and response to the human disturbance gradient (PCA axis 1 from the 1617 site symmetric dataset). Furthermore, REF site index scores were analyzed in relation to catchment area (with linear regression) to determine if waterbody size contributed to variability in the GLIMPSS.

The independent dataset was used to validate the classification efficiency (CE), or ability of GLIMPSS to correctly assign sites to either reference or stress categories (Southerland et al. 2005). This is different from DE in that CE was calculated as the sum of the number of VAL REF sites scoring above the 5th percentile, and the number of VAL STRESS sites scoring below the 5th percentile of the development reference distribution, divided by the total number of sites. VAL sites were also plotted in relation to the PCA axis 1 disturbance gradient.

7.5.2 Precision

Measurement error is introduced from both natural (*e.g.*, patchiness of habitat and associated macroinvertebrates) and methodological (both field and lab methods) sources of variability. This measurement error is most commonly estimated using repeat or duplicate samples which are collected on the same day, or within one index period. We estimated component metric and GLIMPSS measurement error and associated precision (a type of performance measure) from same-day replicates collected at 90 sites. Component metric precision was estimated using all 90 sites. GLIMPSS precision was estimated within individual strata. We followed methods reported by Stribling et al. (2008) for evaluating index and metric performance: 90% confidence intervals (90% CI), coefficient of variation (reported as %CV), and relative percent difference (RPD). An analysis of variance (ANOVA) was used only to determine the within-samples mean square error (MSE or variance). The square root of the MSE (the RMSE) provides the standard deviation, which is an estimate of measurement error. The standard deviation of the all observations was then used to calculate 90% CI. %CV was calculated as the standard deviation divided by the population mean multiplied by 100. RPD was calculated as the absolute difference between two replicate samples divided by the mean of the samples, where lower values indicate better precision.

7.5.3 Temporal Variability at Reference Sites

GLIMPSS scores were also evaluated for annual variability at REF sites. Again, ANOVA was used to determine the within-sample mean square error (MSE or variance). The square root of the MSE (the RMSE) provides the standard deviation, which is an estimate of measurement error. The standard deviation of all observations was then used to calculate 90% CI. % Coefficient of Variation (CV) was calculated as the standard deviation divided by the population mean multiplied by 100. RPD was calculated as the absolute difference between two re-visit samples divided by the mean, where lower values indicate better precision.

8.0 Results and Discussion

8.1 Community Classification

For taxon abundance data using the recommended Bray-Curtis coefficient, MEANSIM showed that ecoregion *x* season (CS=17%) and bioregion *x* season (CS=18%) had the highest classification strengths (Table 5). Moreover, there was minimal difference between these two stratum combinations. The lowest CSs were calculated for Level III ecoregions (10%), and bioregion (12%), indicating the stronger influence that seasonality has on structuring REF benthic communities.

Table 5. Similarity analysis results for strata combinations using Bray-Curtis coefficients from REF sites based on log-transformed abundance data. CS=Classification Strength. All classifications were significant (P<0.0001) based on 10,000 permutations. EcoSeason and BioSeason are abbreviations for combined season and region.

	N	$\overline{W}$	$\overline{B}$	$\overline{W} - \overline{B}$ *(CS)*
Ecoregion	3	0.56	0.46	0.10
Season	2	0.57	0.42	0.15
EcoSeason	6	0.64	0.47	0.17
Bioregion	2	0.56	0.44	0.12
BioSeason	4	0.62	0.44	0.18

Waite et al. (2000) found Bray-Curtis CS was <4% for 6 ecoregions in a study of the Mid-Atlantic Highlands where West Virginia comprises ~30% land area and 3 of the 6 ecoregions. They also found that genus level taxonomy was better than family-level. In the neighboring state of Kentucky, Pond et al. (2003) reported genus-level Bray-Curtis CS as high as 14% using a priori bioregions for streams >5 sq. mi. and 17% for streams <5 sq. mi. At the family level, Gerritsen et al. (2000a) reported <4% for ecoregions but 8.4% for calendar month in West Virginia. These results were similar to a recent study in Virginia (VA DEQ 2006) in which bioregion *x* season gave the best CS (4.7%) at the family level. Our results suggest that combining seasonal and modified ecoregion stratification gave better results with genus-level taxonomy in West Virginia.

In the NMDS analysis, a 3-dimensional solution was fitted after 110 iterations. Axes (*i.e.*, dimensions) 1 and 2 were 100% orthogonal and explained the most variance in the distance between the original distance matrix and final configuration (28% and 23%, respectively) and stress was relatively low (18.9%). Axis 3 explained only 11% variance and is not plotted. Again, "stress" is measured as departure from "goodness of fit" in the relationship between the dissimilarity (distance) in the original p-dimensional space and distance in the reduced k-dimensional ordination space (McCune and Grace 2002). Basically, stress is an important measure of confidence in the interpretability of the ordination. Values of stress less than 20% generally give acceptable ordinations (McCune and Grace 2002). Visual inspection of the NMDS ordinations (Figures 6 and 7) showed comparable results to MEANSIM analysis (*i.e.*, classification strength can be visualized in the ordinations). The best ordination produced was bioregion *x* season (Figure 7b) which had the least scatter across strata and good clustering within strata.

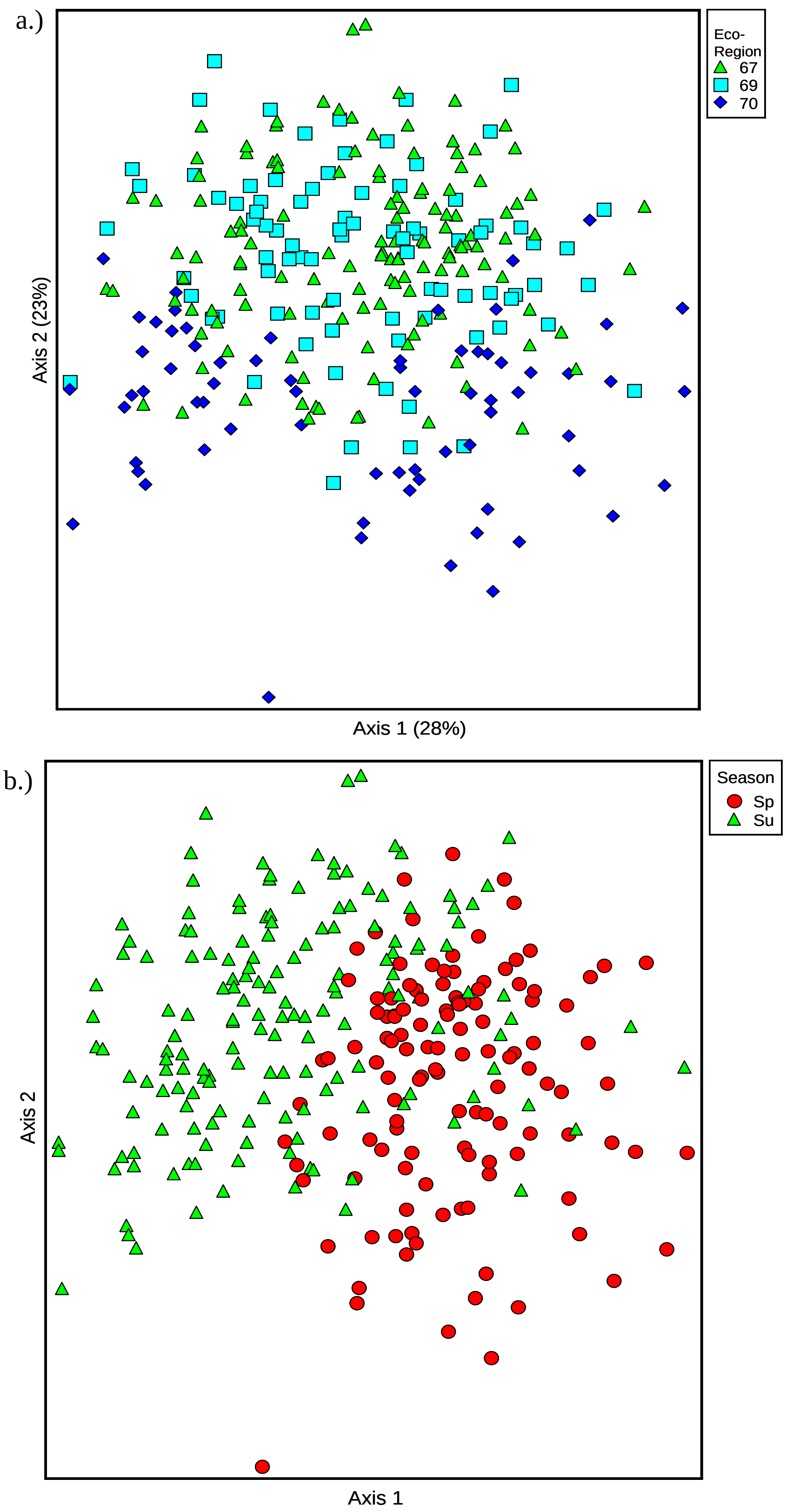


Figure 6. NMDS REF site ordinations for strata combinations used in MEANSIM analysis: (a.) ecoregion, (b.) season.

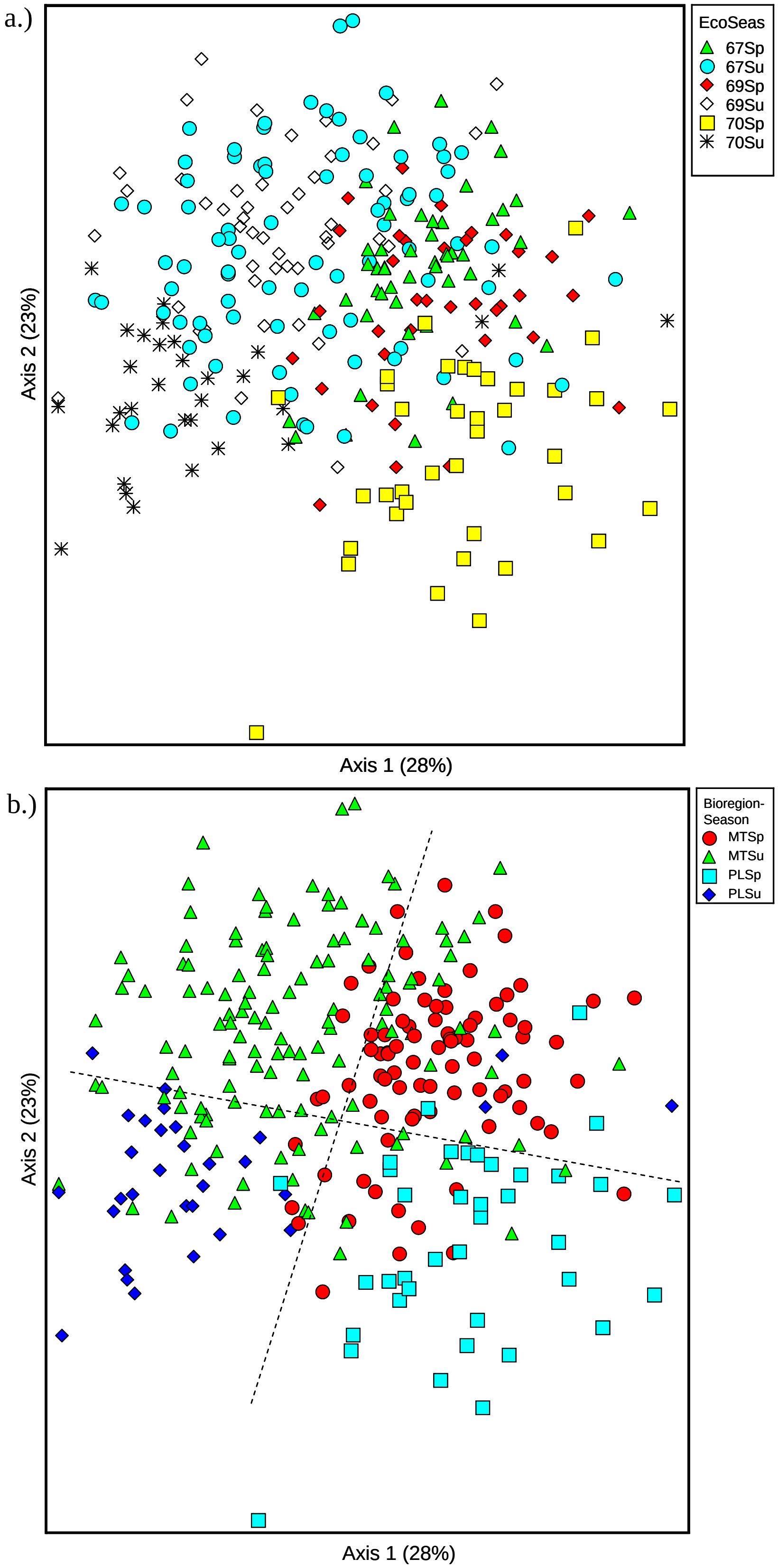


Figure 7. NMDS REF site ordinations for strata combinations used in MEANSIM analysis: (a.) ecoregion *x* season, (b.) bioregion *x* season. Dashed lines in (b) represent approximate demarcation of best classification scheme.

Further analysis revealed that a couple of key environmental factors were more highly correlated with NMDS axis scores (*e.g.*, elevation and sample week number) than others (*e.g.*, longitude, latitude). Intuitively, bioregion *x* season can be largely explained by elevation (*i.e.*, mountain areas of ecoregion 67-69 vs. lower elevation in 70) and week number (spring vs. summer index period). Elevation was found to be a significant contributor to family-level WVSCI scores in a separate study (Griscom et al. in review). The strong influence of calendar week number and temperature on Axis 1 confirms realization that benthic communities change substantially along a seasonal (spring to summer) continuum. Our ordination results indicate that splitting the data into common seasons (*i.e.*, spring and summer) in addition to bioregions provided good separation.

Thus, four (4) classification strata were chosen for index development purposes: Mountain Spring (MT Sp), Mountain Summer (MT Su), Plateau Spring (PL Sp), and Plateau Summer (PL Su). Mean stream width showed some correlation to NMDS axis 1and 2 (Table 6); since this could be biologically meaningful, final index scores within each stratum were further regressed with catchment area in a separate analysis (see Section 8.6 *GLIMPSS Relation to Stream Size*).

Table 6. Pearson correlation coefficients calculated in PC-ORD^{TM} between environmental variables and NMDS axis scores.

	NMDS 1	NMDS 2	NMDS 3
Calendar Week	-0.697	0.358	-0.007
Average Width	-0.209	-0.236	-0.035
Latitude	-0.005	-0.263	-0.176
Longitude	-0.042	0.227	-0.343
Elevation	0.117	0.486	-0.243
pH	-0.282	-0.214	0.005
Sp. Conductance	-0.21	-0.396	-0.024
Temperature	-0.718	-0.154	0.137

8.2 Response to Stress: Constructing a Human Disturbance Gradient

Abiotic data from a total of 1617 sites (combined REF, STRESS, and Non-REF that conformed to a symmetric dataset with the full suite of environmental variables) were analyzed with PCA to construct a synthetic gradient of human disturbance using a suite of habitat, physicochemical, and nutrient data. Figure 8 illustrates a two dimensional PCA ordination of the first 2 axes. In this type of graph, sites are plotted as points and the vectors indicate the strength and direction of an environmental parameter's contribution to the ordination. Table 7 lists the eigenvalues, percent variance explained by each axis, and factor coefficients of each environmental variable on the respective axes. Summary statistics of environmental variables used in the PCA from REF and STRESS sites are reported in Appendix F.

The first PCA axis explained nearly 26% of the variance in the dataset (eigenvalue =5.1) and was deemed suitable for use as a human disturbance gradient. Blocksom (2003) reported 38% variance explained by PCA 1 in her study of MAH streams. Sites spanned a broad gradient along axis 1 (Figure 8). Site coordinates along this principal axis were then used in a correlation analysis with individual stratum-specific macroinvertebrate metrics (see Appendix B). Specific conductance had the highest correlation of chemical variables to axis 1 (+0.47) followed closely by temperature (+0.46), but daily temperatures fluctuate widely; the correlation using instantaneous readings might simply relate to regional and seasonal patterns. Habitat metrics were also strongly correlated to axis 1 (most more than -0.60). Measures of pH and total metals were most correlated to axis 2 (> +0.50), but this axis only accounted for 13% of the total variance.

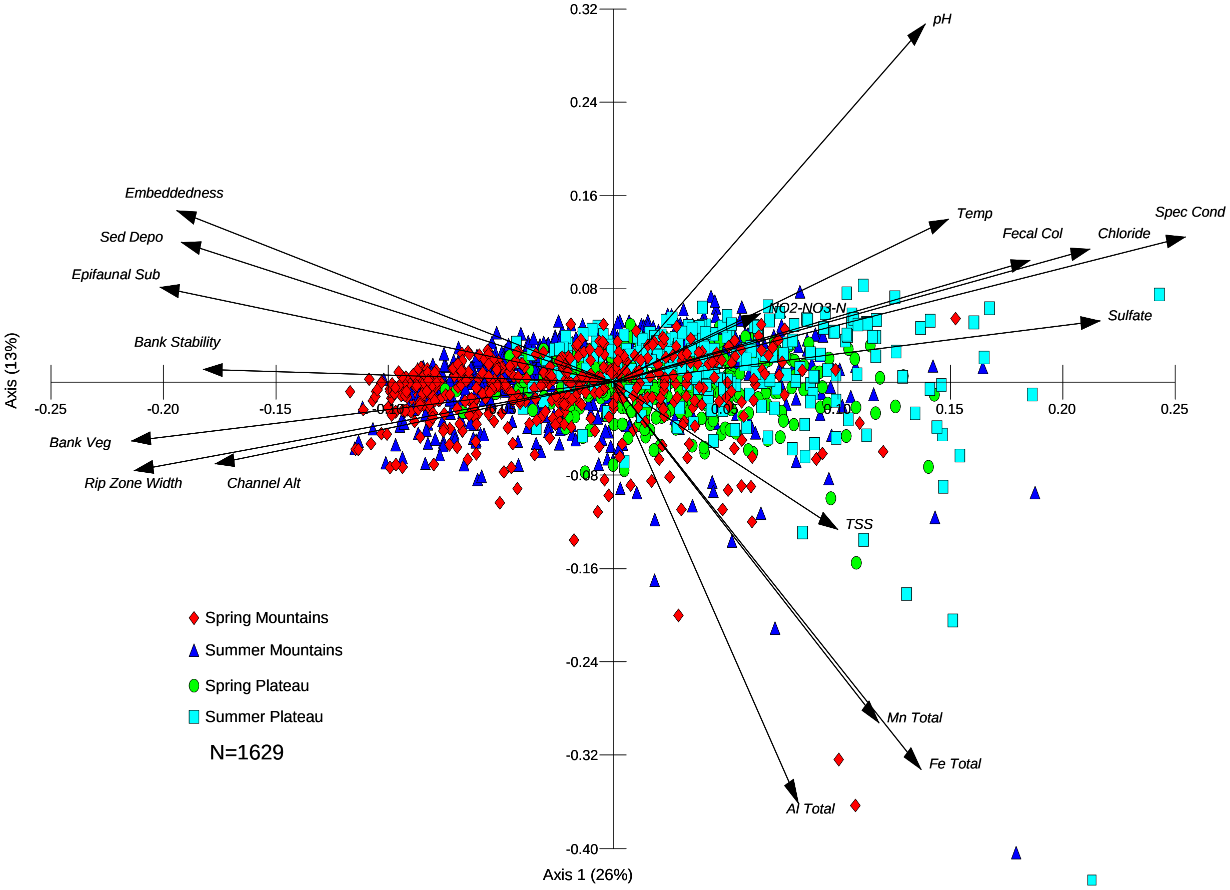


Figure 8. PCA ordination of sites (n=1617) used for metric-stressor response relationships. Short vectors (D.O., Total P) are not plotted.

Table 7. PCA results and factor coefficients of the first three components. Habitat metrics based on RBP scoring procedures.

	Axis 1	Axis 2	Axis 3
Eigenvalues	5.1	2.6	2.1
Percentage	25.9	13.0	10.2
Cum. Percentage	25.9	38.5	48.7

	Axis 1	Axis 2	Axis 3
Temp	0.46	0.31	-0.26
pH	0.42	0.66	-0.14
DO	-0.12	-0.01	0.25
Spec Cond	0.47	0.12	-0.38
Fecal Col	0.19	0.07	-0.09
Sulfate	0.36	0.01	-0.33
Chloride	0.28	0.10	-0.21
TSS	0.23	-0.25	-0.17
P Total	0.06	-0.10	-0.22
NO2-NO3-N	0.23	0.11	-0.34
Al Total	0.20	-0.64	-0.22
Fe Total	0.20	-0.53	-0.18
Mn Total	0.25	-0.59	-0.27
Epifaunal Sub	-0.63	0.15	-0.37
Embed	-0.62	0.27	-0.26
Channel Alt	-0.58	-0.16	-0.30
Sed Dep	-0.60	0.23	-0.30
Bank Stab	-0.58	0.00	-0.42
Bank Veg	-0.69	-0.12	-0.45
Rip Zone Width	-0.68	-0.19	-0.34

8.3 Metric Evaluation and Selection

In the selection process, an effort was made to include at least one metric each from the list of richness, composition, tolerance, dominance, and feeding/habit attributes (Karr and Chu 1999). However, this was not always possible as some feeding/habit metrics had unacceptable DEs, had insufficient range, or were too variable. After initially testing metrics for the highest DE and the highest correlation to human disturbance (PCA axis 1) within each stratum, we chose metrics that lacked redundancy, had sufficient range, and good scope of impairment. Rather than select metrics that could be used statewide, we were primarily interested in using metrics that could best distinguish stream status within each bioregional and seasonal stratum. Appendices B through D provide supporting information for the metric evaluation process.

The GLIMPSS uses ten metrics in both MT strata, while GLIMPSS uses 8 metrics in the PL Spring, and 9 metrics in PL Summer stratum. Table 8 shows the best metrics chosen for GLIMPSS by bioregion/season, sorted primarily by %DE and secondarily by their correlation to the human disturbance gradient identified by PCA axis 1. Maximum redundancy values (correlation) are listed for those metrics chosen for the GLIMPSS.

Table 8. List of final GLIMPSS metrics by stratum with %DE, correlation to PCA axis 1 (p<0.05), and maximum correlation value for metric pairs as a measure of redundancy (absolute value). MT Sp= Mountain Spring, MT Su=Mountain Summer, PL Sp=Plateau Spring, PL Su=Plateau Summer. Some metric names are abbreviated. Metric categories listed as richness (Rich), composition (Comp), tolerance (Tol), habit, dominance (Dom), and feeding group (FFG).

	MT Sp	DE	PCA	Redun
Tol	No. Intol Taxa (<4)	92.1	-0.61	0.75
Comp	% Orthocladiinae	88.8	0.49	0.45
Tol	HBI	84.1	0.62	0.45
Rich	No. Trichoptera Genera	81.0	-0.39	0.59
Rich	No. Ephemeroptera Genera	79.4	-0.36	0.45
Rich	No. Plecoptera Genera	77.8	-0.53	0.75
Habit	No. Clinger Genera	74.6	-0.47	0.66
Comp	% Ephemeroptera	74.6	-0.31	0.26
Dom	% Dominant 5 Genera	73.0	0.35	0.64
FFG	No. Scraper Genera	68.3	-0.42	0.62

	MT Su	DE	PCA	Redun
Tol	HBI	95.9	0.66	0.74
Tol	No. Intol Taxa (<4)	95.0	-0.68	0.74
Rich	No. Plecoptera Genera	91.0	-0.62	0.74
Comp	% EPT (minus *Cheumatopsyche)*	90.1	-0.59	0.74
Habit	No. Clinger Genera	78.8	-0.39	0.69
Rich	No. Ephemeroptera Genera	77.9	-0.43	0.39
Dom	% Dominant 5 Genera	72.1	0.26	0.67
FFG	No. Shredder Genera	71.2	-0.45	0.47
Rich	No. Total Genera	67.6	-0.31	0.69
Comp	% Orthocladiinae	67.1	0.34	0.41

	PL Sp	DE	PCA	Redun
Comp	% EPT (minus *Cheumatopsyche)*	82.6	-0.36	0.66
Tol	No. Intol Taxa (<4)	81.9	-0.49	0.67
Tol	HBI	80.4	0.41	0.75
Rich	No. Plecoptera Genera	79.0	-0.38	0.65
Tol	% Tolerant Taxa (>6)	76.8	0.24	0.75
Comp	% Chironomids+Annelids	75.4	0.31	0.74
Rich	No. Ephemeroptera Genera	65.2	-0.35	0.59
Habit	No. Clinger Genera	60.9	-0.38	0.67

	PL Su	DE	PCA	Redun
Tol	No. Intol Taxa (<3)	97.6	-0.53	0.70
Tol	HBI	94.0	0.55	0.75
Comp	% EPT (minus *Cheumatopsyche)*	85.7	-0.54	0.75
Habit	No. Clinger Genera	83.4	-0.54	0.64
Dom	% Dominant 5 Genera	74.5	0.32	0.62
Rich	No. Total Genera	74.1	-0.37	0.64
Rich	No. Ephemeroptera Genera	68.9	-0.52	0.46
Comp	% Chironomidae	64.9	0.52	0.53
FFG	No. Scraper Genera	62.9	-0.43	0.44

Several core metrics were chosen consistently in all strata (*e.g.*, HBI, No. Clinger Genera, No. Ephemeroptera Genera) or in at least 3 of the 4 strata (*e.g.*, No. Intolerant Taxa <4, No. Plecoptera Genera, %Dominant 5 Genera). However, there were metrics unique to 1 or 2 strata (*e.g.,* No. Shredder Genera, %Orthocladiinae, No. Scraper Taxa). The EPT richness metric had a high %DE and correlation to stress in all strata. However, for diagnostic purposes, independent measures of Ephemeroptera and Plecoptera, or Trichoptera richness were chosen in favor of EPT (see Appendices B-D). There was one selected metric that was enhanced by removing a key taxon (*i.e.*, %EPT less *Cheumatopsyche*) and was selected for use in 3 strata.

Among these best metrics, some pairs (*e.g.*, No. Plecoptera and No. Intolerant <4 in MT Sp) had a correlation of 0.75. After viewing scatterplots of these metric pairs, it was apparent to include both metrics in the GLIMPSS because of the large scatter of one metric at any given value of the other metric (see example in Figure 9). For example, in particular collections having 4 Plecoptera genera, No. Intolerant (<4) ranged from 8 to 15 genera. For collections having 5 Plecoptera genera, No. Intolerant (<4) ranged from 11 to 19 genera.

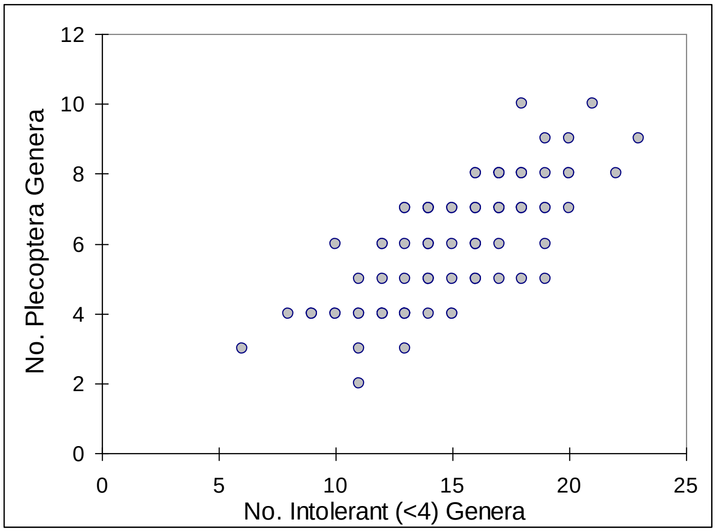


Figure 9. Scatterplot of No. Plecoptera Genera vs. No. Intolerant (<4) Genera in MT Sp. The correlation was 0.75, but there was considerable scatter to warrant inclusion of both metrics in the GLIMPSS. Points represent multiple observations (half of the observations are hidden beneath other points).

8.4 Metric Aggregation and Index Performance

Within individual strata, metric BSVs and WSVs based on the 95th or 5th percentile of the CAL dataset were calculated and all metrics were scored for each site in the CAL dataset. BSVs for each stratum are shown in Table 9. The aggregate 100-point GLIMPSS was calculated as the average metric score. See Appendix D for example scoring procedures.

Table 9. Best standard values (BSVs) and Worst Standard Values (WSVs), as ceilings and floors, for metrics (by stratum) used for CAL scoring purposes.

	MT Sp (n=488)		MT Su (n=1071)		PL Sp (n=457)		PL Su (n=600)	
	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor
No. Total Genera			38	14			34	14
No. Intolerant Genera <4	19	1	15	0	15	1		
No. Intolerant Genera <3							7	0
No. Ephemeroptera Genera	8	1	9	0	10	1	7	0
No. Plecoptera Genera	8	1	7	0	7	0		
No. Trichoptera Genera	7	1						
No. Clinger Genera	20	4	19	5	17	3	15	4
HBI (Genus level)	6.14	2.26	6.27	2.89	6.54	2.52	6.24	3.73
% Dominant 5 Genera	91.8	48.4	91.7	51			91.5	53.3
% Tolerant (>6)					65.0	0.0		
% Ephemeroptera	58.1	0.5						
% mEPT (minus *Cheumatopsyche)*			85.4	5.3	91.2	2.6	67.1	1.4
% Orthocladiinae	50.9	0.5	36.8	0.4				
% Chironomidae							69.1	4.0
% Chironomidae+Annelida					83.5	1.8		
No. Scraper Genera	8	0					7	1
No. Shredder Genera			5	0				

Box plots and %DE were calculated as a means to view the sensitivity of GLIMPSS within individual strata. Figure 10 demonstrates that the GLIMPSS could discriminate between REF and STRESS sites with a high degree of efficiency (>75%). It also indicates that Non-REF sites fell into a presumed intermediate position between REF and STRESS, with respect to scores. DE was greatest in the PL Su stratum where 89% of the STRESS sites fell below the 5th percentile of the REF distribution. In addition, interquartile ranges for each stratum were relatively low (ranging from 13 to 18 points) suggesting low variability of the reference condition.

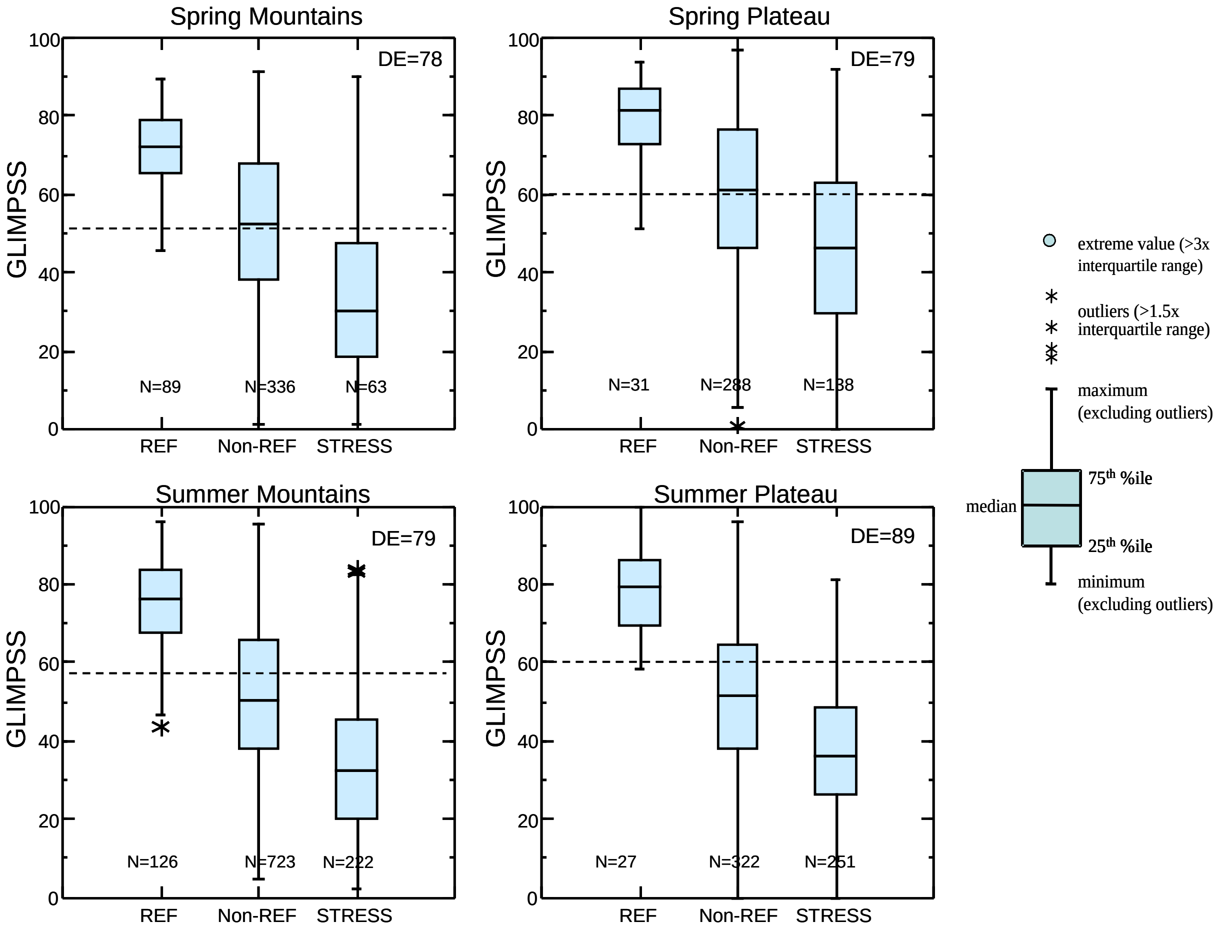


Figure 10. Box plots of REF vs. Non-REF and STRESS site GLIMPSS scores for CAL dataset showing %DE. Dotted line represents approximate 5th percentile of CAL REF distribution. Box plot interpretation is shown at far right.

CAL GLIMPSS scores were also evaluated for response to the human disturbance gradient (PCA axis 1). Recall that this PCA axis was mostly driven by a gradient of habitat quality (RBP habitat metric scores) and ionic and organic strength (chloride, sulfate, fecal coliform). In all strata, the index was negatively correlated to PCA axis 1, and best fit regression lines showed that each stratum had a similar response to increasing stress (Figure 11). Note that acidic deposition sites scored low on the GLIMPSS but had excellent habitat, low ions, low fecal coliform bacteria, and low pH.

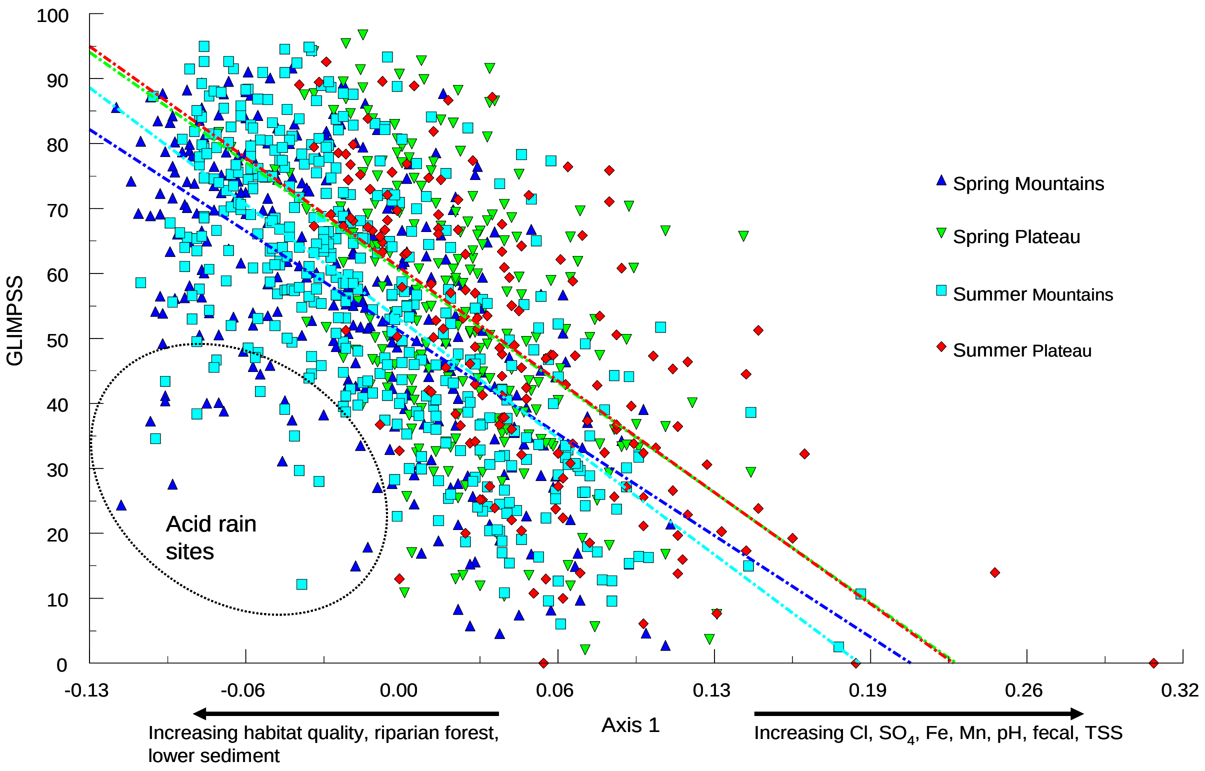


Figure 11. Scatter plot of CAL GLIMPSS scores (labeled by stratum) along human disturbance gradient (PCA axis 1) with best fit linear regression lines. Key abiotic variables are shown along the stressor axis.

8.5 Index Validation

The independent validation dataset (n=118 REF sites, 288 STRESS sites) was applied to the metric scoring criteria (BSVs and WSVs) and the CAL 5th percentile of reference GLIMPSS within each stratum. CE (*i.e.*, % of correctly classified REF and STRESS sites) was excellent for all strata (Figure 12), indicating successful validation. CE ranged from 89 to 95%. This analysis clearly shows that the GLIMPSS could reliably distinguish between reference quality and stressed streams and there was comparably low variability of GLIMPSS scores at validation REF sites within each stratum.

Validation GLIMPSS scores also showed similar response to the PCA Axis 1 stressor gradient (n=491) as with the CAL plot. The slopes of the best fit linear trends among strata were also similar (Figure 13).

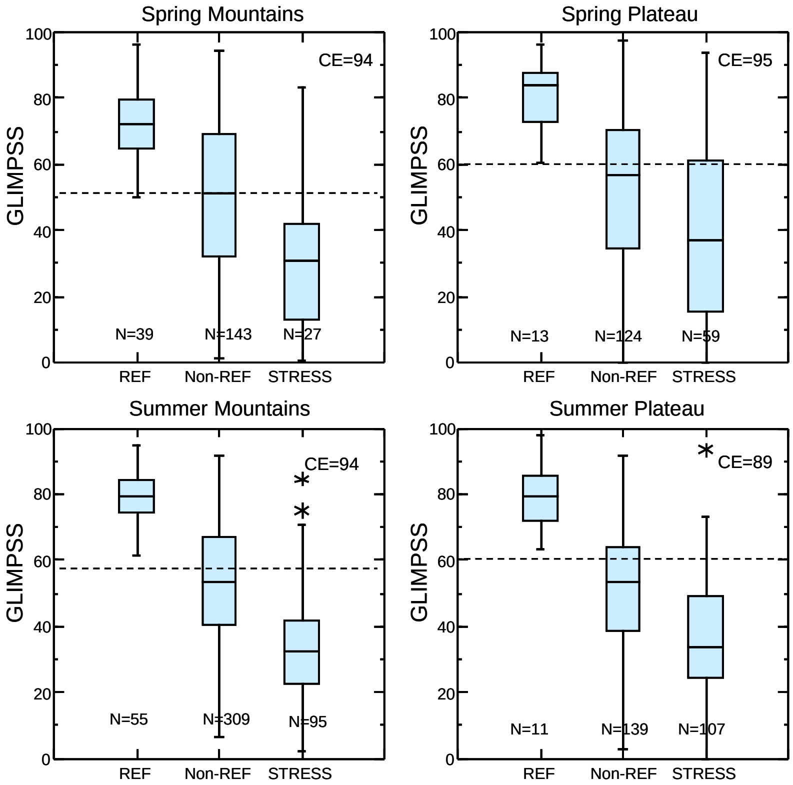


Figure 12. Box plots of REF vs. Non-REF and STRESS site GLIMPSS scores for validation (VAL) dataset showing %classification efficiency (CE). Dotted line represents approximate 5th percentile of the calibration reference distribution for each stratum.

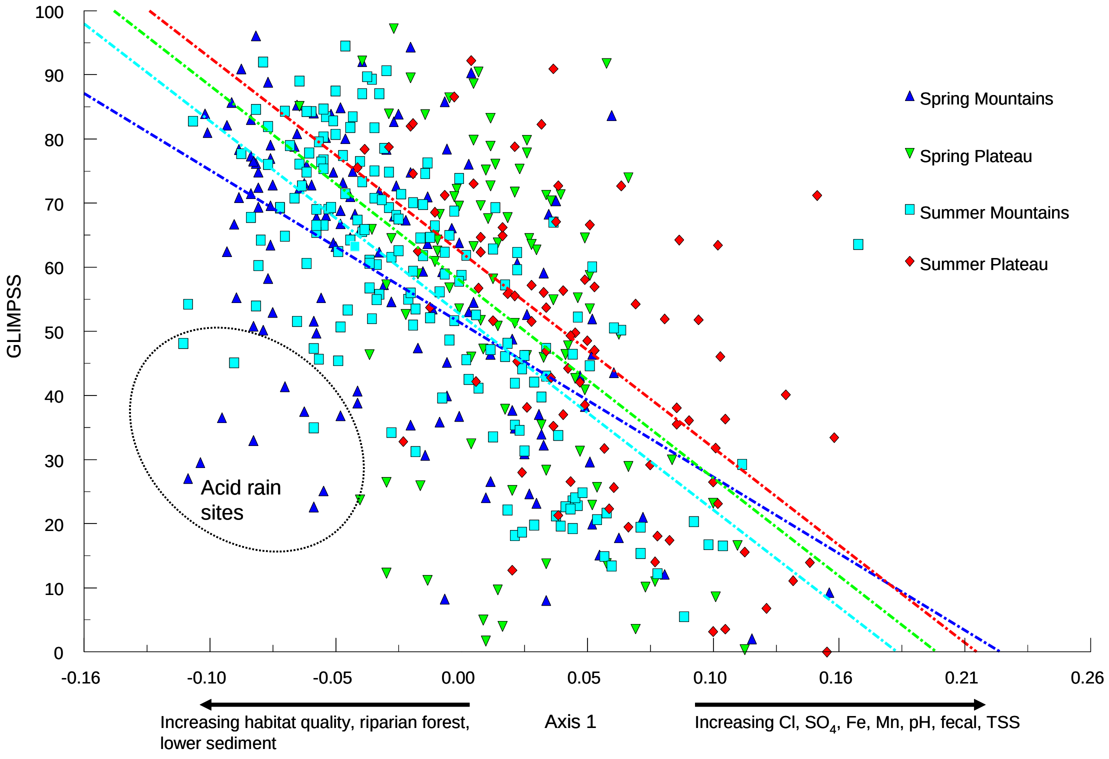


Figure 13. Scatter plot of VAL GLIMPSS scores and PCA axis 1 (as in Figure 11).

8.6 GLIMPSS Relation to Stream Size

Catchment areas were not available for all sites in the WAB database. Out of roughly 2000 sites with calculated areas, we found that average stream width was linearly and strongly correlated to catchment area (r =0.89). Median stream width across all CAL REF and STRESS sites was 3.0 m and 3.2 m, respectively. At VAL sites, median REF and STRESS stream width was also similar (REF=3.0 m, STRESS=3.1 m). Recall that NMDS ordination axes 1 and 2 were somewhat correlated to stream width (r= -0.21, -0.24, respectively), indicating that REF taxonomic community structure was partially driven by stream size. Although catchment areas were not available for all REF sites, a sufficient number were available for analysis. REF GLIMPSS scores were compared to log catchment area with simple linear regressions: MT Sp had 92 sites, MT Su had 109 sites, PL Sp had 33 sites, and PL Su had 29 sites. Although certain metrics (and taxa) might respond to increasing stream size, the aggregate index as a whole generally did not. Catchment area (in sq. mi.) showed no significant relationship to GLIMPSS within individual MT strata or PL Sp (Figure 14). In general, GLIMPSS scores at REF sites >20 sq. mi. were not distinctly different than sites draining less than 1 sq. mi. However, sites in the PL Su stratum showed a trend of decreasing GLIMPSS score with increasing stream size (r^2=0.24, p<0.05). We found that 3 of the 9 metrics in PL Su (# Intolerant (<3) Taxa, HBI, and % mEPT) were negatively related to catchment area but this relationship was driven by only 4 sites. We believe that this actually indicates that REF site quality was reduced at these 4 sites, rather than indicating a true catchment area effect on metric values. However, due to this small sample size, more data will be needed to predict catchment area effects in the PL bioregion.

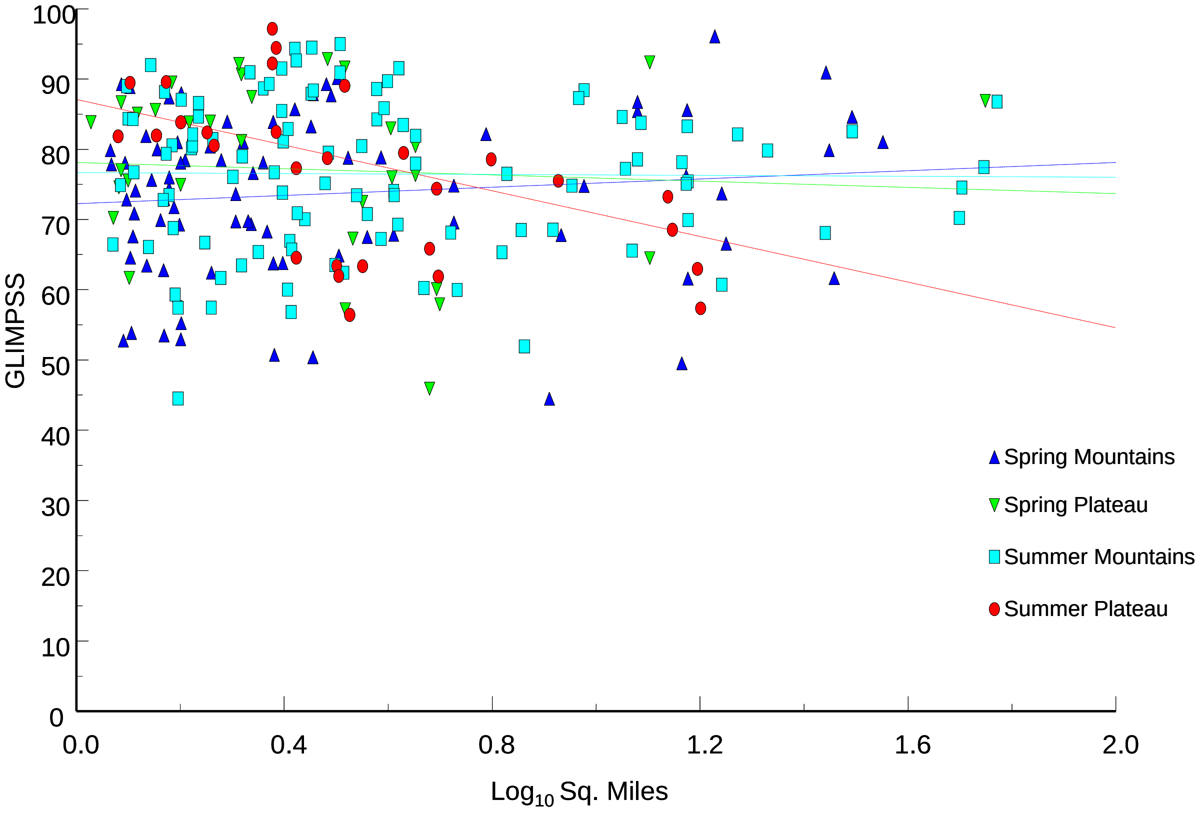


Figure 14. Scatter plots of combined CAL and VAL REF site GLIMPSS scores vs. catchment area (square miles) with linear regression lines.

8.7 Winter Index Period Development

Although WVDEP does not routinely sample macroinvertebrates during the winter months (*i.e.*, December to late-February), an effort was made to define the biological reference condition for this index period to accommodate special programmatic studies (*e.g.*, TMDLs, spill response) and to give the regulated community or other agencies an opportunity to assess data they collect during this time. A total of 29 Mountain Winter (MT Win) REF sites and 18 Plateau Winter (PL Win) REF sites were used in this analysis with the original spring REF sites. Macroinvertebrate genera excluding rare taxa (*i.e.*, occurring at <2.5% of all sites) from REF sites were ordinated using NMDS to further explore classification of communities among strata with the inclusion of the winter index period. NMDS produced a 3-dimensional solution after 108 iterations and a stress of 19.3%. Axis 2 and 3 explained the greatest variance in the ordination and are plotted in Figure 15. This shows that winter communities were somewhat distinct from spring index period (*i.e.*, minimal overlap with spring points), within respective MT and PL bioregions.

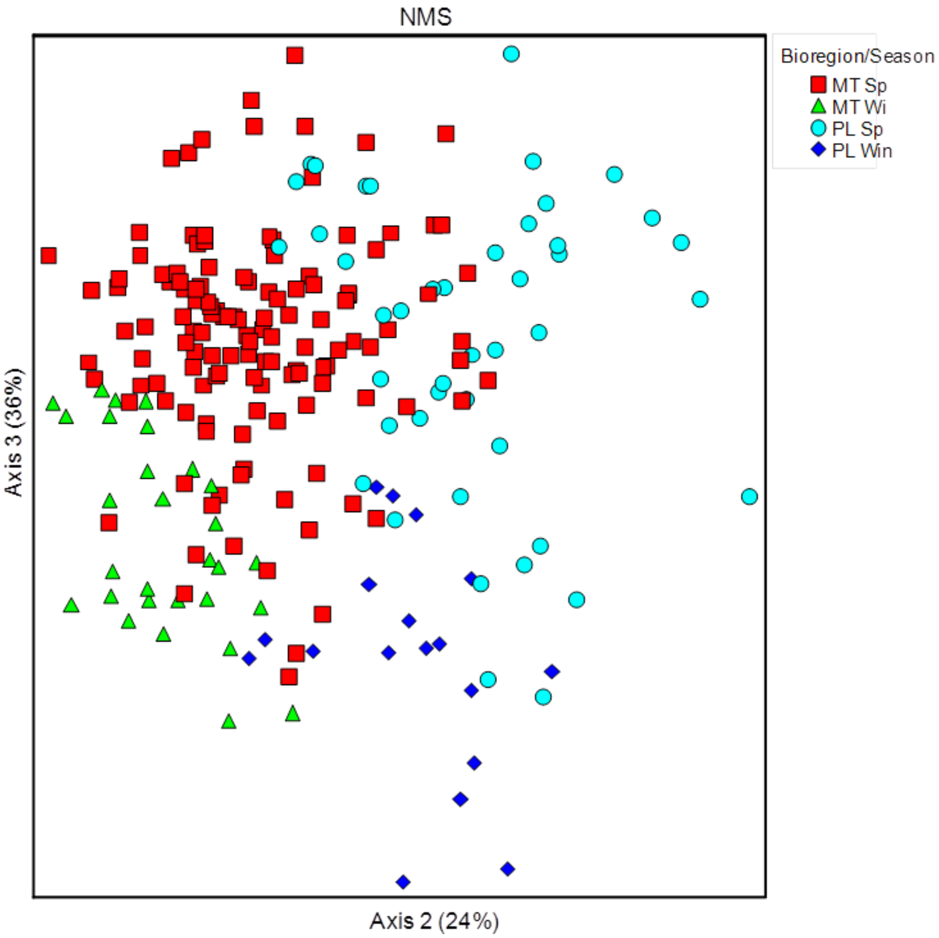
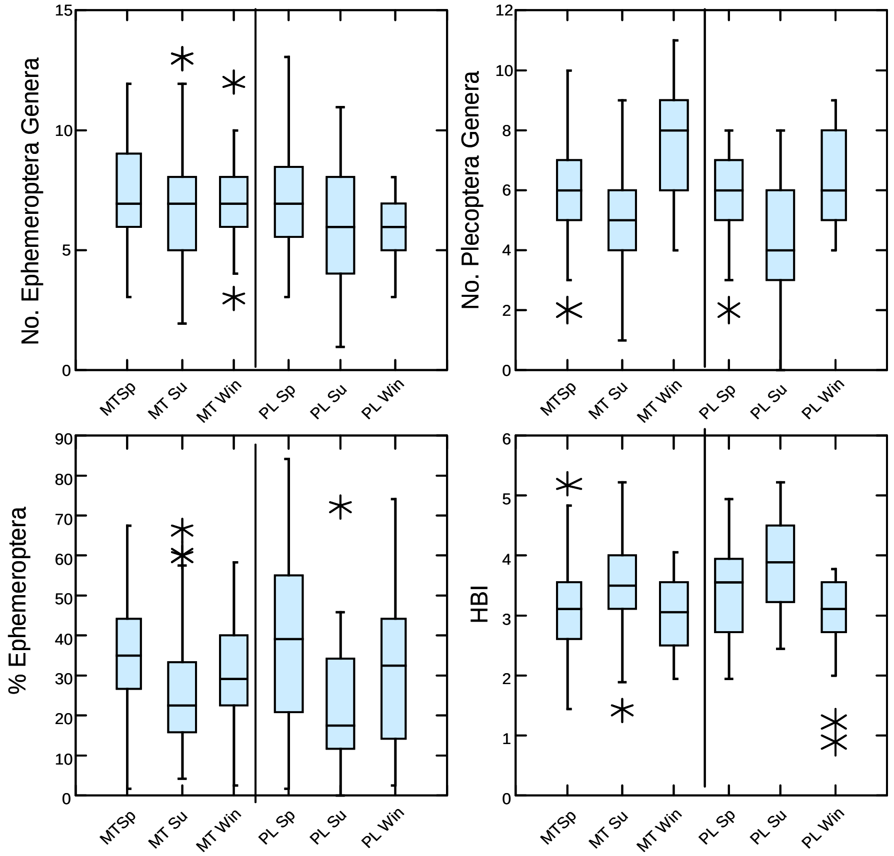


Figure 15. NMDS ordination of Spring-Winter REF sites grouped by bioregion-season. Stress=19.3 for a 3-dimensional solution. Axis 2 and 3 accounted for 60% of the cumulative variance.

Despite relatively distinctive differences in community structure between winter and spring (Figure 15), we compared metric values from REF site macroinvertebrate data collected in the winter (December to late-February) to metric distributions from spring and summer. Because few STRESS sites were available for direct testing in this index period (WVDEP WAB database), we were primarily interested in developing GLIMPSS winter criteria based on selected indicator metrics proven to be effective in other strata. For comparative purposes, Figure 16 shows examples of metric distributions across seasons and bioregions. Although there was considerable overlap, it was apparent that winter REF metrics were most similar to spring samples compared to summer samples. Thus, metric suites from MT Sp and PL Sp were applied to the winter dataset and SVs (*i.e.*, ceilings and floors) were calculated from all combined sites within each respective bioregion.



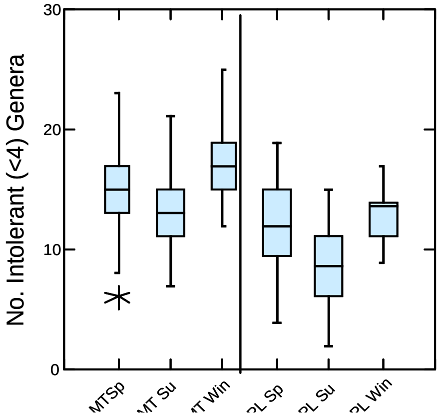


Figure 16. Boxplots of example metrics (No. Ephemeroptera Genera, No. Plecoptera Genera, % Ephemeroptera, HBI, and No. Intolerant (<4) Genera) among bioregions and seasons at REF sites. The vertical lines separate MT from PL bioregions.

8.8 Final Index Best Standard Values and GLIMPSS Reference Percentile Distributions

Data from the calibration and validation sets were combined with all other sites in the database and BSVs and WSVs were recalculated to produce final GLIMPSS scoring benchmarks for six strata (including winter index period). Table 10 shows ceiling and floor values for all metrics in each stratum. Overall, there were only minor adjustments made from the calibration SVs. A few metrics showed substantial variation across seasons (*e.g.*, % EPT (minus *Cheumatopsyche*) and No. Intolerant Genera in PL Sp versus PL Su) or bioregions (*e.g.*, No. Intolerant Genera <4 in MT Sp versus PL Sp); however, some metrics were not considerably different across strata. Table 11 shows calculated percentiles for reference GLIMPSS scores within each stratum. Separate percentile calculations were done with MT Su REF sites >60 sq. mi. (modified MT Su GLIMPSS) using SVs from all MT Su sites >60 mi.^{2} (see Appendix E). Figure 15 indicates the sensitivity of the GLIMPSS using combined calibration and validation datasets.

Table 10. Final (CAL and VAL) standard values (ceiling and floor) used for GLIMPSS. MT Su >60 sq. mi. SVs are found in Appendix E.

	MT Sp&Win (n=732)		MT Su (n=1530)		PL Sp&Win (n=692)		PL Su (n=858)	
	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor
No. Total Genera			38	14			34	14
No. Intolerant Genera <4	19	1	15	0	15	1		
No. Intolerant Genera <3							7	0
No. Ephemeroptera Genera	10	1	9	0	10	1	7	0
No. Plecoptera Genera	8	0	7	0	7	0		
No. Trichoptera Genera	7	1						
No. Clinger Genera	20	4	19	5	17	3	15	4
No. Scraper Genera	8	0					7	1
No. Shredder Genera			5	1				
HBI (Genus level)	6.18	2.23	6.20	2.79	6.64	2.49	6.32	3.82
% Dominant 5 Genera	92	48	91.4	51			91.5	52.8
% Tolerant (>6)					69.5	0.0		
% Ephemeroptera	59.7	0.5						
% mEPT (minus *Cheumatopsyche)*			86.0	5.2	90.8	2.5	67.1	1.3
% Orthocladiinae	52.7	0.5	37.1	0.0				
% Chironomidae							68.8	3.3
% Chironomidae+Annelida					84.6	1.8		

Table 11. Final GLIMPSS percentiles based on all REF sites within each stratum.

	5th %	10th %	25th %	Median	75th %
Spring Mountains (n=128)	52.7	59.6	66.2	73.1	79.3
Spring Plateau (n=44)	60.9	63.6	65.8	75.0	83.2
Summer Mountains (n=181)	55.2	59.9	67.2	73.7	79.3
Summer Plateau (n=38)	57.3	62.7	65.4	73.5	80.1
Winter Mountains (n=29)	62.6	66.8	71.7	76.2	82.4
Winter Plateau (n=18)	65.0	66.1	73.9	80.4	83.2

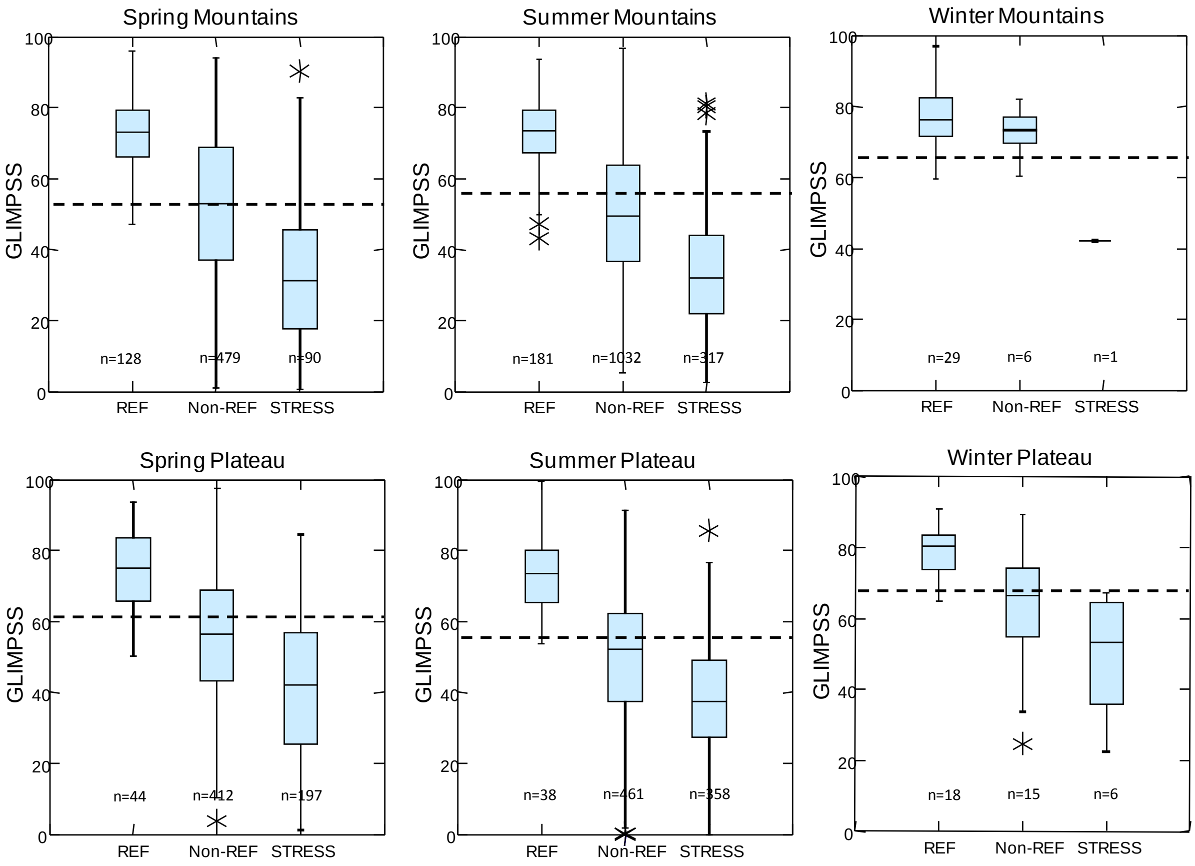


Figure 17. Boxplots of combined calibration and validation GLIMPSS REF vs. Non-REF and STRESS scores using final SVs for 6 strata (MT Su >60 sq. mi. shown in Appendix E). Dashed horizontal lines denote the approximate 5th percentile of the reference distribution.

8.9 Index Precision

8.9.1 Same Day Replicates

Table 12 shows metric and GLIMPSS precision results generated from the 96 replicated sites. Precision of GLIMPSS metrics and the overall aggregate score was satisfactory based on same day replicate sampling. Plecoptera and Shredder richness and several compositional metrics (*e.g.*, % Orthocladiinae, % Tolerant (>6)) generally had higher variability while HBI and % Dominant 5 Genera had the lowest. We believe this only represents a general sense of metric precision, since component metrics were compared using all 96 sites, rather than using individual strata. We assume that metric variability is less within each respective stratum, since low variability was a factor considered in initial metric selection. Final GLIMPSS scores were scored with respect to stratum type. Estimates of precision are essential to the WVDEP WAB as it helps identify sampling method and laboratory processing variability for quality assurance purposes, and can be used to compare precision among field teams to identify areas for improvement. Estimates of measurement error also provide a means to compare GLIMPSS scores between single site observations.

Table 12. Statistics for repeated samples including mean square error (MSE), standard deviation (SD), population mean of all sites, one-tailed 90% confidence limits for a single sample, and % coefficient of variation (%CV). Metric statistics were calculated for all sites regardless of bioregion/season stratification.

n=96 pairs	MSE	SD	Population Mean	One-tailed 90% C.I.	CV (%)
No. Total Genera	14.5	3.8	25.1	4.9	15.1
No. Intolerant (<4) Genera	3.2	1.8	6.0	1.8	29.7
No. EPT Genera	3.4	1.8	10.7	1.8	17.3
No. Ephemeroptera Genera	1.0	1.0	4.4	1.3	22.8
No. Plecoptera Genera	0.8	0.9	2.5	1.2	36.0
No. Trichoptera Genera	1.2	1.1	3.7	1.4	29.7
No. Clinger Genera	4.6	2.2	11.0	2.7	19.5
No. Scraper Genera	1.3	1.1	3.6	1.4	31.2
No. Shredder Genera	0.7	0.9	2.5	1.1	34.6
HBI (Genus)	0.06	0.24	4.63	0.31	5.1
% Tolerant (>6)	9.8	3.1	8.7	4.0	36.0
% Dominant 5 Genera	67.8	8.2	64.1	10.6	12.8
% Ephemeroptera	54.6	7.4	20.1	9.5	36.8
% mEPT (minus *Cheumatopsyche*)	44.7	6.7	48.1	8.5	13.9
% Orthocladiinae	22.5	4.7	12.2	6.1	38.8
% Chironomidae+Annelida	42.4	6.5	24.7	8.3	26.4
% Chironomidae	45.6	6.8	23.9	8.7	28.3

Sources of GLIMPSS variability were examined closer, by separating out strata and general scoring ranges. PL Su had the highest variability of all strata with respect to %CV and RPD (but these were highly dependent on the low population mean). Similarly, higher variation was noted in the lowest scoring range but %CV and RPD were highly dependent on the low population mean. Overall, estimates of measurement error (SD) were fairly stable across strata and scoring ranges (Table 13). Stribling et al. (2008) recommended a %CV of 10% and RPD 15%, for multi-metric index measurement quality objectives for Montana's stream assessment. Our results were on average, higher than those values found in Montana but %CV and RPD values were driven by the low mean of the study sites. Note that in the higher scoring range (60-90, n=28 site pairs), %CV and RPD were 12.6% and 15.9%, respectively.

Table 13. GLIMPSS precision estimates and statistics (including Mean Relative Percent Difference (RPD)) given for each stratum and the combined dataset. Statistics also shown for generalized scoring ranges of good quality (60-90), medium quality (30-59), and low quality (0-29).

GLIMPSS by Stratum	MSE	SD	Population Mean	One-tailed 90% C.I.	CV (%)	RPD
MT Sp (n=17)	63.6	7.9	54.1	10.2	14.7	19.1
MT Su (n=61)	61.1	7.8	47.4	10.0	16.5	21.0
PL Sp (n=9)	31.1	5.6	48.8	7.1	11.4	23.3
PL Su (n=17)	42.9	6.6	38.7	8.4	16.9	31.1
All Strata (=96)	53.7	7.3	47.1	9.4	15.6	22.2
GLIMPSS by Score Ranges						
60 to 90 (n=28)	71.1	8.4	66.9	10.8	12.6	15.9
30 to 59 (n=46)	58.5	7.6	44.6	9.8	17.2	20.0
0 to 29 (n=22)	28.7	5.4	22.3	6.9	24.0	36.9

8.9.2 Annual Revisits at REF Sites

GLIMPSS scores at REF sites were also evaluated for precision, by comparing annual revisits to 37 sites within the same index period (Table 14). These sample events ranged from 1 to 8 years apart. Here, we assumed that true REF site environmental condition was unchanged, so the MMI should not vary considerably from year to year. There are many caveats to this assumption: timing within an index period, preceding weather patterns, and uncertainty surrounding any unknown human impacts that may have occurred within the catchment between sample years. Despite these factors, SD, CV (%) and RPD were good. Individual stratum precision results are provided in Table 14; however, due to low sample size in each stratum, the estimates from all combined strata should be applied. These estimates incorporate natural spatial, method and temporal variability and can be used to detect long-term trends in GLIMPSS scores at individual sites through time, after correcting for any changes due to natural and method variability.

Table 14. GLIMPSS precision estimates and statistics based on REF annual revisits.

REF Revisits	MSE	SD	Population Mean	One-tailed 90% C.I.	CV (%)	RPD
GLIMPSS All Strata (n=37)	39.2	6.3	74.6	8.0	8.4	18.9
MT Sp (n=14)	46.7	6.8	76.2	8.7	8.9	20.4
MT Su (n=12)	33.6	5.8	75.4	7.4	7.7	17.5
PL Sp (n=5)	77.8	8.8	83.3	11.3	10.6	17.9
PL Su (n=6)	31.7	5.6	77.2	7.2	7.3	18.8

8.10 Taxonomic Composition at Reference Sites

Although there was some degree of overlap between taxa and metrics occurring in each stratum, both bioregional and seasonal differences were observed indicating that indeed taxonomic composition varies across the State both temporally and spatially. Table 15 lists the top 20 most frequently occurring genera (excluding Chironomidae) collected at REF sites within individual strata. Chironomid midges were omitted for this comparison because this group is highly ubiquitous in all benthic samples and we wanted to focus on other taxa that were more indicative of the reference condition.

Several ubiquitous genera could be found across all strata. Namely, the stoneflies *Leuctra* and *Acroneuria*, and mayflies *Paraleptophlebia*, *Baetis* and *Maccaffertium*, the crane fly *Hexatoma*, and the caddisfly *Diplectrona* made the top 20 list in at least 5 of the 6 strata. Some genera were fairly restricted to a particular season or bioregion. For example, the stonefly *Amphinemura* and the mayfly *Ephemerella* were in the top 5 of MT and PL sites during the spring index period, but were absent from the top 20 lists in the summer. Similarly, the stonefly *Isoperla* and the mayfly *Ameletus* were absent from samples taken from the summer index period. The mayfly *Cinygmula* was frequent only in the MT bioregion during the spring but was taken sporadically in MT summer and winter and was entirely absent from all seasons in the PL. The elmid beetle *Optioservus* made the top 10 list only at PL sites (all seasons) but occurred much less frequently at MT sites. During the summer, warmwater affiliates such as the caddisfly *Ceratopsyche* were more common at both MT and PL sites, but did not make the top 20 list in the MT spring or the PL spring category. Other important caddisflies included *Rhyacophila*, which made the top 20 list in all seasons and regions except the PL during the summer, and *Dolophilodes*, which was common in all seasons in the MT region but in the PL, it only was common in the PL Su, and absent or taken sporadically in other PL seasons.

It is well known that macroinvertebrate life history phenologies are responsible for their seasonal distribution patterns (Sweeney 1984). These patterns are linked to adaptations to changing abiotic conditions such as increasing temperatures and chemical concentrations associated with summer weather, changes in food availability, as well as direct competition for space and food resources with other species whose life histories may overlap. For example, late fall and winter months bring about large populations of certain leaf-shredding stonefly genera (*e.g.*, *Allocapnia*, *Taeniopteryx*, *Prostoia*) that have low critical thermal maximums (*e.g.*, cold “stenotherms”) or respond to high allochthonous inputs and other phenological cues, yet these taxa are absent from streams in late-spring through late-summer. The absence of taxa from samples in different seasons indicated that these taxa are in the egg stage, or occur in deeper sediments as diapausing, or resting individuals.

Unlike fishes, macroinvertebrate spatial distribution across ecoregions or other geographical strata is far less documented. Thus, an effort was made here to establish seasonal and geographical occurrences of taxa and to further corroborate findings in the ordination analyses. While the multi-metric GLIMPSS is designed to use summary biological attributes of the community (*i.e.*, metrics) to detect impacts to stream ecosystem health, it is often useful to compare taxa lists from new sites to those taxa known to occur frequently at reference sites. This is different and less complex than predictive modeling (*e.g.*, RIVPACS) which estimates the probabilities of capturing individual taxa based on a gradient of quantitative physical attributes of sites (Hawkins et al. 2000). Instead, these frequency lists simply help to provide further interpretation of the fauna expected to occur at new sites in relation to the reference condition.

Table 15. List of top 20 frequently occurring taxa (excluding Chironomidae) found at all REF sites within each stratum. Taxa are sorted by relative frequency (%).

MT Win	%	MT Sp	%	MT Su	%	PL Win	%	PL Sp	%	PL Su	%
Epeorus	96.6	*Leuctra*	94.2	*Leuctra*	93.5	*Epeorus*	88.9	*Amphinemura*	95.7	*Optioservus*	92.5
Diplectrona	82.8	*Paraleptophlebia*	90.5	*Baetis*	89.2	*Neophylax*	83.3	*Paraleptophlebia*	82.6	*Leuctra*	87.5
Paraleptophlebia	82.8	*Amphinemura*	88.3	*Paraleptophlebia*	83.2	*Cheumatopsyche*	77.8	*Isoperla*	76.1	*Baetis*	85.0
Rhyacophila	82.8	*Ephemerella*	87.6	*Dolophilodes*	82.7	*Maccaffertium*	77.8	*Leuctra*	76.1	*Cheumatopsyche*	82.5
Dolophilodes	72.4	*Epeorus*	86.9	*Rhyacophila*	65.4	*Optioservus*	72.2	*Ephemerella*	73.9	*Hexatoma*	80.0
Leuctra	72.4	*Baetis*	80.3	*Maccaffertium*	65.4	*Allocapnia*	66.7	*Epeorus*	67.4	*Psephenus*	75.0
Acroneuria	69.0	*Diplectrona*	78.8	*Diplectrona*	64.3	*Sweltsa*	66.7	*Ameletus*	63.0	*Maccaffertium*	72.5
Hexatoma	69.0	*Rhyacophila*	74.5	*Hexatoma*	58.4	*Isoperla*	61.1	*Baetis*	63.0	*Acroneuria*	67.5
Maccaffertium	69.0	*Hexatoma*	60.6	*Acroneuria*	57.3	*Paraleptophlebia*	61.1	*Rhyacophila*	58.7	*Sweltsa*	67.5
Neophylax	69.0	*Haploperla*	56.2	*Epeorus*	57.3	*Acroneuria*	55.6	*Haploperla*	56.5	*Stenelmis*	65.0
Paracapnia	69.0	*Isoperla*	55.5	*Ceratopsyche*	51.9	*Ameletus*	55.6	*Maccaffertium*	50.0	*Nigronia*	57.5
Sweltsa	65.5	*Cinygmula*	54.7	*Cheumatopsyche*	51.9	*Amphinemura*	55.6	*Optioservus*	50.0	*Stenonema*	52.5
Cheumatopsyche	62.1	*Oulimnius*	52.6	*Pteronarcys*	49.2	*Ephemerella*	55.6	*Stenelmis*	50.0	*Polycentropus*	45.0
Polycentropus	58.6	*Dolophilodes*	50.4	*Dicranota*	44.9	*Rhyacophila*	55.6	*Hexatoma*	45.7	*Ceratopsyche*	42.5
Prosimulium	58.6	*Acroneuria*	48.9	*Oulimnius*	44.9	*Chimarra*	50.0	*Acentrella*	43.5	*Dolophilodes*	42.5
Baetis	55.2	*Sweltsa*	45.3	*Simulium*	44.9	*Hexatoma*	50.0	*Tipula*	43.5	*Acentrella*	40.0
Dicranota	55.2	*Polycentropus*	43.8	*Cambarus*	43.2	*Taeniopteryx*	50.0	*Neophylax*	41.3	*Paraleptophlebia*	40.0
Ephemerella	51.7	*Cambarus*	43.1	*Optioservus*	40.5	*Tipula*	50.0	*Cheumatopsyche*	39.1	*Simulium*	40.0
Bezzia/Palpomyia	48.3	*Pteronarcys*	42.3	*Sweltsa*	40.5	*Leuctra*	44.4	*Cambarus*	37.0	*Ectopria*	37.5
Ceratopsyche	48.3	*Drunella*	40.1	*Polycentropus*	38.9	*Prostoia*	44.4	*Psephenus*	30.4	*Cambarus*	32.5
Pteronarcys	48.3	*Neophylax*	40.1	*Acentrella*	38.4	*Diplectrona*	38.9	*Acroneuria*	26.1	*Diplectrona*	32.5

8.11 Relationships between WVSCI and GLIMPSS

Although there is evidence that family-level taxonomy provides sufficient information for detecting obvious impacts to wadeable streams (Gerritsen et al. 2000, Burton and Gerritsen 2003, Bailey et al. 2001), finer-scaled taxonomic resolution (*e.g.*, genus or species) produces a more comprehensive and detailed evaluation of the benthic assemblage (Guerold 2000, Lenat and Resh 2001, Vinson and Hawkins 2003). Genus level taxonomy allowed for better characterization of seasonal and regional patterns. Genus-level taxonomy allowed for more range in richness metrics, and refinement of other metrics (*e.g.*, %EPT vs. %EPT minus facultative to tolerant *Cheumatopsyche*) and tolerance values (*i.e.*, a family containing several genera may have a wide range of tolerance to stress compared to the individual genera). Moreover, genus-level taxonomy better represents the "aquatic life" for designating Aquatic Life Use under the CWA.

Figure 18 shows the relationship between family-level WVSCI and GLIMPSS. Although the two indices were highly correlated in all strata, there was 18% disagreement in assessing impairment. For example, data points in the lower-right quadrant of each stratum's scatterplot (Figure 18), indicate that the WVSCI appeared to underestimate impairment (*i.e.*, many sites score above the WVSCI threshold but fall below the GLIMPSS thresholds). Thus, out of 3737 assessed sites, there were 584 (~16%) disagreements where GLIMPSS showed impairment and WVSCI did not, with the majority of these disagreements found in the MT Su stratum (>20%). There were fewer instances where WVSCI indicated impairment (upper left quadrant) but GLIMPSS did not (77 of 3737, or 2.1%).

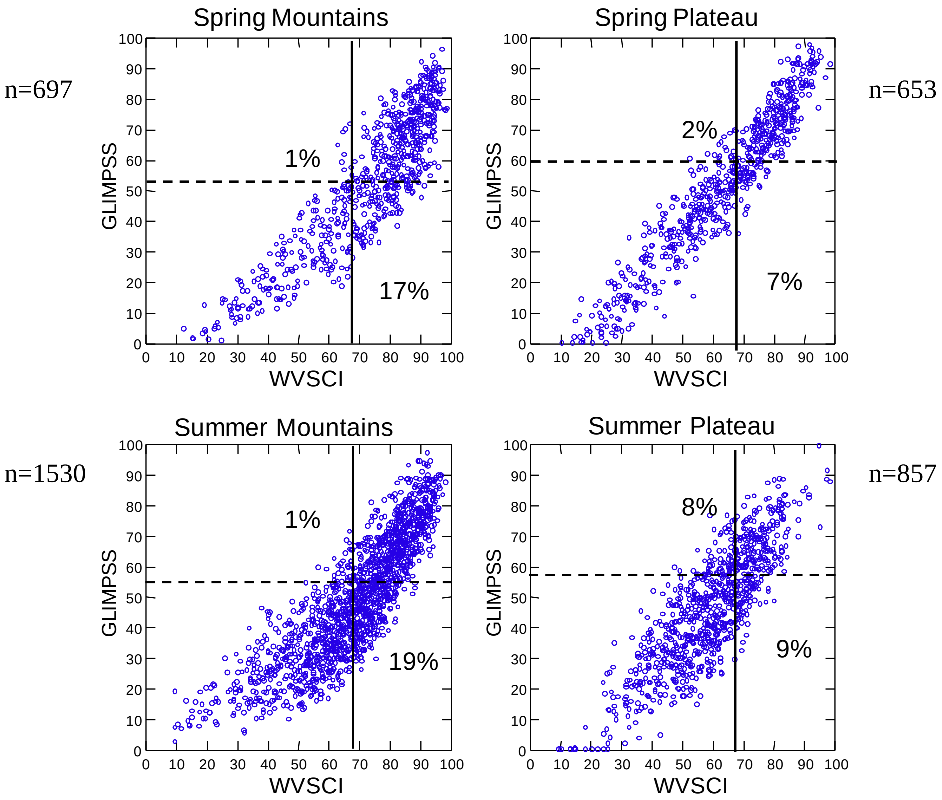


Figure 18. Scatterplots of GLIMPSS scores versus WVSCI scores by stratum. Horizontal and vertical dashed lines represent approximate 5th percentile impairment thresholds for each stratum for GLIMPSS and WVSCI, respectively. Total percent disagreement shown.

8.12 GLIMPSS Scoring Criteria

As with the WVSCI, GLIMPSS thresholds to demarcate impaired from non-impaired status are based upon the 5th percentile of the reference distribution. A main difference between the two models is that GLIMPSS applies stratum-specific criteria for assessment purposes. WVDEP may also apply the GLIMPSS index to identify exceptionally healthy streams or to communicate severity of impact to biological integrity. For example, the 25th percentile of the reference distribution can be used as a lower threshold to identify exceptional biological assemblages found in the State. Similarly, GLIMPSS values below the impairment threshold can be partitioned to provide categories that reflect increased stress to biological communities (*e.g.*, degraded, severely degraded). Table 16 presents stratum-specific GLIMPSS scores indicating the 5th and 25th percentile values and equal bisection of the impairment range.

Table 16. GLIMPSS scoring criteria for all strata. Scores are rounded to nearest whole number. MT Sp = Mountain Spring, MT Su = Mountain Summer, MT Su>60 = Mountain Summer >60 Sq. Mi., MT Win = Mountain Winter, PL Sp = Plateau Spring, PL Su = Plateau Summer, and PL Win = Plateau Winter.

	MT Sp	MT Su	MT Su>60	MT Win	PL Sp	PL Su	PL Win
25th Percentile	66	67	66	72	66	65	74
5th Percentile	53	55	52	63	61	57	65
Impairment Threshold							
Increased Severity ↓	26-52	27-54	26-51	31-62	30-60	28-56	32-64
of Impact	<26	<27	<26	<31	<30	<28	<32

9.0 Conclusions and Recommendations

The GLIMPSS is a powerful yet practical tool for evaluating point and nonpoint source impacts to water quality and stream habitat in wadeable, riffle-run streams that are not ephemeral and will be used by WVDEP for 305(b) reporting, 303(d) listing and de-listing, water quality enforcement cases, or to help identify new high quality streams in need of further protection. Improvements over the family-level WVSCI were noted as benthic assessments using genus-level taxonomy provided WVDEP with distinct seasonal and geographical classification strata to help refine aquatic life uses and ecological expectations across the State. It also allowed for the selection of indicator metrics that best track stressors during different seasons and within particular bioregions. Although the GLIMPSS did not vary considerably with catchment area, care should be applied when assessing sites larger than 60 sq. mi. since fewer REF sites were available in this category. In summer samples of the mountain ecoregions (67 and 69), MT Su >60 sq. mi. criteria will be used for these larger sites. Future work will focus on the development of reference conditions and expectation criteria in larger wadeable and non-wadeable streams and rivers throughout the State.

The following list gives methodological and implementation requirements for the GLIMPSS to be effective. Non-agency personnel using the GLIMPSS must also adhere to these requirements for all monitoring and assessment applications.

- Sample methodology (*e.g.*, sampling gear and sample area) –Identical sampling area (4–0.25m^2) and gear (0.5 m rectangular kicknet with 500µm- mesh) should be used in riffle/run habitat. In limited circumstances, 0.3 m D-frame nets with comparable mesh size can be used as long as 1 m^2 total area is sampled.

- Laboratory subsampling–Samples in which more than the target subsample size was picked (200 ±20%) should be re-sorted to obtain the preferred number of organisms. As a rule-of thumb, samples containing less than 100 organisms should be scrutinized by qualified biologists for comparability before applying the GLIMPSS. These sites may be heavily impacted, or were recently subjected to drought or scour events.

- Taxonomic Resolution–Genus-level taxonomy (including Chironomidae) is required. Some taxa left at higher group levels are acceptable (*e.g.*, Nematoda). If higher taxonomy is necessary (*e.g.*, early instar or damaged specimens), then these taxa should not be counted in richness metrics unless they are believed to be distinct from other genera identified in the sample. WVDEP WAB should be consulted for exact taxonomic resolution of some groups.

- Seasonality–Although the delineation of seasons is relatively straightforward, professional judgment should be applied when sample dates fall close to season cutoffs. For example, after a cooler than normal spring, sites at higher elevations may exhibit spring-like communities well into June. WVDEP recommends applying a 2-3-week buffer between seasons to remove seasonal uncertainties and improve assessment performance with the GLIMPSS. Sampling should not occur between mid-October through November. If samples collected in early-October yield >10% winter stoneflies (Taeniopterygidae <u>or</u> *Allocapnia*), then those samples should be scrutinized by qualified biologists for comparability to summer metrics and scoring criteria. Furthermore, if spring samples contain >50% *Prosimulium*, then those samples should also be scrutinized for overall comparability.

- Tolerance Values, Functional Feeding Groups, and Habit codes—Taxon-specific tolerance values differ somewhat across state programs. GLIMPSS metrics that rely on tolerance values (HBI), FFGs (No. Scraper or Shredder Genera), or habits (No. of Clinger Genera) are specifically calibrated to those used by WVDEP and these specific tolerance values or guild designations (as recognized by WVDEP) must be used.

- GLIMPSS calculations—Use only those BSVs and WSVs and stratum specific metrics found in Table 10 (also see Appendix D for scoring examples). MT Su sites >60 sq. mi. must use SVs found in Appendix E. Inclusion of non-GLIMPSS metrics by outside users of this index is forbidden and should only be used as supplementary information.

- Comparing GLIMPSS scores between strata—Because scoring is based on stratum-specific metrics and metric SVs, in order to compare scores between samples collected in different seasons (or regions), the "percent of threshold" can be easily calculated to further standardize scores for comparability (see Appendix D).

- While the GLIMPSS performs well in limestone -influenced streams (streams with minor carbonate geology), refinements to the index will be needed for streams with significant limestone geology (true limestone streams) and associated karst hydrology. These streams naturally maintain high alkalinities and relatively low temperatures annually, and generally display low diversity and high dominance by few taxa. This phenomenon does not typically occur in ecoregion 69, but is sporadic in ecoregion 66, 67 and 70.

10.0 Literature Cited

Alexander R.B., E.W. Boyer, R.A. Smith, G.E. Schwarz, and R.B. Moore. 2007. The role of headwater streams in downstream water quality. Journal of the American Water Resources Association 43:41–59.

Arnwine, D.H. and G.M. Denton. 2001. Development of regionally-based numeric interpretations of Tennessee's narrative biological integrity criterion. Tennessee Department of Environment and Conservation, Nashville, TN.

Arscott, D.B., J.K. Jackson, and E.B. Kratzer. 2006. Role of rarity and taxonomic resolution in a regional and spatial analysis of stream macroinvertebrates. Journal of the North American Benthological Society 25:977–997.

Astin, L.E. 2006. Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River basin, USA. Ecological Indicators 6:664-685.

Bady, P., S. Doledec, C. Fesl, S. Gayraud, M. Bacchi, and F. Scholl. 2005. Use of invertebrate traits for the biomonitoring of European large rivers: the effects of sampling effort on genus richness and functional diversity. Freshwater Biology 50:159-173.

Bailey, R.C., R.H. Norris, and T.B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society 20:280-286.

Barbour, M.T. and J. Gerritsen. 1996. Subsampling of benthic samples: a defense of the fixed-count method. Journal of the North American Benthological Society 15:386-391.

Barbour, M.T., J. Gerritsen, B.D. Snyder, 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. US Environmental Protection Agency, Office of Water, Washington, DC.

Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M.L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185-211.

Beauchard, O., J. Gagneur, and S. Brosse. 2003. Macroinvertebrate richness patterns in North African streams. Journal of Biogeography 30:1821-1833.

Black, R. W., M. D. Munn, and R. W. Plotnikoff. 2004. Using macroinvertebrates to identify biota-land cover optima at multiple scales in the Pacific Northwest, USA. Journal of the North American Benthological Society 23:340–362.

Blocksom, K.A. 2003. A Performance Comparison of Metric Scoring Methods for a Multimetric Index for Mid-Atlantic Highlands Streams. Environmental Management 31:670-682.

Blocksom, K.A., and B.R. Johnson. 2009. Development of a regional macroinvertebrate index for large river bioassessment. Ecological Indicators 9:313-328.

Botts, W. 2009. An index of biotic integrity for "true" limestone streams. Pennsylvania Department of Environmental Protection, Harrisburg, PA. Available at: http://www.portal.state.pa.us/portal/server.pt?open=512&objID=14295&mode=2&PageID=590867

Burton, J, and J. Gerritsen. 2003. A Stream Condition Index for Virginia Non-Coastal Streams. Report prepared for Virginia DEQ and US EPA by Tetra-Tech, Inc. http://www.deq.virginia.gov/watermonitoring/pdf/vastrmcon.pdf

Butcher, J.T., P.M. Stewart, and T.P. Simon. 2003. A benthic community index for streams in the Northern Lakes and Forests Ecoregion. Ecological Indicators 3:181-193.

Carter, J.L., S.V. Fend, and S.S. Kennelly. 1996. The relationships among three habitat scales and stream benthic invertebrate community structure. Freshwater Biology 35:109-124.

Chalfant, B. 2007. A benthic index of biotic integrity for wadeable freestone streams in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg, PA.

Childs, R.A. 2005. Bureau of Business and Economic Research. College of Business and Economics. West Virginia University, Morgantown, WV.

Dasher, G.R. 2001. The geology of Pendleton County. Pages 17-36 *In* G.R. Dasher (ed.) The Caves and Karst of Pendleton County, West Virginia. West Virginia Speleological Survey Bulletin 15:17-36.

Deshon, J.E. 1995. Development and application of the invertebrate community index (ICI). Pages 217-243 *in* W.S. Davis and T.P. Simon (eds.) Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, FL.

Dobrin, M, and D.J. Giberson. 2003. Life history and production of mayflies, stoneflies, and caddisflies (Ephemeroptera, Plecoptera, and Trichoptera) in a spring-fed stream in Prince Edward Island, Canada: evidence for population asynchrony in spring habitats. Canadian Journal of Zoology 81:1083-1095.

Doledec, S., B. Statzner, and M. Bournard. 1999. Species traits for future biomonitoring across ecoregions: patterns along a human-impacted river. Freshwater Biology 42:737-758.

Dow, C.L. and R.A. Zampella. 2000. Specific conductance and pH as indicators of watershed disturbance in streams of the New Jersey Pinelands, USA. Environmental Management 26:437-446.

Gerritson, J., J. Burton, and M.T. Barbour. 2000a. A stream condition index for West Virginia wadeable streams. Tetra Tech, Inc., Owing Mills, MD.

Gerritson, J., M.T. Barbour, and K. King. 2000b. Apples, oranges, and Ecoregions: on determining pattern in aquatic assemblages. Journal of the North American Benthological Society 19:487-496.

Griscom B., R. Brooks, G. Constantz, W. Myers, A. McQueen, G. Rocco, M. Easterling, J. Bishop. In review. Classification of watersheds in the Mid-Atlantic Highlands to assess condition and vulnerability. Environmental Monitoring and Assessment.

Guerold, F. 2000. Influence of taxonomic determination level on several community indices. Water Research 34:487-492.

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

Hawkins, C.P. and R.H. Norris. 2000. Performance of different landscape classifications for aquatic bioassessments: introduction to the series. Journal of the North American Benthological Society 19:367-369.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31-47 *In* W.S. Davis and T.P. Simon (eds.) Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, FL.

Jezerinac, R.F., G.W. Stocker, and D.C. Tarter. 1995. The crayfishes (Decapoda: Cambaridae) of West Virginia. Ohio Biological Survey Bulletin 10. 193pp.

Johnson, R.K., W. Goedkoop, and L. Sandin. 2004. Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. Freshwater Biology 49:1179-1194

Karr, J.R. and E.W. Chu. 1999. Restoring life in running waters: Better biological monitoring. Island Press, Washington, DC.

Kerans, B.L. and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. Ecological Applications 4:768-785.

Klemm, D.J., K.A. Blocksom, F.A. Fulk, A.T. Herlihy, R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, W.T. Thoeny, M.B. Griffith, and W. S. Davis. 2003. Development and evaluation of a macroinvertebrate Biotic Integrity Index (MIBI) for regional assessing Mid-Atlantic Highland streams. Environmental Management 31:656-669.

Kratzner, E.B., J.K. Jackson, D.B. Arscott, A.K. Aufdenkampe, C.L. Dow, L.A. Kaplan, J.D. Newbold, B.W. Sweeney. 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City drinking-water-supply watersheds**.** Journal of the North American Benthological Society 25:954–976.

Lenat, D.R., and V.H. Resh. 2001. Taxonomy and stream ecology – the benefits of genus- and species-level identifications. Journal of the North American Benthological Society 20:287-298.

Ludwig, J.A., and J.F. Reynolds. 1988. Statistical ecology: a primer on methods and computing. John Wiley and Sons, New York, NY.

Maxted, J.R., M.T. Barbour, J. Gerritsen, V. Poretti, N. Primrose, A. Silvia, D. Penrose, and R. Renfrow. 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. Journal of the North American Benthological Society 14:440-450.

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

Merritt, R.W. and K.W. Cummins (eds.). 1996. An introduction to the aquatic insects of North America. 3rd ed. Kendall/Hunt, Dubuque, IA.

Meyer, J.L. and J.B. Wallace. 2001. Lost Linkages and Lotic Ecology: Rediscovering Small Streams. Pages 295-317 *In* N.J. Huntly and S. Levin (eds.) Ecology: Achievement and Challenge. M.C. Press, Blackwell Science, Oxford, England.

Norris, R.H., and A. Georges. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. Pages 234-286 *In* D.M. Rosenberg and V.H. Resh, (eds.) Freshwater Biomonitoring and Benthic Macroinvertebrates. Kluwer Academic Publishers, Norwell, MA

OH EPA (Ohio Environmental Protection Agency). 2002. Field evaluation manual for Ohio's primary headwater habitat streams. Final Version 1.0. Division of Surface Water, Ohio Environmental Protection Agency, Columbus, Ohio. Available at: http://www.epa.state.oh.us/dsw/wqs/headwaters/PHWHManual_2002_102402.pdf

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118-125.

Paul, M.J., and J.L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333–365.

Perrin, C.J., and J.S. Richardson. 1997. N and P limitation of benthos abundance in the Nechako River, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences, 54: 2574-2583.

Pond, G.J. 2004. Effects of surface mining and residential land use on headwater stream biotic integrity in the eastern Kentucky coalfield region. Kentucky Department for Environmental Protection, Division of Water, Frankfort, KY.

Pond, G.J., S.M. Call, J.F. Brumley and M.C. Compton. 2003. The Kentucky macroinvertebrate bioassessment index: derivation of regional narrative ratings for wadeable and headwater streams. Kentucky Department for Environmental Protection, Division of Water, Frankfort, KY.

Resh, V.H., and E.P. McElravy. 1993. Contemporary quantitative approaches to biomonitoring using benthic invertebrates. Pages 159-194 *In* D.M. Rosenberg and V.H. Resh (eds.) Freshwater Biomonitoring and Benthic Macroinvertebrates. Kluwer Academic Publishers, Norwell, MA.

Resh, V.H. and J.K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. Pages 195-233 *In* D.M. Rosenberg and V.H. Resh (eds.) Freshwater biomonitoring and benthic macroinvertebrates. Kluwer Academic Publishers, Norwell, MA.

Rikard, M., and S. Kunkle. 1990. Sulfate and conductivity as field indicators for detecting coal-mining pollution. Environmental Monitoring and Assessment 15:49-58.

Rosenberg, D.M., and V.H. Resh. 1993. Introduction to freshwater biomonitoring and benthic macroinvertebrates. Pages 1-9 *In* D.M. Rosenberg and V.H. Resh (eds.) Freshwater Biomonitoring and Benthic Macroinvertebrates. Kluwer Academic Publishers, Norwell, MA.

Rueda, J., A. Camacho, F. Mezquita, R. Hernandez, and J.R. Roca. 2002. Effect of episodic and regular sewage discharges on the water chemistry and macroinvertebrate fauna of a Mediterranean stream. Water, Air, and Soil Pollution 140:425-444.

Southerland, M.T., G.M. Rogers, R.J. Kline, R.P. Morgan, D.M. Boward, P.F. Kazyak, R.J. Klauda and S.A. Stranko. 2007. Improving biological indicators to better assess the condition of streams. Ecological Indicators 7:751-767.

Sovell, L.A. and B. Vondracek. 1999. Evaluation of the fixed-count method for Rapid Bioassessment Protocol III with benthic macroinvertebrate metrics. Journal of the North American Benthological Society 18:420-426.

Stanford, J.A., and J.V. Ward. 1983. Insect species diversity as a function of environmental variability and disturbance in stream systems. Pages 265-278 *In* J.R. Barnes and G.W. Minshall, (eds.) Stream Ecology: Application and Testing of General Ecological Theory. Plenum Press, New York.

Stark, B.P., S.W. Szczytko, and C.R. Nelson. 1998. American stoneflies: a photographic guide to the Plecoptera. The Caddis Press, Columbus, OH.

Stephenson, S.L. 1993. An introduction to the upland forest region. Pages 1-9 *In* S. L. Stephenson (ed.) Upland Forests of West Virginia. McClain Printing Co., Parsons, West Virginia.

Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. American Geophysical Union Transactions 38:913-920.

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 27:58-67.

Sweeney, B.W. 1984. Factors influencing life-history patterns of aquatic insects. Pages 56-100 *In* V.H. Resh and D.M. Rosenberg (eds.) The ecology of aquatic insects. Praeger, New York, NY.

Thompson, R.M., and C.R. Townsend. 2000. Is resolution the solution?: the effect of taxonomic resolution on the calculated properties of three stream food webs. Freshwater Biology 44:413-422.

Townsend, C.R., S. Doledec, R. Norris, K. Peacock, and C. Arbuckle. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. Freshwater Biology, 48:768-785.

U.S. EPA. 2000. Level III ecoregions of the continental United States (revision of Omernik 1987). Corvallis, Oregon, USEPA-National Health and Environmental Effects Laboratory. Map-M-1, various scales.

Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. Journal of Agricultural, Biological and Environmental Statistics 2:370-388.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130–137.

Vinson, M. R. and C. P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15:392-399.

Vinson, M.R., and C.P. Hawkins. 2003. Broad-scale geographical patterns in local stream insect genera richness. Ecography 26:751-767.

Virginia Department of Environmental Quality (VADEQ). 2006. Using Probabilistic Monitoring Data to Validate the Non-Coastal Virginia Stream Condition Index. Water Quality Monitoring, Biological Monitoring and Water Quality Assessment Programs, Richmond, Virginia. VDEQ Technical Bulletin WQA/2006-001. Available at: http://www.deq.virginia.gov/probmon/pdf/scival.pdf

Waite, I.R., A.T. Herlihy, D.P. Larsen, and D.J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. Journal of the North American Benthological Society 19:429-441.

Waite, I.R., A.T. Herlihy, D.P. Larsen, N.S. Urquhart, and D.J. Klemm. 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, USA. Freshwater Biology 49:474-489.

West Virginia Department of Environmental Protection (WVDEP). 2011. Watershed Branch 2011 Standard Operating Procedures. West Virginia Department of Environmental Protection, Division of Water and Waste Management, Watershed Assessment Branch. Charleston, WV. Available at: http://www.dep.wv.gov/WWE/watershed/Pages/WBSOPs.aspx

Whittier, T.R., R.M. Hughes, J.L. Stoddard, G.A. Limnicky, D.V. Peck, and A.T. Herlihy. 2007. A structured approach for developing indices of biotic integrity: three examples from streams and rivers of the western USA. Transactions of the American Fisheries Society 136:718-735.

Woods, A.J., J.M. Omernik, D.D. Brown, and C.W. Kiilsgaard. 1996. Level III and IV Ecoregions of Pennsylvania and the Blue Ridge Mountains, the Ridge and Valley, and the Central Appalachians of Virginia, West Virginia, and Maryland. EPA/600R-96/077. USEPA, ORD, Corvallis, OR.

Appendix A. List and definitions of metrics evaluated for the GLIMPSS

Table A - 1. List and definitions of metrics evaluated for the GLIMPSS

Category	**Metric**	**Description**	**Response to Disturbance**
Richness	No. Total Genera	Total No. of Distinct Genera	Decrease
	No. EPT Genera	No. Mayfly+Stonefly+Caddisfly Genera	Decrease
	No. Ephemeroptera Genera	No. Mayfly Genera	Decrease
	No. Plecoptera Genera	No. Stonefly Genera	Decrease
	No. Trichoptera Genera	No. Caddisfly Genera	Decrease
Composition	% Ephemeroptera	Rel. Abundance of Mayflies	Decrease
	% Ephem (no Baetis)	Rel. Abundance of Mayflies (less *Baetis*)	Decrease
	% Plecoptera	Rel. Abundance of Stoneflies	Decrease
	% Trichoptera	Rel. Abundance of Caddisflies	Decrease
	% EPT	Rel. Abundance of Mayflies+Stoneflies+Caddisflies	Decrease
	% mEPT (minus Cheum)	Rel. Abundance of Mayflies+Stoneflies+Caddisflies (less *Cheumatopsyche*)	Decrease
	% mEPT2 (minus Cheum+Baetis)	Rel. Abundance of Mayflies+Stoneflies+Caddisflies (less *Cheumatopsyche+Baetis*)	Decrease
	% Chironomidae	Rel. Abundance of Midges	Increase
	% Chir+Annel	Rel. Abundance of Midges+Worms	Increase
	% Annelida	Rel. Abundance of Worms	Increase
	% Chironomini	Rel. Abundance of Midges in the Tribe Chironomini	Increase
	% Orthocladiinae	Rel. Abundance of Midges in the Subfamily Orthocladiinae	Increase
	% Tanytarsini	Rel. Abundance of Midges in the Tribe Tanytarsini	Decrease
	% Hydropsychidae	Rel. Abundance of Netspinner Caddisflies	Increase
	% Non-Insects	Rel. Abundance of Non-Insect Taxa	Increase
Tolerance	No. Intol <3 Genera	No. Intolerant Genera w/ Tol. Value <3	Decrease
	No. Intol <4 Genera	No. Intolerant Genera w/ Tol. Value <4	Decrease
	% Tolerant (>6)	Rel. Abundance of Tolerant Taxa w/ Tol. Value >6	Increase
	HBI (Genus)	Hilsenhoff Biotic Index (Genus level)	Increase
Dominance	% 5 Dominant Taxa	Rel. Abundance of Top 5 Dominant Genera	Increase
Functional	No. Scraper	No. of Scraper Genera	Decrease
Group	No. Shredder	No. of Shredder Genera	Decrease
Richness	No. Filterer	No. of Filter Feeder Genera	Increase
	No. Predator	No. of Predator Genera	Decrease
	No. Collector	No. of Collector-Gatherer Genera	Increase
Functional	% Scraper	Rel. Abundance of Scrapers	Decrease
Group	% Shredder	Rel. Abundance of Shredders	Decrease
Composition	% Filterer	Rel. Abundance of Filter Feeders	Increase
	% Predator	Rel. Abundance of Predators	Decrease
	% Collector	Rel. Abundance of Collector-Gatherers	Increase
Habit Richness	No. Clinger Genera	No. of Clinger Genera	Decrease
Habit	% Sprawlers	Rel. Abundance of Sprawlers	Increase
Composition	% Swimmers	Rel. Abundance of Swimmers	Increase
	% Climbers	Rel. Abundance of Climbers	Increase
	% Burrowers	Rel. Abundance of Burrowers	Increase
	% Clinger	Rel. Abundance of Clingers	Decrease

Note: Higher classification (e.g., class, order, family) was used for taxa where genus-level could not be identified; care was taken to not "double count" such that a higher classification was included if it reperesented a distinct taxon within the sample.

Appendix B. Metric Selection: Rationale for Selection of GLIMPSS Metrics

Stratum-specific metric selection involved evaluating metrics by running them through a series of performance tests. The sequence of testing was identical in each stratum: Discrimination efficiency>PCA correlation>Redundancy>Scope of Impairment>Range. Throughout each iteration, effort was made to retain metrics that fell into each of the 6 ecological categories (richness, composition, tolerance, habit and trophic guilds, and dominance). The following tables (B1-B4) list all metrics (sorted by %DE), which ones were selected or rejected, and why.

Discrimination Efficiency (%DE)

Calculated DEs for all metrics within each stratum were used as a means to quantify sensitivity. Metrics with DE less than 65% in a particular stratum were automatically dropped from further analysis (except in a special circumstance in PL Sp and PL Su where fewer metrics passed other tests and the Habit or Feeding Group ecological category was forced). While there was consistent performance observed for many metrics, some metrics showed wide ranges of sensitivity across strata, indicating differences in biological potential from seasonal and geographic factors or sensitivity to region-specific stressors. For example, the % Dominant 5 Taxa metric was very sensitive in the PL Su (DE=74.8), but was rejected from further consideration in PL Sp (DE=47.1). By comparison, % Orthocladiinae was very sensitive in the MT Sp (DE=87.3) but had much less discrimination ability in the PL Su (DE=60.2). Although No. of EPT Genera had high %DE in all strata, we chose to use independent measures of No. Ephemeroptera and No. Plecoptera Genera, where possible, in order to benefit from the known diagnostic capability of these two individual metrics (Karr and Chu 1999). Moreover, the No. Trichoptera Genera metric was not sensitive in many strata (except MT Sp). These results confirm the need to explore metrics in the context of a classification scheme such as bioregion and season.

Correlation of Metrics Values to a Human Disturbance Gradient (PCA)

While it is important that metrics can discriminate between known undisturbed and degraded sites, they should also respond predictably to measurements of increasing water quality or habitat stress. Some investigators (Klemm et al. 2003) have relied on testing metrics in relation to synthetic gradients built from the linear combinations of multiple stressors using principal components analysis (PCA). In our PCA analysis, abiotic variables included pH, and log transformed specific conductance, temperature, D.O., total suspended solids, sulfate, chloride, manganese, aluminum, iron, total phosphorus, nitrate-nitrite, fecal coliform concentration, and seven individual RBP habitat metrics. Here, PCA axis 1 represented the strongest significant gradient and individual site coordinates along the axis were correlated (Pearson coefficients) to all metrics within each stratum. Overall, stronger relationships were seen with metrics in the MT bioregion compared to the PL (Tables B1-B4).

Like in the DE analysis, there were observable differences between metrics across strata. The most notable of these was No. Scraper Genera which was significantly related to increasing disturbance in the MT Sp (r= -0.45) but was much weaker in the MT Su (r= -0.24). This could be due to the fact that

Scraper richness naturally declines in summer when streams are more fully canopied, rather than depicting a loss of sensitivity to stress. Several metrics (*e.g.*, HBI, No. Intolerant Genera, No. Clinger Genera) consistently showed high correlation to PCA axis 1 in all strata. Although No. EPT Genera was strongly correlated to the disturbance axis in all strata, we chose to use independent measures of No. Ephemeroptera and No. Plecoptera Genera, where possible.

Paired Reference Metric Correlations for Redundancy Analysis

Pearson correlation coefficients were calculated for metric pairs from the CAL REF site dataset within each stratum. The rationale for examining redundancy is so that the multi-metric index does not contain 2 or more similar metrics that essentially provide the same information. While >0.75 was chosen as a cutoff value to screen for redundancy, metric pairs that approached or slightly exceeded this value were examined with scatter biplots (not shown, but refer to Figure 9 in Section 8.3 for an example). If there was considerable scatter of data points, or if nonlinear patterns were revealed, then both metrics were investigated further for possible inclusion in the index.

After DE and stressor relationships were tested, redundant metrics were removed from the candidate list. Overall, there were few highly correlated pairs and most metrics appeared to offer somewhat different information as denoted by having correlations well below 0.75. In very few instances did metrics that had high DE and stressor responsiveness show redundancy. For example, in the PL Su, No. Plecoptera Genera was highly redundant with No. Intolerant Genera <3 (0.93); however, since No. Intolerant <3 had a much higher range and %DE, it was selected as the preferred metric. For the chosen metrics, maximum redundancy magnitudes (within each stratum) are reported in Tables B1-B4.

Scope of Impairment (SOI)

SOI (modified from Blocksom and Johnson 2009) represents the variability of REF metric values in relation to the metric's range of detectable impairment. This was calculated as the interquartile range (25th percentile to 75th percentile) divided by the range of zero (or 100 depending on metric direction) to the nearest quartile. A value >1 indicated too much variability and a reduced ability to detect deviance of degraded sites from the reference condition. Tables B1 to B4 indicate those metrics that passed %DE and PCA tests but failed the SOI test. For example % Scrapers had good DE (79%) and PCA correlation in the MT Sp, but had a SOI of 1.42. In the MT Su stratum, % Ephemeroptera (minus *Baetis*) had a similarly good DE (77%) and PCA correlation, but a SOI of 1.37. In these cases, comparably sensitive metrics with acceptable SOIs were chosen.

Table B - 1. Mountain Spring metric selection.

Metric	Selected?	DE	MT Sp PCA	Max Redun	Reason for Selection or Rejection
No. Intolerant Genera <4	**X**	**92**	**-0.61**	**0.75**	**Good %DE and PCA correlation, low redundancy; captures richness of all sensitive taxa combined**
No. Intolerant Genera <3		89	-0.58		Less range than the comparable #Intol <4 metric; highly redundant with #Intol <4 metric (0.96)
No. EPT Genera		89	-0.46		Preference for individual diagnostic measure of E, P, and T
% Orthocladiinae	**X**	**89**	**0.49**	**0.45**	**Good %DE and PCA correlation, low redundancy; represents composition of a relatively tolerant, short-lived group of Chironomidae**
HBI	**X**	**84**	**0.62**	**0.45**	**Good %DE and PCA correlation, low redundancy; weights taxon pollution tolerance and abundance**
No.Trichoptera Genera	**X**	**81**	**-0.39**	**0.59**	**Good %DE and PCA correlation, low redundancy; represents moderately sensitive group of insects**
% Scrapers		79	-0.34		Interquartile range : zero to quartile ratio >1
No. Ephemeroptera Genera	**X**	**79**	**-0.36**	**0.45**	**Good %DE and PCA correlation, low redundancy; group has diagnostic sensitivity to known toxins**
No. Plecoptera Genera	**X**	**78**	**-0.53**	**0.75**	**Good %DE and PCA correlation, low redundancy; represents cool/cold water obligates**
% EPT minus Cheumatopsyche		76	-0.49		%P and T had low individual discrimination (e.g., poor performance of %Hydropsychidae and regionally facultative stonefly *Amphinemura* drives P; opted for independent measure of E.
No. Clinger Genera	**X**	**75**	**-0.47**	**0.66**	**Good %DE and PCA correlation, low redundancy; represents organisms adapted to living on stable substrates**
% Ephemeroptera	**X**	**75**	**-0.31**	**0.26**	**Good %DE and PCA correlation, low redundancy; group has diagnostic sensitivity to known toxins**
% EPT minus Cheumatopsyche+Baetis		73	-0.55		%P and T had low individual discrimination (e.g., poor performance of %Hydropsychidae and regionally facultative stonefly *Amphinemura* drives P; opted for independent measure of E.
% EPT		73	-0.46		%P and T had low individual discrimination (e.g., poor performance of %Hydropsychidae and regionally facultative stonefly *Amphinemura* drives P; opted for independent measure of E.
% Chironomidae+Annelida		73	0.49		Redundant with %Orthocladiinae which has better %DE; Annelida not well-represented in MT Sp dataset
% 5 Dominant Genera	**X**	**73**	**0.35**	**0.64**	**Good %DE , low redundancy; some measure of dominance sought as an important category for assessment**
No. Total Genera		71	-0.32	0.79	Redundant with No. Clinger Genera whch had higer %DE and PCA correlation
% Ephemeroptera Minus Baetis		71	-0.18		%Ephemeroptera with higher %DE and PCA correlation
% Tolerant (>6)		70	0.34		Redundant with %Orthocladiinae
% Chironomidae		68	0.39		Redundant with %Orthocladiinae; %Orthocladiinae has better %DE and correlation PCA
No. Scraper Genera	**X**	**68**	**-0.42**	**0.62**	**Good %DE and PCA correlation, low redundancy; represents richness of taxa requiring high quality algal food; FFG metric sought for assessment**
No. Shredder Genera		66	-0.36		Lower %DE and PCA correlation than other FFG metric (No. Scraper Genera)
% Clingers		64	-0.28		DE <65%
% Trichoptera		63	-0.10		DE <65%
% Plecoptera		62	-0.30		DE <65%
No. Predator Genera		61	-0.24		DE <65%
% Collectors		56	0.31		DE <65%
% Shredders		51	-0.17		DE <65%
% Predators		48	-0.10		DE <65%
% Annelida		44	0.22		DE <65%
% Hydropsychidae		44	0.11		DE <65%
% Tanytarsini		41	-0.33		DE <65%
No. Collector Genera		41	0.00		DE <65%
% Sprawlers		40	-0.07		DE <65%
No. Filterer Genera		38	-0.20		DE <65%
% Non-Insect		38	0.07		DE <65%
% Burrowers		29	-0.01		DE <65%
% Chironomini		28	0.23		DE <65%
% Swimmers		22	0.07		DE <65%
% Filterers		22	-0.05		DE <65%
% Climbers		22	0.06		DE <65%

Table B - 2. Mountain Summer metric selection.

Metric	Selected?	DE	MT Su PCA	Max Redun	Reason for Selection or Rejection
HBI	**X**	**96**	**0.66**	**0.74**	**Good %DE and PCA correlation, low redundancy; weights taxon pollution tolerance and abundance**
No. Intolerant Genera <3		95	-0.68		Less range than the comparable #Intol <4 metric
No. Intolerant Genera <4	**X**	**95**	**-0.68**	**0.74**	**Good %DE and PCA correlation, low redundancy; captures richness of all sensitive taxa combined**
No. EPT Genera		95	-0.55		Preference for individual diagnostic measure of E and P; T has low %DE
No. Plecoptera Genera	**X**	**91**	**-0.62**	**0.74**	**Good %DE and PCA correlation, low redundancy; represents richness of cool/cold water obligates**
% EPT minus Cheumatopsyche	**X**	**90**	**-0.59**	**0.74**	**Good %DE and PCA correlation, low redundancy; represents abundance of relatively sensitive orders of insects but excludes 1 frequently hyperdominant taxa**
% Plecoptera		90	-0.41		Interquartile range : zero to quartile ratio >1
% EPT minus Cheumatopsyche+Baetis		86	-0.48		Redundant with %EPT minus Cheumatopsyche which had higher %DE
No. Clinger Genera	**X**	**79**	**-0.39**	**0.69**	**Good %DE and PCA correlation, low redundancy; represents organisms adapted to living on stable substrates**
No. Ephemeroptera Genera	**X**	**78**	**-0.43**	**0.39**	**Good %DE and PCA correlation, low redundancy; group has diagnostic sensitivity to known toxins**
% Ephemeroptera Minus Baetis		77	-0.27		Interquartile range : zero to quartile ratio >1; low PCA corr.
% Ephemeroptera		75	-0.30		Interquartile range : zero to quartile ratio >1
% EPT		75	-0.52		Redundant with and lower DE and PCA correlation than % EPT minus Cheumatopsyche
% 5 Dominant Genera	**X**	**72**	**0.26**	**-0.67**	**Good %DE , low redundancy; Some measure of dominance sought as an important category for assessment**
No. Shredder Genera	**X**	**71**	**-0.45**	**0.41**	**Good %DE and PCA correlation, low redundancy; Represents taxa requiring sufficient CPOM resources**
No. Total Genera	**X**	**68**	**-0.31**	**0.69**	**Good %DE and PCA correlation, low redundancy; Represents total diversity; easiest to communicate**
% Orthocladiinae	**X**	**67**	**0.34**	**0.41**	**Good %DE and PCA correlation, low redundancy; represents composition of a relatively tolerant, short-lived group of Chironomidae**
% Shredders		65	-0.30		DE <65%
% Tolerant (>6)		64	0.29		DE <65%
% Chironomidae+Annelida		62	0.39		DE <65%
No. Scraper Genera		61	-0.24		DE <65%
% Chironomidae		61	0.38		DE <65%
No.Trichoptera Genera		51	0.08		DE <65%
% Predators		50	-0.11		DE <65%
No. Predator Genera		49	-0.26		DE <65%
No. Collector Genera		49	-0.15		DE <65%
% Scrapers		47	-0.03		DE <65%
% Filterers		45	0.18		DE <65%
% Annelida		44	0.11		DE <65%
% Hydropsychidae		41	0.23		DE <65%
% Tanytarsini		39	0.02		DE <65%
% Trichoptera		38	0.08		DE <65%
% Clingers		36	-0.03		DE <65%
% Non-Insect		34	0.13		DE <65%
% Burrowers		33	0.15		DE <65%
% Collectors		31	0.04		DE <65%
% Chironomini		28	0.26		DE <65%
% Sprawlers		23	-0.12		DE <65%
% Climbers		22	0.13		DE <65%
No. Filterer Genera		21	0.09		DE <65%
% Swimmers		15	-0.14		DE <65%

Table B - 3. Plateau Spring metric selection.

Metric	Selected?	DE	PL Sp PCA	Max Redun	Reason for Selection or Rejection
% EPT minus Cheumatopsyche	**X**	**83**	**-0.36**	**0.66**	**Good %DE and PCA correlation, low redundancy; represents abundance of relatively sensitive orders of insects but excludes 1 frequently hyperdominant and facultative taxon**
No. Intolerant Genera <4	**X**	**82**	**-0.49**	**0.67**	**Good %DE and PCA correlation, low redundancy; captures richness of all sensitive taxa combined**
% EPT		82	-0.34		Redundant with and lower PCA correlation than % EPT minus Cheumatopsyche
No. Intolerant Genera <3		82	-0.45		Less range than the comparable #Intol <4 metric
HBI	**X**	**80**	**0.41**	**0.75**	**Good %DE and PCA correlation, low redundancy; weights taxon pollution tolerance and abundance**
% Plecoptera		80	-0.28		Interquartile range : zero to quartile ratio >1
No. Plecoptera Genera	**X**	**79**	**-0.38**	**0.65**	**Good %DE and PCA correlation, low redundancy; represents diversiy of cool/cold water obligates**
% Tolerant (>6)		**77**	**0.24**	**0.75**	**Good %DE and low redundancy; captures abundance of very tolerant taxa**
% Orthocladiinae		76	0.19		Low PCA correlation
% EPT minus Cheumatopsyche+Baetis		75	-0.36		Redundant with and lower DE than % EPT minus Cheumatopsyche
% Chironomidae+Annelida	**X**	**75**	**0.31**	**0.74**	**Good %DE and PCA correlation, low redundancy; representes composition of relatively tolerant family of Diptera+Annelid worms and leaches**
No. EPT Genera		74	-0.46		Preference for individual measure of E and P; T has low %DE (30%)
% Chironomidae		74	0.30		Redundant with %Chironomidae+Annelida
% Ephemeroptera		68	-0.23		Low %DE and PCA correlation compared to other more favorable metrics
% Shredders		67	-0.18		Low PCA correlation
No. Ephemeroptera Genera	**X**	**66**	**-0.35**	**0.59**	**Good %DE and PCA correlation, low redundancy; group has diagnostic sensitivity to known toxins**
% Ephemeroptera Minus Baetis		64	-0.23		DE <65%
No. Clinger Genera	**X**	**61**	**-0.38**	**0.67**	**Although lower %DE, good PCA correlation, low redundancy; some measure of habit sought for inclusion**
% Climbers		52	0.22		DE <65%
% Predators		52	0.02		DE <65%
% Chironomini		50	0.14		DE <65%
No. Shredder Genera		49	-0.20		DE <65%
No. Predator Genera		47	-0.24		DE <65%
% 5 Dominant Genera		47	0.25		DE <65%
% Trichoptera		47	0.03		DE <65%
No. Scraper Genera		46	-0.35		DE <65%
% Scrapers		44	-0.06		DE <65%
% Clingers		44	-0.10		DE <65%
% Filterers		42	0.13		DE <65%
% Tanytarsini		41	0.17		DE <65%
No. Total Genera		40	-0.27		DE <65%
% Annelida		37	0.15		DE <65%
% Hydropsychidae		33	0.10		DE <65%
No.Trichoptera Genera		30	-0.28		DE <65%
% Non-Insect		28	0.02		DE <65%
% Burrowers		23	0.13		DE <65%
No. Collector Genera		21	-0.11		DE <65%
% Collectors		14	-0.04		DE <65%
% Sprawlers		12	-0.14		DE <65%
No. Filterer Genera		7	-0.03		DE <65%
% Swimmers		4	-0.12		DE <65%

Table B - 4. Plateau Summer metric selection.

Metric	Selected?	DE	PL Su PCA	Max Redun	Reason for Selection or Rejection
No. Intolerant Genera <3	**X**	**98**	**-0.53**	**0.70**	**Good %DE and PCA correlation, low redundancy; captures richness of all sensitive taxa combined**
No. Intolerant Genera <4		96	-0.54		Lower %DE than No. Intolerant <4; redundant with HBI (0.76)
HBI	**X**	**95**	**0.55**	**-0.72**	**Good %DE and PCA correlation, low redundancy; weights taxon pollution tolerance and abundance**
% Plecoptera		94	-0.51		Redundant with preferred HBI (r=0.86); Interquartile range : zero to quartile ratio >1; some REF sites with no P
No. EPT Genera		90	0.53		Redundant with No. Intolerant <3 (0.79)
No. Plecoptera Genera		89	-0.49		Redundant with No. Intolerant <3 (0.93); No. Intolerant <3 had higher range
% EPT minus Cheumatopsyche+Baetis		86	-0.53		Redundant with % EPT minus Cheumatopsyche (r=0.93) which had similar %DE, but higher correlation to PCA
% EPT minus Cheumatopsyche	**X**	**86**	**-0.54**	**-0.75**	**Good %DE and PCA correlation, low redundancy; represents abundance of relatively sensitive orders of insects but excludes 1 frequently hyperdominant taxa**
No. Clinger Genera	**X**	**83**	**-0.54**	**0.64**	**Good %DE and PCA correlation, low redundancy; represents organisms adapted to living on stable substrates**
% 5 Dominant Genera	**X**	**75**	**0.32**	**-0.62**	**Good %DE and PCA correlation, low redundancy; Some measure of dominance sought as an important category for assessment**
% Ephemeroptera Minus Baetis		75	-0.47		Interquartile range : zero to quartile ratio >1
No. Total Genera	**X**	**74**	**-0.37**	**-0.64**	**Good %DE and PCA correlation, low redundancy; Represents total diversity; easiest to communicate**
% Ephemeroptera		73	-0.47		Interquartile range : zero to quartile ratio >1
No. Ephemeroptera Genera	**X**	**69**	**-0.52**	**0.46**	**Good %DE and PCA correlation, low redundancy; group has diagnostic sensitivity to known toxins**
% EPT		68	-0.49		Redundant with % EPT minus Cheumatopsyche which had much higher %DE
% Chironomidae	**X**	**65**	**0.52**	**0.53**	**Good %DE and PCA correlation, low redundancy; represents composition of relatively tolerant family of Diptera**
No. Scraper Genera	**X**	**63**	**-0.43**	**0.44**	**Although lower %DE, good PCA correlation, low redundancy; represents richness of taxa requiring high quality algal food; FFG metric sought for assessment**
% Chironomidae+Annelida		61	0.52		DE <65%
% Filterers		60	0.08		DE <65%
% Orthocladiinae		60	0.30		DE <65%
% Tolerant >6		58	0.44		DE <65%
No. Predator Genera		56	-0.17		DE <65%
% Shredders		54	-0.14		DE <65%
% Chironomini		54	0.37		DE <65%
% Predators		52	0.09		DE <65%
% Hydropsychidae		51	0.10		DE <65%
No.Trichoptera Genera		44	-0.17		DE <65%
No. Collector Genera		43	-0.21		DE <65%
% Scrapers		43	-0.33		DE <65%
% Non-Insect		42	0.11		DE <65%
% Clingers		38	-0.27		DE <65%
No. Shredder Genera		37	-0.31		DE <65%
% Tanytarsini		37	0.08		DE <65%
% Climbers		37	0.11		DE <65%
% Collectors		30	0.13		DE <65%
% Burrowers		29	0.43		DE <65%
% Annelida		24	0.06		DE <65%
No. Filterer Genera		21	-0.10		DE <65%
% Trichoptera		18	0.06		DE <65%
% Swimmers		16	-0.33		DE <65%
% Sprawlers		13	-0.02		DE <65%

Appendix C. Reference Site Metric Summary Statistics for Mountain and Plateau Bioregions

Calculation of summary statistics was done to document the distribution of metric values across strata, but also allowed us to screen for metric ranges and minimum values at reference (REF) sites. The criterion for metric range was set at 5 taxa for richness metrics, and 10% for abundance metrics. No richness metrics were rejected with this criterion; however, a few of the habit metrics (*e.g.*, % Burrowers and % Climbers) failed this relative abundance test.

Table C - 1. Summary Statistics for REF sites in Mountain Spring.

MT Spring

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	128	17	29	32	35	44	27
No. Intolerant Genera (<3)	128	6	10	12	13	17	11
No. Intolerant Genera (<4)	128	9	13	15	17	23	14
No. EPT Genera	128	12	16	18	20	26	14
No. Ephemeroptera Genera	128	3	6	7	9	12	9
No. Plecoptera Genera	128	2	5	6	7	10	8
No. Trichoptera Genera	128	0	4	5	6	11	11
No. Clinger Genera	128	6	14	16	18	25	19
% Ephemeroptera	128	7.1	26.8	35.9	44.7	67.9	60.8
% Ephemeroptera (minus *Baetis*)	128	5.9	20.3	30.2	38.8	65.5	59.7
% Plecoptera	128	2.8	12.4	20.6	30.8	70.8	67.9
% Trichoptera	128	0.0	6.0	9.1	15.2	53.2	53.2
% Hydropsychidae	128	0.0	1.9	4.4	8.0	47.7	47.7
% EPT	128	41.3	58.7	72.0	81.1	95.2	54.0
% mEPT (minus *Cheumatopsyche*)	128	36.3	58.7	71.3	79.8	95.2	59.0
% mEPT (minus *Cheumatopsyche+Baetis*)	128	30.3	53.7	65.6	75.7	91.5	61.2
HBI (Genus)	128	1.48	2.65	3.10	3.57	4.85	3.37
% Tolerant (>6)	128	0.0	0.5	1.4	4.3	16.3	16.3
% 5 Dominant Genera	128	41.9	53.3	60.1	65.6	86.0	44.1
% Chironomidae	128	1.3	6.7	12.6	21.8	47.9	46.6
% Orthocladiinae	128	0.0	1.5	2.8	5.2	30.7	30.7
% Chironomiini	128	0.0	0.5	1.8	5.2	18.8	18.8
% Tanytarsini	128	0.0	0.9	4.0	7.0	42.3	42.3
% Chironomidae+Annelida	128	1.3	7.2	14.6	22.7	47.9	46.6
% Annelida	128	0.0	0.0	0.0	0.5	9.4	9.4
% Non-Insect	128	0.0	0.4	0.9	2.4	43.5	43.5
No. Scraper Genera	128	2	4	5	6	10	8
No. Shredder Genera	128	1	4	5	6	8	7
No. Collector Genera	128	3	8	9	10	18	15
No. Filterer Genera	128	1	4	5	6	12	11
No. Predator Genera	128	1	6	7	9	16	15
% Scrapers	128	2.0	12.4	21.5	30.8	53.0	51.0
% Shredders	128	1.5	11.2	19.7	29.9	71.2	69.8
% Collectors	128	4.4	15.2	23.1	31.8	60.9	56.5
% Filterers	128	1.0	8.5	14.6	22.2	50.9	50.0
% Predators	128	0.9	7.3	9.8	14.0	45.4	44.5
% Clingers	128	9.9	34.1	47.6	59.7	83.9	74.0
% Sprawlers	128	0.9	10.6	18.9	27.7	75.8	74.9
% Swimmers	128	0.0	7.1	12.3	20.5	56.5	56.5
% Climber	128	0.0	2.3	6.1	11.6	46.5	46.5
% Burrower	128	0.0	0.9	1.5	2.9	31.3	31.3

Table C - 2. Summary Statistics for REF sites in Mountain Summer.

MT Summer

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	181	20	27	31	35	43	23
No. Intolerant Genera (<3)	181	4	8	10	12	16	12
No. Intolerant Genera (<4)	181	7	11	13	15	21	14
No. EPT Genera	181	9	15	17	19	25	14
No. Ephemeroptera Genera	181	2	5	7	8	13	11
No. Plecoptera Genera	181	1	4	5	6	9	8
No. Trichoptera Genera	181	2	4	5	6	11	9
No. Clinger Genera	181	7	14	16	18	23	16
% Ephemeroptera	181	4.4	16.1	22.9	33.5	66.7	62.3
% Ephemeroptera (minus *Baetis*)	181	1.3	10.1	15.6	25.3	59.6	58.3
% Plecoptera	181	0.5	12.8	22.7	34.2	67.1	66.6
% Trichoptera	181	0.9	11.1	19.7	29.1	67.0	66.1
% Hydropsychidae	181	0.0	3.6	9.8	19.8	64.2	64.2
% EPT	181	27.5	61.4	74.6	82.1	93.6	66.1
% mEPT (minus *Cheumatopsyche*)	181	22.2	58.3	70.6	80.5	93.6	71.4
% mEPT (minus *Cheumatopsyche+Baetis*)	181	19.8	51.1	64.6	74.5	92.3	72.5
HBI (Genus)	181	1.44	3.12	3.53	4.03	5.24	3.80
% Tolerant (>6)	181	0.0	0.5	1.1	3.6	31.6	31.6
% 5 Dominant Genera	181	39.0	54.1	61.3	68.9	86.6	47.6
% Chironomidae	181	0.4	6.4	11.4	18.5	47.2	46.7
% Orthocladiinae	181	0.0	1.3	3.2	5.2	30.7	30.7
% Chironomiini	181	0.0	0.4	1.4	4.3	19.1	19.1
% Tanytarsini	181	0.0	0.7	1.9	4.7	31.0	31.0
% Chironomidae+Annelida	181	0.4	6.6	11.8	19.1	47.6	47.2
% Annelida	181	0.0	0.0	0.0	0.5	18.1	18.1
% Non-Insect	181	0.0	0.4	0.9	2.5	58.0	58.0
No. Scraper Genera	181	1	4	5	6	10	9
No. Shredder Genera	181	1	3	4	5	9	8
No. Collector Genera	181	3	7	8	10	15	12
No. Filterer Genera	181	1	4	5	6	11	10
No. Predator Genera	181	2	6	7	9	13	11
% Scrapers	181	1.5	7.6	12.8	20.9	49.2	47.7
% Shredders	181	0.5	9.5	16.9	28.3	70.7	70.3
% Collectors	181	2.8	12.5	19.4	29.9	62.1	59.3
% Filterers	181	1.4	11.6	21.2	32.7	67.5	66.0
% Predators	181	1.9	7.3	11.3	17.0	40.3	38.4
% Clingers	181	5.9	36.6	52.4	64.9	90.6	84.7
% Sprawlers	181	1.4	10.4	18.0	28.9	71.2	69.8
% Swimmers	181	0.0	6.8	11.3	18.2	52.5	52.5
% Climber	181	0.0	1.7	4.1	8.4	31.9	31.9
% Burrower	181	0.0	0.6	1.4	2.8	20.3	20.3

Table C - 3. Summary Statistics for REF sites in Plateau Spring.

PL Spring

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	44	16	23	28	33	41	25
No. Intolerant Genera (<3)	44	3	7	8	11	16	13
No. Intolerant Genera (<4)	44	4	10	12	15	19	15
No. EPT Genera	44	8	13	15	18	24	16
No. Ephemeroptera Genera	44	3	6	7	8	13	10
No. Plecoptera Genera	44	2	5	6	7	8	6
No. Trichoptera Genera	44	0	1	3	4	7	7
No. Clinger Genera	44	4	11	12	16	21	17
% Ephemeroptera	44	2.4	21.5	39.9	53.8	84.9	82.6
% Ephemeroptera (minus *Baetis*)	44	2.4	19.4	30.1	38.7	69.3	67.0
% Plecoptera	44	5.2	20.6	25.5	44.8	60.8	55.6
% Trichoptera	44	0.0	1.0	3.3	6.9	27.3	27.3
% Hydropsychidae	44	0.0	0.0	0.5	2.4	7.3	7.3
% EPT	44	20.8	66.0	80.0	86.7	94.6	73.8
% mEPT (minus *Cheumatopsyche*)	44	18.9	65.8	79.4	86.3	94.3	75.5
% mEPT (minus *Cheumatopsyche+Baetis*)	44	18.4	51.4	67.0	81.2	94.2	75.8
HBI (Genus)	44	1.96	2.75	3.59	3.92	4.97	3.01
% Tolerant (>6)	44	0.0	0.9	2.6	4.4	45.5	45.5
% 5 Dominant Genera	44	44.1	60.0	69.6	78.9	89.3	45.2
% Chironomidae	44	0.9	3.2	7.9	19.4	52.7	51.8
% Orthocladiinae	44	0.0	0.9	2.4	8.2	23.9	23.9
% Chironomiini	44	0.0	0.0	0.8	2.3	29.6	29.6
% Tanytarsini	44	0.0	0.0	0.5	1.2	7.6	7.6
% Chironomidae+Annelida	44	0.9	4.1	8.4	19.4	52.7	51.8
% Annelida	44	0.0	0.0	0.4	0.9	4.3	4.3
% Non-Insect	44	0.0	0.5	0.9	3.4	45.0	45.0
No. Scraper Genera	44	1	4	5	7	11	10
No. Shredder Genera	44	1	3	4	4	8	7
No. Collector Genera	44	3	6	8	10	16	13
No. Filterer Genera	44	0	1	2	4	8	8
No. Predator Genera	44	2	5	7	9	12	10
% Scrapers	44	3.2	12.5	22.5	28.3	54.2	51.0
% Shredders	44	2.2	12.8	20.0	32.8	50.0	47.8
% Collectors	44	4.9	16.6	28.0	35.9	70.8	65.8
% Filterers	44	0.0	1.1	3.6	7.4	67.0	67.0
% Predators	44	2.4	7.7	13.4	21.3	57.4	55.0
% Clingers	44	19.0	34.3	44.3	58.4	84.6	65.6
% Sprawlers	44	1.8	14.4	21.5	34.5	46.8	45.0
% Swimmers	44	0.8	5.6	13.5	31.8	57.1	56.3
% Climber	44	0.0	0.5	1.2	3.7	29.0	29.0
% Burrower	44	0.0	0.5	1.4	3.3	16.8	16.8

Table C - 4. Summary Statistics for REF sites in Plateau Summer.

PL Summer

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	38	21	27	29	33	44	23
No. Intolerant Genera (<3)	38	1	5	7	9	14	13
No. Intolerant Genera (<4)	38	2	6	9	11	15	13
No. EPT	38	8	11	13	16	23	15
No. Ephemeroptera Genera	38	1	4	6	8	11	10
No. Plecoptera Genera	38	0	3	4	6	8	8
No. Trichoptera Genera	38	1	3	4	5	7	6
No. Clinger Genera	38	8	12	13	15	21	13
% Ephemeroptera	38	0.5	12.4	18.1	33.7	72.4	71.9
% Ephemeroptera (minus *Baetis*)	38	0.0	7.5	13.4	31.5	63.4	63.4
% Plecoptera	38	0.0	11.5	27.0	37.8	68.1	68.1
% Trichoptera	38	0.5	5.1	10.2	18.8	48.7	48.2
% Hydropsychidae	38	0.0	3.4	6.8	17.5	40.5	40.5
% EPT	38	13.9	51.9	66.4	79.5	89.0	75.1
% mEPT (minus *Cheumatopsyche*)	38	11.3	44.3	55.4	72.8	89.0	77.7
% mEPT (minus *Cheumatopsyche+Baetis*)	38	10.8	40.1	50.6	69.1	83.8	73.0
HBI (Genus)	38	2.44	3.24	3.91	4.49	5.25	2.81
% Tolerant (>6)	38	0.0	0.9	1.9	3.2	11.2	11.2
% 5 Dominant Genera	38	44.6	59.5	63.9	67.6	86.5	41.9
% Chironomidae	38	1.8	8.0	12.3	22.8	39.3	37.5
% Orthocladiinae	38	0.0	0.9	2.4	5.0	19.3	19.3
% Chironomiini	38	0.0	1.6	2.6	4.4	17.1	17.1
% Tanytarsini	38	0.0	1.3	2.6	5.1	29.1	29.1
% Chironomidae+Annelida	38	3.2	9.0	13.2	24.0	40.2	37.0
% Annelida	38	0.0	0.0	0.5	1.0	7.5	7.5
% Non-Insect	38	0.0	0.5	1.0	2.3	25.3	25.3
No. Scraper Genera	38	3	4	5	6	9	6
No. Shredder Genera	38	1	2	3	4	9	8
No. Collector Genera	38	4	6	7	10	17	13
No. Filterer Genera	38	1	4	5	7	11	10
No. Predator Genera	38	3	7	8	9	12	9
% Scrapers	38	3.4	11.2	22.0	30.7	48.2	44.8
% Shredders	38	2.5	8.0	20.5	29.8	65.7	63.2
% Collectors	38	3.4	9.4	17.0	23.5	40.6	37.2
% Filterers	38	0.5	7.8	15.2	25.1	55.3	54.8
% Predators	38	2.1	10.6	16.0	19.7	35.7	33.6
% Clingers	38	21.9	37.9	44.5	63.4	77.6	55.7
% Sprawlers	38	3.0	12.9	20.0	35.6	70.6	67.6
% Swimmers	38	0.5	4.1	10.0	17.1	47.8	47.4
% Climber	38	0.0	2.2	4.5	11.9	26.7	26.7
% Burrower	38	0.4	1.3	2.6	4.0	13.2	12.8

Table C - 5. Summary Statistics for REF sites in Mountain Winter.

MT Winter

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	29	24	32	33	38	47	23
No. Intolerant Genera (<3)	29	8	10	12	14	18	10
No. Intolerant Genera (<4)	29	12	15	17	19	25	13
No. EPT	29	14	18	20	23	26	12
No. Ephemeroptera Genera	29	3	6	7	8	12	9
No. Plecoptera Genera	29	4	6	8	9	11	7
No. Trichoptera Genera	29	4	5	6	7	9	5
No. Clinger Genera	29	9	15	17	19	23	14
% Ephemeroptera	29	3.0	22.8	29.5	40.8	58.3	55.3
% Ephemeroptera (minus *Baetis*)	29	3.0	22.5	28.6	40.3	56.0	53.0
% Plecoptera	29	10.4	15.2	19.7	28.6	41.9	31.5
% Trichoptera	29	4.7	11.4	14.6	25.1	45.9	41.2
% Hydropsychidae	29	0.0	3.9	7.2	8.7	25.7	25.7
% EPT	29	42.1	65.8	78.6	82.6	91.8	49.6
% mEPT (minus *Cheumatopsyche*)	29	39.1	64.5	74.4	81.1	90.9	51.8
% mEPT (minus *Cheumatopsyche+Baetis*)	29	39.1	62.7	71.6	79.2	90.9	51.8
HBI (Genus)	29	1.98	2.53	3.07	3.60	4.05	2.07
% Tolerant (>6)	29	0.0	1.4	2.4	5.4	9.9	9.9
% 5 Dominant Genera	29	42.0	50.0	55.2	62.4	75.3	33.3
% Chironomidae	29	1.7	7.6	8.7	12.0	41.5	39.8
% Orthocladiinae	29	0.0	3.2	4.7	5.9	17.0	17.0
% Chironomiini	29	0.0	0.0	1.0	3.1	13.7	13.7
% Tanytarsini	29	0.0	1.4	2.0	4.2	13.5	13.5
% Chironomidae+Annelida	29	1.7	7.9	9.4	13.3	41.5	39.8
% Annelida	29	0.0	0.0	0.0	0.5	4.5	4.5
% Non-Insect	29	0.0	0.0	0.5	3.3	42.7	42.7
No. Scraper Genera	29	3	4	5	7	8	5
No. Shredder Genera	29	3	5	6	7	10	7
No. Collector Genera	29	3	7	9	11	14	11
No. Filterer Genera	29	3	4	5	6	8	5
No. Predator Genera	29	4	6	8	10	13	9
% Scrapers	29	6.1	16.9	21.8	34.7	42.0	35.9
% Shredders	29	4.6	13.2	18.0	21.9	27.2	22.6
% Collectors	29	1.3	17.9	23.4	29.3	55.2	53.9
% Filterers	29	6.1	10.0	15.5	19.5	49.2	43.1
% Predators	29	4.5	7.2	11.7	16.0	20.5	16.0
% Clingers	29	28.9	46.7	54.8	61.0	76.7	47.7
% Sprawlers	29	6.6	12.8	16.7	22.7	31.2	24.6
% Swimmers	29	2.1	9.1	11.8	17.1	42.1	40.0
% Climber	29	0.4	1.9	3.3	7.5	18.5	18.0
% Burrower	29	0.0	1.4	2.4	3.7	10.4	10.4

Table C - 6. Summary Statistics for REF sites in Plateau Winter.

PL Winter

Metric	N	Min	25th %ile	Median	75th %ile	Max	Range
No. Total Genera	18	16	23	27	31	42	26
No. Intolerant Genera (<3)	18	3	8	9	10	13	10
No. Intolerant Genera (<4)	18	9	11	14	14	17	8
No. EPT	18	11	14	16	19	22	11
No. Ephemeroptera Genera	18	3	5	6	7	8	5
No. Plecoptera Genera	18	4	5	6	8	9	5
No. Trichoptera Genera	18	0	3	4	5	10	10
No. Clinger Genera	18	8	11	14	17	22	14
% Ephemeroptera	18	2.9	14.2	33.2	44.1	74.4	71.5
% Ephemeroptera (minus *Baetis*)	18	2.9	14.2	33.2	43.7	74.4	71.5
% Plecoptera	18	15.3	25.1	31.1	45.9	79.3	64.0
% Trichoptera	18	0.0	5.1	8.2	19.3	37.5	37.5
% Hydropsychidae	18	0.0	1.4	3.2	8.3	15.4	15.4
% EPT	18	57.3	76.9	86.5	89.8	94.5	37.3
% mEPT (minus *Cheumatopsyche*)	18	51.7	71.1	83.1	87.0	94.5	42.8
% mEPT (minus *Cheumatopsyche+Baetis*)	18	51.7	71.1	83.1	87.0	94.5	42.8
HBI (Genus)	18	0.90	2.74	3.16	3.54	3.79	2.89
% Tolerant (>6)	18	0.0	0.8	2.1	6.2	23.4	23.4
% 5 Dominant Genera	18	49.2	59.4	72.4	78.6	90.4	41.3
% Chironomidae	18	0.9	3.5	9.9	12.5	39.2	38.3
% Orthocladiinae	18	0.4	1.7	7.0	11.0	30.4	30.0
% Chironomiini	18	0.0	0.0	0.5	0.9	3.4	3.4
% Tanytarsini	18	0.0	0.0	0.5	0.9	5.7	5.7
% Chironomidae+Annelida	18	1.0	4.6	10.3	13.8	39.2	38.3
% Annelida	18	0.0	0.0	0.2	0.6	7.3	7.3
% Non-Insect	18	0.0	0.0	0.5	1.0	9.0	9.0
No. Scraper Genera	18	3	4	5	7	8	5
No. Shredder Genera	18	3	4	5	6	7	4
No. Collector Genera	18	1	5	6	7	10	9
No. Filterer Genera	18	1	2	4	5	8	7
No. Predator Genera	18	1	5	6	6	10	9
% Scrapers	18	3.8	11.3	23.2	37.9	70.8	67.0
% Shredders	18	9.8	18.0	25.7	41.9	76.4	66.7
% Collectors	18	1.0	7.1	16.0	25.3	32.2	31.2
% Filterers	18	0.5	4.0	8.8	15.5	29.6	29.1
% Predators	18	1.0	4.7	8.6	11.8	15.6	14.7
% Clingers	18	24.7	60.0	67.6	72.3	90.9	66.2
% Sprawlers	18	2.7	7.2	20.5	24.7	62.4	59.6
% Swimmers	18	0.0	1.9	2.4	5.8	28.1	28.1
% Climber	18	0.0	0.4	1.0	1.9	7.5	7.5
% Burrower	18	0.0	0.5	1.4	2.1	6.2	6.2

Appendix D. Examples of Stratum-Specific Metric Scoring

Metrics scores are derived using simple equations that standardize metric values by the Best Standard Values (BSVs) and Worst Standard Values (WSVs) which are based upon the 95th and 5th percentile (ceiling and floor), depending on metric direction, from all sites within the stratum. These BSVs and WSVs basically serve to standardize the metrics in the equations below. Metrics scoring greater than 100 are corrected to the maximum value of 100. The GLIMPSS is simply the average of the metric scores for the site. Tables D1 and D2 show example calculations for MT Sp and PL Sp, respectively. Refer to Table 8 for ceiling and floor values for scoring additional strata.

Since metric scoring and the GLIMPSS reference distributions differ across strata (seasons or regions), it is impossible to directly compare GLIMPSS scores between the strata without further standardization. This is easily remedied by calculating a "percent of threshold" value for each sample. Examples of this simple procedure are provided in Table D-3. When compared to the5th percentile of the reference distribution within a particular stratum, a percent of threshold value >100% is unimpaired, while a score <100% is impaired. Other applications of this method to interpret relative site ratings could be done by calculating the percent of other benchmarks found in Table 15 (*e.g.*, the 25th percentile of REF, or any further downward dissections from the 5th percentile).

Table D - 1. Example scoring procedure for the 10-metric MT Sp GLIMPSS. West Fork Pond Fork is a stressed site sampled here for the MT Sp stratum.

Metric	Ceiling	Floor	Equation	Example for West Fork of Pond Fork	Metric Score
No. Intol<4	19	1	$\frac{\textbf{\#Intol} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{2} - 1}{19 - 1}$ X 100	5.6
No. Ephemeroptera Genera	10	1	$\frac{\textbf{\# Ephem} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{1} - 1}{10 - 1}$ X 100	0
No. Plecoptera Genera	8	0	$\frac{\textbf{\# Plecop} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{3} - 0}{8 - 0}$ X 100	37.5
No. Trichoptera Genera	7	1	$\frac{\textbf{\# Trichop} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{3} - 1}{7 - 1}$ X 100	33.3
No. Clinger Genera	20	4	$\frac{\textbf{\# Clingers} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{8} - 4}{20 - 4}$ X 100	25
HBI (Genus)	6.18	2.23	$\frac{\text{Ceiling} - \textbf{HBI}}{\text{Ceiling - Floor}}$ X 100	$\frac{6.18 - \mathbf{5.55}}{6.18 - 2.23}$ X 100	15.9
% Ephemeroptera	59.7	0.5	$\frac{\textbf{\% Ephem} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{17.5} - 0.5}{59.7 - 0.5}$ X 100	28.7
% Orthocladiinae	52.7	0.5	$\frac{\text{Ceiling} - \textbf{\% Orthclad}}{\text{Ceiling - Floor}}$ X 100	$\frac{52.7 - \mathbf{12.9}}{52.7 - 0.5}$ X 100	76.2
% 5 Dominant Taxa	92	48	$\frac{\text{Ceiling} - \textbf{\% Dom5}}{\text{Ceiling - Floor}}$ X 100	$\frac{92 - \mathbf{63.6}}{92 - 48}$ X 100	64.5
No. Scraper Genera	8	0	$\frac{\textbf{\# Scraper} - \text{Floor}}{\text{Ceiling - Floor}}$ X 100	$\frac{\mathbf{1} - 0}{8 - 0}$ X 100	12.5
				GLIMPSS (Ave. Score) =	**29.9**

Table D - 2. Example scoring procedure for the 8-metric PL Sp GLIMPSS. Bear Fork is a reference site sampled here the in PL Sp stratum.

Metric	Ceiling	Floor	Equation	Example for Bear Fork	Metric Score
No. Intol<4	15	1	$\frac{\textbf{\#Intol} - \text{Floor}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{\textbf{13} - 1}{15 - 1} \text{ X 100}$	85.7
No. Ephemeroptera Genera	10	1	$\frac{\textbf{\# Ephem} - \text{Floor}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{\textbf{10} - 1}{10 - 1} \text{ X 100}$	100
No. Plecoptera Genera	7	0	$\frac{\textbf{\# Plecop} - \text{Floor}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{\textbf{8} - 0}{7 - 0} \text{ X 100}$	100
No. Clinger Genera	17	3	$\frac{\textbf{\# Clingers} - \text{Floor}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{\textbf{16} - 3}{17 - 3} \text{ X 100}$	92.9
HBI (Genus)	6.64	2.49	$\frac{\text{Ceiling} - \textbf{HBI}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{6.64 - \textbf{3.60}}{6.64 - 2.49} \text{ X 100}$	73.2
% EPT (minus *Cheumatopsyche*)	90.8	2.5	$\frac{\textbf{\% mEPT} - \text{Floor}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{\textbf{86.2} - 2.5}{90.8 - 2.5} \text{ X 100}$	94.7
% Chironomidae + Annelida	84.7	1.8	$\frac{\text{Ceiling} - \textbf{\% Chiro + Annelid}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{84.7 - \textbf{9.9}}{84.7 - 1.8} \text{ X 100}$	90.1
% Tolerant (>6)	69.5	0	$\frac{\text{Ceiling} - \textbf{\% Tolerant}}{\text{Ceiling - Floor}} \text{ X 100}$	$\frac{69.5 - \textbf{0}}{69.5 - 0} \text{ X 100}$	100
				GLIMPSS (Ave. Score) =	**92.1**

Table D - 3. Calculation of Percent of Threshold as a means to compare GLIMPSS scores across different strata. Comparisons are made here between multiple seasons in Camp Creek, a Non-REF site, and then between adjacent REF and STRESS sites in MT Winter versus MT Summer. Actual GLIMPSS scores are rounded to whole numbers.

Example for Single Site	**Score**	**Threshold (5^{th} percentile)**	**Equation**	**% of Threshold**
Camp Creek-MT Spring	76	53	$\frac{76}{53}X100$	143.4
Camp Creek-MT Summer	65	55	$\frac{65}{55}X100$	118.2
Camp Creek -MT Winter	72	63	$\frac{72}{63}X100$	114.3
Example for REF vs. STRESS				
White Oak Branch -MT Winter (REF)	74	63	$\frac{74}{63}X100$	117.4
Beech Creek-MT Summer (STRESS)	22	55	$\frac{22}{55}X100$	40.0

Appendix E. Modified GLIMPSS Scoring for MT Su Sites >60 sq. mi.

Previous analyses found that larger mountain streams with catchment areas > 60 sq. mi. often behaved differently compared to smaller streams, especially in the summer index period. We chose to evaluate these streams separately so that larger sites were not unfairly scored. First, we examined metric distributions among large river (>60 sq. mi.) REF sites (n= 53 using modified reference screening), Non-REF (n=242), and those that were deemed STRESS (n=19) with stressor screening criteria set forth in Section 7.0. Since some rivers spanned more than one bioregion, the watershed's dominant bioregion was used.

Modified REF screening criteria were established for these 53 MT Su sites. Calculated catchment areas were not available for all sites and some were generally categorized as >60 sq. mi. for obvious large river sites. Those that had catchment areas available ranged from 62 to 6,536 sq. mi. (mean= 411) at 36 REF sites. Larger Non-REF sites with available catchment areas in the database (n=128) ranged from 63 to 7017 sq. mi. (mean=527) and STRESS sites (n=12) ranged from 61 to 862 sq. mi. (mean=206). Because large rivers are much fewer across the State compared to individual streams, this dataset is not free from pseudoreplication (*i.e.*, multiple sites along the same river were used in the distributions); however, we controlled for some pseudoreplication by omitting samples collected from the same sites that were visited <5 years apart. This possibly can be remedied in the future as more large rivers and streams are sampled.

We drew from the metrics chosen for use in the MT Su and PL Su strata, and compared distributions across REF, Non-REF, and STRESS categories. We found that three metrics (# of Shredder Taxa, # of Plecoptera Taxa and % Orthocladiinae) failed DE, range, and SOI tests and so were eliminated. We selected Total No. EPT Genera to provide for maximum range of values. All metrics passed the redundancy test with the highest being No. EPT Genera and No. Clinger Genera (r=0.74). Figure E-1 shows distributions of 7 metrics for consideration for the >60 sq. mi. index.

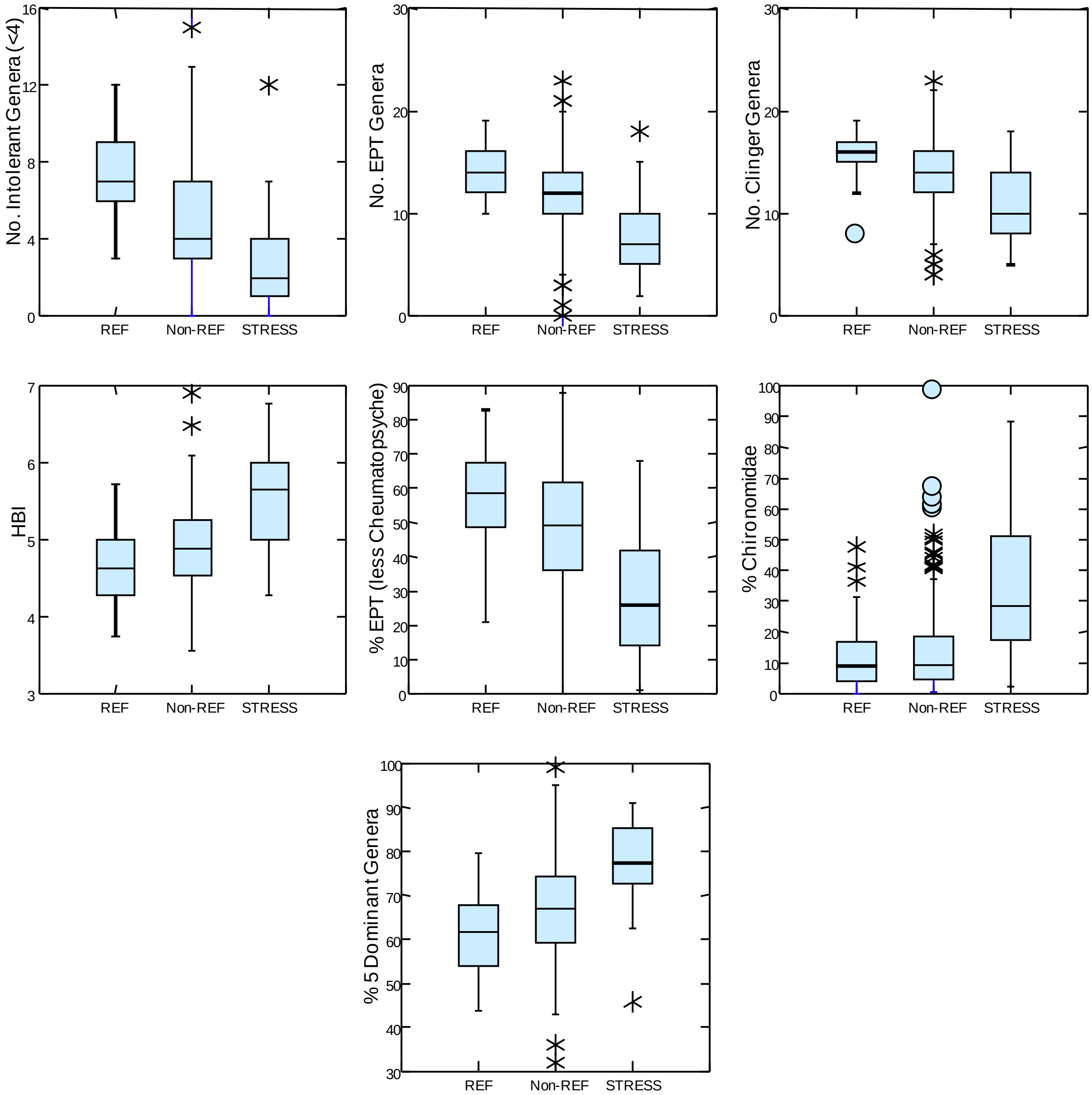


Figure E - 1. Boxplots showing metric distributions across REF, Non-REF, and STRESS categories for MT Su >60 sq. mi. sites.

Because these larger sites were deemed to represent a different benthic community, we established separate standard values and REF GLIMPSS percentiles from a total of 317 and 53 sites, respectively, based on the 7 metrics. These BSVs and WSVs are shown in Table E-1. A modified GLIMPSS score of 51.7 represents the 5th percentile of the 53 REF sites and was used as the impairment threshold for this stream stratum. The modified MT Su GLIMPSS for sites >60 sq. mi. performed reasonably with a DE of 77% (Figure E-2). Moreover, REF site scores did not appear to correspond to increasing drainage area (Figure E-3).

Table E - 1. Standard Values (ceiling and floor) for metrics chosen for the MT Su >60 sq. mi. GLIMPSS. REF percentiles also provided.

	MT Su >60 Sq. Mi. (n=317)	
	Ceiling	Floor
No. Intolerant (<4) Genera	11	1
No. EPT Genera	18	5
No. Clinger Genera	19	8
HBI (Genus)	5.87	4.06
% EPT (minus *Cheumatopsyche*)	76.9	13.8
% Chironomidae	46.1	1.5
%Dominant 5 Genera	86.2	49.0

	5th %ile	10th %ile	25th %ile	Median	75th %ile
MT Su (>60)	51.7	58.1	66.0	71.7	78.1

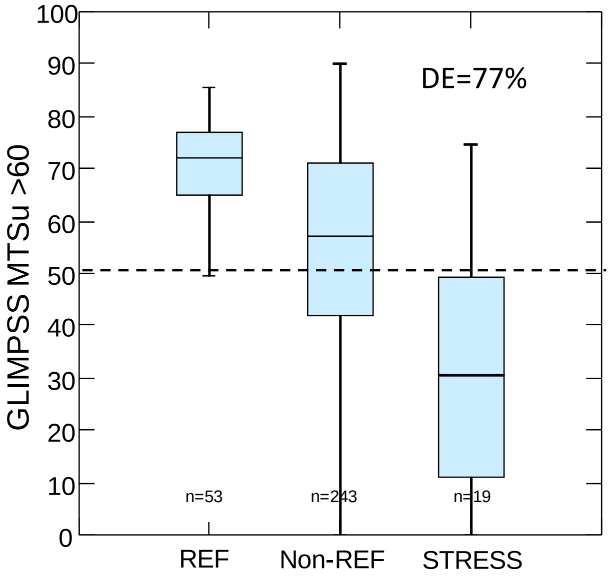


Figure E - 2. REF, Non-REF, and STRESS sites with the modified MT Su GLIMPSS for sites >60 sq. mi. The dashed horizontal line represents the approximate 5th percentile of the REF site distribution.

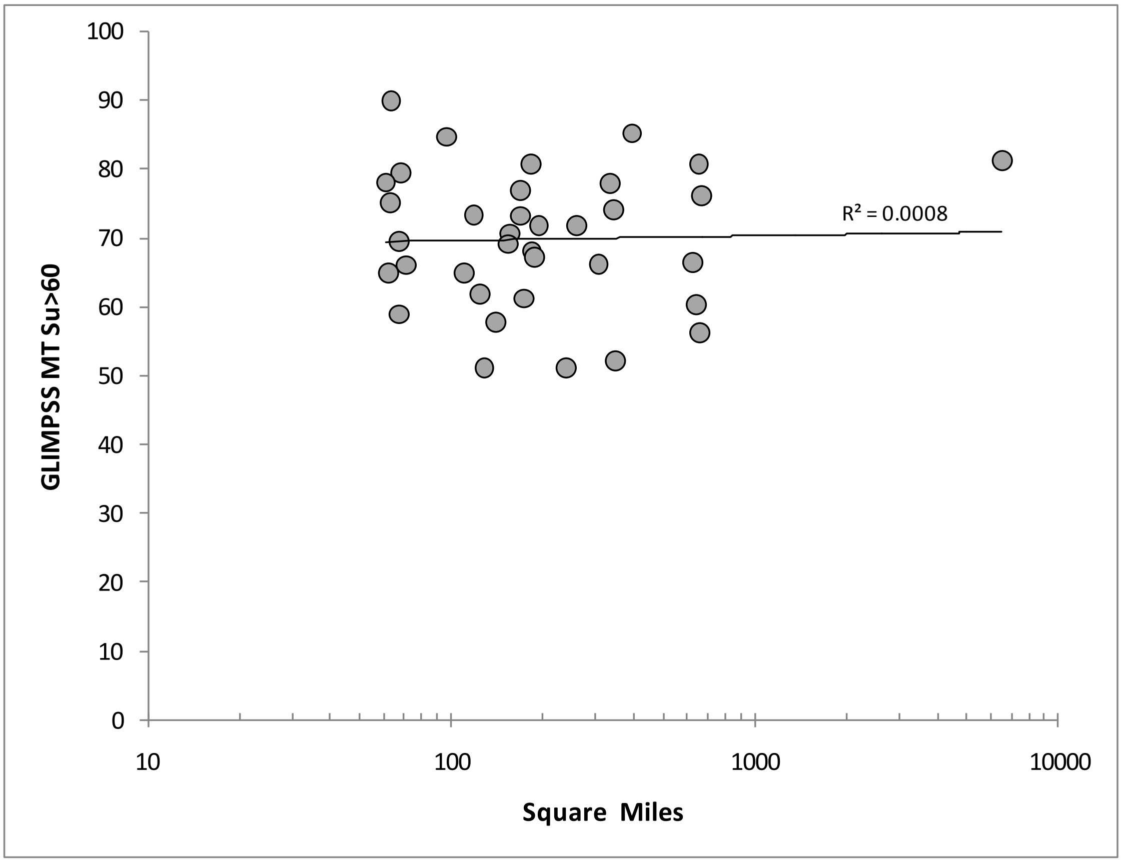


Figure E - 3. Scatterplot of modified MT Su >60 sq. mi. GLIMPSS scores from REF sites versus log catchment area (n=36).

Appendix F. Reference and Stressed Site Summary Statistics for Selected Abiotic Variables

Table F - 1. Calculation of basic descriptive statistics for selected abiotic variables was simply done to characterize the REF and STRESS site environmental conditions found across West Virginia. Only those REF (n=248) and STRESS (n=326) sites that conformed to the symmetric PCA dataset were used (all bioregions were combined).

REF	N	Minimum	25th %ile	Median	75th %ile	Maximum	STRESS	N	Minimum	25th %ile	Median	75th %ile	Maximum
Temperature (°C)	248	5.43	11.39	14.17	17.17	24.83	Temperature (°C)	326	7.56	15.05	18.06	21.6	30.75
pH (SU)	248	5.98	6.63	6.97	7.37	8.27	pH (SU)	326	2.4	7.1	7.57	7.95	10.07
D.O. (mg/l)	248	6	8.74	9.57	10.27	13.69	D.O. (mg/l)	326	2.21	8.15	9.245	10.32	15.53
Specific Conductance (uS/cm)	248	15.4	33.7	57.5	106.0	502.0	Specific Conductance (uS/cm)	326	22	160	303	1014	11227
Sulfate (mg/l)	248	1	5	7.31	13.90	65.70	Sulfate (mg/l)	326	0.94	17.2	38.5	232	6000
Chloride (mg/l)	248	1	1	2.00	2.99	37.50	Chloride (mg/l)	326	0.26	4	7	18	1673
P Total (mg/l)	248	0.01	0.02	0.02	0.03	1.28	P Total (mg/l)	326	0.01	0.02	0.02	0.04	1.28
Fecal Coliform (col/100 ml)	248	0	2	10.00	44	2050	Fecal Coliform (col/100 ml)	326	0	40	238	1600	91000
NO2+NO3 N (mg/l)	248	0.01	0.185	0.30	0.50	1.33	NO2+NO3 N (mg/l)	326	0.01	0.1	0.203	0.4	30
Al Total (mg/l)	248	0.008	0.05	0.10	0.15	2.44	Al Total (mg/l)	326	0.02	0.081	0.12	0.26	53.7
Fe Total (mg/l)	248	0.009	0.05	0.11	0.23	1.63	Fe Total (mg/l)	326	0.02	0.15	0.318	0.56	148
Mn Total (mg/l)	248	0.002	0.008	0.01	0.02	0.55	Mn Total (mg/l)	326	0.002	0.027	0.05605	0.12	18.2
TSS (mg/l)	248	1	3	3.14	5	38	TSS (mg/l)	326	1	3	5	8	59.2
Embeddedness Score	248	9	14.5	16	18	20	Embeddedness Score	326	0	8	10	13	20
Sediment Deposition Score	248	10	13	15	17	20	Sediment Deposition Score	326	0	6	9	11	19
Channel Alteration Score	248	11	18	19	20	20	Channel Alteration Score	326	0	11	14	16	20
Total Bank Stability Score	248	10	16	17	18	20	Total Bank Stability Score	326	1	9	12	15	20
Total Riparian Zone Width Score	248	11	17	18	20	20	Total Riparian Zone Width Score	326	0	3	6	11	20
Total RBP Score	248	132	156	165	171.5	197	Total RBP Score	326	55	101	110	122	185

Appendix G. Derivation of a Modified GLIMPSS (CF) that Excludes Genus-Level Chironomidae

Rationale

The midge family Chironomidae is very diverse in WV streams with over 100 genera found throughout the State. Bioassessment practitioners outside of the WVDEP WAB (*e.g.*, consultants, academia, and local watershed groups) often do not have sufficient expertise to correctly identify these organisms to the genus-level; therefore, we developed a Modified GLIMPSS that omits this often taxonomically difficult group of insects (called the GLIMPSS (Chironomidae Family), or GLIMPSS (CF)). Furthermore, the WAB may wish to employ this modified GLIMPSS (CF) when assessments must be done quickly, or if taxonomic resources become limited. In the neighboring states of PA and VA, chironomids are routinely identified to the family-level, while in adjacent OH, KY, and MD, they are routinely identified to genus-level. In this section, we re-evaluated GLIMPSS metrics from the initial dataset while collapsing genus-level chironomid identifications to the family-level.

Methods

The identical dataset used to calibrate and validate the GLIMPSS (CAL and VAL; REF, Non-REF, STRESS) was re-analyzed. First, all chironomid genera in the benthic enumeration table were converted (collapsed) to family level. Stratum-specific GLIMPSS metrics that involve chironomid genera include:

No. Total Genera (MT Su, PL Su)
No. Intolerant Genera <4 (MT Sp/Win, MT Su, PL Sp/Win)
No. Intolerant Genera <3 (PL Su)
No. Clinger Genera (MT Sp/Win, MT Su, PL Sp/Win, PL Su)
HBI (MT Sp/Win, MT Su, PL Sp/Win, PL Su)
% Tolerant >6 (PL Sp/Win)
% Dominant 5 (MT Sp/Win, MT Su, PL Su)
No. Scraper Genera (MT Sp/Win, PL Su)
No. Shredder Genera (MT Su)

These metrics were then recalculated with chironomids collapsed to the family level. The % Orthocladiinae metric (a chironomid sub-family metric) was automatically omitted from the Modified GLIMPSS (CF). All other metrics that comprise the GLIMPSS were retained and unmodified (*e.g.*, No. Ephemeroptera Genera, % EPT (less Cheumatopsyche).

Metric re-testing and confirmation involved calculating the %DE, PCA correlation, redundancy magnitude, range, and SOI of the newly modified GLIMPSS metrics from the CAL dataset (as described in Section 7.3 and Appendix B). In the case of % Orthocladiinae, two highly comparable analogs (% Chironomidae and % Chironomidae+Annelida) were tested as a potential replacement metric.

BSVs and WSVs (as ceilings and floors) for metrics were recalculated from the full dataset and GLIMPSS scores were derived for all CAL and VAL sites. As in the development of the full GLIMPSS, the Modified GLIMPSS (CF) was similarly evaluated for sensitivity, discrimination efficiency, relation to stress, and precision. Finally, the full GLIMPSS model and the modified GLIMPSS (CF) models were compared using contingency tables and simple linear regression.

Results

Metric Testing and Selection

Table G-1 shows the metrics that make up the Modified GLIMPSS (CF) by bioregion/season, sorted primarily by %DE and secondarily by their correlation to the human disturbance gradient identified by PCA axis 1. Maximum redundancy magnitudes (correlation) are listed for those metrics chosen for the Modified GLIMPSS (CF).

Table G - 1. List of Modified GLIMPSS (CF) metrics by stratum with %DE, correlation to PCA axis 1 (p<0.05), and maximum correlation value for metric pairs as a measure of redundancy (absolute value). MT Sp= Mountain Spring, MT Su=Mountain Summer, PL Sp=Plateau Spring, PL Su=Plateau Summer.

	MT Sp	DE	PCA	Redun		MT Su	DE	PCA	Redun
Tol	No. Intol Taxa (<4)	93.4	-0.63	0.76	Tol	No. Intol Taxa (<4)	95.4	-0.67	0.72
Habit	No. Clinger Genera	85.2	-0.52	0.68	Tol	HBI	95.0	0.66	0.74
Tol	HBI	83.6	0.61	0.41	Rich	No. Plecoptera Genera	91.0	-0.62	0.72
Dom	% Dominant 5 Genera	81.9	0.47	0.61	Comp	% EPT (minus *Cheumatopsyche)*	90.1	-0.59	0.74
Rich	No. Trichoptera Genera	81.0	-0.39	0.58	Habit	No. Clinger Genera	85.8	-0.47	0.79
Rich	No. Ephemeroptera Genera	79.4	-0.36	0.42	Dom	% Dominant 5 Genera	82.1	0.41	0.59
Rich	No. Plecoptera Genera	77.8	-0.53	0.76	Rich	No. Total Genera	79.7	-0.45	0.79
Comp	% Chironomids+Annelids	73.4	0.49	0.41	Rich	No. Ephemeroptera Genera	77.9	-0.43	0.48
Comp	% Ephemeroptera	74.6	-0.31	0.39	FFG	No. Shredder Genera	76.1	-0.49	0.54
FFG	No. Scraper Genera	67.2	-0.43	0.42	Comp	% Chironomids+Annelids	61.5	0.39	0.4
	PL Sp	DE	PCA	Redun		PL Su	DE	PCA	Redun
Comp	% EPT (minus *Cheumatopsyche)*	82.6	-0.36	0.68	Tol	No. Intol Taxa (<3)	97.6	-0.54	0.70
Tol	No. Intol Taxa (<4)	81.8	-0.49	0.72	Tol	HBI	94.4	0.60	0.77
Rich	No. Plecoptera Genera	79.0	-0.38	0.65	Dom	% Dominant 5 Genera	90.4	0.54	0.57
Comp	% Chironomids+Annelids	75.9	0.31	0.68	Comp	% EPT (minus *Cheumatopsyche)*	85.7	-0.54	0.77
Tol	HBI	75.2	0.40	0.55	Habit	No. Clinger Genera	84.4	-0.53	0.74
Habit	No. Clinger Genera	68.1	-0.46	0.70	Rich	No. Total Genera	80.5	-0.54	0.74
Rich	No. Ephemeroptera Genera	65.2	-0.35	0.60	Rich	No. Ephemeroptera Genera	68.9	-0.52	0.57
					Comp	% Chironomidae	64.9	0.52	0.53
					FFG	No. Scraper Genera	63.3	-0.46	0.54

By excluding chironomid genera, metric %DE, PCA correlation, and redundancy generally increased compared to metrics comprising the full GLIMPSS model. DE increased between ~1% (*e.g.*, No. Scraper Genera) to 11% (*e.g.*, No. Clinger Genera). Unmodified metrics (those not involving chironomids) were not re-tested. In the PL Sp, the % Tolerant >6 metric was rejected for further use, indicating that by collapsing chironomid genera to family, this metric lost substantial sensitivity. Among these best metrics, some pairs (*e.g.*, No. Total Genera and No. Clinger Genera in MT Su) had a correlation of 0.79. After viewing scatterplots of these metric pairs, it was still reasonable to include both metrics in the GLIMPSS (CF) because of the large scatter of one metric at any given value of the other metric (see example in Figure G-1). For example, in particular collections having 23 Total Genera, No. Clinger Richness ranged from 9 to 17 genera. For collections having 26 Total Genera, Clinger Richness ranged from 11 to 20 genera.

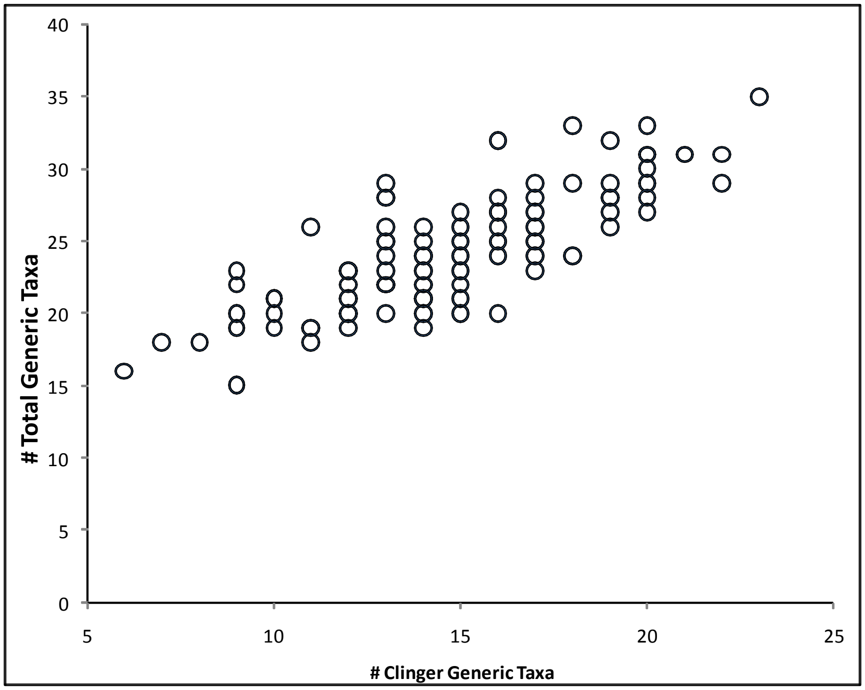


Figure G - 1. Scatterplot of No. Total Genera and No. Clinger Genera. The correlation was 0.79, but there was sufficient scatter to warrant inclusion of both metrics in the GLIMPSS (CF). Points represent multiple observations (half of the observations are hidden beneath other points).

Comparison of GLIMPSS versus GLIMPSS (CF) Metric Performance

We compared mean %DE, correlation PCA, and redundancy magnitude across metrics within each stratum. Table G-2 shows that the modified GLIMPSS (CF) metrics performed similarly to the original GLIMPSS most of the time. It indicates that the GLIMPSS (CF) metrics are highly suitable for GLIMPSS assessment purposes when necessary.

Table G - 2. Mean values for Calibration %DE, PCA, and Redundancy between stratum-specific GLIMPSS metrics and modified GLIMPSS (CF) metrics. Bolded values highlight the highest performing set of metrics by stratum.

	DE	DE (CF)	PCA	PCA (CF)	Redun	Redun (CF)
Spring Mountains	79.4	**79.7**	0.45	**0.47**	0.56	**0.54**
Summer Mountains	80.7	**83.5**	0.47	**0.52**	**0.63**	0.65
Spring Plateau	75.3	**75.4***	0.37	**0.39***	0.69	**0.65***
Summer Plateau	78.5	**81.1**	0.49	**0.54**	**0.64**	0.67
* Based on 7 metrics (% Tolerant >6 metrics was rejected for use in Spring Plateau)						

Metric Aggregation and Index Performance

Within individual strata, metric BSVs and WSVs based on the 95th or 5th percentile of the CAL and VAL datasets were calculated and all metrics were scored for each site in the dataset. BSVs for each stratum are shown in Table G-3. The aggregate 100-point GLIMPSS (CF) was then calculated as the average metric score. Refer to Appendix D for example scoring formulae.

Table G - 3. Best standard values (BSVs) and Worst Standard Values (WSVs), as ceilings and floors, for metrics (by stratum) used for Final scoring purposes of GLIMPSS (CF).

	MT Sp&Win (n=732)		MT Su (n=1530)		MT Su >60 (n=317)		PL Sp&Win (n=692)		PL Su (n=858)	
	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceilin	Floor
No. Total Genera			30	8					25	8
No. Intolerant Genera <4	18	1	15	0	10	1	14.5	1		
No. Intolerant Genera <3									7	0
No. EPT Genera					18	5				
No. Ephemeroptera Genera	10	1	9	0			10	1	7	0
No. Plecoptera Genera	8	0	7	0			7	0		
No. Trichoptera Genera	7	1								
No. Clinger Genera	19.5	3.5	18	4	18	7	16.5	3	14	3
No. Scraper Genera	8	0							7	1
No. Shredder Genera			4	0						
HBI (Genus level)	5.87	2.19	5.90	2.80	5.75	4.03	5.94	2.45	5.98	3.84
% Dominant 5 Genera	96.7	55.5	96.7	57.7	92.1	55.1			97.5	64.4
% Ephemeroptera	59.7	0.5								
% mEPT (minus *Cheumatopsyche)*			86.0	5.2	76.9	13.8	90.8	2.5	67.1	1.3
% Chironomidae					46.1	1.5			68.8	3.3
% Chironomidae+Annelida	75.2	2.8	65.0	2.7			84.6	1.8		

Box plots and %DE were calculated as a means to view the sensitivity of GLIMPSS (CF) within individual strata. Figure G-2 demonstrates that the modified index could discriminate between REF and STRESS sites with a high degree of efficiency (≥75% of STRESS sites score below 5th percentile of REF distribution). It also indicates that Non-REF sites fell into a presumed intermediate position between REF and STRESS, with respect to scores. DE was greatest in the PL Su stratum where ~90% of the STRESS sites fell below the 5th percentile of the REF distribution. In addition, interquartile ranges for each stratum were relatively low (ranging from 15 to 20 points) suggesting low variability of the reference condition.

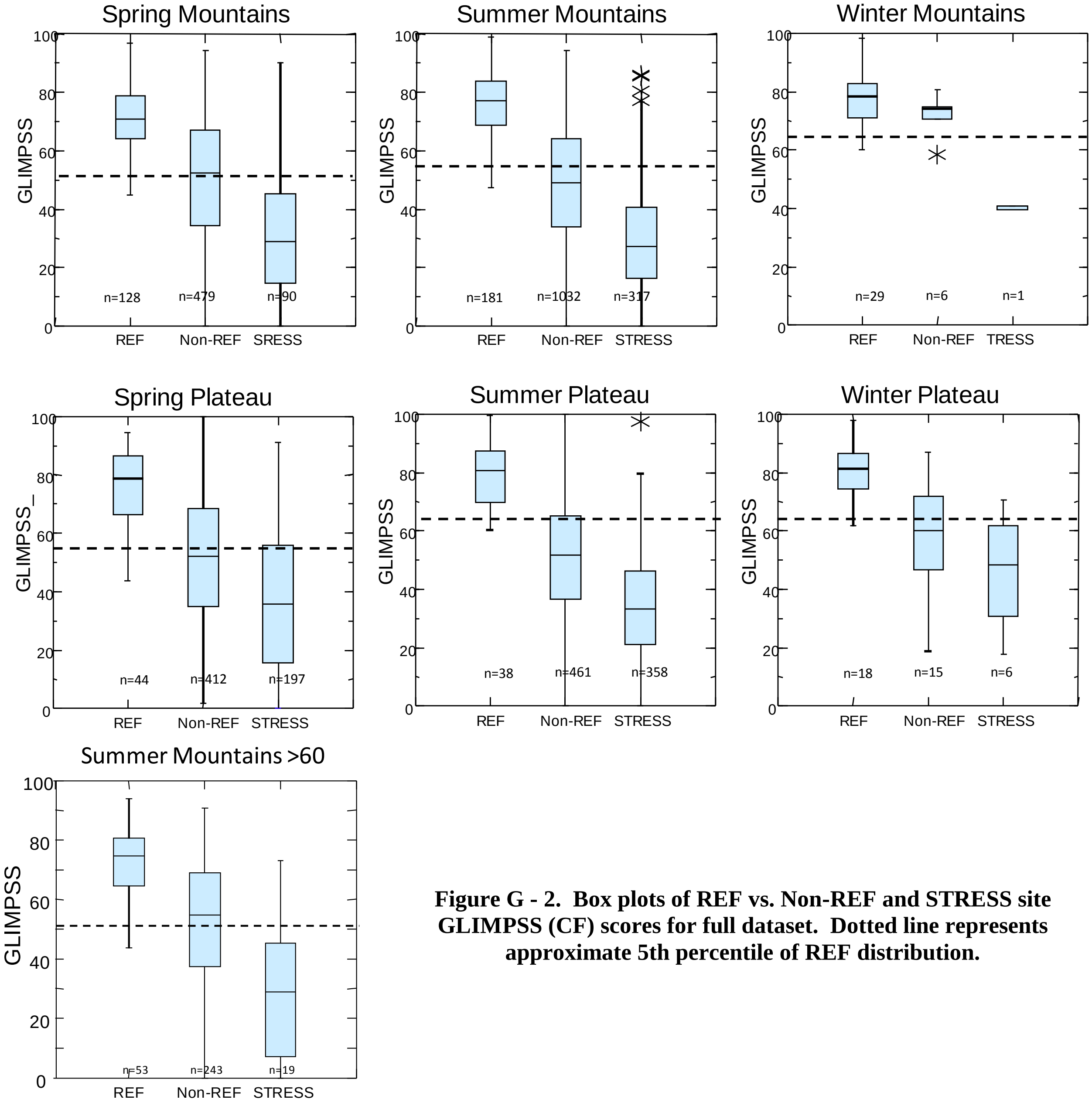


Figure G - 2. Box plots of REF vs. Non-REF and STRESS site GLIMPSS (CF) scores for full dataset. Dotted line represents approximate 5th percentile of REF distribution.

Relationships between GLIMPSS and GLIMPSS (CF)

Although there was not a perfect 1:1 relationship, the 2 models were highly correlated (Figure G-3). Slightly more scatter was seen in MT Su and PL Su compared to other strata. In most strata, the full GLIMPSS tended to score higher than GLIMPSS (CF) as depicted by points lying above the 1:1 line. However, further comparisons of REF site mean and SD indicate that the 2 models are quite similar (Table G-3). Mean GLIMPSS (CF) scores generally declined (compared to the full GLIMPSS) across all strata and REF, Non-REF, and STRESS categories while SD increased. One notable difference was observed in the PL Su where mean REF site score increased by 6.4 points, while mean STRESS score decreased by 2.1 points. This difference led to a greater number of sites in PL Su scoring below the threshold compared to the full GLIMPSS (Figure G-3, PL Su). This suggests that removal of Chironomidae increased index performance.

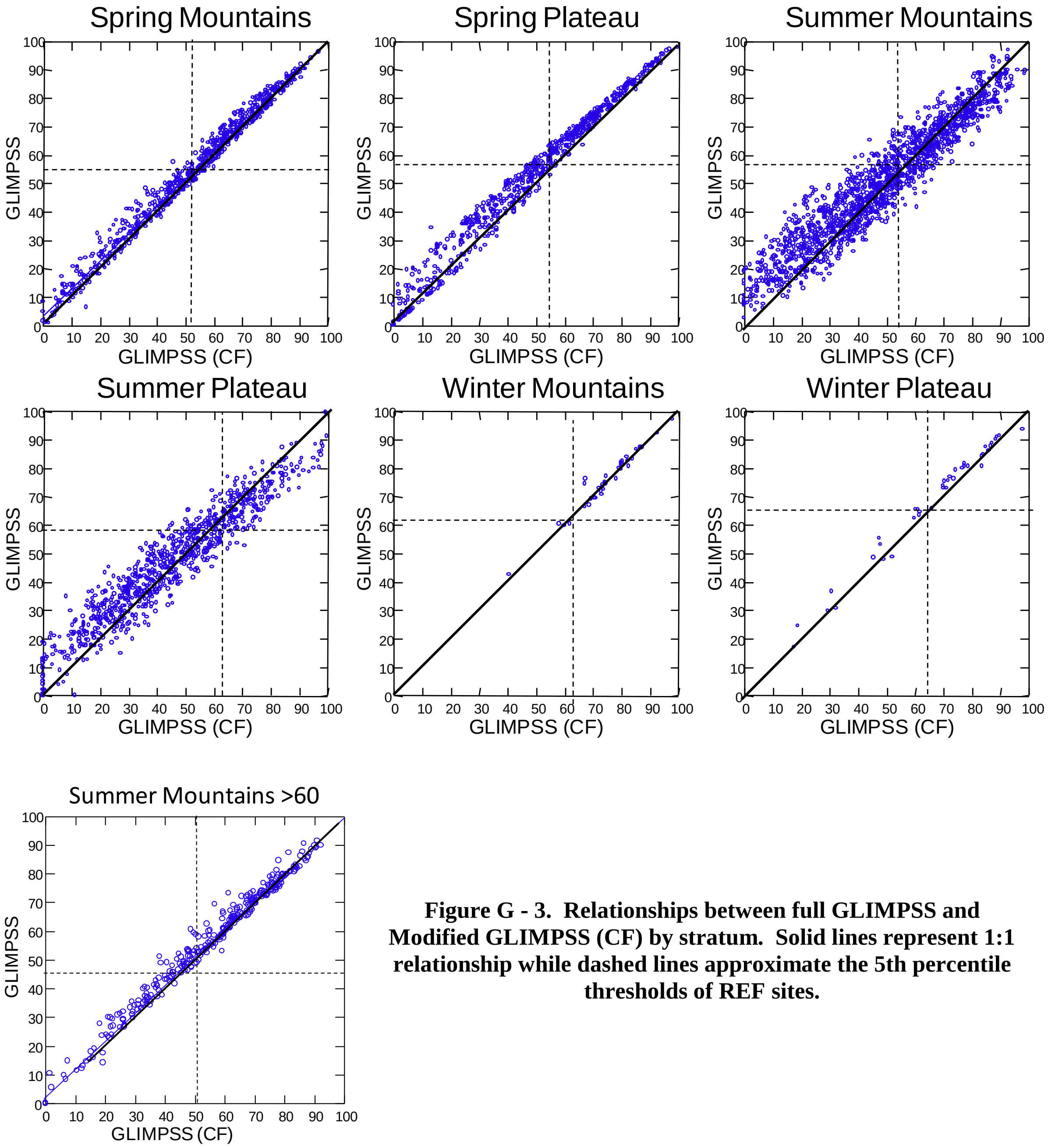


Figure G - 3. Relationships between full GLIMPSS and Modified GLIMPSS (CF) by stratum. Solid lines represent 1:1 relationship while dashed lines approximate the 5th percentile thresholds of REF sites.

Table G - 4. Comparison of mean and SD of GLIMPSS and GLIMPSS (CF) scores between REF, Non-REF, and STRESS datasets among individual strata.

			GLIMPSS	GLIMPSS (CF)				GLIMPSS	GLIMPSS (CF)
MT Sp	REF	Mean	72.5	71.2	MT Su	REF	Mean	73.0	75.6
		SD	10.2	10.9			SD	9.9	11.5
	Non-REF	Mean	51.9	50.5		Non-REF	Mean	50.2	48.8
		SD	20.9	21.6			SD	18.1	20.7
	STRESS	Mean	33.8	31.4		STRESSS	Mean	33.7	29.4
		SD	20.0	19.9			SD	15.6	18.0
PL Sp	REF	Mean	78.4	76.2	PL Su	REF	Mean	73.6	80.0
		SD	11.5	12.8			SD	10.2	11.4
	Non-REF	Mean	54.7	51.2		Non-REF	Mean	49.9	50.0
		SD	22.2	23.0			SD	18.0	21.0
	STRESS	Mean	40.7	36.7		STRESS	Mean	38.7	34.6
		SD	23.4	23.4			SD	16.7	19.2
MT Wi	REF	Mean	77.2	77.0	MT Su >60	REF	Mean	72.6	73.2
		SD	8.8	8.8			SD	10.6	11.9
	Non-REF	Mean	72.6	72.1		Non-REF	Mean	54.5	52.7
		SD	7.4	7.5			SD	19.8	20.8
	STRESS	Mean	42.5	40.7		STRESS	Mean	30.8	28.5
		SD	NA	NA			SD	25.3	24.5
PL Wi	REF	Mean	81.6	80.1					
		SD	8.3	9.3					
	Non-REF	Mean	60.4	57.9					
		SD	20.4	20.3					
	STRESS	Mean	48.8	46.3					
		SD	20.4	19.4					

As with the WVSCI and GLIMPSS, Modified GLIMPSS (CF) thresholds to demarcate impaired from non-impaired status are based upon the 5th percentile of the reference distribution. The 25th percentile of the reference distribution can be used as a lower threshold to identify exceptional biological assemblages found in the State. Similarly, GLIMPSS values below the impairment threshold can be partitioned to provide categories that reflect increased stress to biological communities (*e.g.*, degraded, severely degraded). Table G-5 presents stratum-specific GLIMPSS (CF) scores indicating the 5th and 25th percentile values and equal bisection of the impairment range.

Table G - 5. Modified GLIMPSS (CF) scoring criteria for all strata. Scores are rounded to nearest whole number. MT Sp = Mountain Spring, MT Su = Mountain Summer, MT Su>60 = Mountain Summer >60 Sq. Mi., MT Win = Mountain Winter, PL Sp = Plateau Spring, PL Su = Plateau Summer, and PL Win = Plateau Winter. Criteria are rounded to whole numbers.

	MT Sp	MT Su	MT Su>60	MT Win	PL Sp	PL Su	PL Win
25th Percentile	64	69	65	71	66	70	74
5th Percentile	51	54	51	64	57	62	65
Impairment Threshold							
Increased Severity ↓	25-50	27-53	25-50	32-63	28-56	30-61	32-64
of Impact	<25	<27	<25	<32	<28	<30	<32