A BENTHIC MACROINVERTEBRATE MULTIMETRIC INDEX FOR

LARGE SEMI-WADEABLE RIVERS:

TECHNICAL REPORT

Prepared by:

Dustin Shull
PA Department of Environmental Protection
Office of Water Programs
Bureau of Clean Water
11th Floor: Rachel Carson State Office Building
Harrisburg, PA 17105

2018

Acknowledgements

Many individuals across several state, interstate, federal, and academic organizations have contributed a tremendous amount of work toward field collection of data, method development, and peer review of this report. Great credit must first be given to the individuals (below) that have spent many hours collecting and processing one of the most robust and highest quality datasets available for large rivers in the Mid-Atlantic region to date.

Much appreciation must also go to the members of the peer review committee (below) who were integral in developing and refining the analyses in this report, and for assisting with the determination of the final assessment methodology.

Name	Affiliation	Office
Aaron Henning	Susquehanna River Basin Commission	Harrisburg, PA
Matthew Shank	Susquehanna River Basin Commission	Harrisburg, PA
Luanne Steffy	Susquehanna River Basin Commission	Harrisburg, PA
Zachary Smith	Interstate Commission on the Potomac River	Rockville, MD
Gordon Selkmann	Interstate Commission on the Potomac River	Rockville, MD
Geoff Smith	Pennsylvania Fish and Boat Commission	Harrisburg, PA
Mark Hartle	Pennsylvania Fish and Boat Commission	Bellefonte, PA
Thomas Shervinskie	Pennsylvania Fish and Boat Commission	Bellefonte, PA
Dave Lieb	Pennsylvania Fish and Boat Commission	Bellefonte, PA
Greg Pond	United States Environmental Protection Agency	Wheeling, WV
Louis Reynolds	United States Environmental Protection Agency	Wheeling, WV
Mike Bilger	Susquehanna University	Selinsgrove, PA
Robert Limbeck	Delaware River Basin Commission	Trenton, NJ
Kelly Maloney	United States Geological Survey	Leetown, WV
Daniel Spooner	United States Geological Survey	Wellsboro, PA
John Jackson	Stroud Water Research Center	Avondale, PA
Theo Light	Shippensburg University	Shippensburg, PA
Jeff Butt	PADEP	Harrisburg, PA
Tim Wertz	PADEP	Harrisburg, PA
Josh Lookenbill	PADEP	Harrisburg, PA
Gary Walters	PADEP	Harrisburg, PA
Rodney Kime*	PADEP	Harrisburg, PA
Travis Stoe	PADEP	Harrisburg, PA
Timothy Wertz	PADEP	Harrisburg, PA
Mark Hoger	PADEP	Harrisburg, PA
Justin Lorson	PADEP	Harrisburg, PA
Erika Bendick	PADEP	Harrisburg, PA
Shawn Miller	PADEP	Harrisburg, PA
Kristen Bardell	PADEP	Harrisburg, PA
Rick Spear	PADEP	Pittsburg, PA
Joe Brancato	PADEP	Meadville, PA
Dave Rebuck	PADEP	Williamsport, PA
Steve Means	PADEP	Williamsport, PA
Walter Holtsmaster	PADEP	Wilkes-Barre, PA
Alan Everett	PADEP	Norristown, PA
James Grazio	PADEP	Erie, PA

*Retired PADEP staff.

iii

Table of Contents

Executive Summary

Macroinvertebrate Collection Method for Large Semi-wadeable Rivers

Benthic macroinvertebrates have a long-standing tradition of being used as indicators of water quality in wadeable streams. However, the use of benthic macroinvertebrate in large non-wadeable rivers has become more common in recent years. Large non-wadeable rivers likely require more complex collection methods to represent the macroinvertebrate community with sufficient accuracy and precision. These modifications are required due to the great size, variable discharges, and increased depth of large rivers. However, stream size alone may not be the best indicator of when macroinvertebrate collection methods should transition from wadeable to non-wadeable. This transition is not well-defined, but previous recommendations include separating streams with drainages greater than 2,600 km^2 or 6th order streams and larger. More detailed definitions exist; however, without the influence of impounds, this is most likely a continuous rather than a discrete transition with a high degree of variability depending on the individual river.

The Pennsylvania Department of Environmental Protection (PADEP) has developed several macroinvertebrate assessment methodologies to account for the large diversity of wadeable stream types that exist within the Commonwealth of Pennsylvania. These methods have proven effective at discriminating between good and poor water quality for much of the 138,400 stream kilometers in Pennsylvania; however, there are limitations to using these methods for the remaining streams, which are predominantly large rivers (drainage areas >2,600 km^2). Approximately half of these large rivers (≈ 3,000 stream kilometers) could be considered semi-wadeable in the Commonwealth of Pennsylvania. Semi-wadeable rivers are defined as predominantly free-flowing systems with drainage areas >2,600 km^2, and have physical characteristics that allow for riffle and run sections to occur with relative frequency. These river systems tend to lack a well-defined and navigable U-shaped channel for any significant distance and frequently present difficulties for both wadeable and non-wadeable macroinvertebrate data collection methodologies. Well over half of the large rivers within the Commonwealth are considered semi-wadeable.

Several studies have now shown that large semi-wadeable rivers may not mix major water influences (i.e., major tributaries) across the width of the river. These factors can potentially increase the variability of macroinvertebrate collection methods to unacceptable levels. To account for these issues, PADEP began increasing the number of macroinvertebrate samples collected across the width of rivers. One implemented approach was termed the "transect method". The transect method was a modification of

the wadeable 6D-200 (6 one minute kicks using a D-frame net, and subsampling macroinvertebrates down to 200 individuals) method where the number of macroinvertebrate samples collected was dictated by the number of significant water quality differences across the width of a river. For instance, at Rockville, PA this resulted in three separate 6D-200 macroinvertebrate samples representing each of the three major influences (Juniata, West Branch Susquehanna, and mainstem Susquehanna Rivers). The goal of this work was to determine whether major tributary influences consistently influence macroinvertebrate community composition at a single location, and to determine whether the transect method can reduce method variability to acceptable levels for index development.

The Susquehanna River was chosen to test whether macroinvertebrate communities responded to water influence differences across the width of the river, because it represented the most complex semi-wadeable river system in the Commonwealth. Results of this work suggested that the macroinvertebrate communities do respond to water influences across the width of a river (Figure 1). It is critical to note that these distinctions between macroinvertebrate communities were observed even with all other sampling errors being inherently incorporated into the analyses. Most importantly, however, the results suggested that the transect method substantially reduced macroinvertebrate index score variability by 45% on average when compared to the wadeable collection method (Table 1). From this information, it was concluded that the transect method can be used to create an Aquatic Life Use assessment tool.
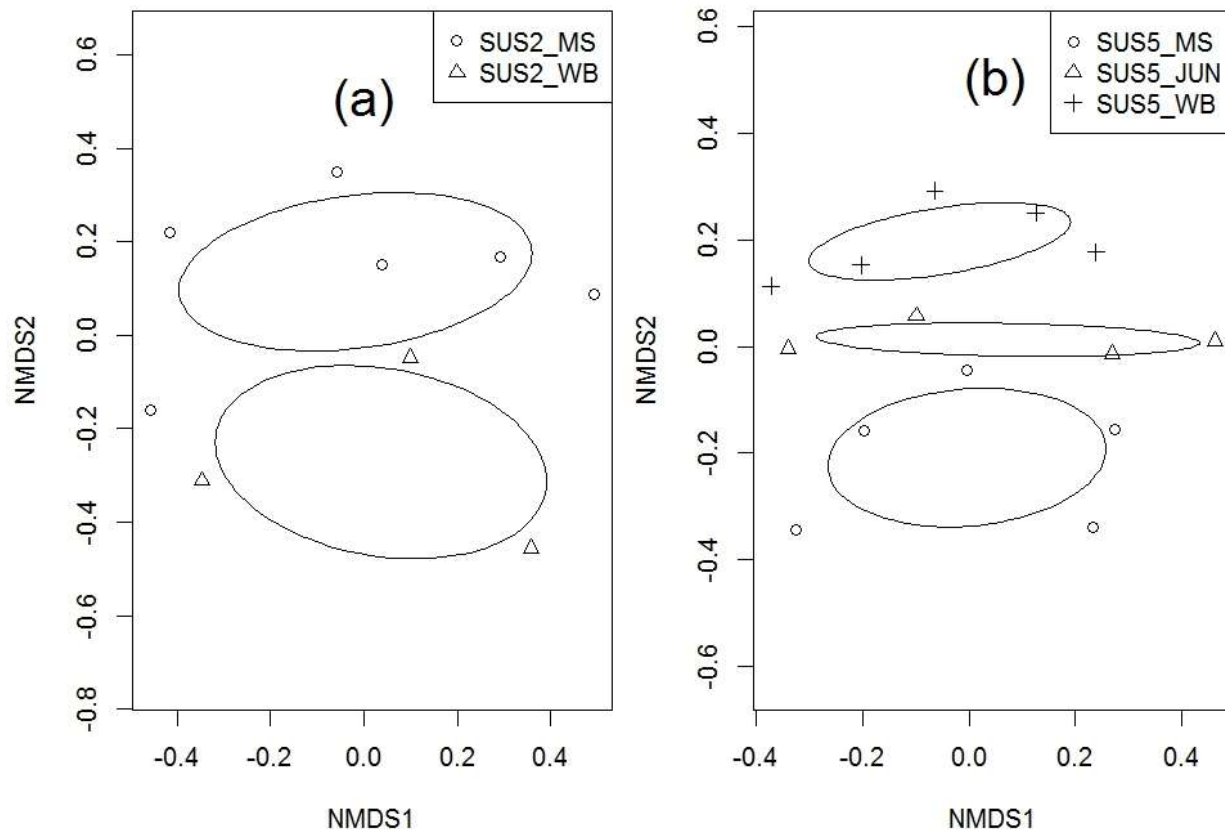
Figure 1. (a) Non-metric multidimensional scaling (NMDS) (dimensions = 2, stress = 0.1) of macroinvertebrate samples collected in the Susquehanna River at Sunbury, PA based on the transect method, where water quality transects suggest unique and consistently different influences. Sample sizes were 2SUS_MS (n=6) and 2SUS_WB (n=3). Ellipses represent the standard deviation of points in each group. (b) NMDS (dimensions = 2, stress = 0.2) of macroinvertebrate samples collected in the Susquehanna River at Rockville, PA based on the transect method, where water quality transects suggest unique and consistently different influences. Sample sizes were 5SUS_JUN (n=4), 5SUS_MS (n=5), and 5SUS_WB (n=5). Ellipses represent the standard deviation of points in each group.

Table 1. Method variability comparisons using macroinvertebrate index score coefficients of variation (CV) at sites where both datasets were available.

River	Location Description	Original Wadeable Method CV	Transect Method CV
Delaware River	Morrisville, PA	29.5%	9.5%
Susquehanna River	Sunbury, PA	15.6%	11.0%
Susquehanna River	Danville, PA	18.3%	13.6%
West Branch Susquehanna River	Milton, PA	16.2%	9.2%

French Creek	Franklin, PA	21.7%	12.1%

Briefly described, the transect method begins with collecting at least one water quality transect consisting of multiple discrete field meter readings at intervals of 5 to 10% of the wetted width of the river. The four common field meter parameters of water temperature, pH, specific conductance, and dissolved oxygen are typically sufficient to determine differential influences; however, if possible, measuring turbidity can also indicate important water quality differences. Percent differences of 10% or greater for any parameter indicate that an additional 6D-200 sample is required. Differentiating between major and minor influences is also important; therefore, influences that span more than 10% of total width should be characterized with an additional 6D-200 sample. If only one transect can be conducted before macroinvertebrate sampling takes place, it should be completed at base flow conditions to ensure macroinvertebrates are being collected in the correct area across the width of the river. However, multiple transects before macroinvertebrate collection is highly recommended.

Semi-wadeable Macroinvertebrate Multimetric Index

One of the most important steps in the development of a multimetric index (MMI) is to evaluate anthropogenic effects and develop a stressor gradient based on abiotic factors before biological communities are evaluated. Once the stressor gradient is built and sites are classified on natural factors, macroinvertebrate communities are evaluated in response to anthropogenic stress. Benthic macroinvertebrates are especially good indicators of water quality because many groups are sensitive to pollution, and they can readily recolonize habitats under improving environmental conditions. The MMI development process uses direct quantification of biological attributes along the stressor gradient. Proper MMI development not only selects macroinvertebrate metrics that respond well to stress, but also goes through a series of redundancy and precision checks to ensure each metric is viable. Once metrics are evaluated and selected, the final index standardizes and composites metrics into one score, which measures the overall extent to which anthropogenic activities compromise a stream's ability to support healthy aquatic communities.

This semi-wadeable large river MMI (SWMMI) development incorporated over 400 macroinvertebrate samples from 178 sites across the Mid-Atlantic Region (Figure 2). An abiotic stressor gradient was applied to all sites (Table 2). Thresholds for each abiotic parameter were selected based on a combination of literature, empirical data, Aquatic Life Use (ALU) criteria, quantile analysis, and expert consensus. Sites were then classified based on macroinvertebrate community differences in the least disturbed condition. Results suggest that seasonality was the strongest natural classification for macroinvertebrate communities (Figure 3); hence, SWMMI development should

proceed with two distinct metric evaluations for the summer and the fall macroinvertebrate communities.
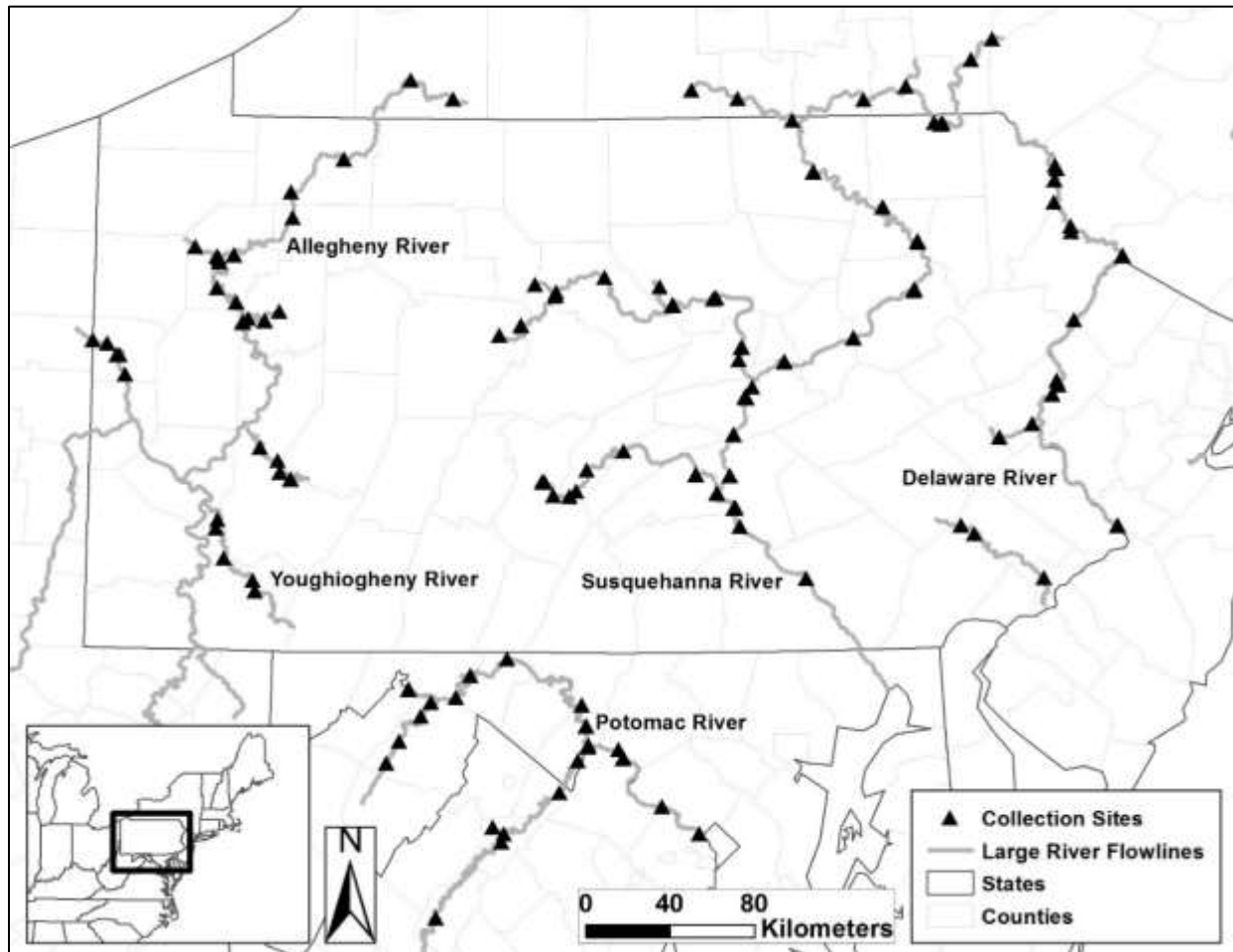
Figure 2. Study area with collection sites separated by calibration and validation datasets.

Table 2. Environmental condition gradient criteria.

Stressor	Good 0	Fair 1	Poor 2
% Developed land use in the watershed	< 5	5 to 10	> 10
% Agriculture land use in the watershed	< 10	10 to 30	> 30
Total Phosphorus (mg/L)	< 0.1	0.1 to 0.2	> 0.2
Total Nitrogen (mg/L)	< 1.0	1.0 to 2.0	> 2.0
Specific Conductance (µS/cm)	< 300	300 to 600	> 600
Total Chloride (mg/L)	< 15	15 to 200	> 200
Total Sulfate (mg/L)	< 20	20 to 200	> 200
Total Ammonia (mg/L)	< 0.03	0.03 to 0.05	> 0.05
Instream Habitat (Total of instream cover, epifaunal substrate, and embeddedness)	> 45	45 to 30	< 30
Within 16 km of a major impoundment			X
Within 1.6 km of a minor impoundment			X

Figure 3. NMDS analysis (dimensions = 3, stress = 0.2) plot of first two dimensions using season as groups. Ellipses represent the standard deviation of points in each group.

Index development results suggest that both the summer and fall SWMMI are effective at distinguishing between least disturbed and stressed conditions (Figure 4). Each SWMMI is therefore an applicable and defensible tool for the Commonwealth, and potentially, the greater Mid-Atlantic region. Results also indicate that each SWMMI is acceptably precise, further supporting that this method can pick up on ecological changes due to anthropogenic stress or restoration efforts over time. Overarching results of this work indicate that the summer and fall SWMMI are not only very responsive to anthropogenic stress, but also show that two unique macroinvertebrate communities exist between seasons in large semi-wadeable rivers. Thus, each tool should be used separately when making Aquatic Life Use assessments.

(a)   (b)
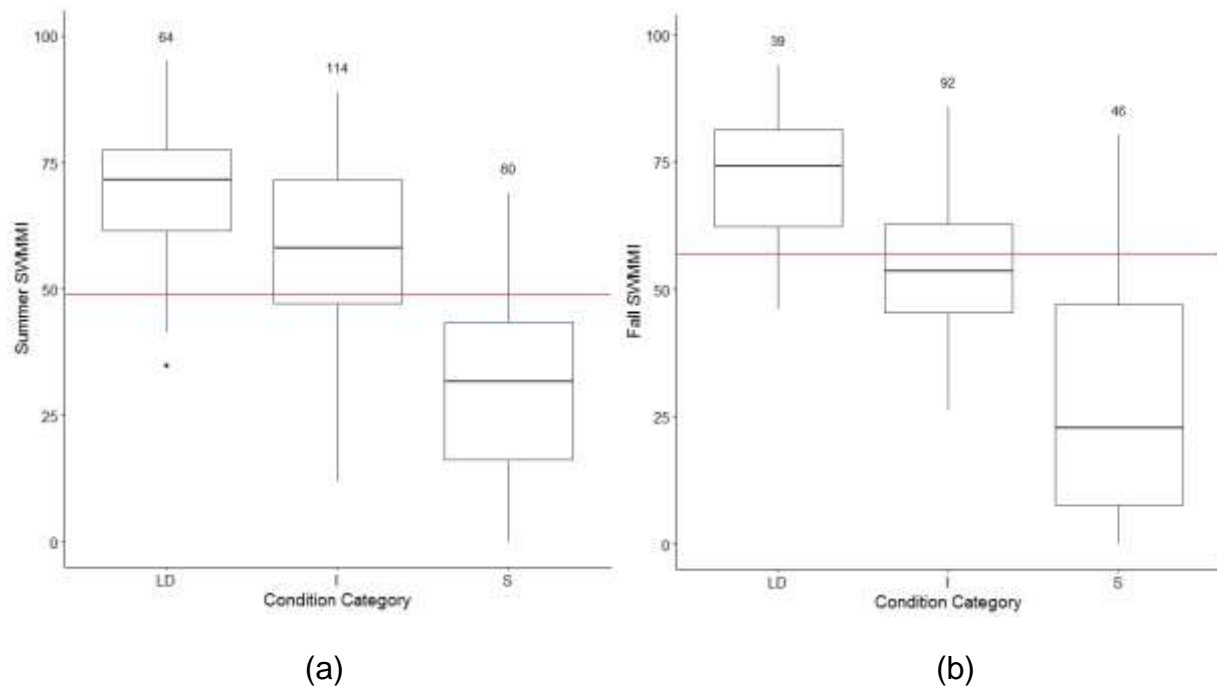
Figure 4. Performance of each SWWMI, where LD indicates least disturbed conditions, I indicates intermediate conditions, and S indicates stressed conditions. Numbers above each boxplot indicate sample size in each condition category. (a) Summer SWMMI calculated from the final combined dataset versus condition category. Red line indicates impairment threshold (49). (b) Fall SWMMI calculated from the final combined dataset versus condition category. Red line indicates impairment threshold (57).

To date, this index development process is the most exhaustive and rigorous that PADEP has completed. This intense effort greatly increases the confidence in using these indices as tools for evaluating impacts to semi-wadeable large river water quality. The sampling method also provides the opportunity to source track major issues in large semi-wadeable rivers while making Aquatic Life Use assessment determinations simultaneously. This greatly increases work efficiency and reduces overall cost.

However, one unique aspect to this sampling method and assessment tool is that it can create multiple assessment points across the width of a river with potentially different outcomes. Therefore, how this tool is applied to assessment decisions will also be important. Ideally, assessment decisions resulting from SWMMI scores will understand and compensate for the complexity of the biological communities that exist in these rivers, as well as conform to the conventional reporting format within the Integrated Reports required by the U.S. Environmental Protection Agency (USEPA).

Application for Aquatic Life Use Determination

PADEP spent several years developing and refining the transect collection method, and because of this method, each SWMMI can not only be used to make one assessment determination that is reflective of water quality and biology of a river, but also produce results that retain the unique aspects of water quality variations. The first critical point to understand is that the transect method can produce multiple assessment zones at any given location based on the number of major water influences discovered during transect data collection. Just over 10% of a river's width may seem like an insignificant amount of area to be concerned with, especially if the remaining portion of the river suggests Aquatic Life Use attainment; however, the great spatial scale that some of these major influences have must be accounted for. For example, one impaired SWMMI score representing just over 10% of a 1.0 km wide river is approximately 100 m. Yet, due to the nature of semi-wadeable rivers, that measured degradation in the biological community may persist for 105 km. Therefore, the actual scale of impact on that river is 10.5 km^2, which is roughly equivalent to the area of a moderately sized town. Biological impairment of a large region within a river complex can have far reaching and unforeseen consequences; therefore, 10% of a river's width is significant and should not be ignored.

Each SWMMI result is also independently applicable when making Aquatic Life Use determinations. This is based on USEPA guidance, which says that all biological communities PADEP has assessment methods for must be evaluated on a stand-alone basis. Consequently, each SWMMI is functionally equivalent to having two completely different biological assessment tools (e.g., fish MMI and a macroinvertebrate MMI). Therefore, it is not appropriate to average both the summer and fall SWMMI scores to obtain an overall result. It is also not appropriate to favor the results of one SWMMI over the other. Given the potential issues that can arise due to this reality, it must be noted that PADEP will always strive to collect as much information as possible to make the most accurate assessment decisions. However, based on independent applicability, it is also understood that only one SWMMI is required to make an Aquatic Life Use determination for a semi-wadeable river segment.

There will be situations when data resulting from SWMMI scores are not used in making Aquatic Life Use determinations. In fact, PADEP uses macroinvertebrate wadeable methods for several other purposes, including, but not limited to cause and effect surveys and incremental improvement reports. These surveys can collect biological information in areas that are not appropriate for making Aquatic Life Use determinations. All Aquatic Life Use assessments on semi-wadeable rivers should acknowledge the upstream to downstream scale that each macroinvertebrate sample represents. If a macroinvertebrate sample is determined to be more representative of a local scale impact, then compliance or corrective actions may be more appropriate than making an Aquatic Life Use impairment determination. Additionally, SWMMI temporal precision estimates may also be used to evaluate whether conditions are degrading or improving at a given site over time. The purpose of studies like this are to provide valuable information about incremental change in the biological community and may not target locations appropriate for making Aquatic Life Use determinations.

Ultimately, the combination of the each SWWMI impairment threshold and the assessment framework described above will create a better representation of overall conditions in a river. It will also more appropriately address ecological concerns in critical habitats for many important species. Understanding and accounting for major water influences in impairment decisions introduces the understanding of continuity between factors causing impairment (i.e., impaired major tributaries, and local scale impacts) and the greater riverine ecosystem at a regulatory level, which has been lacking in the past.

Chapter 1. Field Method Assessment

Introduction

Benthic macroinvertebrates have a long-standing tradition of being used as indicators of water quality in wadeable streams (Hilsenhoff 1987, Barbour et al. 1999, Astin 2007, Chalfant 2012). However, the use of benthic macroinvertebrate bioassessment tools in large non-wadeable rivers has become more common in recent years (Royer et al. 2001, Applegate et al. 2007, Wessell et al. 2008, Blocksom and Johnson 2009, Weigel and Dimick 2011). Large non-wadeable rivers likely require passive methods that use artificial substrates (Cairns 1982, De Pauw et al. 1986) or a combination of collection methods to represent the community with sufficient accuracy and precision (Flotemersch et al. 2006). These modifications are required due to the great size, variable discharges, and increased depth of large rivers. Due to the challenges associated with large rivers, it is argued that river characteristics decrease the ability to characterize aquatic communities using wadeable collection methods (Bartsch et al. 1998, Poulton et al. 2003, Blocksom and Flotemersch 2005). Additionally, many studies suggest the collection of physical and chemical data to inform macroinvertebrate collections is valid in large rivers (Applegate et al. 2007, Blocksom and Johnson 2009, King et al. 2014, PADEP 2014). Stream size alone may not be the best indicator of when field methods should transition from wadeable to non-wadeable. This transition is not well-defined, but previous recommendations include separating streams with drainages greater than 2,600 km^2 (Ohio EPA 1989) or 6th order streams and larger (Johnson et al. 1995). More detailed definitions exist (Wessell et al. 2008), however, without the influence of impoundments, this is most likely a continuous rather than a discrete transition (Vannote et al. 1980) with a high degree of variability depending on the individual river.

The Pennsylvania Department of Environmental Protection (PADEP) has developed several macroinvertebrate assessment methodologies to account for the large diversity of wadeable stream types that exist within the Commonwealth of Pennsylvania (here after referred to as the Commonwealth, PADEP 2007, Botts 2009, Chalfant 2012). These methods have proven effective at discriminating between good and poor water quality for much of the 138,400 stream kilometers in Pennsylvania; however, there are limitations to using these methods for the remaining streams, which are predominantly large rivers (drainage areas > 2,600 km^2). Approximately half of the large rivers (≈ 3,000 stream kilometers) could be considered semi-wadeable in the Commonwealth. For this work, semi-wadeable rivers are defined as predominantly free-flowing systems with drainage areas ≥ 2,600 km^2, and physical characteristics that allow for riffle and run sections to occur with relative frequency. These river systems tend to lack a well-defined and navigable U-shaped channel for any significant distance and frequently

present difficulties for both wadeable and non-wadeable field methodologies, although both methods have been used in large rivers for the Commonwealth (Chalfant 2012, Guild et al. 2014, King et al. 2014, PADEP 2014). The Susquehanna River is an excellent example of this situation. Even at a drainage size of 60,000 km^2, the Susquehanna River can be sampled using wadeable methodologies targeting a single habitat – riffles and runs – under varying flow conditions with the assistance of boats or kayaks and prior knowledge of the reach. Additionally, many stream kilometers outside of Pennsylvania, including rivers such as the Potomac and Allegheny Rivers could fall into this semi-wadeable category.

The current wadeable method developed by Chalfant (2012) composites 6 one-minute kicks in a single habitat using a D-frame net within a 100 m reach, and then sub-samples the collected macroinvertebrates down to 200 (± 40) organisms in the laboratory (6D-200). This method was initially applied to large rivers because, given the habitat complexity, a single habitat collection method could potentially reduce habitat variability, thereby facilitating observed biological response to water quality stressors. However, during preliminary assessment of this wadeable method in large rivers, which included 29 semi-wadeable large river sites from PADEP's Water Quality Network (WQN, PADEP 2012), it was determined that approximately half of those sites had index of biotic integrity (IBI) score coefficients of variation (CV) over recommended limits (10-15% CV, Stribling et al. 2008). Numerous sites with high and low variability and with no apparent links to water quality or natural factors suggested systemic error rather than error attributed to a drainage or group of collectors. Upon further analysis, it was determined that these samples had one common theme; they were typically collected along the shoreline of the river, because collectors did not have the equipment necessary to spread out the 6 one-minute kicks across the entire width of these rivers. During quality assurance reviews, it was confirmed that shoreline collections were inconsistently applied at certain sites, which then resulted in increased index variability. These issues demonstrated the need for clarification of macroinvertebrate collection methods on large rivers.

For PADEP, one of the most important goals when creating field collection methods is to ensure the method captures holistic water quality conditions rather than focusing on minor impacts that may not be representative. Several studies suggest that direct application of a wadeable benthic macroinvertebrate method along the shoreline of a large river may be acceptable (Merritt et al. 2005, Angradi 2006). However, previous work by PADEP (2014) suggested that shoreline habitats were consistently influenced by water quality from minor upstream tributaries and/or point and nonpoint source discharges for many kilometers. Consequently, if the goal of collecting macroinvertebrates in rivers is to assess water quality of the river holistically, then shoreline collection methods may not be sufficient, especially in semi-wadeable rivers.

Furthermore, it has been well documented, that major tributary influences in semi-wadeable rivers do not mix across the width for many kilometers downstream of the respective confluence (Guild et al. 2014, PADEP 2014, Shull and Pulket 2015). For example, the West Branch Susquehanna River and upper mainstem Susquehanna River influences do not mix from the confluence at Sunbury, PA to York Haven, PA (approximately 110 km). This is easily evidenced through aerial photography at Rockville, PA (Figure 1-1). To analyze and confirm these observations, PADEP (2014) began collecting field chemistry data at regular intervals across the width of the river at many large river locations, including Rockville, PA (Figure 1-2). PADEP (2014) documented that spring and summer specific conductance (µS/cm) at the Susquehanna River at Rockville, PA was, on average, approximately 15% higher in the Juniata River influence compared to the West Branch influence, which completely separated the Juniata and upper mainstem Susquehanna Rivers. The upper mainstem Susquehanna River specific conductance was, on average, approximately 11% higher when compared to the West Branch influence. Additionally, the upper mainstem Susquehanna River influence was often more turbid than the West Branch even during base flow conditions, which created a relatively distinct visual separation. However, specific conductance signals relating to the three major tributaries were far more mixed and difficult to delimit approximately 45 km downstream at Marietta, PA (Figure 1-3). Percent differences in specific conductance across the width of the river were reduced to approximately 5% across the width of the river. The reduction in specific conductance differences (mixing of river influences) between Rockville, PA and Marietta, PA was attributed to the large and somewhat unique impoundment near York Haven, PA (Figure 1-4).

Figure 1-1. Aerial photography of the Susquehanna River at Rockville, PA. Vertical red lines indicate significant water quality transition points across the width of the river, which separate the river into 3 major influences.
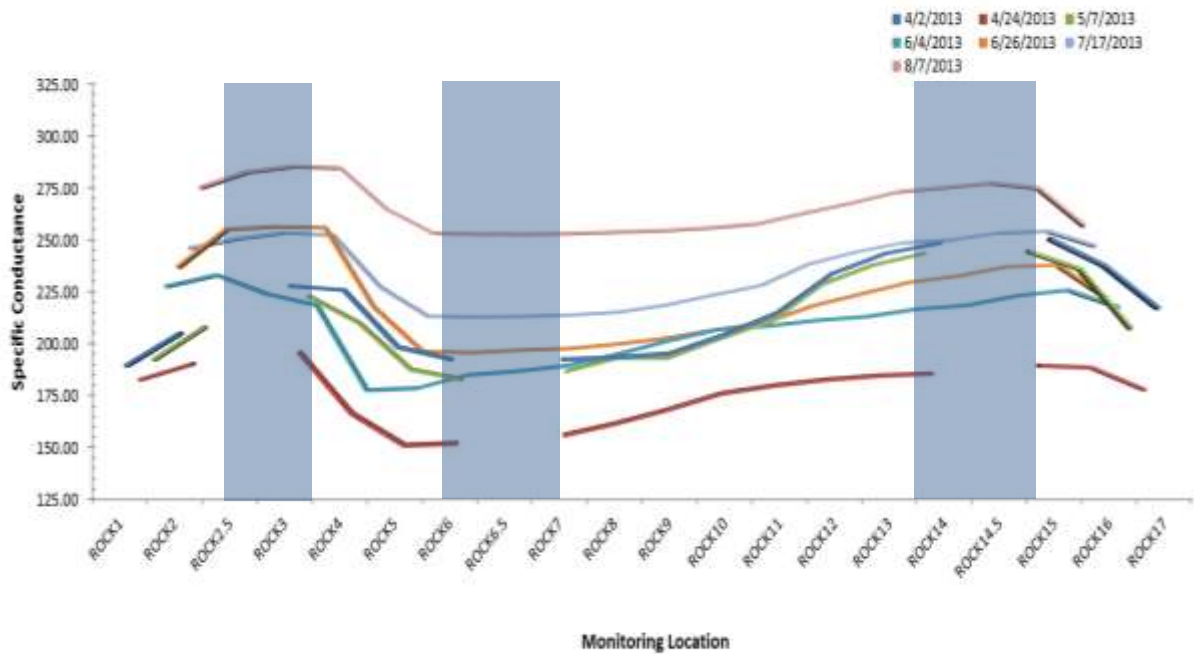
Figure 1-2. Modified from Shull and Pulket (2015). Water quality transect data on the Susquehanna River at Rockville, PA through the 2013 study. Blue rectangles represent the spatial area where separate macroinvertebrate samples were collected in each water influence. The Juniata River water quality is associated with the area around ROCK2.5, West Branch Susquehanna River with ROCK6.5, and upper mainstem with ROCK14.5.
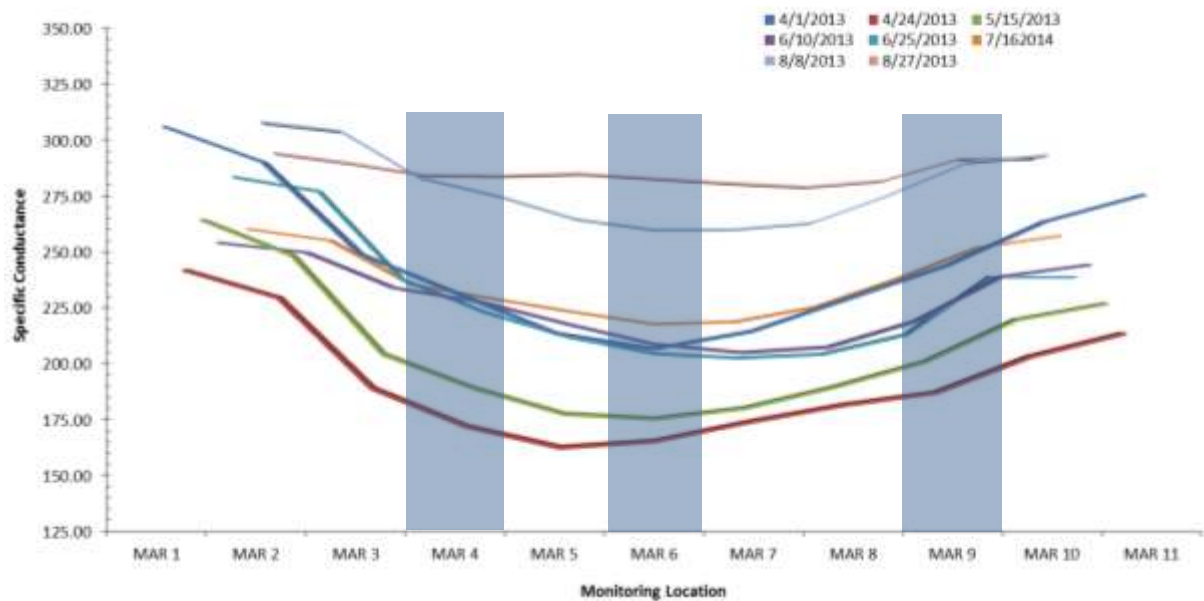
Figure 1-3. Modified figure from Shull and Pulket (2015). Water quality transect data on the Susquehanna River at Marietta, PA through the 2013 study. Blue rectangles represent the spatial areas where separate macroinvertebrate samples were collected in this reach of the river.

Figure 1-4. Imagery of the impoundment at York Haven, PA illustrating the constriction and mixing of the water quality influences between Rockville and Marietta. This impoundment is somewhat unique in that it not only pools the water, which allows for mixing, but it forces the release of water unevenly in the lateral (across the width of the river) scale.

Given the complexities associated with water influences across the width of these large semi-wadeable rivers, in 2012, PADEP began increasing the number of macroinvertebrate samples collected across the width of rivers. One implemented approach was termed the "transect method". The transect method was a modification of the wadeable 6D-200 method where a certain number of 6D-200 samples collected were dictated by the number of significant water quality differences across the width of a river (Figure 1-5). For instance, at Rockville, PA this resulted in three separate 6D-200 samples to represent each of the three major influences. One objective of this study was to determine if this method modification depicts expected changes along the longitudinal gradient (upstream to downstream), without being confounded by the additional samples collected across the width of the river. This was an important first step in determining if the method can serve water quality assessment goals in large semi-wadeable rivers. Another goal was to determine whether major tributary influences

consistently influence macroinvertebrate community composition at a single location. If tributary influences do not greatly or consistently influence the macroinvertebrate community, then simply increasing sampling effort in semi-wadeable rivers may be sufficient. Nonetheless, the critical goal of this study was to determine if high macroinvertebrate index variability previously observed can be reduced to acceptable levels when the transect method is used. Understanding these aspects in large semi-wadeable rivers is important as large river methods continue to develop in the future.
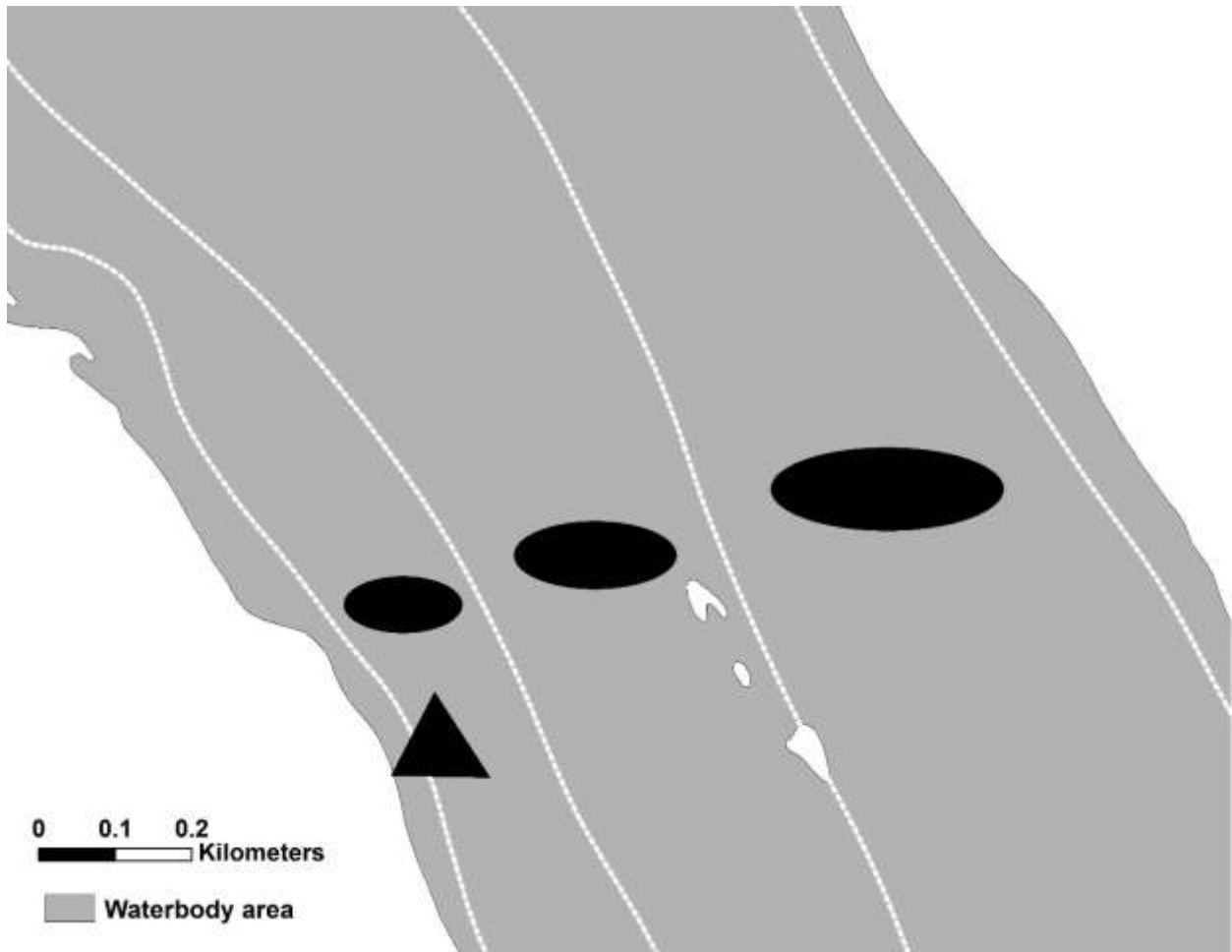
Figure 1-5. Implementation of the 6D-200 methods on a large semi-wadeable river. Vertical white lines within the waterbody area indicate water influence transitions across the width of the waterbody. These lines separate the waterbody area into five sections representing two minor influences along the left and right margins, and three major influences in the center of the waterbody. The triangle designates the 6D-200 collection area for original wadeable collections, where all 6 kicks were taken along the shoreline. Ovals indicate 6D-200 collection areas that conform to the transect method, where all 6 kicks are composited within each major influence.

Methods

The Susquehanna River was selected as the focus of this collection method analysis because it represents the most complex and largest semi-wadeable river system in the Commonwealth (Figure 1-6). Hence, there should be greater chance to observe and account for potential differences in macroinvertebrate communities. Results from analysis in the Susquehanna River should also provide evidence for a method that successfully reduces macroinvertebrate index variability in all semi-wadeable rivers. The Susquehanna River watershed drains approximately 71,000 km^2. This river originates from Otsego Lake in Cooperstown, New York and flows south through Pennsylvania and Maryland to become the major source of fresh water for the Chesapeake Bay (Brown et al. 2005).
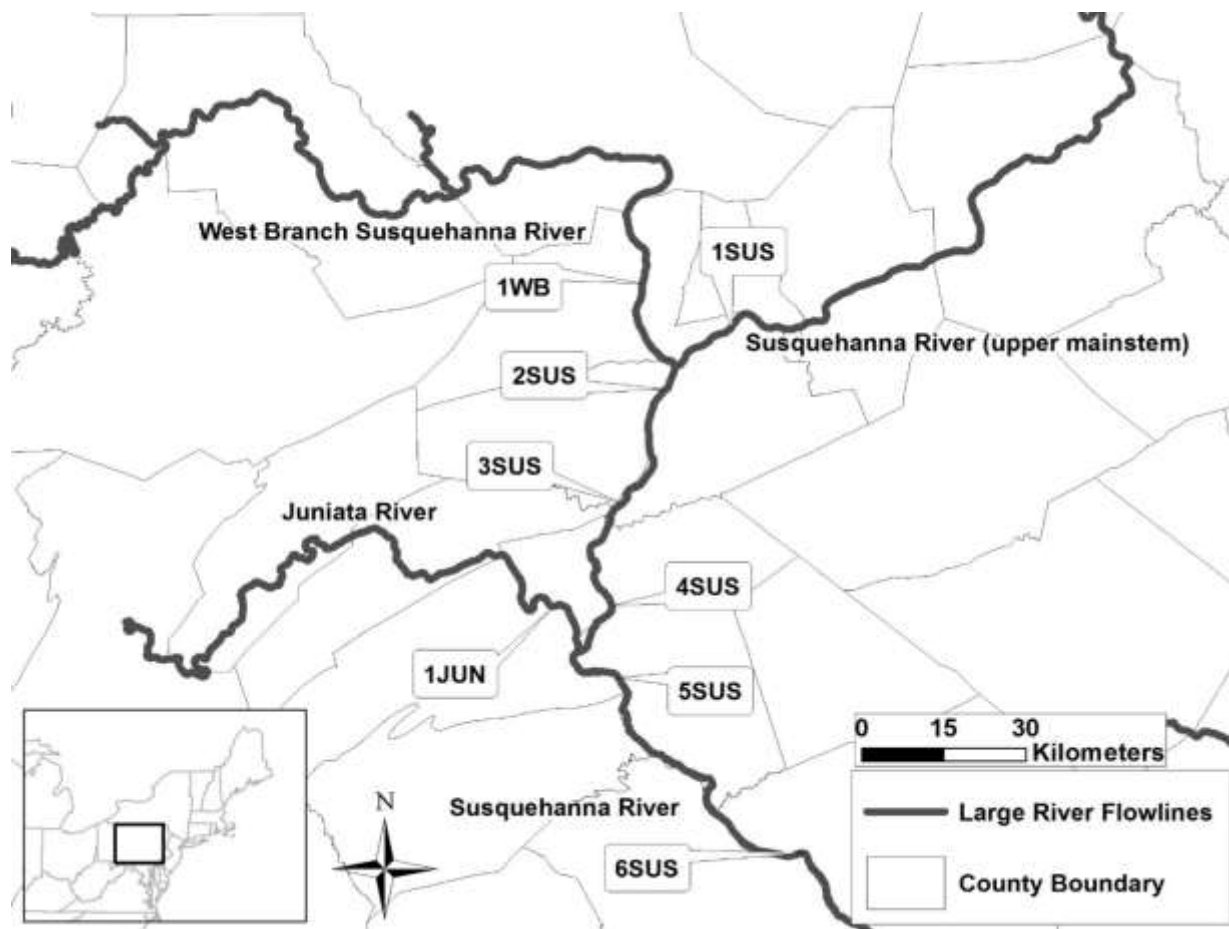
Figure 1-6. Map of the Susquehanna River study area. Site codes were: 1WB = West Branch Susquehanna River at Milton, PA, 1SUS = Susquehanna River at Danville, PA, 2SUS = Susquehanna River at Sunbury, PA, 3SUS = Susquehanna River at Browns Island, 4SUS = Susquehanna River at Clemson Island, 5SUS = Susquehanna River at Rockville, PA, 6SUS = Susquehanna River at Marietta, PA, and 1JUN = Juniata River at Newport, PA.

9

At each site, transects of discrete measurements were conducted to delimit water quality differences across the width of the river. These transects consisted of multiple discrete measurements taken at regular intervals (distance between each point was approximately 5 to 10% of the wetted width) across the width of the river. Discrete measurements were collected using YSI® 6-series sondes as field meters. Sondes were maintained and calibrated using standard PADEP protocols (Shull and Lookenbill 2015). Parameters collected by the field meters were water temperature (°C), pH, dissolved oxygen (mg/L), specific conductance (µS/cm), and turbidity (NTU). These parameters were used to delineate significant water quality differences resulting from the non-mixing of major and minor influences (constituent differences of 10% or greater were typically considered significant). Influences that were wider than 10% of the wetted width were considered major influences.

There were 64 macroinvertebrate samples collected within the Susquehanna River and major tributaries between 2012 and 2016, which were used for transect method analysis. Samples were collected in both summer (August through September) and fall (November through December). A programmatic decision was made to avoid winter and spring sampling due to typically higher flow conditions. All macroinvertebrate samples were collected using 6D-200 field collection and laboratory subsampling procedures detailed in PADEP protocols (Chalfant 2012). Briefly, the original wadeable macroinvertebrate collection protocol consisted of 6 one-minute kicks evenly distributed across the entire width of the waterbody using a 500 µm mesh D-frame net. With the transect method modification, an increased number of 6D-200 samples were collected in accordance with the number of major water quality influences discovered during water quality transect data collection. For instance, if three distinctly different water influences were discovered, then three 6D-200 samples were collected, with each sample composited across the entire width of the respective water influence. Water chemistry grab samples corresponding to each macroinvertebrate sample were collected in the middle of each major influence at mid-depth. Boats and kayaks were often used to navigate to targeted reaches and were used to navigate between individual D-net collections.

In addition to macroinvertebrate and water chemistry collections, PADEP's version of the rapid physical habitat assessment initially developed by and modified from Plafkin et al. (1989), who originally created nine habitat assessment parameters divided into three scoring ranges of 20-0, 15-0, and 10-0. Subsequent modifications added one more habitat parameter to each of the three original categories, bringing the total parameters to 12. The scoring ranges for each parameter were also increased to 20-0. This habitat protocol has undergone several more iterations (Barbour et al, 1999), but PADEP currently uses the 12 criteria and 20-point scoring habitat assessment method (Chalfant 2012). For this analysis, however, only the instream parameters instream cover,

epifaunal substrate, and embeddedness were evaluated, because they evaluate stream conditions in the immediate vicinity of the benthic macroinvertebrate sampling point. Instream cover evaluates the percent makeup of the substrate (boulders, cobble, other rock material) and submerged objects (logs, undercut banks) that provide refuge for fish. Epifaunal Substrate evaluates riffle quality, i.e. areal extent relative to stream width and dominant substrate materials that are present. Embeddedness estimates the percent (vertical depth) of the substrate interstitial spaces filled with fine sediments. These three instream habitat measurements were summed to provide a possible range of 0 (indicating worst possible instream conditions) to 60 (indicating best possible instream conditions) points at each sampling site. Evaluation of instream parameters was restricted to the area where the macroinvertebrate sample was taken.

Data were analyzed using R version 3.1.2 software (R Core Team 2015). Using the R package, vegan, non-metric multidimensional scaling (NMDS) illustrated separations between the macroinvertebrate communities (Oksanen et al. 2016). For NMDS analysis biological data were transformed using Wisconsin square root transformation. Bray-Curtis dissimilarity distance was used to measure community significance and plot sample/species points on the NMDS plots (Bray & Curtis 1957). The R package ggplot2, was used for other graphical illustrations (Wickham 2009). A non-parametric analysis of similarity (the adonis function in the vegan package) was used to test statistical significance and magnitude of the observed macroinvertebrate community dissimilarity (McArdle and Anderson 2001).

A large scale intersite analysis using NMDS was conducted to determine if this method modification can show expected changes in the macroinvertebrate community along the longitudinal gradient. This analysis included 52 samples from the following sites: West Branch Susquehanna River at Milton, PA (1WB), Susquehanna River at Danville, PA (1SUS), Susquehanna River at Sunbury, PA (2SUS), Susquehanna River at Browns Island (3SUS), Susquehanna River at Clemson Island (4SUS), Susquehanna River at Rockville, PA (5SUS), and Juniata River at Newport, PA (1JUN). Multiple macroinvertebrate samples collected using the transect method were analyzed individually in the NMDS, but plotted based on the location they were sampled to simplify the legend. Environmental factors were log base-10 transformed and then overlaid to evaluate biological community and abiotic relationships.

Community composition using NMDS was also evaluated on an intrasite basis for the Sunbury, PA (2SUS) and Rockville, PA (5SUS) sites separately. Both sites had enough repeat samples collected consistently through time and space, thus reducing the potential for longitudinal differences to confound results. Each sample site at Rockville, PA (three sample sites in total) was specifically positioned within the Juniata River (5SUS_JUN), West Branch Susquehanna River (5SUS_WB), and mainstem Susquehanna River (5SUS_MS) influences based off the transect data. There were

also two sample sites at Sunbury, PA, which corresponded to the water quality differences between the West Branch Susquehanna River (2SUS_WB) and the Mainstem Susquehanna River (2SUS_MS). A comparison between the two sites was conducted to see if the transect method can consistently show expected community differences related to water quality. To validate the concept behind the transect method, NMDS on 12 macroinvertebrate samples collected at Marietta, PA (6SUS) was also completed. Due to the mixing of water influences at Marietta, PA, it was hypothesized that the macroinvertebrate community would fail to show a community pattern in the absence of distinct water influence differences. Consequently, multiple repeat samples were collected consistently within three evenly spaced areas across the width of the river. Since macroinvertebrate samples were not sampled within any distinct water influence, the sites were simply coded as 6SUS_1, 6SUS_2, and 6SUS_3 from west to east respectively.

Analyses were then expanded to multiple drainages across the Commonwealth to determine whether sample variability was successfully reduced when using the transect method. Where collections using the original wadeable method and the new transect method existed, variability was compared. Variability analysis was conducted using %CV of index of biotic integrity scores – index described by Chalfant (2012) – among multiple samples collected at the same site over time, using the same method. Ideally, this analysis would have been conducted using same-day duplicate samples to obtain a more direct measure of method variability; however, there were not enough data available for that analysis. Thus, this variability analysis incorporated both method variability and temporal variability. However, temporal variability was assumed to be relatively reduced because sample collections were within similar time periods for each year and all collections only spanned a few years with few to no major weather incidents or known anthropogenic changes. Sites that had greater than 10% CV during preliminary PADEP method analysis were specifically included in this analysis to determine the maximum potential variability reduction produced by using the transect method. The CV for each site/method combination was then averaged to obtain an overall measure of performance. Since the transect method created two or three sites where only one site existed using the original method, transect method CVs were averaged across all sites at the same location to obtain a single measure with which to compare to the original method. In addition to CV comparisons between the two methods, the transect method CV was compared to drainage size for all rivers where data were available to ensure method variability was not increasing with river size.

Results

Large Scale Analysis

Large scale intersite community analysis suggested that the upper Susquehanna macroinvertebrate communities (i.e., 1WB, 1SUS, and 2SUS) were different from the middle and lower Susquehanna communities (Figure 1-7a). Dissimilarity based on this large-scale site classification was significant and explained a moderate amount of separation (p= 0.001, R^2 = 0.43). When potential environmental factors were overlaid with the NMDS plot, macroinvertebrate communities appeared to have a strong relationship with habitat, total sulfate, total chloride, total nitrogen, total phosphorous, and season (Figure 1-7b). Total ammonia and specific conductance had weak relationship with the macroinvertebrate community. Nutrient parameters and instream habitat had a positive relationship with the middle and lower Susquehanna sites, while total sulfate and chloride showed a positive relationship with the upper Susquehanna sites. The Juniata River sites had the smallest drainage size, but grouped with the lower Susquehanna sites (generally the largest drainage size sites), suggesting a lack of a drainage size influence on these communities.
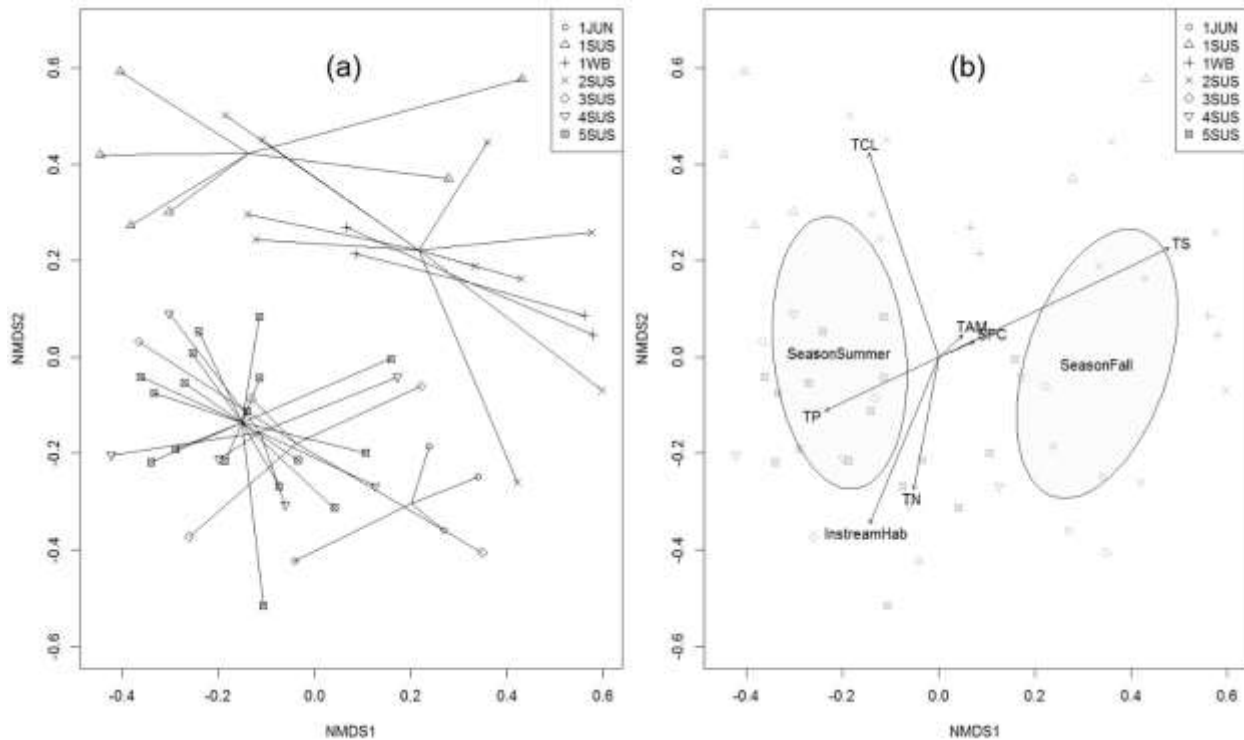
Figure 1-7. (a) NMDS (dimensions = 2, stress = 0.2) of Susquehanna River macroinvertebrate samples collected using the transect method. Each symbol represents a single 6D-200 sample, and samples include those from both summer and fall seasonal classifications over the course of several years. Sample sizes were 1SUS

(n = 6), 2SUS (n = 8), 3SUS (n = 5), 4SUS (n = 5), 5SUS (n = 14), 1WB (n = 4), and 1JUN (n = 4). (b) Susquehanna River macroinvertebrate samples with environmental factors were overlaid to show the perpendicular shift of samples compared to seasonal spread. Environmental codes were: TN = Total Nitrogen, TP = total phosphorous, TCL = total chloride, SPC = specific conductance, TS = total sulfate, TAM = total ammonia, and InstreamHab = the sum of instream cover, embeddedness, and epifaunal substrate scores. Ellipses represent one standard deviation of points in each group.

Intrasite Analysis

When longitudinal scales were reduced to only transect method samples collected at Sunbury, PA (2SUS), results suggest there are two consistently distinct communities between 2SUS_WB and 2SUS_MS (Figure 1-8a). An analysis of similarity based on site and season classification was performed to determine the magnitude and significance of these influences at the intrasite level. Analysis of similarity results for this classification were statistically significant (p = 0.006), and accounted for a large portion of community dissimilarity (R^2 = 0.59). Average instream habitat scores were 35 and 37 for 2SUS_WB and 2SUS_MS respectively, indicating somewhat mediocre, yet consistent results between sites at Sunbury, PA.

When longitudinal scales were reduced to only transect method samples collected at Rockville, PA (5SUS), results suggest there are three consistently distinct macroinvertebrate communities across the width of the river (Figure 1-8b). An analysis of similarity using the site and seasonal classification was performed to determine the magnitude and significance of these combined influences at the intrasite level. Analysis of similarity results for this classification were statistically significant (p = 0.001), and accounted for a large portion of community dissimilarity (R^2 = 0.64). Instream physical habitat parameters were also evaluated at each sampling site to confirm differences in community were not confounded by habitat differentials. The average instream habitat scores were slightly improved from the 2SUS sites and changed little between 5SUS_JUN (44), 5SUS_WB (40), and 5SUS_MS (40).
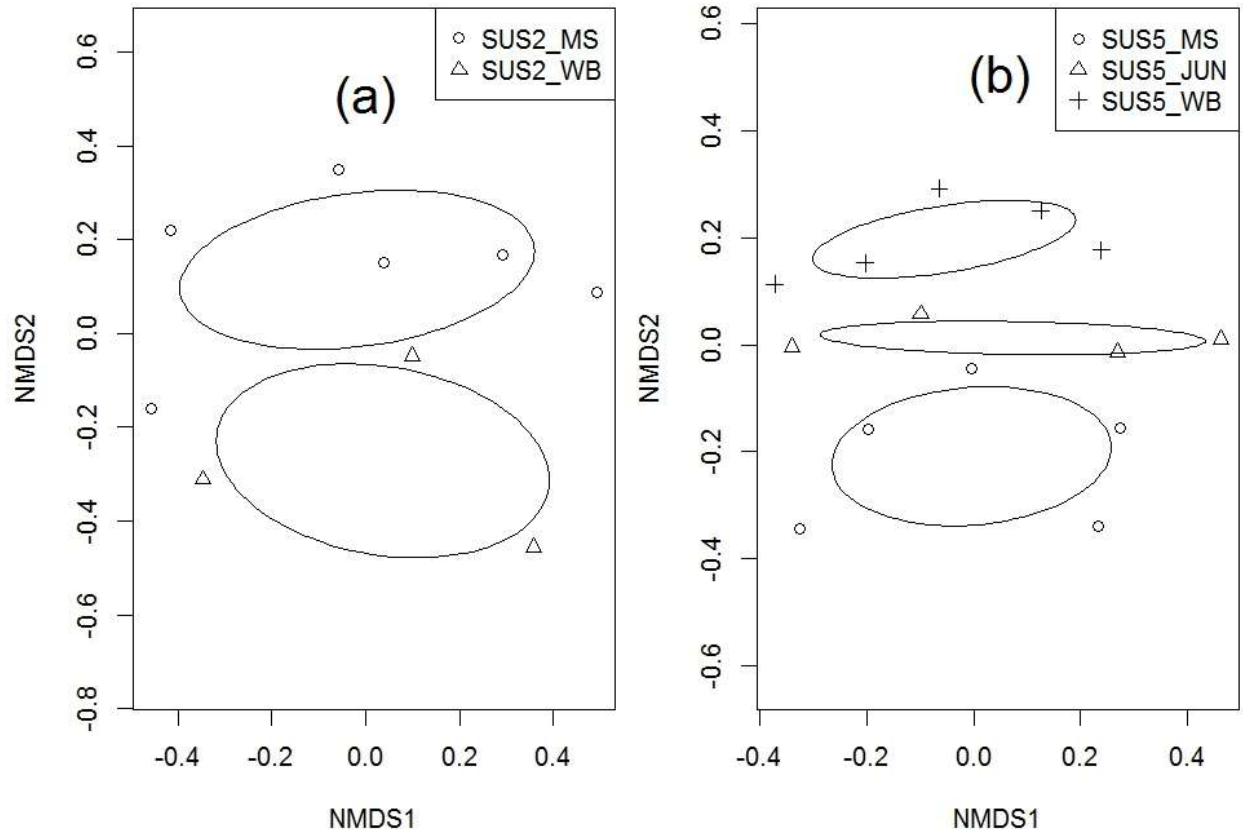
Figure 1-8. (a) NMDS (dimensions = 2, stress = 0.1) of macroinvertebrate samples collected at Sunbury, PA based on the transect method, where water quality transects suggest unique and consistently different influences. Sample sizes were 2SUS_MS (n=6) and 2SUS_WB (n=3). Ellipses represent the standard deviation of points in each group. (b) NMDS (dimensions = 2, stress = 0.2) of macroinvertebrate samples collected at Rockville, PA based on the transect method, where water quality transects suggest unique and consistently different influences. Sample sizes were 5SUS_JUN (n=4), 5SUS_MS (n=5), and 5SUS_WB (n=5). Ellipses represent one standard deviation of points in each group.

Contrary to the results at Sunbury, PA and Rockville, PA, analysis of samples collected repeatedly within the same lateral zone at Marietta, PA (6SUS) suggests communities lack distinction between the three sites across the width of the river (Figure 1-9). The combination of site and season into one classification was not statistically significant (p = 0.33). However, the seasonality classification alone remained significant at the Marietta sites (p = 0.004, R^2 = 0.32). Instream habitat scores did not differ more than 5 points at the Marietta sites (i.e., 6SUS_1 = 44, 6SUS_2 = 39, and 6SUS_3 = 44), which suggested that instream habitat was likely not affecting macroinvertebrate communities at these locations.
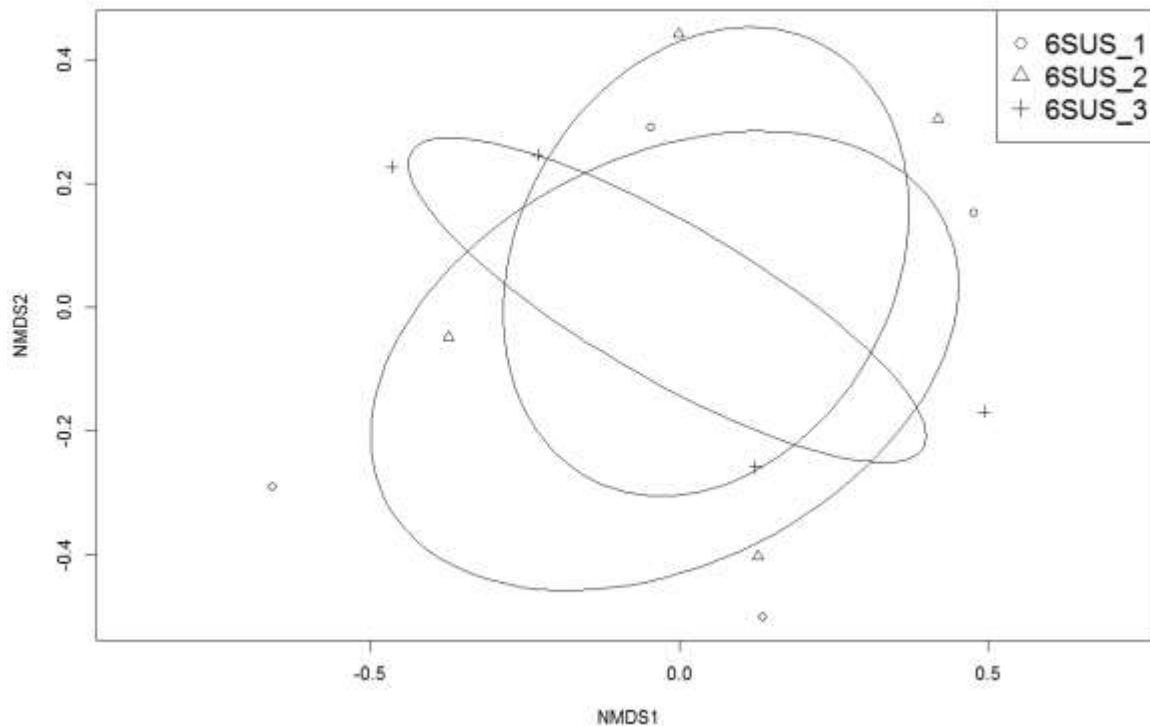
Figure 1-9. NMDS (dimensions = 2, stress = 0.1) of macroinvertebrate samples collected consistently at the same three sites across the width of the river through time at Marietta, PA. Sample sizes were 6SUS_1 (n=4), 6SUS_2 (n=4), and 6SUS_3 (n=4). Ellipses represent one standard deviation of points in each group.

Variability Analysis

For method variability comparisons, there were 5 sites that had samples collected using both the original wadeable method and transect method. These included sites on the lower Delaware River, upper mainstem Susquehanna River, middle Susquehanna River, West Branch Susquehanna River, and French Creek (tributary to the Allegheny River). The transect method successfully decreased sample collection variability at each site when compared to the original method (Table 1-1). On average and among all sites, utilization of the transect method reduced method variability by approximately 8.5%, and in some cases, observed variability was reduced by approximately 10 to 20%. Average transect method CV from all 5 sites was 11%, indicating acceptable variation when compared to recommended levels, even with the inclusion of temporal variability in the analysis. It is important to note that these variability decreases were observed even with the inclusion of some spatial variability when multiple transect samples spanning a river were averaged into single site. Additionally, transect method CV did not increase with increasing drainage size (p = 0.57, Figure 1-10).

16

Table 1-1. Comparison of IBI score coefficients of variation (%CV) between the wadeable method and the transect method at sites where both datasets were available. The original wadeable method collections were conducted between 2010 and 2013, whereas the transect method collections were conducted between 2012 and 2014. Each site/method had 2 samples with which to calculate %CV and compare, except for the Susquehanna River at Danville that had 4 transect method samples and the Susquehanna River at Milton, PA that had 3 original wadeable method samples.

River	Location Description	Original Wadeable Method CV	Transect Method CV
Delaware River	Morrisville, PA	29.5%	9.5%
Susquehanna River	Sunbury, PA	15.6%	11.0%
Susquehanna River	Danville, PA	18.3%	13.6%
West Branch Susquehanna River	Milton, PA	16.2%	9.2%
French Creek	Franklin, PA	21.7%	12.1%

Figure 1-10. Temporal coefficient of variability (CV) of repeat samples collected using the transect method at the same site compared to stream size (log base 10 scale).

Discussion

Large scale results show changes in macroinvertebrate communities along the longitudinal gradient, with relatively strong relationships with changes in water quality due to human impacts. Yet, natural forces such as habitat may still be driving

17

macroinvertebrate community results (e.g., increased habitat scores resulting from higher gradient and more bedrock outcroppings in the lower portions of the Susquehanna River than the upper portion). Results also show a strong divergence in macroinvertebrate community relating to season. Seasonal community differences were clear when October sampling was not conducted for transect method samples. If season truly is a major natural factor that drives macroinvertebrate community differences, then October collections could be problematic for water quality assessment purposes. Chalfant (2012) described October as a transitional period for which best professional judgment should be applied when making water quality assessment decisions. Therefore, exclusion of samples collected during this period should allow for more accurate water quality assessments in the future. Another major result of this work was, when analyses were reduced to single sites, major tributary influences showed a significant effect on the biological community. With these results, it becomes clear that a better understanding of the water quality and biology across the entire width of the river will also be required.

Both the macroinvertebrate community and variability analyses suggest that the 6D-200 may be suitable for large semi-wadeable rivers once major influences are accounted for and samples are collected across the entire width of a river. The NMDS analyses supported the utility of the transect method because it clearly showed differences in season, distinguished between sites on longitudinal scales, and picked up on targeted water quality influences. It is critical to note that these distinctions and relationships were observed even with sampling error being inherently incorporated into the analyses. The sampling error inherent within these results included different field collectors, microhabitat selection variability, temporal variability across several years, and laboratory error. This increased ability to observe biological community preferences in accordance with water quality is most likely tied to the selection of a single habitat collection method combined with strict quality assurance procedures. Most importantly, however, this study demonstrates that the transect method can substantially reduce IBI score variability at sites across several major drainages. This suggests that the transect method would be useful for more than just the Susquehanna River.

Understanding major water quality influences across the width of a large semi-wadeable river and using them for source-tracking impacts should be a goal of biological assessments in the future. These results support the conclusions and recommendations of others that more robust methods are necessary for larger systems (Blocksom and Flotemersch 2005, Guild et al. 2014, King et al. 2014). However, distinction between what is wadeable versus what is truly non-wadeable is important from a monitoring resource standpoint. The results of this study support expansion of a modified 6D-200 wadeable method to large semi-wadeable rivers. The transect method modification does not greatly increase the amount of time and effort required to collect samples

compared to some methods that require passive collection and site revisits to obtain samples. However, it does require an increase in infrastructure and collector evaluation of the site. Boats, kayaks, and canoes were often essential to navigate large distances, find suitable locations, and sample targeted habitat. This is especially true under even slightly elevated discharge conditions, or on high gradient reaches. Many targeted reaches in this study were not accessible via foot from shore. As a result, the modified 6D-200 method requires some physical stamina and working experience of the complex flow dynamics in large waters, both in and while exiting watercraft. Exiting a watercraft in the middle of a 1.5 km wide river can be a humbling experience, so individuals should be able to collect samples with resolve. In addition, the D-framed net is particularly advantageous for sampling large semi-wadeable rivers, because it produces significantly less drag than commonly used – and larger – kick nets.

For samples collected in semi-wadeable watersheds greater than 2600 km^2, it is recommended that at least one water quality transect consisting of multiple discrete readings at intervals of 5 to 10% of the wetted width be collected using a hand-held field instrument before collecting macroinvertebrate samples. The four common physical and chemical parameters of water temperature, pH, specific conductance, and dissolved oxygen are typically sufficient to determine differential influences; however, if possible, measuring turbidity can also indicate important water quality differences. Percent differences of 10% or greater for any parameter may be an indication that an additional 6D-200 sample is required. Differentiating between major and minor influences is also important; therefore, influences that span more than 10% of total width should be characterized with an additional 6D-200 sample. If only one transect can be conducted before macroinvertebrate sampling takes place, it should be completed at base flow conditions to ensure macroinvertebrates are being collected in the correct lateral area of the river. However, additional transects can refine the site selection process. This method should not only facilitate assessments of water quality as a whole, but also provide the necessary information to track sources of major impacts.

This study supports other work that suggests macroinvertebrate communities in larger systems may be driven by variable water quality across the width of a river (King et al. 2014, PADEP 2014). The West Branch Susquehanna River and the Upper Mainstem Susquehanna River tend to show significant differences in concentration and particle size of suspended solids (King et al. 2014). King et al. (2014) observed distinct relationships between filter feeder percent dominance and the concentrations of total suspended solids and Chlorophyll *a* between these two rivers. It is probable that the results of this study are detecting these very same differences, but detailed taxonomic and ecological analyses were not within the scope of this study. Most importantly though, these findings support a position that collections of either water quality or macroinvertebrates in an arbitrary fashion (without understanding the dynamics of these

river systems) can be problematic depending on the goal of the study. For instance, some work has promoted the collection of macroinvertebrate samples along shoreline habitats (Lazorchak et al. 2000, Angradi 2006), yet this work as well as Guild et al. (2014) reported an increase in methodological variability when collection is limited to shoreline habitats. Small tributary influences or point source discharges between sites may have a significant impact on near shore macroinvertebrate communities, even over large distances. Therefore, samples taken along the shore of a semi-wadeable river may be more of a reflection of the closest upstream influence or point source discharge rather than the targeted waterbody, regardless of whether both shorelines are sampled. Consequently, if the goal of macroinvertebrate collections is to measure water quality of a river holistically, then perhaps sampling across the width is more advantageous (Collier et al. 2013, King et al. 2014).

It could be argued that applying a single habitat collection method may significantly reduce the number of available sampling locations throughout a river system. However, the Delaware River Basin Commission has applied a similar modified wadeable method to the Delaware River mainstem from Callicoon, NY downstream to Trenton, NJ (Sildorff and Limbeck 2009). Sildorff and Limbeck (2009) were able to collect approximately 160 samples from 25 unique sites along the mainstem Delaware River, demonstrating additional support for a modified wadeable method in semi-wadeable rivers. It has also been suggested that larger rivers experience less holistic change longitudinally compared to small streams due to local-scale impacts having less weight than the sum of all upstream influences at the point of interest (Blocksom and Johnson 2009). Consequently, this concept would not support the argument for having a high density of sample points along the longitudinal gradient. Additionally, as the geographic size that a macroinvertebrate sample represents increases, it may be appropriate to expand assessment methods to more than one large river (Sildorff and Limbeck 2009) and even beyond government boundaries. Blocksom and Johnson (2009) discussed the utility of expanding a field method and subsequent multimetric index to more than one drainage. The goal would be to capture a higher variation of stressor gradients in order to build a more acceptable assessment method. The semi-wadeable transect method has already been used in the majority of large rivers throughout Pennsylvania, including, but not limited to the Delaware, Susquehanna, Allegheny, and Youghiogheny Rivers. Thus, this method shows promise for successful multimetric index development across a wide spectrum of drainages.

As demonstrated by the results of this study, the transect method modification may acceptably reduce method variability when collecting macroinvertebrates in large rivers for water quality assessments. However, this method may not provide sufficient information for a detailed quantitative macroinvertebrate community analysis in semi-wadeable river systems, because it remains a semi-quantitative collection method.

These results also do not recommend a maximum number of macroinvertebrate samples across the width of a river, but rather, a minimum. Special circumstances may require an additional number of samples to properly describe the waterbody of interest. For example, regarding habitat variability, even though this study used a single habitat approach to standardize habitat, it does not suggest significant habitat differences should be ignored. It is possible that habitat difference across the width would require more macroinvertebrate data to be collected. Results of this study suggest that large river macroinvertebrate sampling requires at least some understanding of major influences upstream and some increase in overall sampling effort. If these additional data are not collected and understood within the context of the river system as a whole, then it is difficult to recognize why macroinvertebrate community differences exist or even to use macroinvertebrates as a tool to measure water quality.

Literature Cited

Applegate, J. M., P. C. Baumann, E. B. Emery, and M. W. Wooten. 2007. First steps in developing a multimetric macroinvertebrate index for the Ohio River. River Research and Applications 23:683–697.

Angradi, T. R. (editor) 2006. Environmental monitoring and assessment program: Great River ecosystems field operations manual. EPA/620/R-06/002. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

Astin, L. E. 2007. Developing biological indicators from diverse data: The Potomac Basin-wide Index of Benthic Integrity (B-IBI). Ecological Indicators 7:895–908.

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

Bartsch L. A., W. B. Richardson, and T. J. Naimo. 1998. Sampling benthic macroinvertebrates in a large floodplain river: considerations of study design, sample size, and cost. Environmental Monitoring and Assessment 52:425–439.

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

Blocksom K. A., and J. E. Flotemersch. 2005. Comparison of macroinvertebrate sampling methods for non-wadeable streams. Environmental Monitoring and Assessment 102: 243–262.

Botts, W. 2009. An Index of Biological Integrity (IBI) for "True" Limestone Streams. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Reg

ulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/An%20Index%20of%20Biological%20Integrity-Limestone%20Streams.pdf

Bray J. R., and J. T. Curtis 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecological Monographies 27:325–349.

Brown, J.J., J. Perillo, T.J. Kwak, and R.J. Horowitz. 2005. Implications of *Pylodictis olivaris* (Flathead Catfish) introduction into the Delaware and Susquehanna drainages. Northeastern Naturalist 12:473–484.

Cairns, J. Jr. (editor) 1982. Artificial Substrates. Ann Arbor Science Publications, Inc. Ann Arbor, Michigan, 279 pp.

Chalfant, B. 2012. A benthic index of biotic integrity for wadeable freestone streams in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Technical%20Documentation/freestoneIBImarch2012.pdf

Collier, K. J., J. E. Clapcott, I. C. Duggan, D. P. Hamilton, M. Hamer, and R. G. Young. 2013. Spatial variation of structural and functional indicators in a large New Zealand river. River Research and Applications 29:1277–1290.

De Pauw, N., D. Roels, and A. P. Fontoura. 1986. Use of artificial substrates for standardized sampling of macroinvertebrates in the assessment of water quality by the Belgian Biotic Index. Hydrobiologia 133:237–258.

Flotemersch, J. E., K. A. Blocksom, J. J. Hutchens Jr., and B. C. Autrey. 2006. Development of a Standardized Large River Bioassessment Protocol (LR-BP) for Macroinvertebrate Assemblages. River Research and Applications 22:775–790.

Guild, K., A. Anthony, M. Bilger, and J. Holt. 2014. Assessment of Passive and Active Macroinvertebrate collection methods in adjacent reaches on the upper main stem of the Susquehanna River. Journal of the Pennsylvania Academy of Science 88:47–56.

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

Johnson, B. L., W. B. Richardson, and T. J. Naimo. 1995. Past, present, and future concepts in large river ecology: How rivers function and how human activities influence river processes. Bioscience 45: 134–141.

King, N. R., M. E. Mctammany, M. J. Wilson, J. C. Chakany, H. N. Coffin, and M. E. Reilly. 2014. Variability in Macroinvertebrate Communities of the Susquehanna River in Central Pennsylvania. Journal of the Pennsylvania Academy of Science 88:67–75.

Lazorchak, J. M., B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (editors). 2000. Environmental Monitoring and Assessment Program -Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-

Wadeable Rivers and Streams U.S. Environmental Protection Agency, Cincinnati OH.

Loucks, D. P. 2003. Managing America's rivers: who is doing it? International Journal of River Basin Management 1: 23–31.

McArdle, B. H. and M. J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology. 82: 290–297.

Merritt, R. W, J. D. Allan, K. W. Cummins, K. J. Wessell, and J. O. Wilhelm. 2005. Qualitative biological and habitat protocols for Michigan's non-wadeable rivers. Michigan Department of Environmental Quality, Lansing, Michigan.

Ohio EPA. 1989. Biological criteria for the protection of aquatic life. Vol. III. Standardized field sampling and laboratory methods for assessing fish and macroinvertebrate communities. Columbus, OH, Ohio EPA, Division of Water Quality Monitoring and Assessment.

Oksanen, J., G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, H. H. Stevens, and H. Wagner. 2016. vegan: Community Ecology Package. R package version 2.3–4. http://CRAN.Rproject.org/package=vegan

PADEP. 2007. Pennsylvania DEP Multihabitat Stream Assessment Protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/Multihabitat%20Protocol.pdf

PADEP. 2012. Pennsylvania's Surface Water Quality Monitoring Network (WQN). Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://www.elibrary.dep.state.pa.us/dsweb/Get/Document-91443/3800-BK-DEP0636.pdf

PADEP. 2014. 2012–13 Susquehanna River Sampling and Assessment Report. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Water%20Quality%20Standards/Susquehanna_Report_2013_FinalDraft.pdf

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

Poulton B. C., M. L. Wildharber, C. S. Charbonneau, J. F. Fairchild, B. C. Mueller, and C. J. Schmitt. 2003. A longitudinal assessment of the aquatic macroinvertebrate

community in the channelized lower Missouri River. Environmental Monitoring and Assessment 85: 23–53.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at http://www.R-project.org/.

Royer, T. V., C. T Robinson, and G. W. Minshall. 2001. Development of macroinvertebrate-based index for bioassessment of Idaho Rivers. Environmental Management. 27:627–636.

Silldorff, E. L., and R. L. Limbeck. 2009. Interim Methodology for Bioassessment of the Delaware River for the DRBC 2010 Integrated Assessment. [Draft]. http://www.nj.gov/drbc/library/documents/10IntegratedList/Bioassessment-draft-July2009rev.pdf

Shull, D. R., and M. J. Lookenbill. 2015. Continuous Instream Monitoring Protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/CIM_PROTOCOL.pdf

Shull, D. R., and M. Pulket. 2015. Causal Analysis of the Smallmouth Bass decline in the Susquehanna and Juniata Rivers. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/SusquehannaRiverStudyUpdates/SMB_CADDIS_Report.pdf

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.

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.

Weigel, B. M., and J. J. Dimick. 2011. Development, validation, and application of a macroinvertebrate based index of biotic integrity for nonwadeable river of Wisconsin. Journal of the North American Benthological Society 30: 665–679.

Wessell, K. J., R. W. Merritt, J. G. O. Wilhelm, J. D. Allan, K. W. Cummins, and D. G. Uzarski. 2008. Biological evaluation of Michigan's non-wadeable rivers using macroinvertebrates. Aquatic Ecosystem Health and Management 11: 335–351

Wickham, H. 2009. Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Chapter 2. Determining least disturbed conditions and stream classification

Introduction

One of the most important steps in the development of a multimetric index (MMI) is to evaluate anthropogenic effects on the aquatic ecosystem and to define good versus poor conditions before biological communities are evaluated. In many cases, environmental stressor gradients are built upon varying types of data, which are then developed into condition categories that are used to differentiate between good and poor water quality (Mattson and Angermeier 2007, Weigel and Robertson 2007, Weigel and Dimick 2011, Chalfant 2012). Typically, when water quality is within an acceptable state, the balance and integrity of biological communities is also within an acceptably balanced state (Novotny 2004). Hence, developing the environmental benchmark that describes acceptable water quality serves as the foundation for developing subsequent biological standards. Yet, biological communities on the good side of the environmental benchmark also have natural variation, spatial variation, and sampling error that must be understood and accounted for (Hughes 1995, Stoddard et al. 2006). Therefore, within the upper ranges of good conditions there should also be an analysis of how these natural variations affect the biological community in order to develop a stronger and more reliable assessment tool (Barbour et al. 1999). There are several approaches for establishing the benchmark condition, including comparing least disturbed sites, using best professional judgment, comparing historical conditions, using empirical data, and evaluating ambient conditions. It is probably most useful to draw on as many of these approaches as possible (Hughes 1995, Stoddard et al. 2006).

Large semi-wadeable rivers present several unique challenges when establishing stressor gradient benchmarks. Not only do large rivers experience significant changes over vast longitudinal gradients (Vannote et al. 1980), but they are also tied to a greater frequency of cumulative anthropogenic impacts than smaller streams and rivers (Johnson et al. 1995, Weigel and Dimick, 2011). Traditionally, biological assessments are based on comparing observed conditions to conditions described as natural or undisturbed (Davies and Jackson 2006, Chalfant 2012). These conditions were traditionally defined as "reference condition" (Karr and Chu 1999). However, natural or undisturbed conditions do not exist in large semi-wadeable rivers, where at least some degradation has occurred at all sites. As a result of these issues, Stoddard et al. (2006) recommended several alternative definitions to better describe how comparisons are made between "reference conditions" and anthropogenically stressed conditions. For large semi-wadeable rivers, the term "least disturbed" (LD) may be most appropriate, as

it inherently accepts that current environmental conditions are no longer natural, but also not degraded enough to lose an acceptable range of conditions within the stressor gradient. Frequently, the stressor gradient is established using abiotic parameters (Barbour et al. 1999). Biological information is typically avoided in the development of the environmental stressor gradient, because it introduces circular reasoning and may allow for preconceived notions regarding the biological community before stressors are fully understood (Bailey et al. 2004). However, there may be some cases where biological data are useful in describing LD conditions versus stressed (S) conditions. Examples may include toxic responses in biology due to chemicals that are not readily measured or have interactive effects, or where non-native species have created significant imbalance in community structure (Stoddard et al. 2006).

Several methods for establishing LD conditions exist, each with potential advantages and disadvantages. However, most approaches create a series of criteria which are designed to capture varying types of stressors. The LD criteria should incorporate both physical and chemical aspects as well as local scale and large scale features. For example, dams are physical alterations that create irregular flow and are usually understood to be local scale impacts. Conversely, particularly for large rivers, land use is considered a large-scale factor that is sufficiently measured when looking at the entire watershed (Hunsaker and Levine 1995, Sliva and Williams 2001). All aspects can affect the biological community to varying degrees; therefore, the final stressor gradient should include as many local scale and large scale impacts as possible (Blocksom and Johnson 2009). As a result of the complexities associated with larger rivers, it may be advantageous to build a stressor gradient based on known water quality and physical condition thresholds rather than an approach that arbitrarily selects a percentile from the stressor gradient.

Equally important is the need to understand and classify natural variability of the LD condition (Stoddard et al. 2006). Biological communities often vary naturally across time and space. Causes for these community shifts may include season, geology, gradient, elevation, climate, and other non-anthropogenic factors. Site classification (based on non-anthropogenic factors) within the LD condition can reduce potential assessment errors associated with natural factors. This process can be difficult in large rivers because there is often a relationship between natural classifications – such as geology or ecoregion – and impacts to water quality (Becher and Root 1981, Hurd 2012). For instance, the dolomitic and limestone rich geologies of the Great Valley also hold some of the highest concentrations of agriculture in the Commonwealth (Stranahan 1993). Stressor gradients that fail to exclude sites with agricultural impacts into the LD condition will inevitability and erroneously show regions such as the Great Valley as unique. Additionally, larger rivers likely pass through and pick up characteristics from multiple geologies. Thus, it becomes increasingly difficult to determine from which

geology or ecoregion a site belongs along a river. Although, analysis of these natural variables will ultimately suggest whether they are important to incorporate or not, great care should still be taken to reduce the likelihood that biological relationships based on natural variation are not confounded by correlated anthropogenic stressors. One of the best ways to achieve this goal is to select sites with the least amount of degradation as possible.

The goals of this work are to establish screening criteria for the environmental stressor gradient, categorize sites using this stressor gradient into LD, intermediate (I), and S conditions, and classify sites within the LD condition according to natural variability. This is a straightforward process described and used by many individuals developing multimetric indices across the United States and other countries (Barbour et al. 1999, Klemm et al. 2003, Hering et al. 2006, Baptista et al. 2007, USEPA 2016). Yet, the challenges of working with larger rivers that do not fit the categories of either wadeable or non-wadeable will likely require a mixture of principal and novel strategies. One important and novel (compared to previous MMIs developed by PADEP) goal for this project is the incorporation of a suite of chemical parameters that are known to affect river conditions within the region. This work in conjunction with the extensive macroinvertebrate collection study described in Chapter 1 should result in a stronger and more reliable biological assessment tool.

Methods

Large semi-wadeable rivers within and surrounding the Commonwealth were selected for this study (Figure 2-1). In total, there were 178 unique sites with sufficient data to perform stressor gradient and classification analyses. Sites within the Potomac River basin – not within the Commonwealth – were included to increase sample size and to provide a more regionally appropriate assessment method for other agencies. Site selection and sample collection were consistent with the definition of large, semi-wadeable rivers discussed in Chapter 1 (i.e., drainage area ≥ 2,600 km^2 and wadeable during summer-fall baseflow conditions). Much of the Delaware, Susquehanna, and Potomac Rivers, as well as major tributaries to the Upper Ohio River (including, but not limited to the Allegheny, Beaver, Mahoning, and Youghiogheny Rivers) met these criteria, totaling approximately 3,200 stream kilometers. The upper range of drainage area captured by this method was approximately 63,000 km^2. Between 2010 and 2016, approximately 415 macroinvertebrate samples were collected at these large river sites. In total, this sampling effort covered four major basins and eight Level III ecoregions (Omernik 1987).
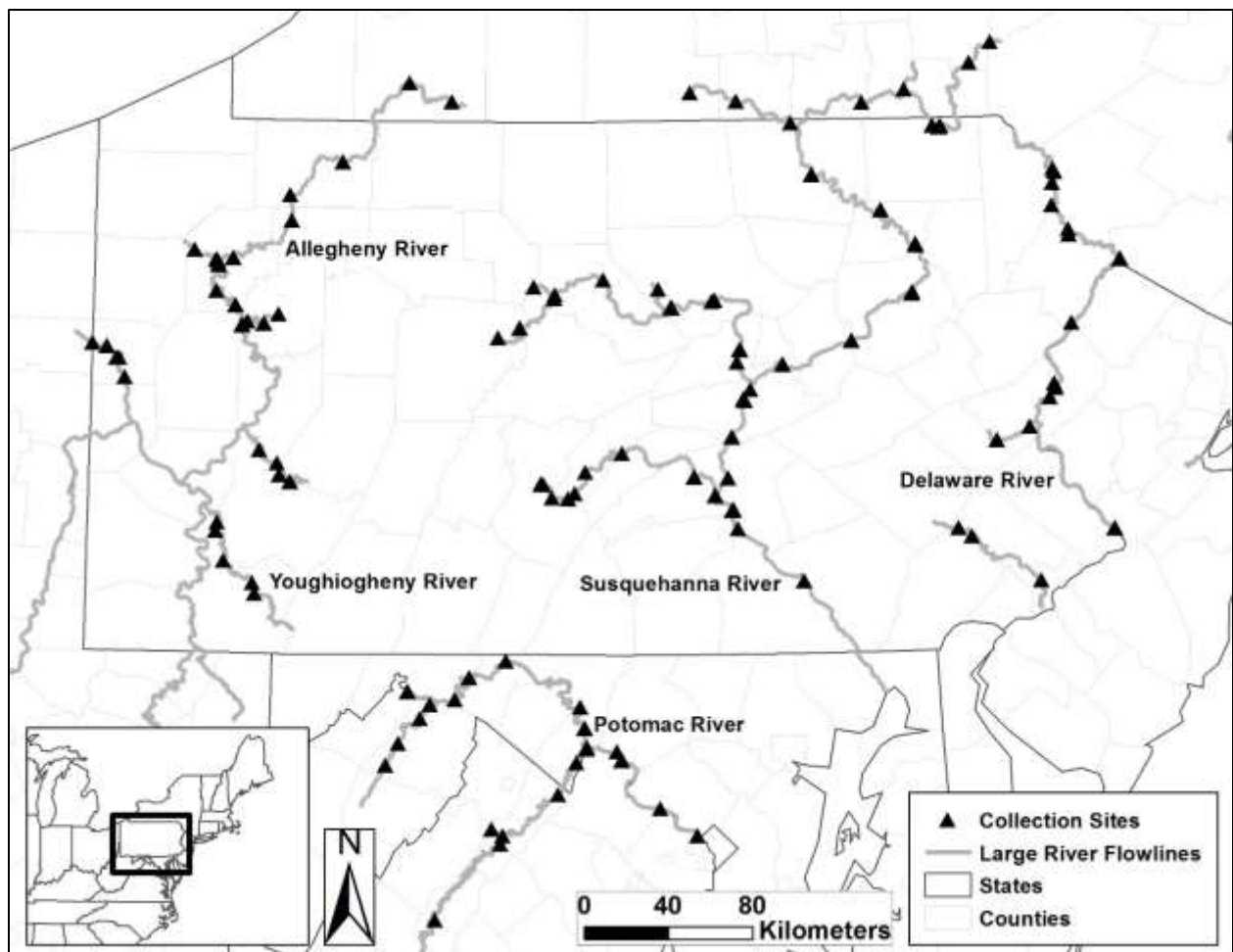
Figure 2-1. Study area and collection sites.

At each site, field data collections consisted of water chemistry, physical habitat, and macroinvertebrate samples. Macroinvertebrate, habitat, and chemistry sampling were conducted using the transect method described in Chapter 1. Quality control measures were implemented to ensure sites with only one macroinvertebrate sample collected was composited across the entire width of the river. Any sample that did not meet these measures was removed from the analyses. In order to reduce the likelihood of autocorrelation, sites were separated on the longitudinal gradient by influences such as dams, towns, and major tributaries. Chemistry grab samples collected at the time of macroinvertebrate collection comprised the majority of the chemical dataset; however, some sites had long records of chemistry analysis, which included both high and low flow conditions. Longer chemistry records were available from the PADEP WQN dataset, but for the purposes of this study, these chemistry records were temporally truncated to the period of 2010 to 2016 (period when biological samples were collected). Due to the relatively large chemical dataset under variable flow conditions, the median value of chemical parameters was used for chemical stressor gradient analysis. During transect data collection, field parameters were collected using PADEP

28

standard operating procedures (Shull and Lookenbill 2015). Habitat parameters were collected using standard PADEP protocols and riffle-run habitat field sheets (PADEP 2013). Instream habitat parameters (instream cover, epifaunal substrate, and embeddedness) were summed into one composite score for each sample collection.

Desktop data collection consisted of watershed delineation at the point of sample collection, land use calculation, and distance calculations to major and minor impoundments. Watershed polygons were created from point data using the ArcGIS Online watershed tool (Scopel 2014). Watershed polygons (containing watershed area data) and the 2011 National Land Cover Database (NLCD, Homer et al. 2015) were then imported into R, version 3.1.2 software (R Core Team 2015) to calculate land use percentages. Watershed percentages of the NLCD land use categories "Developed, Open Space", "Developed, Low Intensity", "Developed, Medium Intensity", and "Developed, High Intensity" were summed into a total percent developed score. Watershed percentages from the NLCD land use categories "Pasture/Hay" and "Cultivated Crops" were summed to obtain a total percent agriculture score. Both composited scores were then used for subsequent stressor gradient analysis. Calculation of major and minor dam distances were completed using a combination of the National Transportation Atlas Database (USDT 2015) and aerial imagery (Esri 2017). Distances were measured from the sample point upstream to the dam. Major dams were defined as structures greater than 15 meters in height and creating some significant divergence from run-of-the-river flow dynamics (i.e., flow peaking, water releases not triggered by natural factors). Minor dams were defined as structures that maintain run-of-the-river flow dynamics and were less than 15 m in height.

A slightly modified version of the criteria based scoring approach – using only abiotic characteristics – was selected to establish a stressor gradient (Roth et al. 1997, Mattson and Andermeier 2007, Botts 2009, Weigel and Dimick 2011, Chalfant 2012). There were a total of 11 abiotic parameters included in the stressor gradient (Table 2-1). Possible scores for each stressor were 0, 1, or 2, where a score of 0 indicated good conditions, 1 indicated fair conditions, and 2 indicated poor conditions. An exception to this scoring was for major and minor dam influence categories where only two scoring possibilities exist: 0 or 2. These parameters covered a large spectrum of anthropogenic impacts including nutrients, ions, land use, instream habitat, and major dams. More specifically, these stressors were chosen, because they are suggested to be leading causes of impaired water quality throughout the United States (USEPA 2016). The composited scoring approach combined many different potential impacts into a more holistic assessment of the respective river site. Individual parameters were also compared to the overall condition categories to obtain a general idea of how each condition parameter performed, and to elucidate the median values that existed within the LD condition. This information is important because it establishes at least some

baseline conditions experienced for the measured period, and may be used in subsequent large river analyses. In addition, Spearman correlations were analyzed on the median chemical values between each site to ensure parameters such as specific conductance and sulfate were not redundant, thereby causing inappropriately inflated condition gradient scores.

Table 2-1. Environmental condition gradient criteria.

Stressor	Good 0	Fair 1	Poor 2
% Developed land use in the watershed	< 5	5 to 10	> 10
% Agriculture land use in the watershed	< 10	10 to 30	> 30
Total Phosphorus (mg/L)	< 0.1	0.1 to 0.2	> 0.2
Total Nitrogen (mg/L)	< 1.0	1.0 to 2.0	> 2.0
Specific Conductance (µS/cm)	< 300	300 to 600	> 600
Total Chloride (mg/L)	< 15	15 to 200	> 200
Total Sulfate (mg/L)	< 20	20 to 200	> 200
Total Ammonia (mg/L)	< 0.03	0.03 to 0.05	> 0.05
Instream Habitat (Total of instream cover, epifaunal substrate, and embeddedness)	> 45	45 to 30	< 30
Within 16 km of a major impoundment			X
Within 1.6 km of a minor impoundment			X

Condition score thresholds for each abiotic parameter were selected based on a combination of literature, empirical data, state and federal Aquatic Life Use (ALU) criteria, quantile analysis, and expert consensus. Land use benchmarks for percent urban and percent agriculture at the watershed scale were consistent across several sources of literature and provided reasonable thresholds for each condition category (Roy et al. 2003, Weigel 2003, Weigel and Dimick 2011, USEPA 2016). There was sufficiently consistent literature for both total nitrogen and total phosphorous thresholds (Weigel and Dimick 2011, USEPA 2016), which generally agreed with empirical data collected within Pennsylvania and expert consensus. For ammonia, the chronic ALU criterion in Table 3 of 25 Pa. Code §93.7(a) was used in conjunction with the understanding that large semi-wadeable rivers frequently reach 30°C and pH levels of 9.0 consistently during the summer (Shull and Pulket 2015). Sulfate and Chloride benchmarks were developed from literature (Elphick et al. 2011), and condition benchmarks from other multimetric index criteria within the region (USEPA 2016). Specific conductance was used as a surrogate for total ion concentration or total dissolved solids. Lower expected benchmarks for specific conductance generally agreed between literature for the region (Cormier et al. 2011, Griffith 2014) and empirical data collected for large rivers within the region. The poor category benchmark

for specific conductance was extrapolated from specific conductance measurements in streams with high carbonate influence within Pennsylvania (Shull et al. 2016), which was used as a theoretical highest potential threshold for large semi-wadeable rivers under somewhat natural conditions. Benchmarks for instream habitat condition were modifications from the condition categories (Optimal, Suboptimal, etc.) provided by Barbour et al. (1999). Major and minor dam benchmarks were developed from a combination of literature (Stevens et al. 1997, Lessard and Hayes 2003), empirical data, and expert consensus.

All stressor scores were summed to obtain an overall stressor gradient score for each site. Possible ranges for the composite stressor gradient scores were between 0 and 22 points, where 0 indicated least stressful conditions and 22 indicated most stressful conditions. Composite stressor gradient scores were then used to categorize sites as either LD, I, or S. For a site to be placed in the LD category, it needed to have an overall stressor gradient score of 3 or less, and no single stressor was permitted to be in the poor category (i.e., single stressor score of 2). An overall stressor gradient score ≥ 7 was defined as a S site. Yet, any site with an overall stressor gradient score of 6 and a single stressor score in the poor category (i.e., score of 2) was also placed in the S condition.

Unless stated otherwise, the R package, vegan, was used throughout macroinvertebrate community analysis (Oksanen et al. 2016). The R package ggplot2 was used for creating boxplots, which illustrated abiotic stressor gradient similarities and differences between sites (Wickham 2009). Scatterplot correlation matrices were created using the R package psych and were used to evaluate whether abiotic stressors were providing unique information to the overall stressor gradient (Revelle 2017). A non-parametric analysis of similarity (adonis in the vegan package) was used to test statistical significance and magnitude of the observed macroinvertebrate community dissimilarity (McArdle and Anderson 2001). Analysis of natural macroinvertebrate community variability was conducted to determine if classifications were required, and if so, how many classifications existed in the dataset. Biological community data were analyzed using R version 3.1.2 software (R Core Team 2015). Rare taxa (taxa occurring in the dataset less than 5% of the time) were removed during macroinvertebrate community analysis. It is common practice to remove rare taxa, because it reduces the likelihood of misclassifying sites and avoids the possibility that extremely rare taxa were the result of misidentification (Hawkins et al. 2000).

Traditionally, agglomerative hierarchical cluster analysis is used to suggest an optimum number of classifications; however, this traditional cluster analysis was avoided because it forces a predetermined number of classifications on the dataset without evidence to suggest how many classifications should exist. It could be argued that there is no correct number of classifications in the environment or within biological

communities, because the world is not made up of discrete classes with clear delineations (Oksanen 2014). With the advent of the R statistical software and the wealth of expertise that is now openly shared, more useful statistical tools can facilitate determination of a reasonable number of biological community classifications based on probability instead of rigid assignments. As an alternative to agglomerative hierarchical cluster analyses, a series of three independent statistics were used as multiple lines of evidence to arrive at the most appropriate number of classifications based on the dataset provided.

The first biological community analysis utilized heat maps (using the "tabasco" function in the R package, vegan, Oksanen et al. 2016) to gain a general idea of patterns in the dataset, identify specific taxa or groups of taxa that were driving observed patterns, and to identify any samples that were obvious outliers. The heat map used detrended correspondence analysis to group similar species together and similar samples together. This provided a visually useful illustration within the heat map. As a result, a great amount of information could be processed visually without losing the ability to identify important details within the dataset. Another tool used for determining the appropriate number of classifications was an alternative cluster analysis described by Ruspini (1970). This cluster analysis is unique in that it creates a probability for each sample belonging to a selected number of classifications rather than forcing a predetermined number of classifications on the dataset. This cluster method was useful for determining several possible numbers of classifications. Clustering was conducted using the R package, cluster (Maechler et al. 2014). The final tool used to assist in the validation of how many classifications should exist based on the dataset was the Calinski-Harabasz criterion (Calinski and Harabasz 1974). The Calinski-Harabasz criterion uses the Euclidean metric after a Hellinger transformation (Rao 1995) to suggest an optimum number of classifications. By using these three tools together in a step-wise process, it was expected that a more confident number of classifications would be elucidated.

Once the number of classifications was determined, NMDS was used to validate the likely causes of community differences. For NMDS analysis, data were transformed using Wisconsin square root transformation. Bray-Curtis dissimilarity distance was used to measure community significance and plot sample/species points on the NMDS plots (Bray and Curtis 1957). Many possible causes of community classifications were evaluated, including season, ecoregion, drainage size, major basin, gradient, thermal regime, latitude, and alkalinity. Samples were collected within and categorized into summer and fall classifications. Summer samples were defined as being collected between late July through September and fall samples were defined as samples collected from November through December. There were no samples collected during the month of October, because previous analysis by Chalfant (2012) suggested October

samples may confound results if a multimetric index was developed using a seasonal classification. Results from Chapter 1 suggesting that seasonality may be a significant driver of macroinvertebrate community supported that stance. Ecoregion classifications were based on level III ecoregions (Omernik 1987). Drainage size category, gradient, thermal regime, and alkalinity data were extracted from the Appalachian Stream Classification dataset developed by The Nature Conservancy and funded by the U.S. Fish and Wildlife Service Appalachian Landscape Conservation Cooperative (Olivero Sheldon 2015). All classifications were used as described by the Appalachian Stream Classification. For instance, the Appalachian Stream Classification groups stream size based on medium mainstem rivers (2,600 to 13,000 km^2), large rivers (13,000 to 26,000 km^2), and great rivers (>26,000 km^2). Finally, latitude was refined down to a categorical level corresponding to the whole degree unit where the sample was collected. For example, samples collected within this region of the country were collected roughly between the 40th and 42nd north parallels, which created three categories corresponding to each parallel.

Results

Out of 178 total sites sampled, the environmental stressor gradient analysis identified 48 LD sites, 77 I sites, and 53 S sites. Final condition scores had a moderate amount of variation with sites scoring between 0 and 14 out of a possible 22 stress points. The scoring method clearly separated LD and S sites through abiotic parameters (Figure 2-2). All major drainages were captured within the LD condition. Specifically, LD sites included portions of the Delaware, West Branch Susquehanna, upper mainstem Susquehanna, Potomac, and Allegheny Rivers. In addition, LD sites were not relegated to only the smallest drainage sizes, with many sites exceeding 13,000 km^2 and approaching 26,000 km^2. Within the LD condition, median values for ionic concentration (measured as specific conductance), total chloride, and total sulfate were 167 µS/cm, 15.4 mg/l, and 8.4 mg/l, respectively (Figure 2-3). The median LD values within the nutrient parameters total ammonia, total nitrogen, and total phosphorous were approximately 0.02 mg/l, 0.4 mg/l, and 0.02 mg/l, respectively (Figure 2-4). Median values within the LD condition for the physical and landscape parameters such as percent developed, percent agriculture, and instream habitat score were 4.6%, 14.4%, and 46, respectively (Figure 2-5). There were a few sites that had a total condition score of three or below that were not placed in the LD condition. Many of these sites were automatically rejected from LD condition because of the proximity to major and minor impoundments that placed the site in the poor category for those stressors. The median score for I sites was 5 points, with many sites scoring in the fair category for most stressors. Sites within the S condition had a median stressor gradient score of 8, with generally high levels of nutrients (particularly total nitrogen), ions, and urban land uses.

Some of the highest scoring sites in the final stressor gradient were on rivers such as the Lehigh, Schuylkill, Beaver, Mahoning, and Shenango Rivers.

All individual parameters showed good to moderate separation between LD and S sites. This was particularly true for specific conductance, total sulfate, total nitrogen, and percent developed stressors, which showed the greatest amount of separation between the two groups. Stressors such as total phosphorous and instream habitat had slightly less separation between the two groups. To ensure chemical parameters were not redundant correlation analysis was conducted. Spearman correlation analysis suggested that each parameter evaluated was contributing unique stressor information to the overall score (Figure 2-6). Total sulfate and specific conductance had the highest correlations ($r = 0.74$); however, one parameter would only be able to explain approximately 55% of the other parameter.
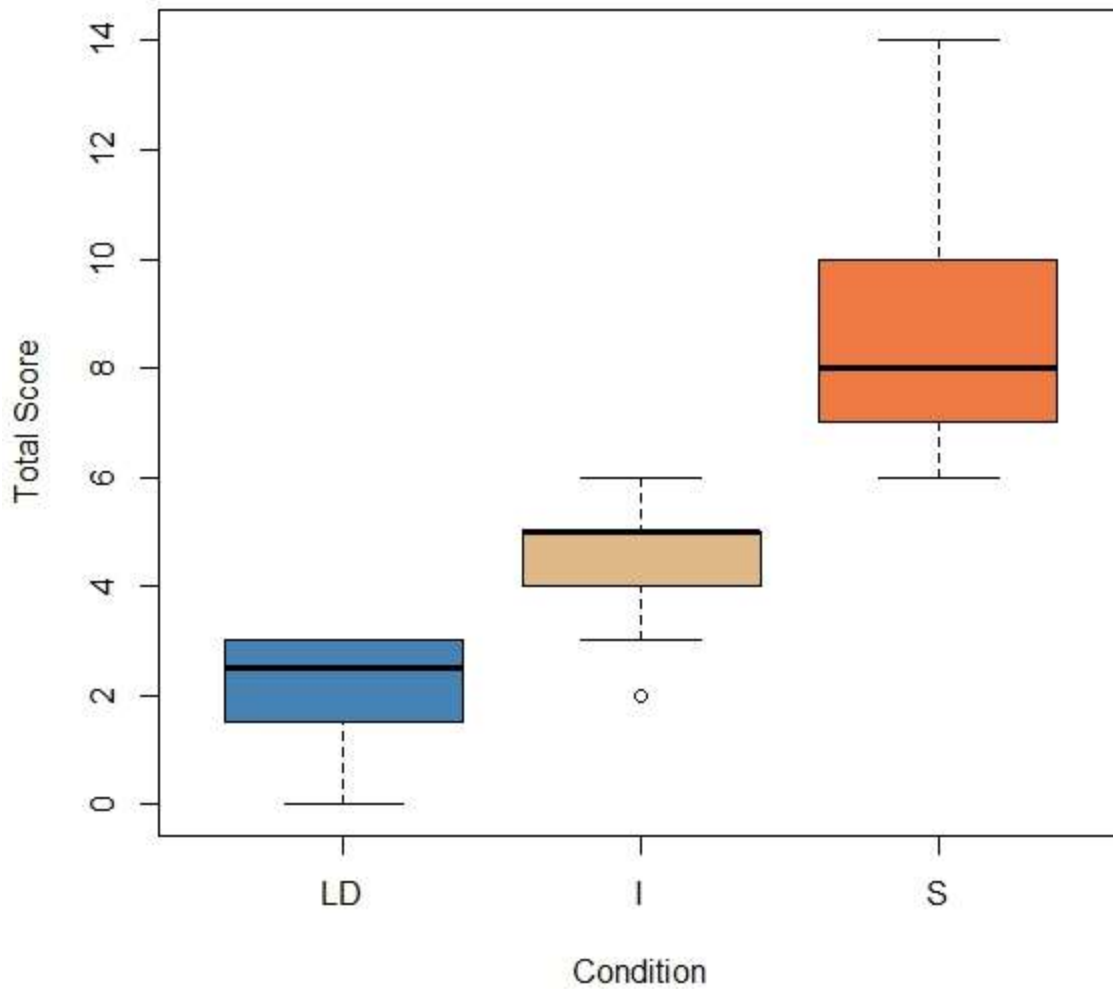
Figure 2-2 Least disturbed (LD, n=48), intermediate (I, n=77), and stressed (S, n=53) classifications with total scores for each site, illustrating clear separation between LD and S sites.
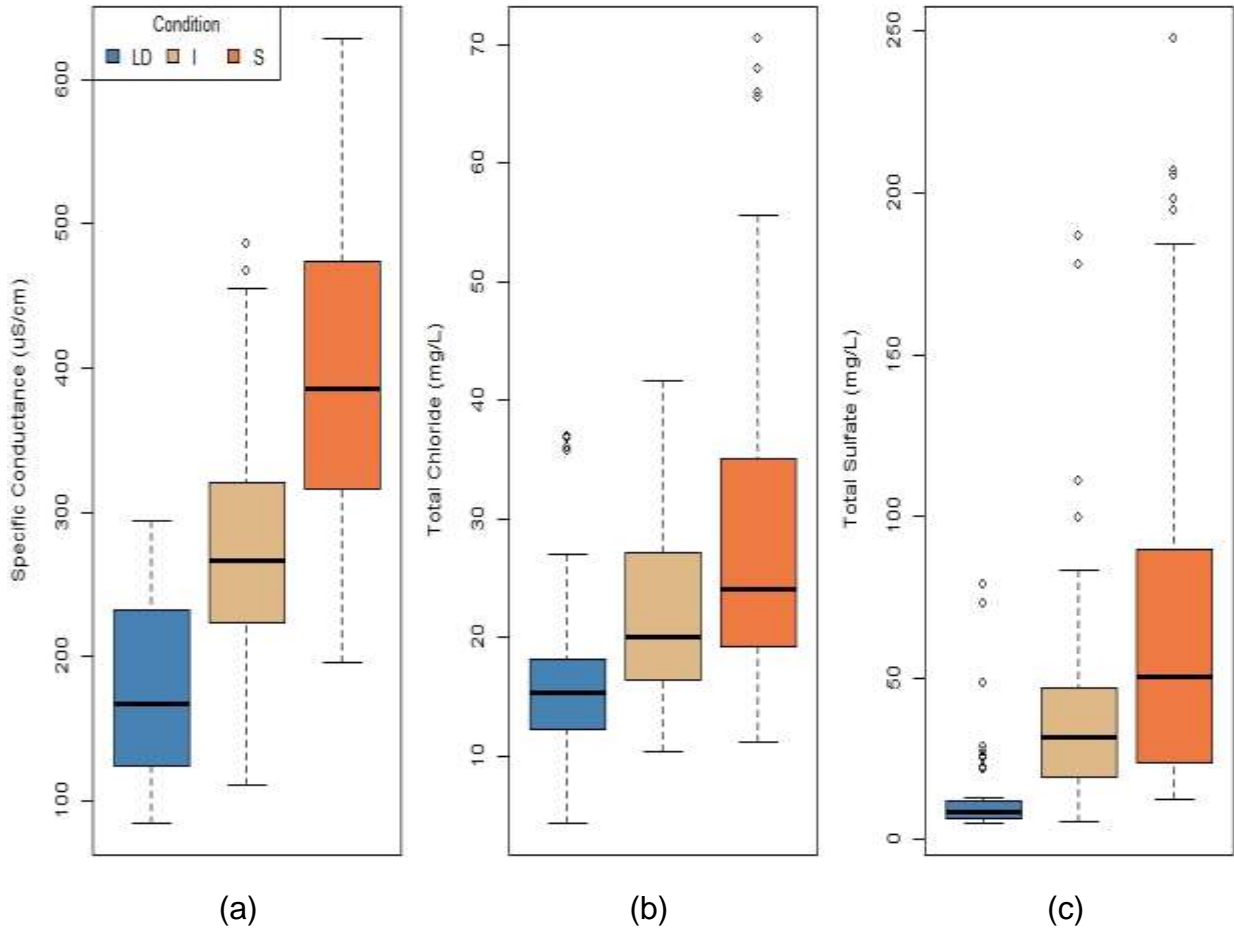
Figure 2-3 Stressor gradient condition classification compared to specific conductance (a), total chloride (b), and total sulfate (c).

(a)   (b)   (c)
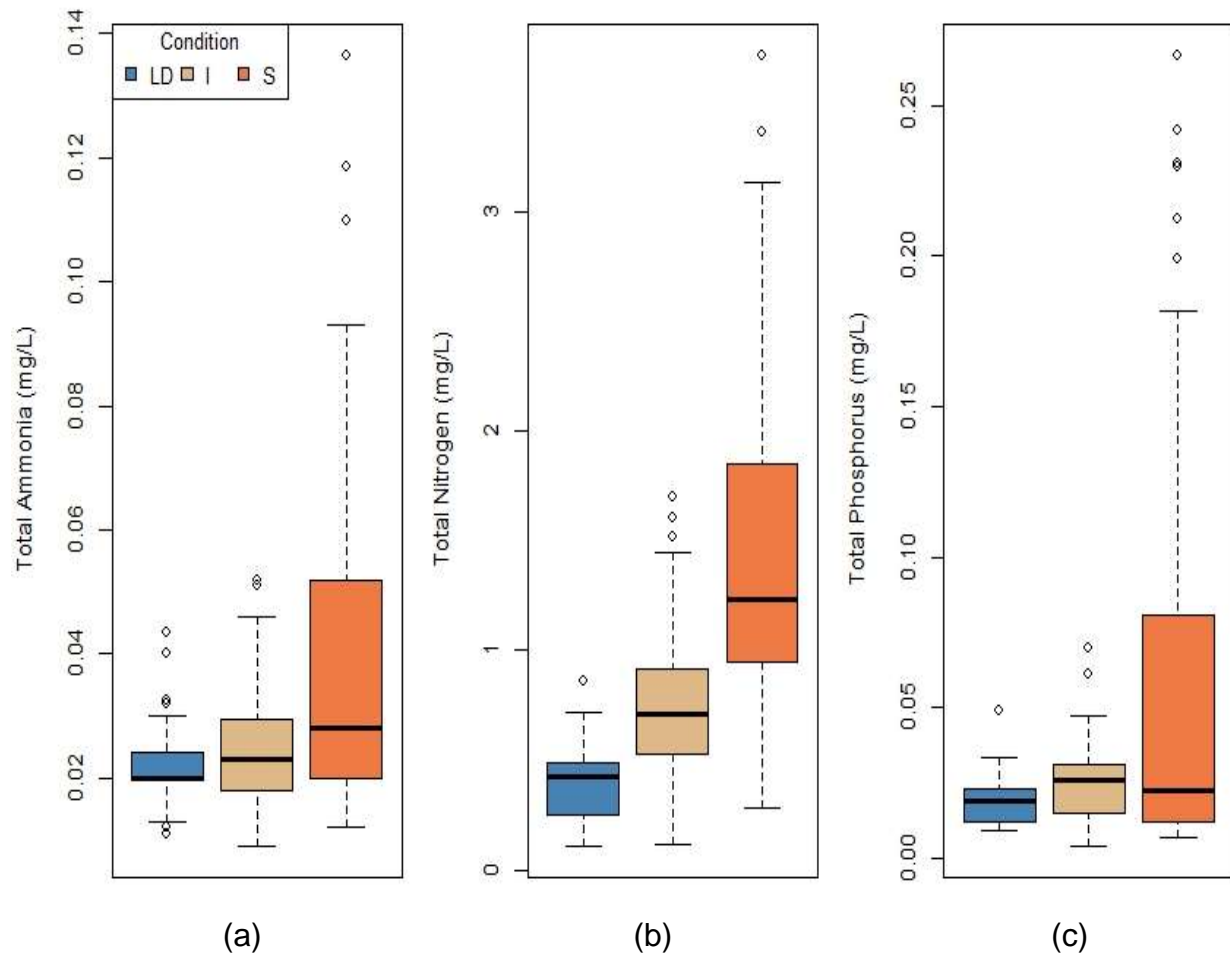
Figure 2-4 Stressor gradient condition classification compared to nutrient parameters such as total ammonia (a), total nitrogen (b), and total phosphorous (c).

(a)   (b)   (c)
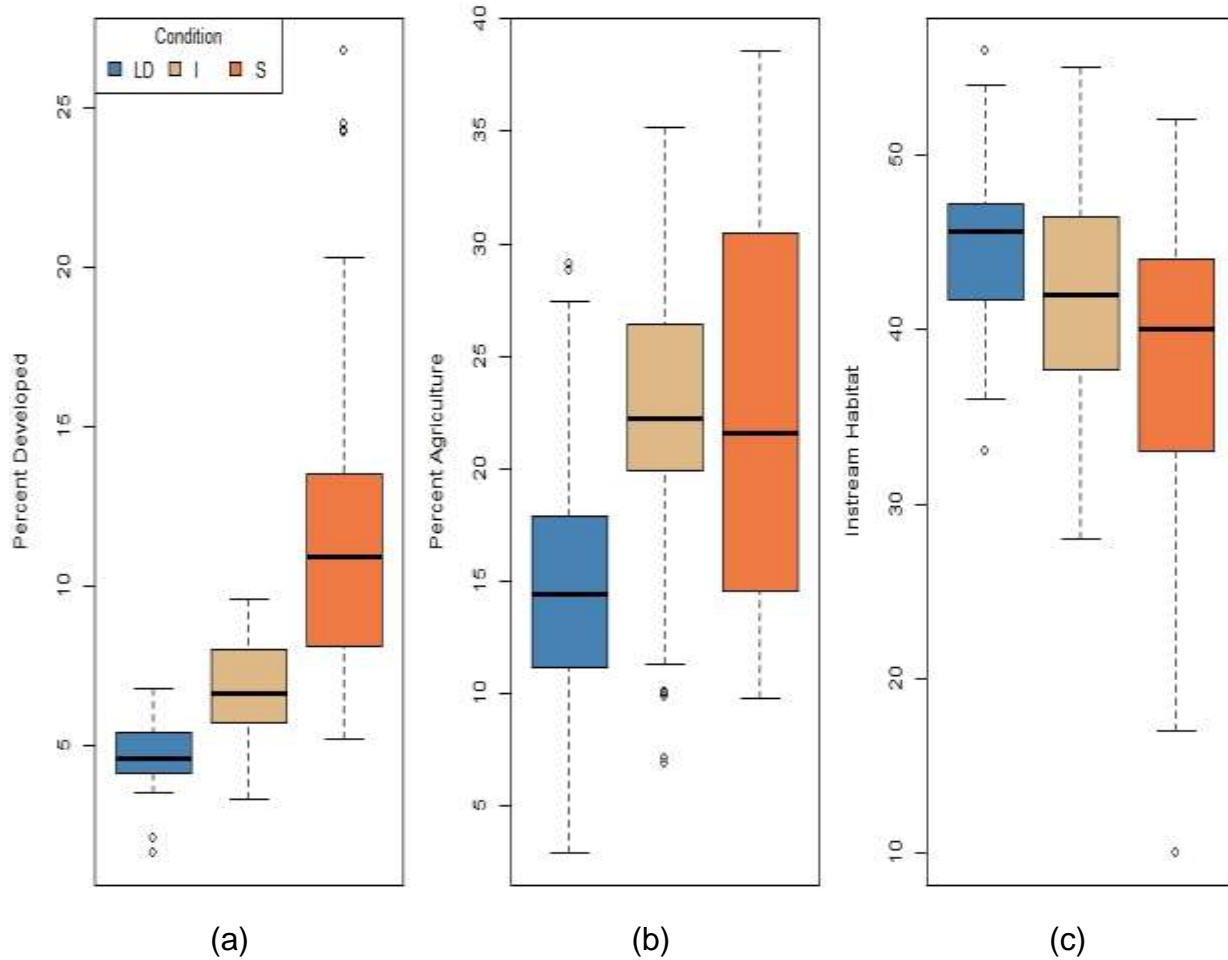
Figure 2-5 Stressor gradient condition classification compared to landscape and
physical parameters such as percent developed (a), percent agriculture (b), and
instream habitat (c).
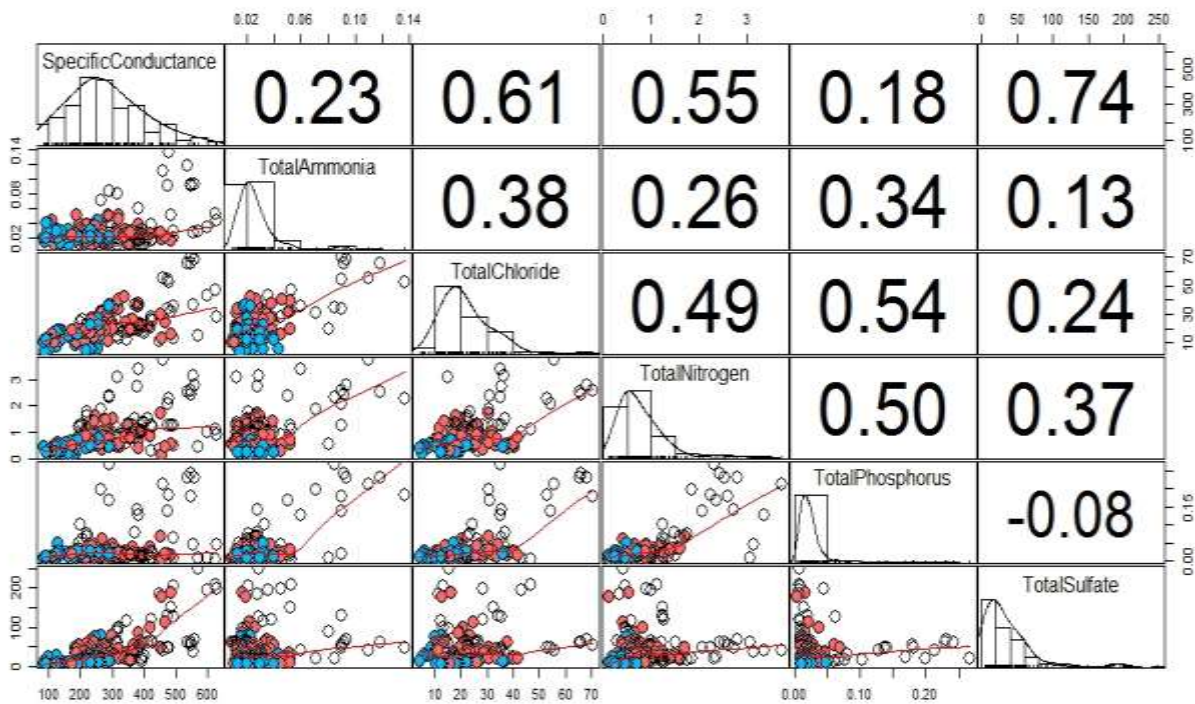
Figure 2-6. Spearman correlation coefficients among chemical stressors. Blue dots in the lower panels indicate LD sites, red dots indicate I sites, and white dots indicate S sites.

There were 105 macroinvertebrate samples collected at sites within the LD condition, which were available for biological community analysis. Initial biological community analysis was conducted with heat maps. Heat map results suggested two potentially distinct communities of macroinvertebrates separating out based on visual assessment of the dataset (Figure 2-7). One group of samples had higher relative abundances of certain genera including, but not limited to, *Baetis* spp., *Chimarra* spp., *Stenelmis* spp., *Simulium* spp., *Teloganopsis* spp., and *Heterocloeon* spp. The other set of samples had higher relative abundances of taxa including, but not limited to, *Allocapnia* spp., *Ephemerella* spp., Chironomidae spp., *Neophylax* spp*., Taeniopteryx* spp*., and Strophopteryx* spp. Initial assessment of these macroinvertebrates suggested seasonality may be driving the observed grouping in heat map analysis. The heat map was also evaluated for obvious outlier samples in the dataset. No single sample or group of samples stood out from the overall group, so no samples were scrutinized further or removed as outliers.
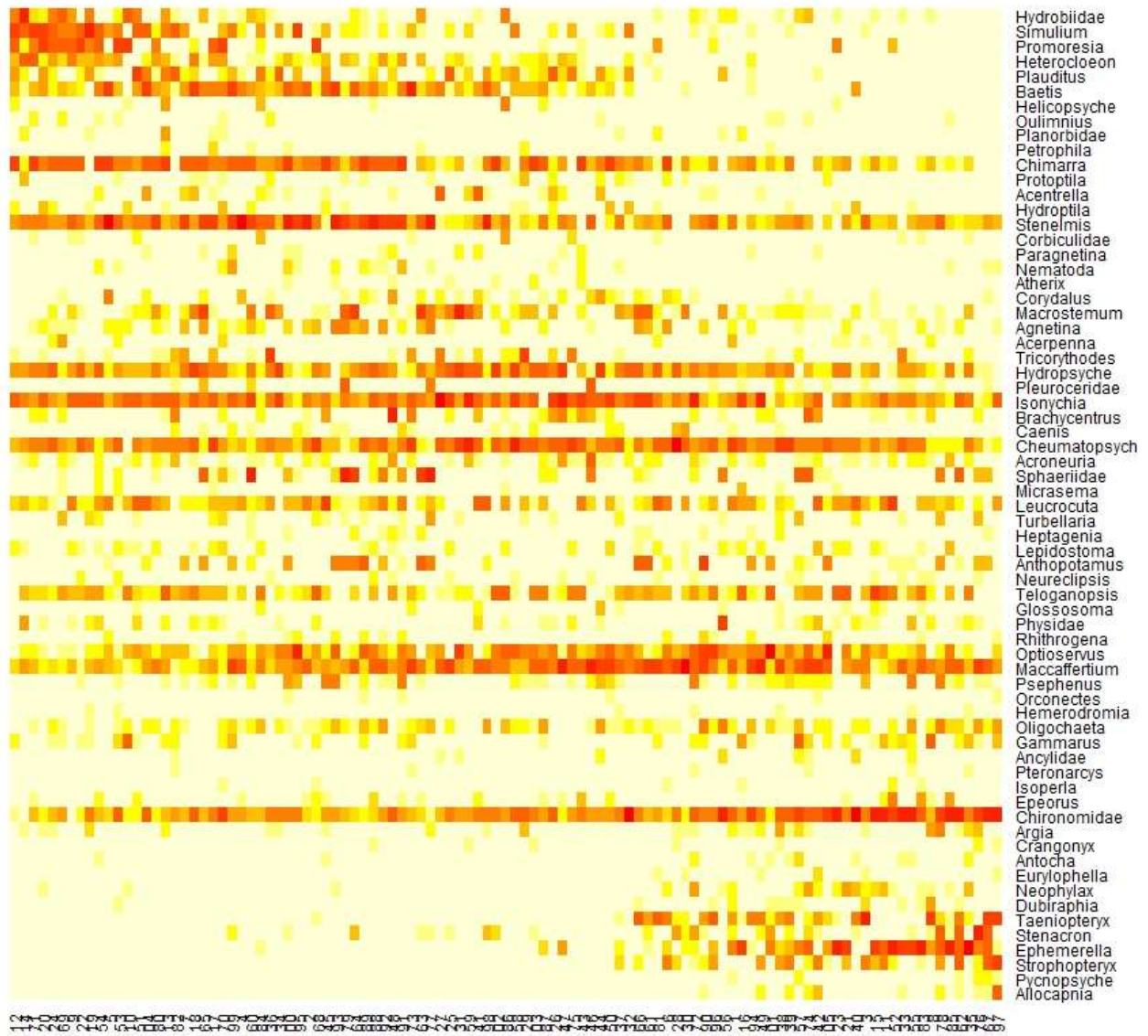
Figure 2-7 Heat map of taxa relative abundance for each sample collected within LD conditions. Darker cells indicated higher relative abundances of taxa among samples. Samples were on the x-axis. Samples and taxa (on the y-axis) were ordered based on the first axis of the detrended correspondence analysis.

Alternative cluster analysis using probability ratios was used to suggest how many classifications the dataset supported, and to which classification a sample belonged. Samples were analyzed based on 2, 3, 4, and 5 classes separately to determine if there was utility in increasing the number of classifications from what the heat map analysis suggested. Classifying samples based on two categories showed clear separation in the first dimension with little uncertainty (overlap) between the two groups (Figure 2-8a). When samples were classified into three (Figure 2-8b) and four (Figure 2-8c)

40

classifications there was also a moderate amount of separation, but this separation was not as clearly defined as two categories. Once samples were classified based on five categories, results showed a lack of distinction for a fifth classification, suggesting the addition of the fifth classification was not supported (Figure 2-8d).
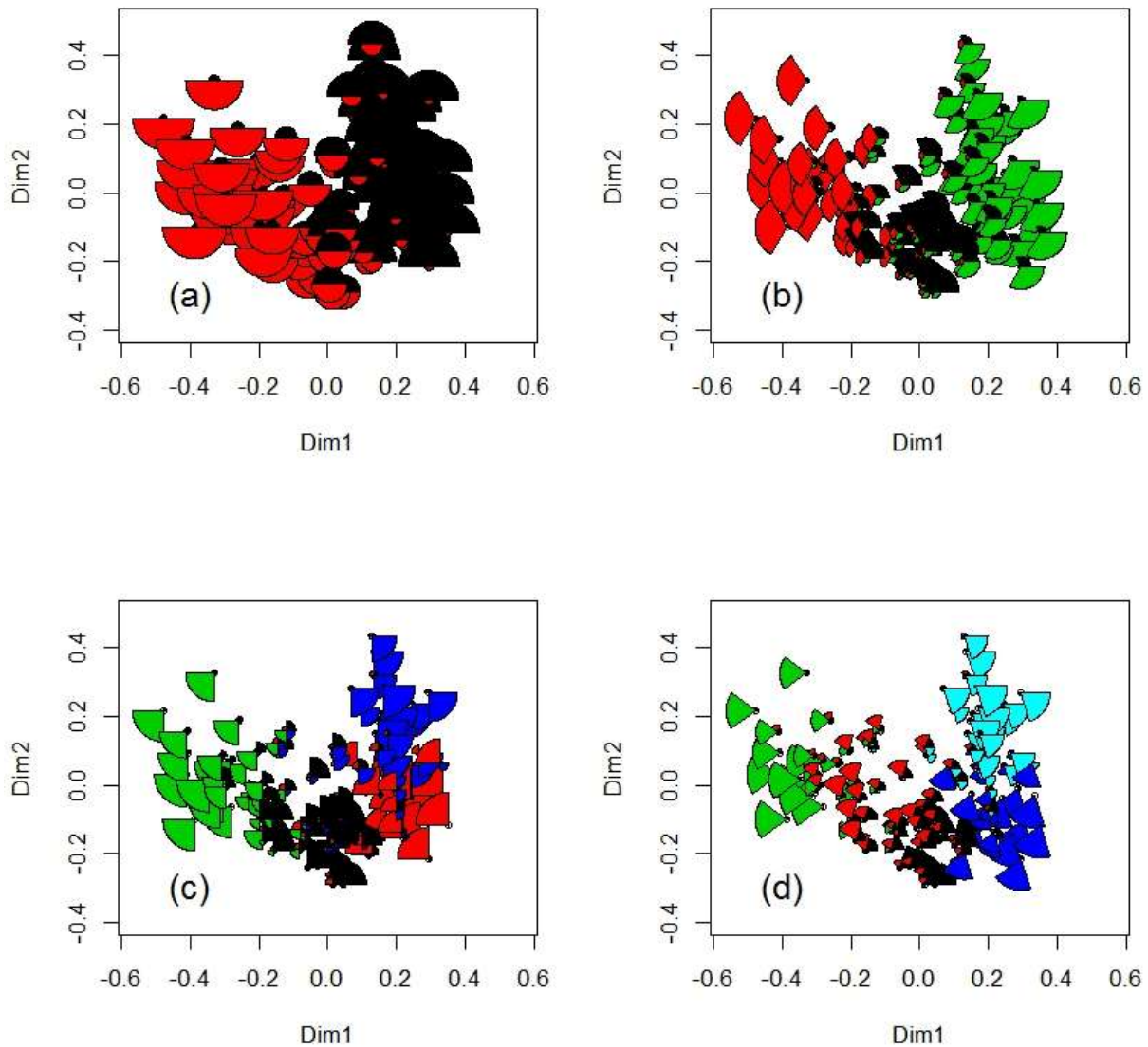
Figure 2-8 Ordinations based on the cluster analysis. Each fraction of the circle indicates the probability of that sample belonging to one of a certain number of unique classifications, where (a) is two possible classifications, (b) is three possible classifications (c) is four possible classifications, and (d) is five possible classifications

The Calinski-Harabasz criterion was used to obtain the optimum number of classifications based on the dataset. This criterion and Mondrian plot (Figure 2-9) was also used to compare against results from the heat map and cluster analyses. A total range of 2 to 15 potential classifications were analyzed using the Calinski-Harabasz criterion. The highest criterion value suggested two classifications were most optimal. As more classes were suggested in the analysis, values quickly decreased and leveled off, indicating that less useful information was available beyond two categories.
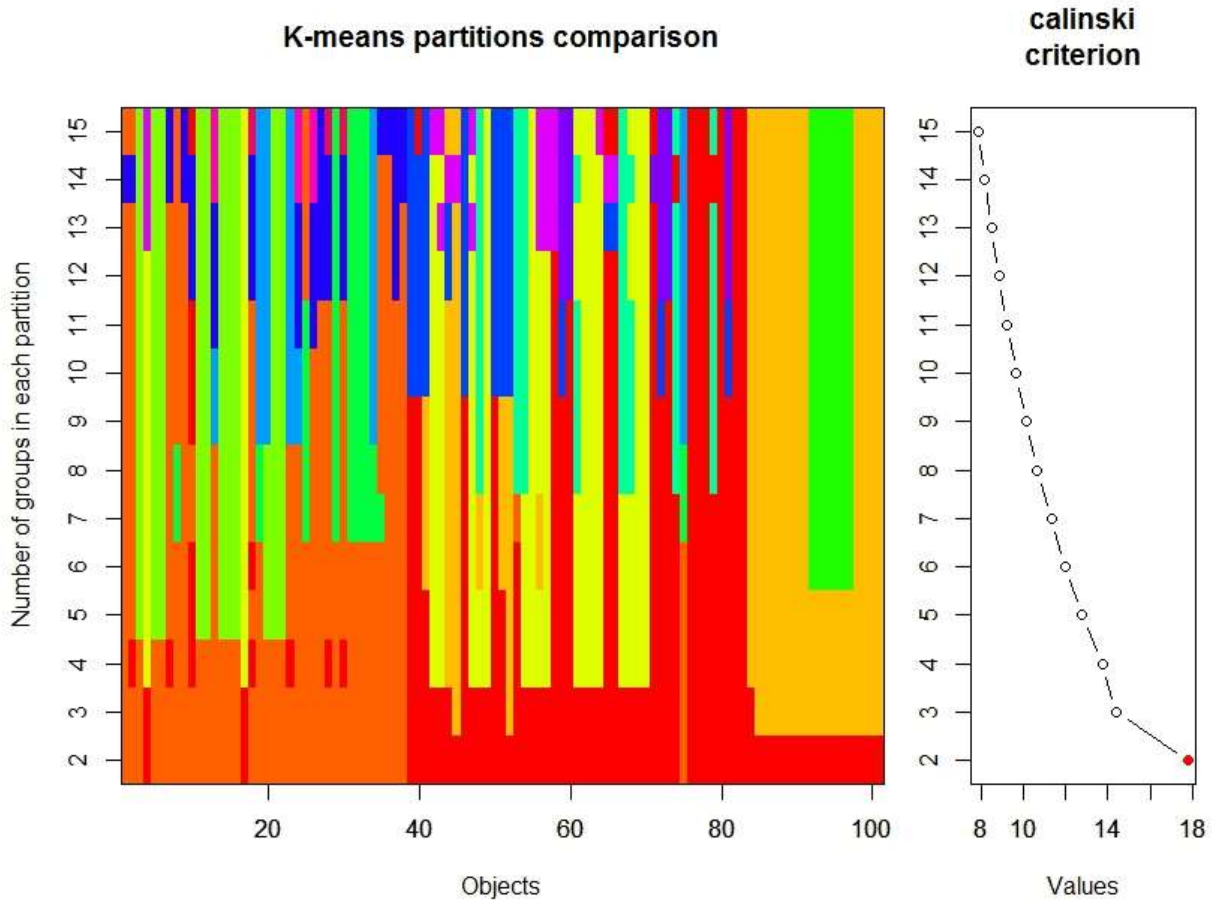
Figure 2-9 Mondrian plot including the K-means partitions comparison and Calinski-Harabasz criterion with value indicators. For K-means, the number of groups along the y-axis was also signified by color. Objects in the x-axis referred to the macroinvertebrate samples. Samples on the x-axis were manually sorted first based on season and then on alkalinity. For the Calinski-Harabasz criterion, the highest criterion value is highlighted in red, indicating the optimum number of classifications according to the statistic.

NMDS was used to evaluate the most reasonable causes for classification schemes. Out of all 9 natural variables evaluated, seasonality was the strongest explanatory

variable and was consistent with previous statistics that recommended two classifications (Figure 2-10). The seasonal classification based on summer and fall classifications was significant (p = 0.01) and accounted for the most amount of community dissimilarity (R^2 = 0.15) compared to the other natural factors. Separation with season was almost completely along the first NMDS dimension, also suggesting this was potentially the strongest natural variable. There were additional natural variables that showed potential as classification factors. Specifically, river basin (Figure 2-11a) showed some community separation, and when river basin was grouped into the three major drainages (Chesapeake Bay, Delaware, and Ohio, Figure 2-11b) results suggest this was the second strongest variable influencing community separation (p = 0.01, R^2 = 0.13). Interestingly, the Delaware and Ohio major drainages were more similar to each other than the Chesapeake Bay drainage even though the Chesapeake Bay drainage completely separated these two other drainages from east to west. Additional spatial variables such as level III ecoregion (Figure 2-11c), and latitude (Figure 2-11d) were also significant (p < 0.05), but explained less community separation ($R^2 \leq 0.10$). Physical characteristics such as river slope (Figure 2-12a), drainage size (Figure 2-12b), and alkalinity (Figure 2-12d) also accounted for little macroinvertebrate community separation (R^2 < 0.1). Thermal regime (Figure 2-12c) was the only natural variable that did not explain community separation at a significant level (p > 0.05).
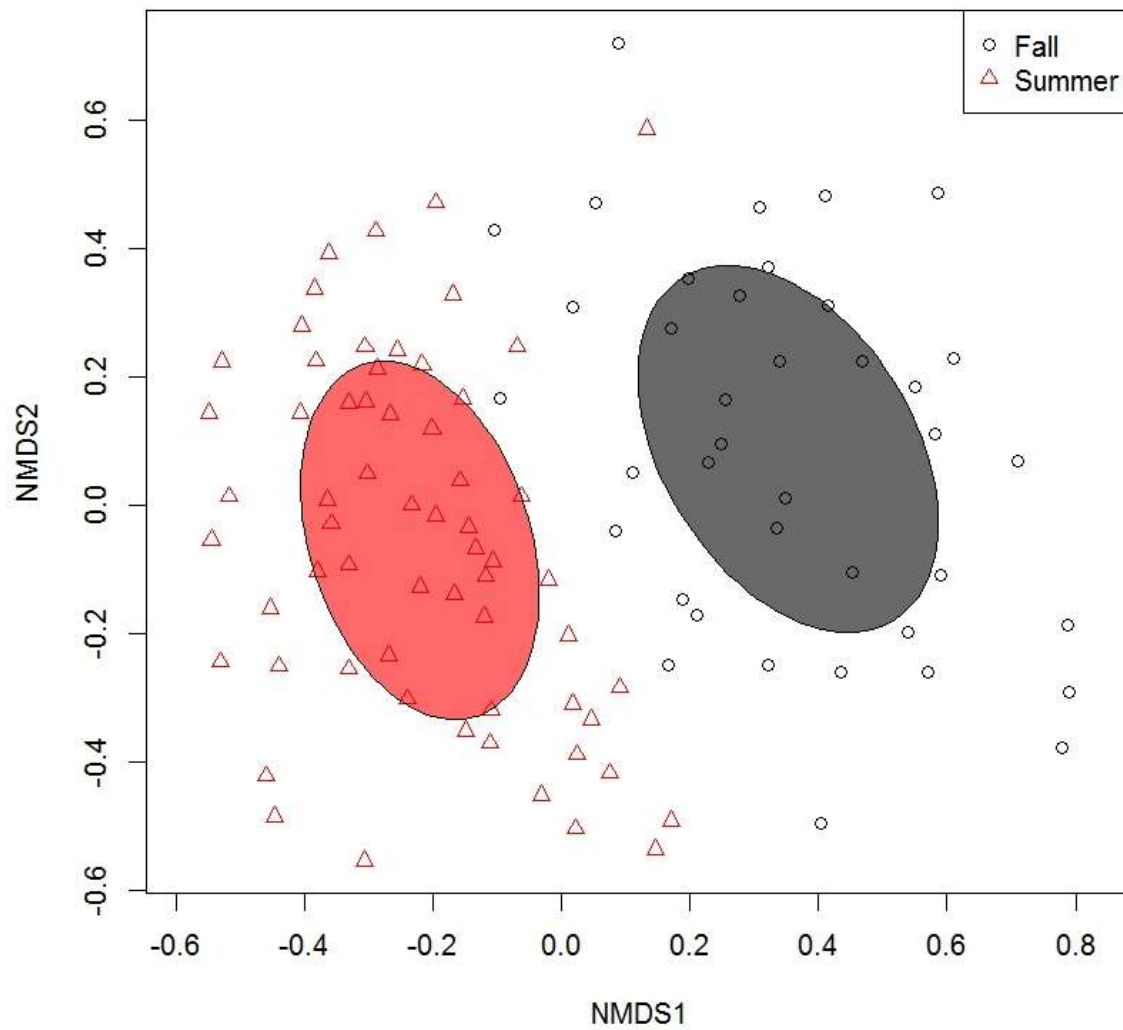
Figure 2-10 NMDS analysis (dimensions = 3, stress = 0.2) plot of first two dimensions using season as groups. Ellipses represent one standard deviation of points in each group.
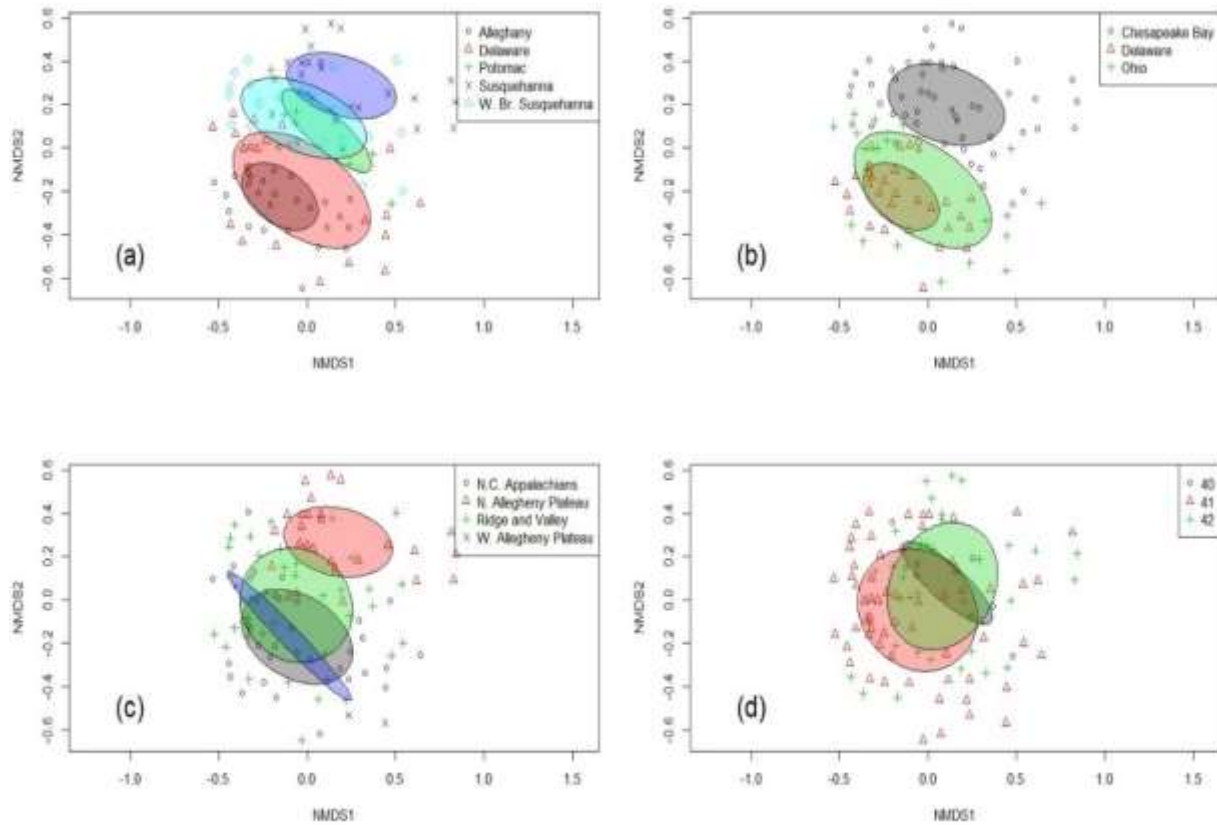
Figure 2-11 NMDS analysis (dimensions = 3, stress = 0.2) plot of first two dimensions with spatial characteristics such as (a) river basin, (b) major drainage, (c) level III ecoregion, and (d) latitude as groups. Ellipses represent one standard deviation of points in each group.
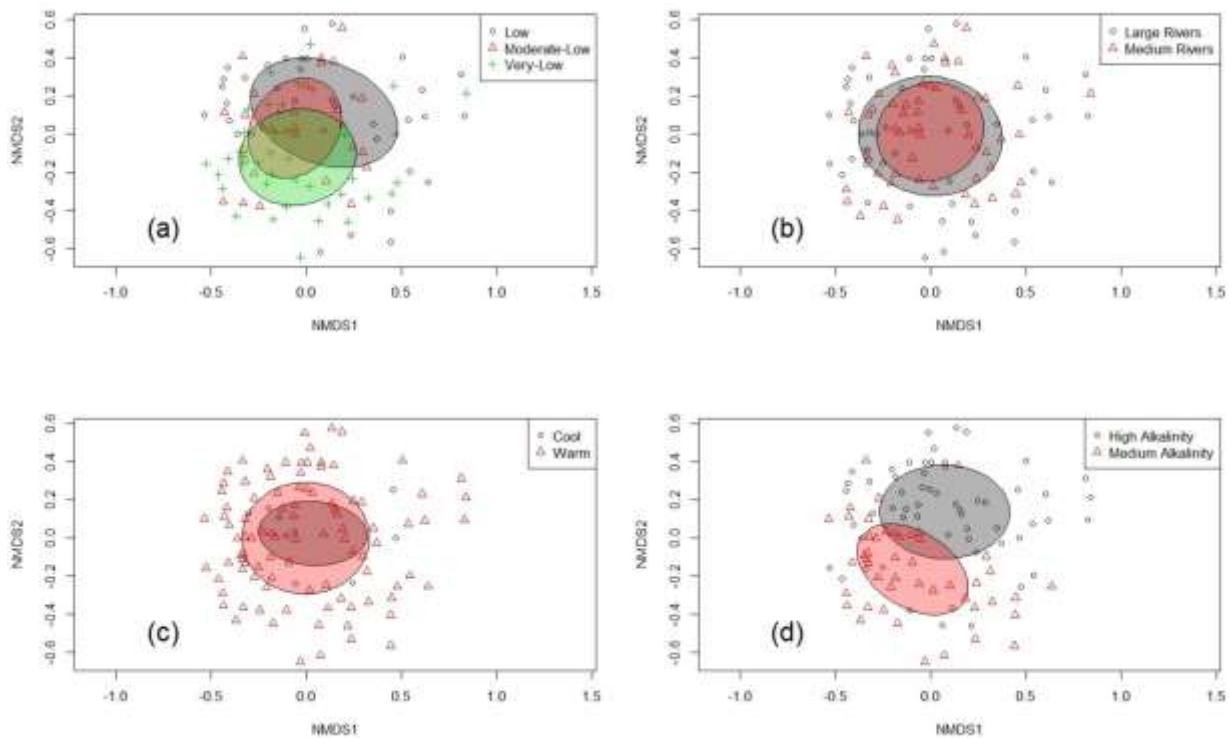
Figure 2-12 NMDS analysis (dimensions = 3, stress = 0.2) plot of first two dimensions with physical characteristics such as; (a) river slope, (b) drainage size, (c) thermal regime, and (d) alkalinity as groups. Ellipses represent one standard deviation of points in each group.

Discussion

Establishing the reference condition is one of the most important aspects of MMI development, because it sets the foundation for subsequent analyses such as natural variability and biological metric development (Stoddard et al. 2006). The environmental stressor gradient established here successfully distinguishes between LD and S sites, indicating that a sufficient range of good to poor water quality sites was collected. A reduced range of stressor gradient scores may have indicated that either an inadequate amount of field sampling occurred, or the stressor gradient chosen lacked responsiveness; however, this was not the case. Each individual stressor selected showed a suitable range of conditions with minimal redundancy, suggesting that each individual stressor will have a relatively equal chance of being an indicator to the biological community changes when metric and index development analyses are conducted. It was also fortunate that a portion of all major drainages made it into the LD condition, because this allowed for a more robust spectrum of sites during classification analysis.

Successful LD classification would not be possible without a robust environmental stressor gradient. It is critically important to capture as many aspects of anthropogenic degradation as possible, because complexity of degradation increases with stream size (Johnson et al. 1995, Weigel and Dimick 2011). It was with these issues in mind that this environmental stressor gradient analysis comprised a sufficient suite of chemical parameters to coincide with physical and landscape aspects. Of the three macroinvertebrate indices developed previously by PADEP (PADEP 2007, Botts 2009, Chalfant 2012), this is the first method to fully incorporate direct measurements of ions and nutrients during the environmental stressor gradient evaluation. This was particularly critical for larger watersheds because of the greater variability of potential stressors that exists compared to smaller watersheds (Johnson et al. 1995). Consequently, this stressor gradient should significantly increase the likelihood of observing community differences along the stressor gradient when metric and index analyses are conducted.

It was important for the stressor gradient to not only capture as many stressors as possible, but also have a stringent set of criteria. Stringent criteria significantly reduce the likelihood that observed natural variation within a biological community is being influenced by anthropogenic stress. This is particularly true when many impact indicators – such as land use – are highly related to ecoregions or physiographic provinces. For example, the Northern Piedmont and Mid-Atlantic Coastal Plain have significantly lower percentages of natural land uses such as forest (approximately 25% and 2.5%, respectively) than other level III ecoregions in the Commonwealth (North Central Appalachians and Blue Ridge have approximately 83% and 77%, respectively). It was also critical that LD threshold criteria are supported by literature from the region. For instance, the specific conductance criterion for LD conditions (300 µS/cm) was developed by a combination of regionally relevant literature that suggested specific conductance in this region rarely exceed 200 µS/cm (Griffith 2014), with evidence that naturally high specific conductance waters are measured around 600-800 µS/cm (Becher and Root 1981, Shull et al. 2016). Therefore, a specific conductance threshold > 300 µS/cm is arguably too lenient for determining LD sites for large rivers in the Mid-Atlantic region and would inevitably include anthropogenically disturbed sites. The observed distribution of specific conductance across sites during this study supports this conclusion. Many LD sites had median specific conductance below 300 µS/cm, and even most S sites had median specific conductance values below 500 µS/cm.

It must be noted that the selection of an active single habitat collection method in large rivers could potentially reduce the observation of natural differences that would be detected by methods designed to collect many habitat types both passively and actively (Cairns 1982, De Pauw et al. 1986, Flotemersch et al. 2006), or methods that identify taxonomic groups such as Chironomidae to higher resolutions (Lenat and Resh 2001,

Pond et al. 2013). This might be problematic if the goal of collecting macroinvertebrates in semi-wadeable rivers is to document and quantify the macroinvertebrate community. However, the purpose of this method is to assess water quality using macroinvertebrates. In addition, there are several studies that suggest increasing cost and effort such as increasing taxonomic resolution may not drastically improve index performance (Waite et al. 2004, Melo 2005, Corbi and Trivnho-Strixino 2006, Mueller et al, 2013). As with any rapid biological assessment, there must be balance between the costs of additional work and the benefits.

Seasonality dictated the majority of natural variability within biological communities at LD sites. This suggests that accounting for season during metric and index development will adequately account for most natural variability in the macroinvertebrate communities, and provide a more accurate and precise tool for water quality assessments. It is possible that a single MMI standardized based on season as was done previously (Chalfant 2012) would adequately compensate for seasonal difference in the macroinvertebrate communities. However, it is also possible that two MMI using different sets of metrics would be more appropriate. Both options may have advantages and disadvantages, but deciding on which is appropriate should be determined during metric and index development. Natural variability analysis also suggested some separation between river basins. It was interesting that the Delaware and the Ohio (upper Allegheny River sites) drainages had macroinvertebrate communities that were more closely related to each other than the Chesapeake Bay drainage sites, which included the Potomac, West Branch Susquehanna, and upper Susquehanna Rivers, and spanned a large distance from north to south. Based on the NMDS results relating to alkalinity, it is possible that this natural factor is manifesting as biological differences between the major drainages. However, the Allegheny River and Delaware River LD sites are both relatively close to major impoundments, which are unnatural factors that tend to change thermal regime and flow dynamics for many kilometers downstream. Therefore, it is difficult to determine whether this biological community dissimilarity is truly due to natural factors. Additionally, macroinvertebrate community difference based on drainage or alkalinity as explanatory variables was relatively weak, suggesting that more complex classifications schemes (e.g., season among alkalinity classes, or season among drainages) would not appreciably increase the accuracy of the subsequent assessment tool. However, these relationships should be further scrutinized as more data are collected and future refinements to this protocol are developed.

Overall, the goal of this work was to analyze and understand abiotic stressors within large semi-wadeable rivers, and to classify macroinvertebrate communities based on natural factors. Understanding these factors within a larger scope than what is encompassed within the Commonwealth provided a much better perspective of

conditions. Few, if any, of these rivers originate and converge within the same governmental jurisdiction; hence, understanding stressors at a large spatial scale may be critical for future restoration efforts. For PADEP, the goal of including these additional sites was simply to build a stronger assessment tool for rivers within the Commonwealth. Yet, with the inclusion of sites outside the Commonwealth, it is suggested that this method can be successfully utilized by other organizations, which has been suggested as a goal for large river assessment methods (Blocksom and Johnson 2009).

Literature Cited

Baptista, D. F., D. F. Buss, M. Engler, A. Giovanelli, M. P. Silveira, and J. L. Nessimian, 2007. A multimetric index based on benthic macroinvertebrates for evaluation of Atlantic Forest streams at Rio de Janeiro State, Brazil. Hydrobiologia 575: 83–94.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, second edition. EPA 841-B-99-002. United States Environmental Protection Agency. Office of Water. Washington, D.C.

Becher, A., and S. Root. 1981, Ground water and Geology of the Cumberland Valley, Cumberland County, Pennsylvania. Pennsylvania Geological Survey. Available online at www.dcnr.state.pa.us/topogeo/pub/water/w050.aspx. Accessed 5 May 2016.

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 Biological Integrity (IBI) for "True" Limestone Streams. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/An%20Index%20of%20Biological%20Integrity-Limestone%20Streams.pdf

Bray J.R., and J.T. Curtis 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecological Monographies 27:325–349.

Calinski, T. and J. Harabasz, 1974. A dendrite method for cluster analysis. Communications in Statistics. 3:1–27.

Chalfant, B. 2012. A benthic index of biotic integrity for wadeable freestone streams in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Technical%20Documentation/freestoneIBImarch2012.pdf

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143.

Corbi, J. J., and S. Trivinho-Strixino. 2006. Influence of taxonomic resolution of stream macroinvertebrate communities on the evaluation of different land uses. Acta Limnologica Brasiliensia 18:469–475.

Cormier, S.M., G.W. Suter, L.L. Yuan & L. Zheng. 2011. A Field Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams. EPA/600/R-10/023F. www.epa.gov/ncea

Elphick, J.R., M. Davies, G. Gilron, E. C. Canaria. Bonnie lo and H. C. Bailey, 2001. An aquatic toxicological evaluation of sulfate: the case for considering hardness as a modifying factor in setting water quality guidelines. Environmental toxicology and chemistry. 30:247–253.

Esri. 2017. World Imagery Basemap. http://server.arcgisonline.com/arcgis/services/World_Imagery/MapServer

Griffith, M.B. 2014. Natural Variation and current reference for specific conductivity and major ions in wadeable streams of the conterminous USA. Freshwater Science. 33(1):1–17.

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 Application 10:1456-1477.

Hering, D., C. K. Feld, O. Moog and T. Ofenbock, 2006. Cook book for the development of a multimetric index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566:311–324.

Homer, C.G., J.A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N.D. Herold, J.D. Wickham and K. Megown, 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing. 81:345–354.

Hughes, R.M., 1995. Defining acceptable biological status by comparing with reference conditions. Chapter 4 in *Biological assessment and criteria: tools for water resource planning and decision making*, W.S. Davis and T.P. Simon, eds. (pp. 31–47). CRC Press, Boca Raton.

Hunsaker, C. T., D. A. Levine, 1995. Hierarchical approaches to the study of water quality in rivers. Bioscience. 45:193–203.

Hurd, T. M., 2012. Determination of Preferential Groundwater Flow Patterns to Cumberland County Springs with Fluorescent Dye Tracing. Pennsylvania Geology. 42(3):3–11.

Johnson, B. L., W. B. Richardson and T. J. Naimo. 1995. Past, present, and future concepts in large river ecology: how rivers function and how human activities influence river processes. BioScience. 45:134–141.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C., USA

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.

Lessard, J.L. and D.B. Hayes. 2003. Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. River Research and Applications. 19: 721–732

Ludwig, J.A. and J.F. Reynolds. 1988. *Statistical ecology.* John Wiley and Sons. New York, New York.

Lyons, J., R. R. Piette, and K. W. Niermeyer. 2001. Development, validation, and application of a fishbased index of biotic integrity for Wisconsin's large warmwater rivers. Transactions of the American Fisheries Society 130:1077–1094.

Maechler, M., P.Rousseeuw, A. Struyf, M.Hubert and K. Hornik. 2014. cluster: Cluster Analysis Basics and Extensions. R package version 1.15.3.

Mattson, K.M. and P.L. Angermeier. 2007. Integrating human impacts and ecological integrity into a riskbased protocol for conservation planning. Environmental Management 39(1):125–138.

Melo, A. S. 2005. Effects of taxonomic and numeric resolution on the ability to detect ecological patterns at a local scale using stream macroinvertebrates. Archiv für Hydrobiologie 164:309–323.

McArdle, B.H. and M.J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology. 82: 290-297.

Mueller, M., J. Pander, and J. Geist. 2013. Taxonomic sufficiency in freshwater ecosystems: effects of taxonomic resolution, functional traits, and data transformation. Freshwater Science *32*(3):762–778.

Novotny, V. 2004. Linking pollution to water body integrity – literature review. Center for Urban Environmental Studies, Northeastern University. Boston, Massachusetts.

Oksanen, J., G. 2014. Cluster Analysis: Tutorial with R. University of Oulu, Oulu. http://cc.oulu.fi/~jarioksa/opetus/metodi/sessio3.pdf

Oksanen, J., G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, H. H. Stevens, and H. Wagner. 2016. vegan: Community Ecology Package. R package version 2.3–4. http://CRAN.Rproject.org/package=vegan

Olivero Sheldon, A., A. Barnett, and M. G. Anderson. 2015. A Stream Classification for the Appalachian Region. The Nature Conservancy, Eastern Conservation Science, Eastern Regional Office. Boston, MA.

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

PADEP. 2013. Instream comprehensive evaluation surveys. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/ICE.pdf

Pond, G. P., J. E. Bailey, B. M. Lowman & M. J. Whitman, 2013. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. Environmental Monitoring and Assessment 185:1515–1540.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at http://www.R-project.org/.

Rao, C. R., 1995. A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. Questiio 19:23–63

Revelle, W. 2017. psych: Procedures for Psychological, Psychometric, and Personality Research. https://cran.r-project.org/web/packages/psych

Roth, N. E., M.T. Southerland, J.C. Chaillou, J.H. Vølstad, S.B. Weisberg, H.T. Wilson, D.G. Heimbuch & J.C. Seibel, 1997, Maryland biological stream survey: ecological status of non-tidal steams in six basin sampled in 1995. Maryland Department of Natural Resources, Chesapeake Bay and Watershed Programs, Monitoring and Non-tidal Assessment, Annapolis, Maryland.

Roy, A. H., A. D. Rosemond, M. J. Paul, D. S. Leigh & J. B. Wallace, 2003. Stream macroinvertebrate response to catchment urbanization (Georgia, U.S.A). Freshwater Biology 48:329–346.

Ruspini, E. H., 1970. Numerical methods for fuzzy clustering. Information Sciences. 2: 319–350.

Scopel, C., 2014. Create watersheds and trace downstream in your ArcGIS Online web map. Environmental Systems Research Institute. Redlands, CA. https://blogs.esri.com/esri/arcgis/2014/12/12/create-watersheds-and-trace-downstream-in-your-arcgis-online-web-map/

Shull, D. R. and M. Pulket, 2015. Causal Analysis of the Smallmouth Bass decline in the Susquehanna and Juniata Rivers. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/SusquehannaRiverStudyUpdates/SMB_CADDIS_Report.pdf

Shull, D. R., and M. J. Lookenbill. 2015. Continuous Instream Monitoring Protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/CIM_PROTOCOL.pdf

Shull, D.R., R.L. Stewart Jr., T. Hurd, and T. Light. 2016. A Case for Unique Habitat Selection by Sigara mathesoni (Hemiptera: Corxidae) in South Central Pennsylvania. Northeastern Naturalist. 23(1): 174–183.

Sliva, L., D. D. Williams, 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Resources 35:3462–3472.

Stevens, L.E., J.P. Shannon and D.W. Blinn. 1997. Colorado river benthic ecology in Grand Canyon, Arizona, USA, tributary and geomorphological influences. Regulated River: Research and Management 13: 129–149.

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

Stranahan, S. Q. 1993. Susquehanna: river of dreams. Johns Hopkins University Press. Baltimore, Maryland.

USDT. 2015. National Transportation Atlas Database. Bureau of Transportation Statistics. Washington, DC. https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2015/point

USEPA. 2016. National Rivers and Streams Assessment 2008-2009 Technical Report. Office of Water and Office of Research and Development. Washington, DC. EPA/841/R-16/008

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.

Wang, L., J. Lyons and P. D. Kanehl. 2003. Impacts of urban land cover on trout streams in Wisconsin and Minnesota. Transactions of the American Fisheries Society 132: 825–839.

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

Weigel, B. M. and D. M. Robertson. 2007. Identifying biotic integrity and water chemistry relations in nonwadeable rivers of Wisconsin: toward the development of nutrient criteria. Environmental Management 40:691–708.

Weigel, B. M. and J. J. Dimick. 2011. Development, validation, and application of a macroinvertebrate based index of biotic integrity for nonwadeable rivers of Wisconsin. Journal of the North American Benthological Society 30(3):665–679.

Wickham, H. 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York

Chapter 3. Multimetric Index Development, Validation, and Precision

Introduction

To meet the objectives outlined in the federal Clean Water Act and Pennsylvania's Clean Streams Law, evaluations of aquatic ecosystem integrity include assessments of physical characteristics, water chemistry, and biological communities. However, chemical evaluations are of limited value in assessing overall ecosystem condition because of the difficulty evaluating every relevant chemical parameter, synergistic chemical effects on ecosystems, and high variability associated with lotic systems. Physical evaluations of streams – although informative – are also limited in value for assessing overall ecosystem integrity. For example, in some acid deposition impacted streams, watershed and instream physical conditions may suggest excellent conditions, but the biotic community is drastically impacted by the acidification. On the contrary, biological monitoring offers the ability to assess long-term, cumulative effects of many ecosystem stressors, including both chemical and physical factors. Organisms living in aquatic environments are intimately associated with and affected by chemical water quality and the physical conditions of streams and watersheds. Consequently, these organisms can be viewed as living indicators of overall ecosystem condition (Chalfant 2012). Benthic macroinvertebrates are especially good indicators because they can readily recolonize habitats under improving environmental conditions. These characteristics make benthic macroinvertebrates well suited for an assessment tool (Pond et al. 2013). In fact, benthic macroinvertebrates have been widely used as indicators of water quality in both small streams and large rivers (Chutter 1972, Hilsenhoff 1987, Barbour et al. 1999, Royer et al. 2001, Applegate et al. 2007, Astin 2007, Wessell et al. 2008, Blocksom and Johnson 2009, Weigel and Dimick 2011, Chalfant 2012).

There are many assessment tools that have been successfully implemented around the world including multimetric indices (MMI), biological condition gradients (BCG), and predictive models of taxa richness (Wright 1994, Charles et al. 2010, Baptista et al. 2011). These assessment types are potentially suitable tools for large semi-wadeable rivers in the Commonwealth, but PADEP a has long record of using MMIs to make ALU assessments (PADEP 2007, Botts 2009, Chalfant 2012). Consequently, using a MMI for assessing large semi-wadeable rivers may be beneficial for this work, because local, state, and federal resource managers as well as the public are already familiar with the development process and are more likely to comprehend results. However, resource managers and the public may be more familiar with PADEP MMIs being called indices of biological integrity (IBI). Both terms refer to using a multimetric index to evaluate biological condition; however, the original intent of the term "index of biological integrity"

was to understand and quantify biological balance and integrity between pristine and stressed conditions within a region (Karr and Dudley 1981). Since conditions in large rivers – and most streams – are no longer natural or pristine, a score of 90 or 100 on a 0-100-point scale may not necessarily be indicative of historic "biological integrity". Therefore, it would be more appropriate to define these indices as simply MMIs rather than IBIs.

The MMI development process uses direct quantification of biological attributes along a gradient of ecosystem conditions (developed in Chapter 2). Assessment of biological conditions using an MMI involves measuring multiple biological metrics of varying type (e.g., richness, composition, function, dominance, and tolerance) and comparing how they differ along a stressor gradient (Davis and Simon 1995, Davies and Jackson 2006, Hawkins 2006, Stoddard et al. 2006). It is ideal to incorporate several different types of metrics to obtain the best picture of biological condition and ecosystem balance (Gerritsen et al. 2000). For instance, indices that include only the best performing metrics may miss important compositional aspects of the biological community, such as the loss of mayflies. Therefore, metrics that best distinguish between minimally impacted conditions and stressed conditions must be evaluated further for potential inclusion in an MMI. Selected metrics must have acceptable limits of range and minimum values within the LD condition (Gerritsen et al. 2000). Additionally, they must not be redundant with each other, in other words, measure the same biological aspect of the community. Scope of impairment (SOI) is yet another important metric measure, which checks that metric variability within the LD condition is not too large when compared to the S condition (Klemm et al. 2003, Blocksom and Johnson 2009, Pond et al, 2013). Once metrics are evaluated and selected, the final index standardizes and composites metrics into one score, which measures the overall extent to which anthropogenic activities compromise a stream's ability to support healthy aquatic communities (Davis and Simon 1995). The newly developed biological assessment tool will help guide and evaluate policy, management strategies, and create incremental progress benchmarks for the Commonwealth's aquatic resources (Davis and Simon 1995, Davies and Jackson 2006, Hawkins 2006).

Traditionally, PADEP macroinvertebrate metrics incorporated pollution tolerance values (PTV) developed by and modified from work done by Hilsenhoff (1977, 1987, 1988). These values have proven very effective in creating defensible MMIs for the Commonwealth; however, PADEP has recently explored using a BCG concept in making water quality assessments (Gerritsen and Jessup 2007). The BCG concepts do have advantages, which include providing a framework to interpret changes in ecological conditions regardless of geography or stream type. The BCG also provides a clear means to communicate existing conditions and potential incremental improvement to the public (USEPA 2016a). Between 2006 and 2008, PADEP conducted a series of

expert workshops to develop a BCG for the small and large wadeable, freestone streams. In conjunction with external taxonomic experts and ecologists from various organizations (e.g., Delaware River Basin Commission, Nature Conservancy, and USEPA), PADEP biologists used the BCG attributes of benthic macroinvertebrates (Table 3-1) to characterize specific changes in community composition (Figure 3-1, Gerritsen and Jessup 2007). At the beginning of the expert workshop, participants assigned a BCG attribute for pollution sensitivity or origination status (i.e., attributes I–V, or VI, respectively) to each macroinvertebrate taxon based on expert knowledge and biological response data. The data used was from sites that spanned a range of condition from minimally disturbed to heavily impacted. Using the BCG level descriptions and the taxa attributes as a guide, the group assigned each site to one of the six BCG levels and developed candidate decision rules (Gerritsen and Jessup 2007). Although PADEP does not use a fully implemented BCG tool for making water quality assessments, components such as the BCG taxa attributes are currently used for wadeable freestone stream assessments (Chalfant, 2012). Since BCG taxa attributes were created for large warm-water streams, they can also be evaluated for inclusion in this large river semi-wadeable MMI (SWMMI) development. Inclusion of taxa metrics using BCG attributes may not only provide a check against the traditionally used PTVs, they may also show better response in large rivers. Inclusion of BCG metrics may also provide an opportunity for individual metrics to become more meaningful to the public; hence, increasing the usefulness of information provided by the overall assessment tool.

Table. 3-1 Taxa attributes used to develop the BCG for PADEP (Modified from USEPA 2016)

Attribute	Description
I	Historically documented, sensitive, long-lived, or regionally endemic taxa
II	Highly sensitive taxa
III	Intermediate sensitive taxa
IV	Intermediate tolerant taxa
V	Tolerant taxa
VI	Non-native or intentionally introduced species
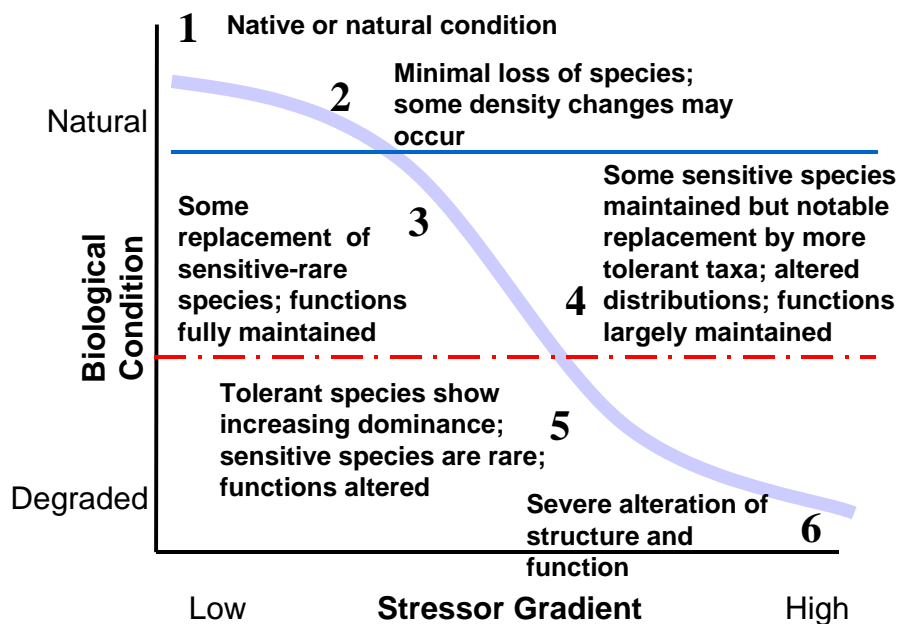
Figure 3-1 Representation of biological condition gradient, showing six defined tiers (Modified from Davies and Jackson. 2006).

The objectives of this work were to develop and validate a MMI for ALU assessment in large semi-wadeable rivers. This SWMMI must account for the natural variation in these rivers (i.e., have good precision) and distinguish between LD and S conditions with a high degree of accuracy. As discussed in Chapter 2, there were two distinctly different macroinvertebrate communities within the LD condition, and these differences were primarily driven by seasonality. Therefore, it is prudent to further analyze these communities for potentially two different SWMMIs, because there is a high likelihood that different metrics will respond better for each community.

The final SWMMI must consistently use biological information to differentiate between good and poor sites to be considered accurate. This is traditionally measured through discrimination efficiency (DE) between LD and S conditions (Barbour et al. 1999, Gerritsen et al. 2000, PADEP 2007, Botts 2009, Chalfant 2012). Accuracy is also measured through validation using an independent or synthetic dataset (Chalfant 2012, Smith et al. 2017). Using a validation dataset, the SWMMI should accurately classify new information into the appropriate environmental stressor condition category, which will demonstrate that the development dataset was not over fitting the index into an environmental stressor condition category. The final SWMMI must also have a reasonable degree of precision to ensure the method is repeatable and can pick up on genuine changes in environmental condition over time. Coefficients of variation (CV)

from replicate sample collections are commonly used to measure method precision (PADEP 2007, Stribling et al. 2008, Chalfant 2012, Pond et al. 2013). This estimate is referred to as intrasite precision and incorporates variability such as differences in microhabitat collection, subsampling variability, and other human error. A CV substantially greater than 10-15% (Stribling et al. 2008) may indicate that too much variability exists within the selected SWMMI and collection method. Another form of precision that should be measured is temporal precision. As the name implies, temporal precision measures all the variability discussed above, along with additional variability associated with sampling in different months and years. A measure of temporal variability that is commonly used and easily related to the SWMMI is 90% confidence interval (CI90). Sample results after a baseline sample or set of samples that exceed the CI90 have a high probability of measuring an actual change in the environment. Consequently, the CI90 has been used to show incremental progress due to restoration efforts of degraded waters (Botts 2009, Chalfant 2012).

Completion of both an accurate and precise biological assessment tool will ultimately support the mission of PADEP and provide valuable information on these important resources. Current understanding of the biological community within large river systems is lacking. Previous attempts to measure biological integrity of large rivers in the Commonwealth (USEPA, 2016b) failed to compensate for the great complexity and natural differences of large rivers by comparing large rivers to much smaller reference streams. Hence, any biological assessment tool that can compensate for the increased complexity of these rivers and maintain the ability to differentiate between good and poor environmental conditions (with acceptable precision) will greatly benefit resource managers and the public.

Methods

There were 457 samples from 178 sites available for this study. However, several sites did not have macroinvertebrate collections consistent with the transect method and were removed for all subsequent analyses. The remaining sites were then separated into a calibration and validation dataset. The calibration dataset consisted of 275 samples from 126 sites and the validation dataset consisted of 106 samples from 42 (25% of the total) sites (Figure 3-2). Replicate samples were removed from both datasets during metric development and validation to reduce potential bias toward certain sites; however, samples collected at the same site on different years were included to retain a sufficient population of samples. Additionally, unlike reference classification analysis in Chapter 2, rare taxa were kept for these analyses. To account for season, each dataset was further refined based on the season in which samples were collected. For the calibration dataset, this resulted in 155 summer samples from 114 sites and 120 fall samples collected at 93 sites. For the validation dataset, this

resulted in 63 summer samples from 42 sites and 43 fall samples from 32 sites. Data were analyzed using R version 3.3.2 software (R Core Team 2016). The R package ggplot2, was used for scatterplot and boxplot graphical illustrations (Wickham 2009). Scatterplot correlation matrices were created using the R packages PerformanceAnalytics (Peterson et al. 2015), and psych (Revelle 2017). Principal components analysis was conducted using the R package FactoMineR (Le et al. 2008). Data manipulation in R was conducted using the packages plyr (Wickham 2011) and dplyr (Wickham and Francois 2016).
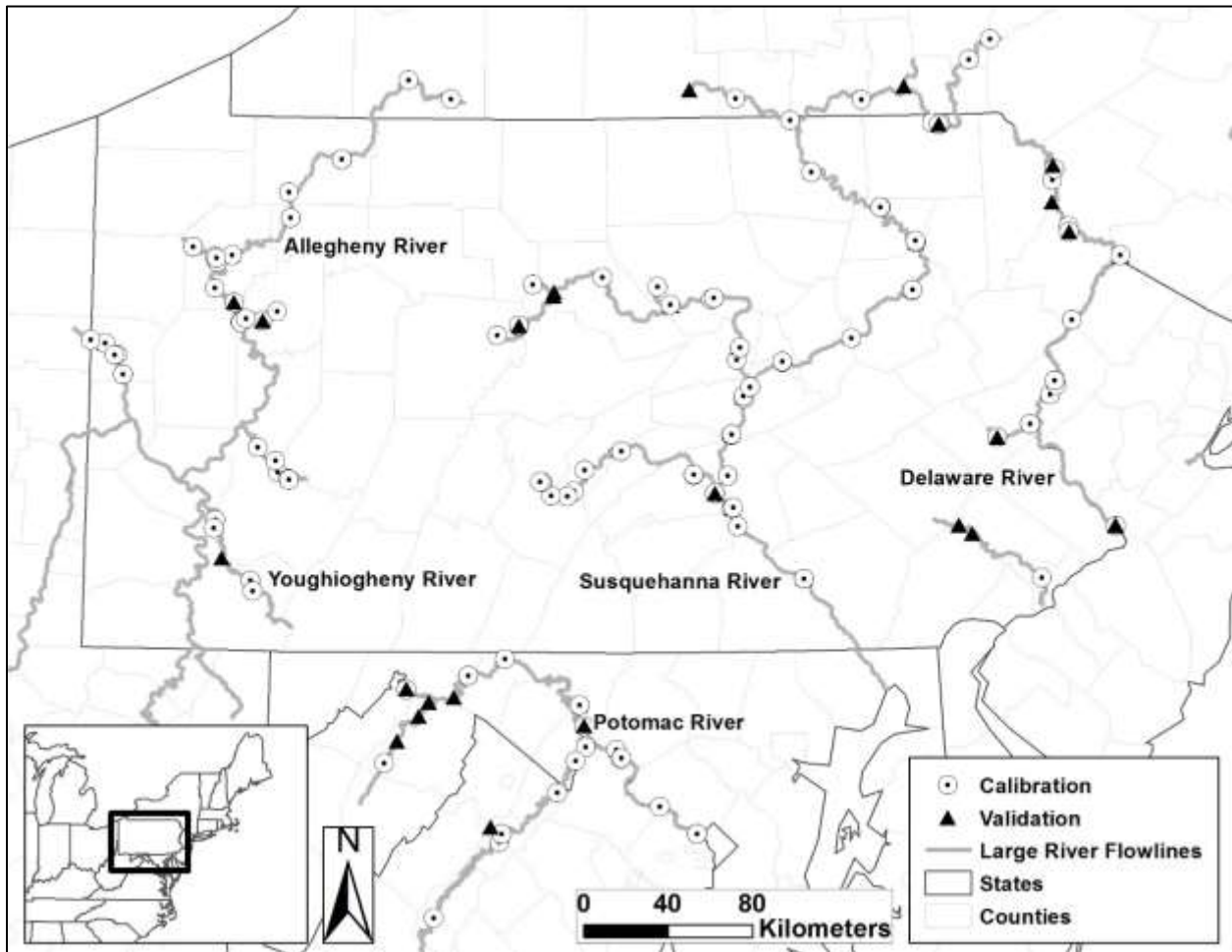
Figure 3-2. Study area with collection sites separated by calibration and validation datasets.

Metric Selection

To develop the SWMMI, the calibration dataset was used to evaluate and refine the list of metrics that met several stringent criteria. There was a total of 167 candidate metrics – originally developed by Chalfant (2012) – evaluated for this MMI development. Initially

60

however, 39 ratio and proportionality metrics were removed from analysis, leaving 128 metrics for evaluation. These metrics were removed because they functionally act like composition metrics (percent metrics); hence, they do not add new insight into the biological community. Each taxon was coded for various traits (PTV, BCG, and functional feeding group) based on published literature (Barbour et al. 1999, Merritt et al. 2008) and further adapted for the Mid-Atlantic region. All metric and index development steps were conducted separately for summer and fall samples. A stepwise process was used to reduce the original number of available metrics to a final set of selected metrics. As the first level of metric refinement, the ability of each candidate metric to discriminate along the environmental stressor gradient was measured using DE. For metrics expected to decrease in value with increasing disturbance (negative-response metrics), the following DE equation was used:

$$DE = \frac{n_{S < \% LD}}{n_{S\ total}} * 100$$

Where:

- DE = discrimination efficiency (%)
- $n_{S < \% LD}$ = the number of samples in the S condition with metric values less than the 25th percentile of LD condition samples
- $n_{S\ total}$ = the total samples in the S condition

For metrics expected to increase in value with increasing disturbance (positive-response metrics), the following DE equation was used:

$$DE = \frac{n_{S > \% LD}}{n_{S\ total}} * 100$$

Where:

- DE = discrimination efficiency (%)
- $n_{S > \% LD}$ = the number of samples in the S condition with metric values greater than the 75th percentile of LD condition samples
- $n_{S\ total}$ = the total samples in the S condition

Metrics with greater than 70% DE were retained and subsequently evaluated for redundancy, which was measured using Spearman rank correlation coefficients. Spearman rank correlations were conducted using only samples collected within LD sites. Metrics with correlation coefficients greater than 0.75 were evaluated further by visually inspecting scatterplots. This was done because high metric correlations can

hide divergences at some point within each metric's range, which indicates that potentially important differences exist. Therefore, candidate metrics with r > 0.75 and with insufficient spread in the scatterplot were removed as candidate metrics.

Initial DE and redundancy evaluations produced a refined subset of candidate metrics for which final metric selection evaluations were conducted. Final evaluations consisted of examining metric characteristics such as minimum values and ranges within the LD condition, responsiveness to the stressor gradient, SOI, and consideration of metric type. Richness metrics with minimum LD values of ≤ 1 and ranges of < 5 taxa were flagged to perform further analysis and determine whether the metric should be removed from consideration. Percent metrics with minimum LD values of < 10% were also flagged. Responsiveness to the stressor gradient was evaluated two ways. First, each metric was visually compared using boxplots across the condition gradient developed in Chapter 2, where sites scores ranged from 0, indicating the least amount of disturbance, to 14, indicating the most amount of disturbance. Responsiveness was also measured by comparing metrics to principal component analysis (PCA) results of chemical and physical environmental stressors. Environmental stressors included measures of instream habitat, total sulfate, total chloride, total specific conductance, total nitrogen, total phosphorous, total ammonia, percent agriculture land use in the watershed, and percent developed land use in the watershed. Pearson correlation coefficients were used to compare the first principal component of the PCA (log transformed environmental parameters along the stressor gradient) and metrics. The SOI was measured by dividing the interquartile range (25-75th percentile, IQR) by the 0-25th percentile (for negative response metrics) or 75-100th percentile range (for positive response metrics) of the LD condition. Ratios >1 indicated increased metric variability within the LD condition and were flagged. Selection of the final set of metrics were ultimately based on a comprehensive evaluation of all metric evaluations. Metrics with flags that could not be resolved were removed from consideration. As a final step in metric selection, special consideration was given to incorporate as many different attributes of the biological community as possible.

Index Development

Once final metrics for each SWMMI were selected, they were standardized and composited into a single score that ranged from 0, indicating worst possible score to 100, indicating best possible score. Metrics were standardized using a floor and ceiling method (Blocksom and Johnson 2009, Pond et al. 2013). For metrics that decrease in value with increasing disturbance (negative-response metrics), standardizations were calculated using the following equation (observed value - floor) / (ceiling - floor) * 100. For metrics that increase in value with increasing disturbance (positive-response metrics) standardizations were calculated using the following equation:

(ceiling - observed value) / (ceiling - floor) * 100. The ceiling and floor values were defined at the 95th and 5th percentiles of the entire sample distribution. Each SWMMI was subsequently calculated by averaging the standardized metrics.

Overall index performance of the calibration dataset was evaluated using DE, responsiveness to the stressor gradient (developed from PCA), and Spearman rank correlation analysis between SWMMI scores within the LD condition and natural factors. Natural factors that were evaluated included month, drainage size (km^2), latitude, longitude, gradient classification, and alkalinity classification. Drainage size, gradient, and alkalinity data were obtained from the Appalachian Stream Classification dataset (Olivero Sheldon 2015). The final set of metrics of both SWMMIs were also compared to determine if the two datasets could be combined and evaluated as one SWMMI. Once it was determined whether the two datasets would be combined or not, impairment thresholds were calculated. Several impairment thresholds were evaluated, including the 5th, 10th, and 25th percentiles of samples in the LD condition as well as the best separation point (BSP, Smith et al. 2017) between samples within the LD and S condition.

Validation

Validation metrics were selected and standardized using the results from the calibration dataset. The summer and fall validation datasets were either grouped or not based on information provided by the preceding calibration dataset analyses (i.e., either one SWMMI for both seasons, or two SWMMIs). Performance of the validation dataset was evaluated using classification efficiency (CE). The CE was determined by calculating the percent of LD and S samples correctly classified using the impairment threshold selected during index development using the calibration dataset. In addition, the validation dataset was evaluated using other MMI performance measures such as responsiveness to the stressor gradient (developed from PCA), and Spearman rank correlation analysis in the LD condition with natural factors. Natural factor analysis was conducted using the same methods as the calibration dataset.

Precision

Once the calibration and validation datasets were evaluated, they were recombined into one dataset to obtain final metric standardization values, impairment thresholds, and to evaluate precision. Both intrasite precision and temporal precision estimates were calculated for the standardized metrics and SWMMI. Intrasite precision was used to determine if the method was sufficiently precise and was calculated using replicate samples (samples collected on the same day at the same site). Temporal precision was used to determine the confidence around a single sample collection. The CV and CI90 were calculated for both intrasite precision and temporal precision; however, CV

$((\sigma/\mu) * 100)$ was used to evaluate method variability and CI90 was used to evaluate the confidence around the collection of a single sample (incremental degradation or improvement measurements). Additionally, two sets of temporal precision estimates were calculated; one set for all sites regardless of the stressor gradient condition category, and one set of estimates of only sites within the LD condition. The CI90 was calculated using the following formula:

$$CI90 = 1.282 * \left(MSE^{0.5} / n^{0.5} \right)$$

Where:

- CI90 = 90% confidence interval
- MSE = analysis of variance mean square error
- n = the number of samples (typically 1 sample)

Results

Comparisons of final metric selection between the summer and fall calibration datasets showed moderate differences. There were six metrics that were selected for each SWMMI, but only two metrics were the same. The summer SWMMI tended to select metrics that focused on tolerance, whereas the fall SWMMI selected more richness and functional metrics. As a result, separate SWMMIs based on season were developed, and summer and fall SWMMI development results are discussed separately in the following sections.

Summer SWMMI

Initial metric DE evaluations using a 70% threshold reduced 128 candidate metrics to 16 candidate metrics. Redundancy analysis using Spearman rank correlation coefficients further reduced the number of viable metrics to 8 candidate metrics (Table 3-2). Metrics with Spearman rank correlations of r > 0.75 were further analyzed using scatterplot matrices. Highest redundancy was observed between the Hilsenhoff index using BCG attributes (BCGindex2) and percent tolerant individuals using BCG attributes (BCGpct5) along with percent intolerant individuals using PTV attributes (PTVpct03) with correlations of r = 0.81 and r= 0.80, respectively. Correlations suggest only approximately 65% of one metric is explaining the other. Additionally, scatterplots showed sufficient variation at the tails of each distribution suggesting these metrics are responding differently depending on environmental conditions (Figure 3-3). Consequently, these metrics were retained for further analysis. Analysis of minimum and range values within the LD condition suggested good performance from all metrics. Percent intolerant Ephemeroptera using BCG attributes (pctEbcg13) was flagged for

having a minimum value lower than 10%; however, only 1 out of 41 samples in the LD condition had a value lower than 10%, while all other values were greater than 13%. This suggested that the minimum value was likely an outlier, so the metric was retained. All metrics except for BCGpct5 and BCGpct456 had SOI ratios < 1; however, these two metrics were very close to 1 and were consequently retained. Additionally, all metrics were adequately correlated (r > 0.4) to the abiotic stressor gradient developed from the first principal component (explaining 49% of variance) of PCA. The main contributors to the PCA stressor gradient were percent development of the watershed, total chloride, total nitrogen, and specific conductance. The second principal of the PCA explained 16% of the variance and was mainly driven by total sulfate and total phosphorus. As a result of these analyses, all 8 candidate metrics passed refinement analysis, but many of these metrics were tolerance metrics that did not add to the body of biological information. Thus, the best performing metrics were not necessarily included in the final metric list; rather, metrics of varying attributes were preferentially selected. The final six metrics included a combination of richness, composition, tolerance, and dominance attributes (Table 3-2), which comprised a diverse set of metrics. The final six metrics had good to excellent DE (Figure 3-4) and responded well when compared to the stressor gradient composite scores developed in Chapter 2 (Figure 3-5).

Table 3-2. Summer metric evaluations and final selection. RTD indicates the direction of metric response to disturbance, MR indicates maximum redundancy observed during Spearman rank correlation analysis, SOI indicates the scope of impairment ratio, and stressor gradient correlation (SGC) indicates the Pearson correlations between metrics and the first principal component of PCA analysis. The min, median, max, range, and MR values were based on the LD data. Values highlighted in gray were flagged for exceeding recommended limits.

Metric Code	Metric Name	Metric Type	RTD	Min	Median	Max	Range	DE	MR	SOI	SGC (PC 1)	Selected
BCGpct456	Percent Tolerant/Invasive Individuals (BCG 4-6)	Tolerance	POS	37.7	61.4	81.5	43.9	92.1	0.78	1.3	0.55	
BCGpct5	Percent Tolerant Individuals (BCG 5)	Tolerance	POS	21.1	41.7	65.0	43.9	92.1	0.81	1.1	0.54	X
PTVpct03	Percent Sensitive Individuals (PTV 0-3)	Tolerance	NEG	12.2	33.2	59.4	47.2	92.1	0.80	0.4	-0.62	X
BCGindex2	Hilsenhoff Index (BCG attributes)	Tolerance	POS	3.7	4.1	4.6	1.0	89.5	0.81	0.7	0.60	X
pctEbcg13	Percent Ephemeroptera (BCG 1-3)	Composition	NEG	5.6	24.0	51.0	45.5	86.8	0.73	0.5	-0.41	X
pctDOM	Percent Dominant Taxon	Dominance	POS	12.7	19.2	34.2	21.5	84.2	0.37	0.7	0.41	X
PTVbeck3	Beck Index (PTV 0-3)	Tolerance	NEG	3	10	18	15	81.6	0.47	0.7	-0.59	
richEPTbcg	EPT Richness (BCG1-3)	Richness	NEG	4	8	13	9	76.3	0.78	0.3	-0.62	X
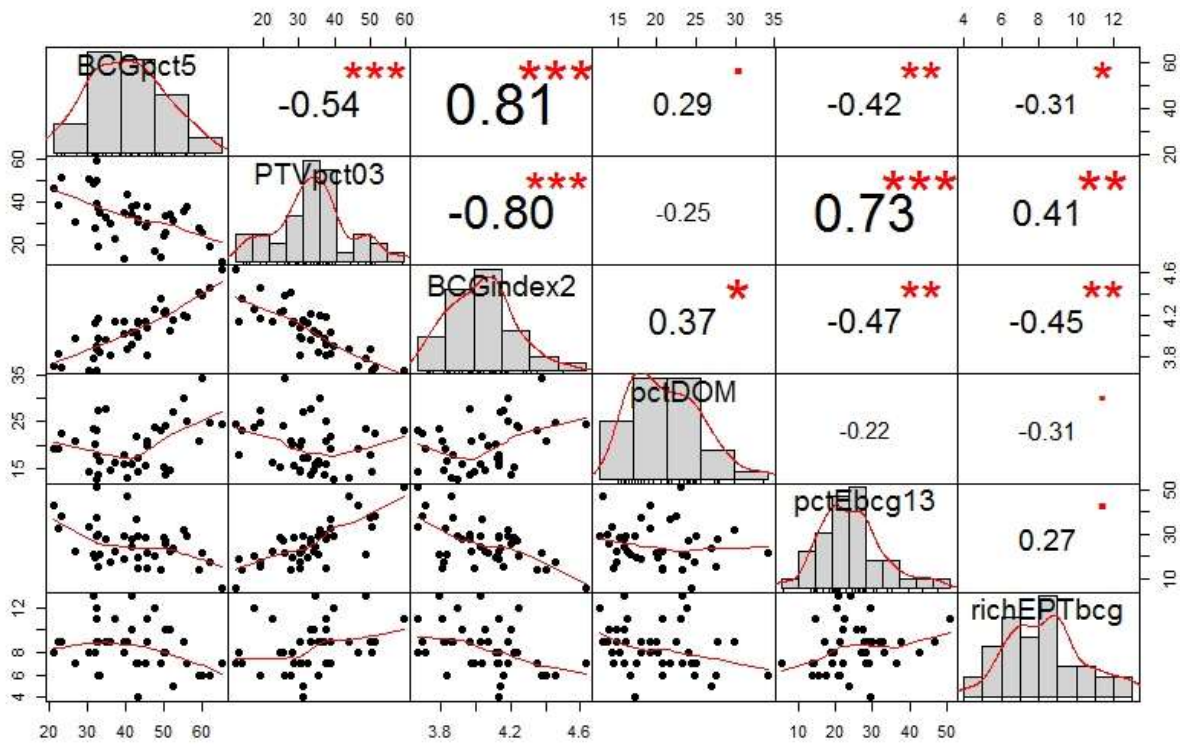
Figure 3-3. Summer Spearman rank correlation coefficients of final selected metrics within the LD condition. * indicates significance at p ≤ 0.05, ** indicates significance at p ≤ 0.01, and *** indicates significance at p ≤ 0.001. Red lines in scatterplots indicate loess model fit.
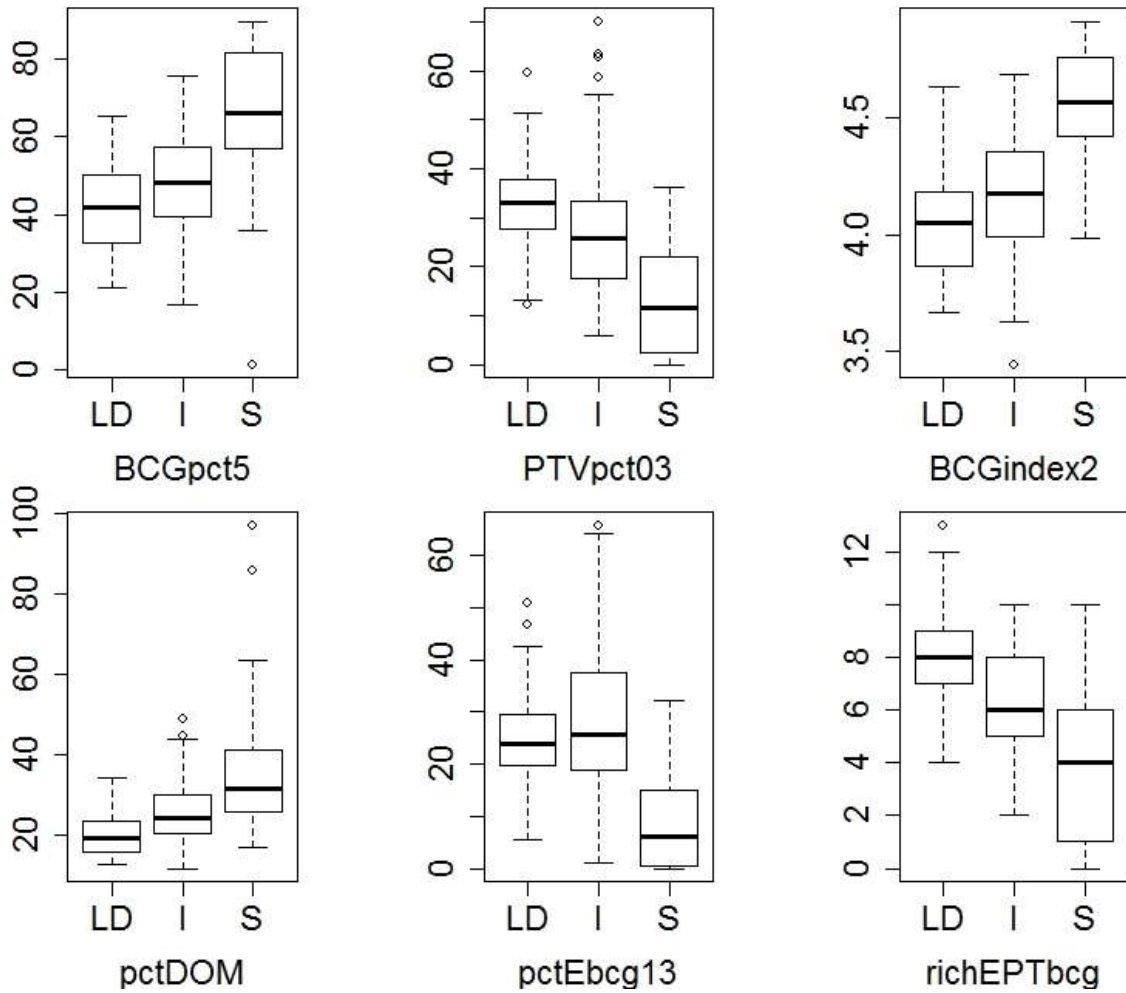
Figure 3-4. Metric boxplots for selected summer metrics. BCGpct5, BCGindex2, and pctDOM were positive response metrics, while PTVpct03, pctEbcg13, and richEPTbcg were negative response metrics.
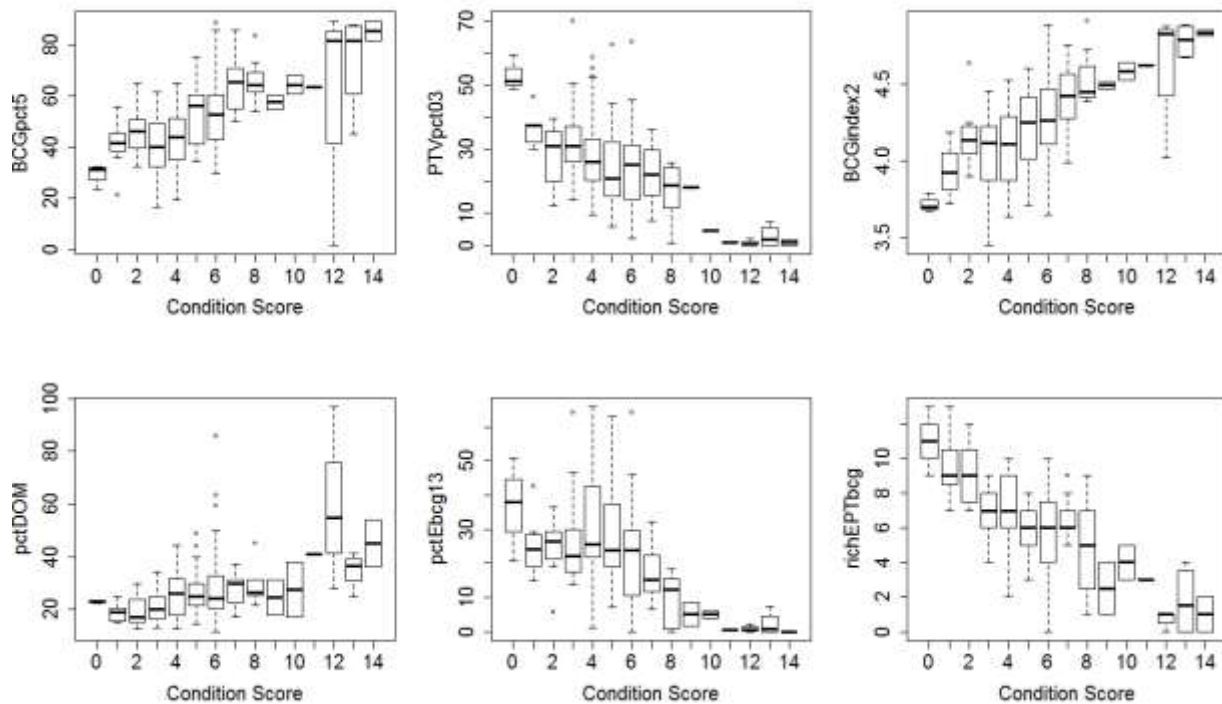
Figure 3-5. Summer metric responses to the stressor gradient composite scores, where 0 indicated the least amount of disturbance and 14 indicated the most amount of disturbance.

The final six metrics were averaged together to obtain a summer SWMMI using the calibration dataset. Evaluation of DE between the LD and S conditions indicated high performance (DE = 97%, Figure 3-6). When the summer SWMMI was related to the principal component of PCA, results suggested that the summer SWMMI was responsive to abiotic stressors ($p < 0.001$, $R^2 = 0.47$, Figure 3-7). Natural factor analysis did not show correlation values that would suggest the summer SWMMI was being greatly affected by natural factors (Table 3-3). Spearman rank correlations were adequately low ($r < 0.50$) and most of the natural parameters measured were not significantly correlated ($p > 0.05$). Since the summer SWMMI was sufficiently responsive and did not respond to natural factors, further evaluation or restandardization of metrics was not necessary. There were four impairment thresholds evaluated including the 5th, 10th, and 25th percentile of LD condition, along with the BSP described by Smith et al. (2017). Resulting impairment thresholds were 44, 49, 61, and 53 respectively. Results from DE analysis suggest a high fidelity of good quality sites being placed into the LD condition with minimal, but some apparent error. As a result, the high confidence 25th and low confidence 5th percentile thresholds were disregarded. Both the 10th percentile and the BSP thresholds closely agreed with each other (Figure 3-6), which suggested that either may serve to adequately reduce the likelihood of

69

making type I and type II error during assessments. Hence, final impairment threshold selection was decided during validation dataset analysis.
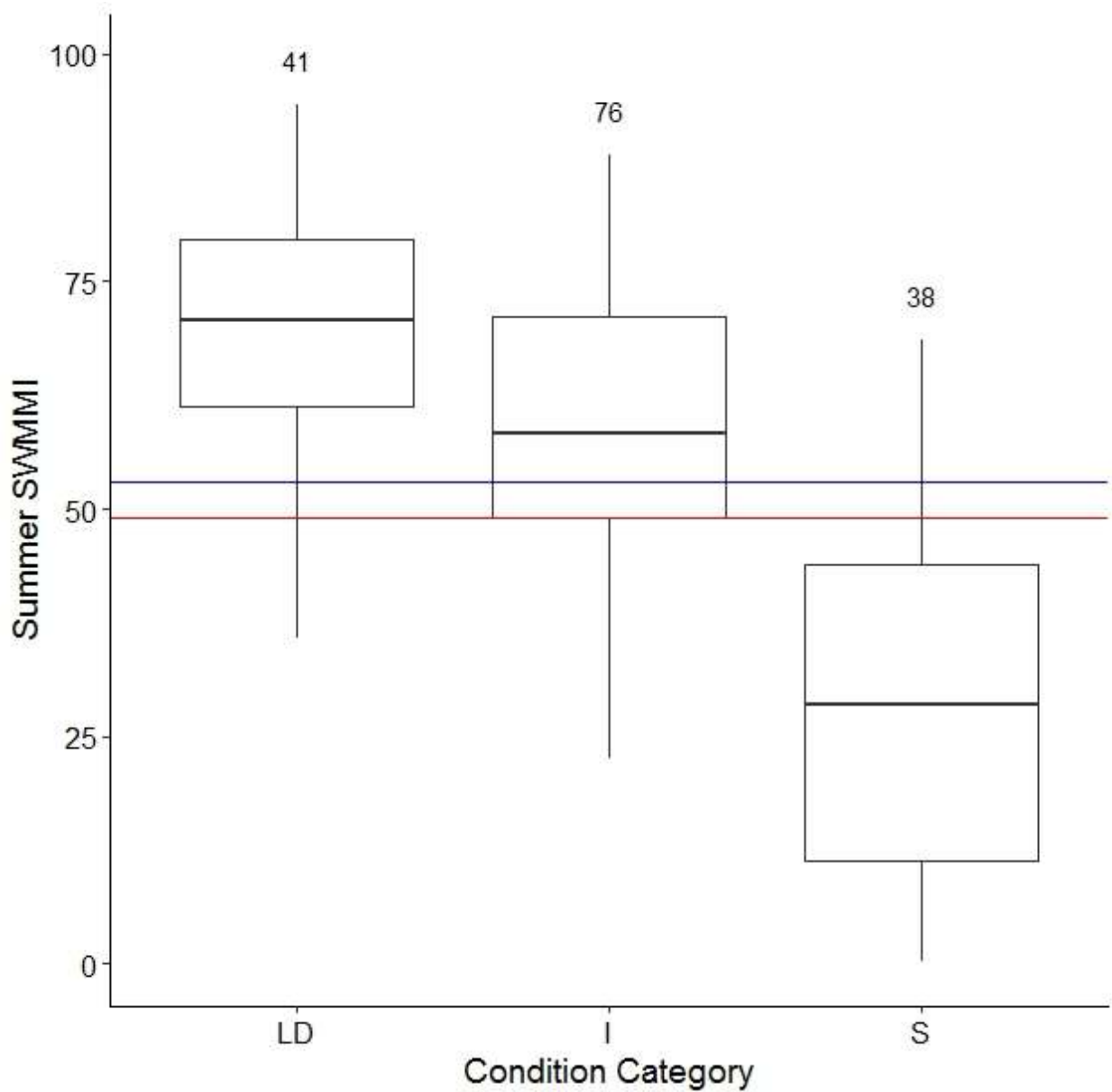
Figure 3-6. Calibration summer SWMMI response to condition categories. Numbers above each boxplot indicate samples size in each condition category. Red line indicates impairment threshold based on the 10th percentile of samples in the LD condition (49), and the blue line indicates the impairment threshold based on the BSP method (53).
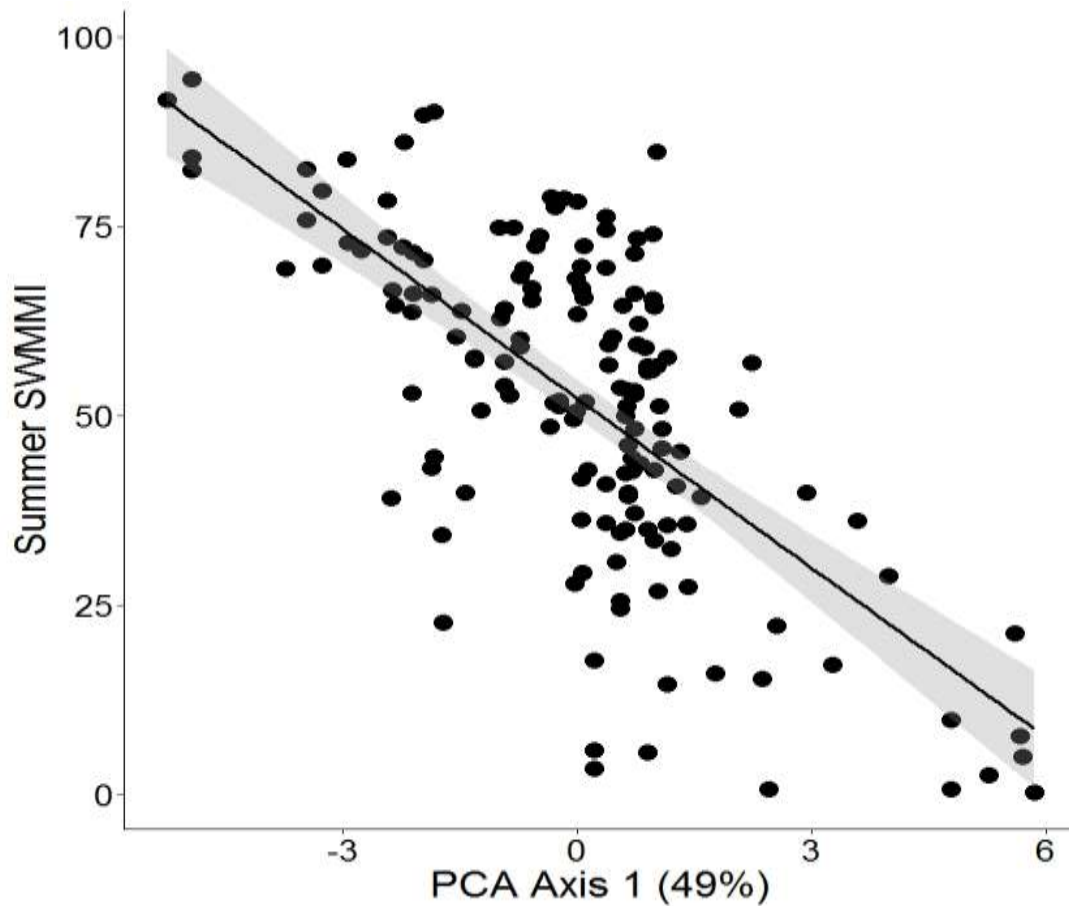
Figure 3-7. Summer SWMMI response to stress. The black line and shaded area indicate linear model fit and standard error, respectively. Based on PCA analysis, higher positive values indicated more stressful abiotic conditions and lower values indicated less stressful abiotic conditions.

Table 3-3. Spearman rank correlation coefficients between natural parameters and summer SWMMI scores within the LD condition of the calibration dataset. * indicates significance at $p \leq 0.05$, and ** indicates significance at $p \leq 0.01$.

Natural Parameters	LD Summer SWMMI Correlations
Month	0.26*
Basin Size (km^2)	-0.15
Latitude	-0.32*
Longitude	0.21
Slope	0.11
Alkalinity	0.10

The validation dataset was imported, metrics were standardized based on values from the calibration data, and standardized metrics were then averaged together to obtain a summer SWMMI for the validation dataset. Evaluation of CE using both impairment thresholds between the LD and S conditions produces good results regardless of the impairment threshold used (Figure 3-8), suggesting the calibration dataset was not over fitted to the stressor gradient. However, the 10th percentile impairment threshold was able to properly classify a higher percentage of LD and S sites (CE = 84%) than the BSP impairment threshold (CE = 82%). Therefore, the 10th percentile of reference may be a more accurate impairment threshold for this dataset. As an additional check, the validation summer SWMMI within the LD condition was compared to natural factors Results suggest that variation within the LD condition of the validation dataset were not being driven by natural factors (Table 3-4). All Spearman rank correlations were adequately low (r ≤ 0.50) and most of the natural parameters measured were not significantly correlated ($p > 0.05$).
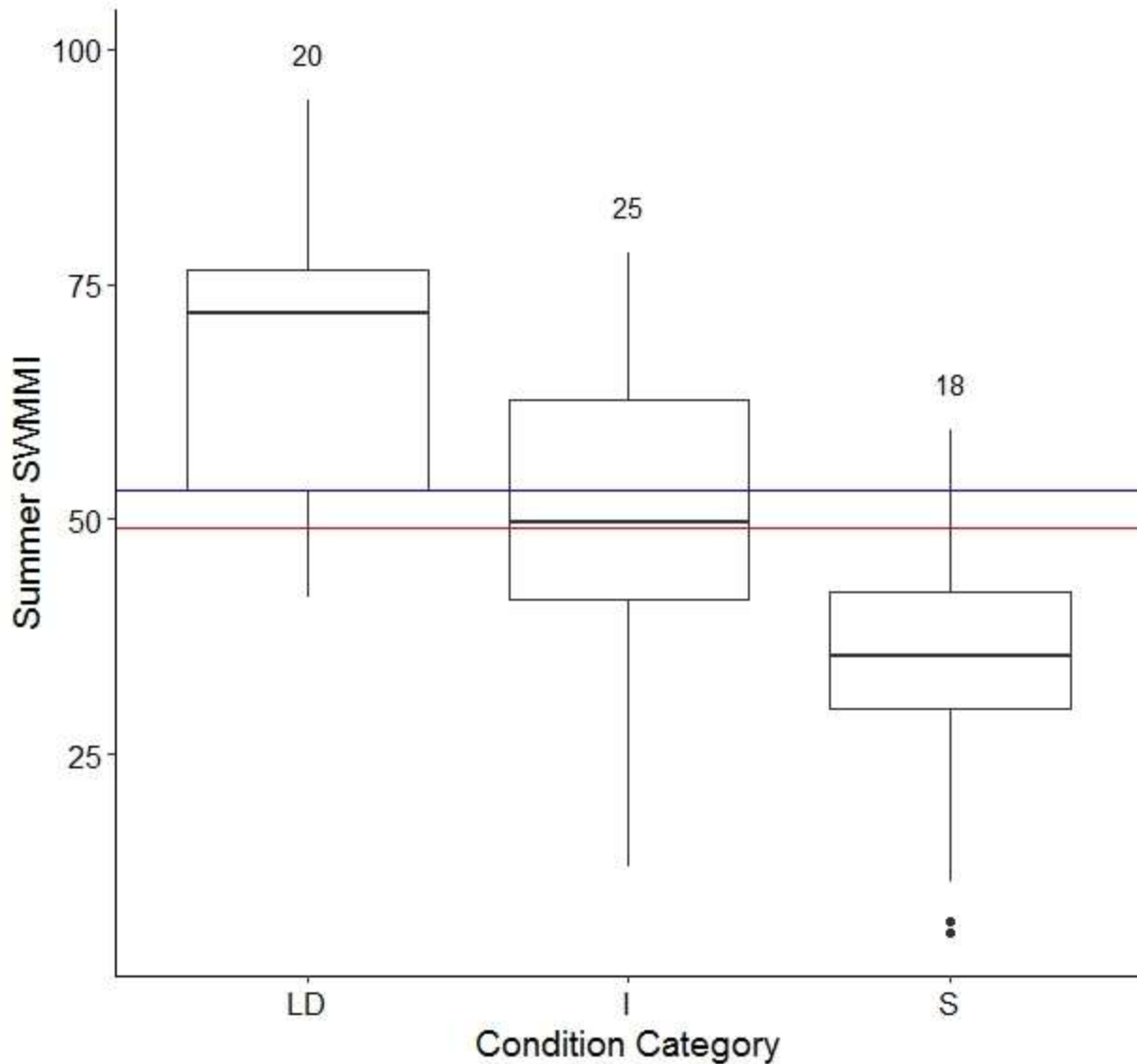
Figure 3-8. Summer SWMMI calculated from the validation dataset versus condition category. Numbers above each boxplot indicate sample size in each condition category. Red line indicates impairment threshold based on the 10th percentile of samples in the LD condition (49), and the blue line indicates the impairment threshold based on the BSP method (53).

Table 3-4. Spearman rank correlation coefficients between natural parameters and summer SWMMI scores within the LD condition of the validation dataset. * indicates significance at p ≤ 0.05.

Natural Parameters	LD Summer SWMMI Correlations
Month	0.39*
Basin Size (km^2)	-0.19
Latitude	0.38*
Longitude	0.28
Slope	0.26
Alkalinity	-0.50*

Both datasets were subsequently combined to obtain the final metric standardization values, the final summer SWMMI impairment threshold, and to calculate intrasite and temporal precision. The final summer SWMMI had a DE of 98%, CE of 85%, and a final impairment threshold of 49 (Figure 3-9). There were 24 samples from 12 sites that were available for intrasite precision estimation. Method precision measured by the CV of replicate samples suggested that the summer SWMMI was adequately precise (CV = 8.8%, Table 3-5). Temporal precision (CI90 of 1 sample) of all stressor gradients and only the LD category was 14.7 (Table 3-6) and 14.2 (Table 3-7) points, respectively. Consequently, summer SWMMI scores that vary 15 points or more can be interpreted as true changes to the biological condition with a relative degree of confidence.
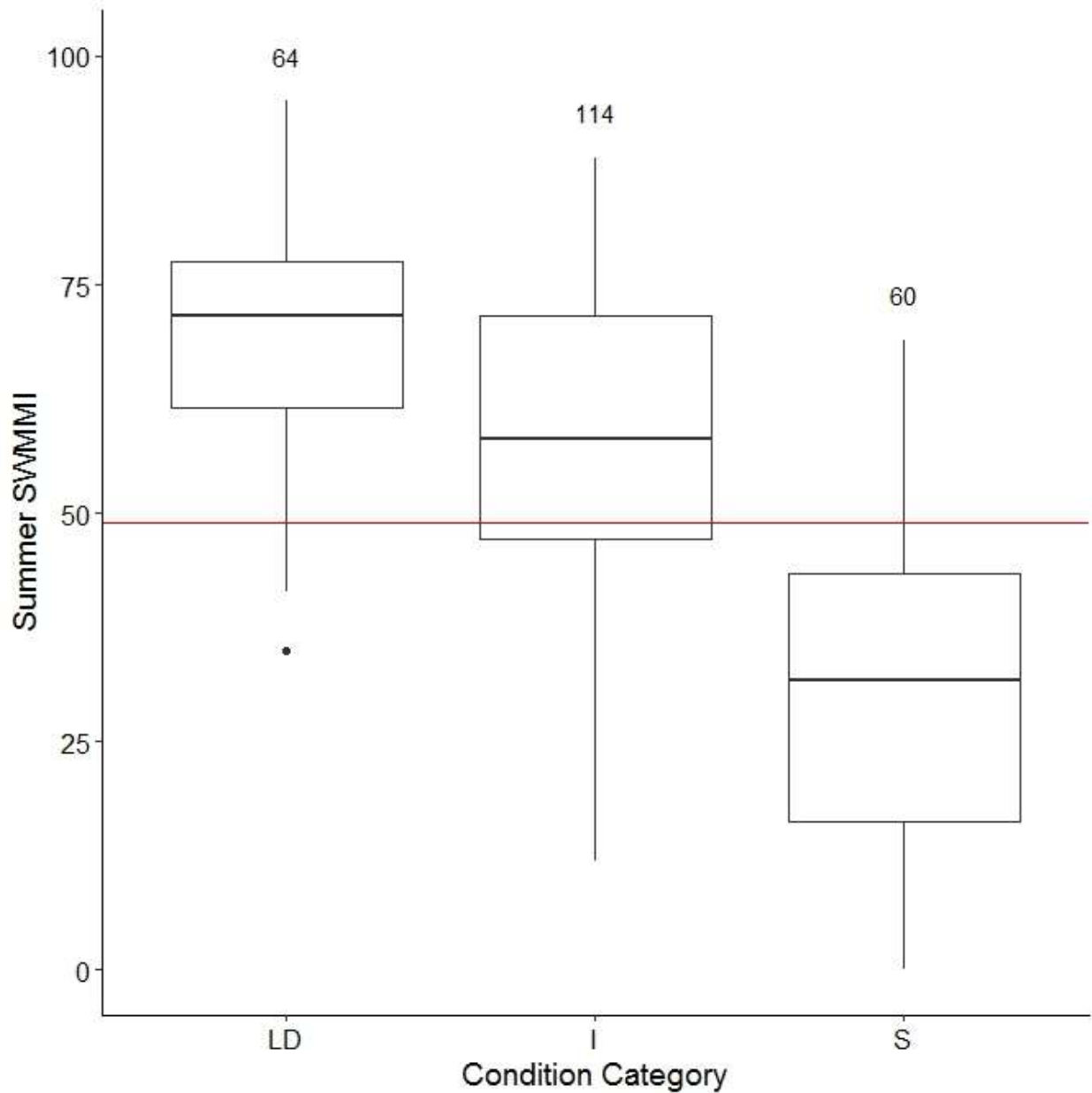
Figure 3-9. Summer SWMMI calculated from the final combined dataset versus condition category. Numbers above each boxplot indicate sample size in each condition category. Red line indicates impairment threshold based on the 10th percentile of samples in the LD condition (49).

Table 3-5. Summer SWMMI intrasite precision estimates using replicate samples (samples collect at the same site on the same day). MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. SWMMI precision estimates are highlighted in gray.

		12 Sites 24 Samples			
Metric	MSE	CI90 (1 sample)	Mean	sd	CV
SWMMI	34.4	7.5	52.9	4.6	8.8
BCGpct5	133.9	14.8	62.0	7.3	15.8
PTVpct03	46.4	8.7	42.1	5.8	21.0
BCGindex2	163.8	16.4	53.1	7.9	18.4
pctDOM	150.5	15.7	72.9	7.7	12.5
pctEbcg13	43.6	8.5	43.5	5.1	22.4
richEPTbcg	119.4	14.2	43.7	9.0	28.8

Table 3-6. Summer SWMMI temporal precision estimates using samples collected over several years across all condition gradients. MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. Summer SWMMI precision estimates are highlighted in gray.

		43 sites 100 samples			
Metric	MSE	CI90 (1 sample)	Mean	sd	CV
SWMMI	130.9	14.7	60.5	9.7	23.3
BCGpct5	270.0	21.1	63.7	13.4	27.2
PTVpct03	317.4	22.8	53.2	14.8	37.3
BCGindex2	291.0	21.9	60.8	13.7	31.4
pctDOM	649.0	32.7	69.9	20.0	36.1
pctEbcg13	365.9	24.5	52.8	14.4	35.6
richEPTbcg	170.7	16.7	62.7	10.7	22.4

Table 3-7. Summer SWMMI temporal precision estimates using samples collected over several years, but only within the LD condition category. MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. SWMMI precision estimates are highlighted in gray.

Metric	MSE	CI90 (1 sample)	Mean	sd	CV
		16 sites 36 samples			
SWMMI	123.8	14.2	72.1	8.8	12.8
BCGpct5	119.0	14.0	76.1	7.7	11.1
PTVpct03	253.3	20.4	65.8	12.1	20.5
BCGindex2	259.8	20.7	73.6	11.1	16.7
pctDOM	221.1	19.1	84.7	11.6	14.8
pctEbcg13	518.8	29.2	50.9	16.7	32.0
richEPTbcg	140.7	15.2	81.5	9.0	12.3

Fall SWMMI

Metric DE evaluations using a 70% threshold reduced 128 candidate metrics to 30 candidate metrics. Redundancy analysis using Spearman rank correlation coefficients further reduced the number of viable metrics to 13 candidate metrics (Table 3-8). No candidate metrics were redundant (r < 0.75); therefore, no additional analysis was needed to determine whether metrics should be retained. Analysis of minimum and range values within the LD condition showed several metrics including percent *Ephemerella* spp. (pctEPHMRLA), richness of sensitive taxa with a BCG attribute 2 (BCGrich2), and percent sensitive individuals with PTV attributes of 0-2 (PTVpct02) failed initial screening criteria. All metrics with the exception of pctEPHMRLA and PTVpct02 had SOI ratios < 1. Since these metrics also were flagged for minimum and range value exceedances, they were no longer considered for inclusion in the fall SWMMI. All metrics except for percent sensitive individuals with BCG attribute 3 (BCGpct3) were acceptably correlated (r > 0.4) to the stressor gradient developed from the first principal component (explaining 49%) during PCA analysis. As with the summer metrics selection, final selection of six metrics for standardization and index development was based on incorporating varying types of metrics. A total of 9 metrics passed refinement analysis, but many of these metrics consisted of richness metrics that did not appreciably add to the body of biological information. Therefore 9 metrics were reduced to 6 based on metric type evaluation. The final set of metrics showed acceptable correlation and spread within the scatterplots (Figure 3-10), had ample DE (Figure 3-11), and responded well when compared to the stressor gradient composite scores developed in Chapter 2 (Figure 3-12).

Table 3-8. Fall metric evaluations and final selection. RTD indicates metric response to disturbance, MR indicates maximum redundancy observed during Spearman rank correlation analysis, SOI indicates the scope of impairment ratio, and stressor gradient correlation (SGC) indicates the Pearson correlations between metrics and the first principal component of PCA analysis. The min, median, max, range, and MR values were based on the LD data. Values highlighted in gray were flagged for exceeding recommended limits.

Metric Code	Metric Name	Metric Type	RTD	Min	Median	Max	Range	DE	MR	SOI	SGC (PC 1)	Selected
PTVbeck3	Beck Index (PTV 0-3)	Tolerance	NEG	3.0	11.0	24.0	21.0	97.0	0.70	0.6	-0.64	X
richEPTptv	EPT Richness (PTV 0-4)	Richness	NEG	7.0	12.5	18.0	11.0	90.9	0.74	0.3	-0.62	X
PTVpct03	Percent Sensitive Individuals (PTV 0-3)	Tolerance	NEG	23.6	46.6	69.1	45.5	84.8	0.69	0.8	-0.63	X
PTVrich05	Richness (PTV 0-5)	Richness	NEG	14.0	18.0	23.0	9.0	81.8	0.70	0.2	-0.55	
pctEbcg13	Percent Sensitive Ephemeroptera (BCG 1-3)	Composition	NEG	11.4	34.1	58.5	47.0	78.8	0.69	0.7	-0.47	X
PTVpct02	Percent Sensitive Individuals (PTV 0-2)	Tolerance	NEG	7.3	21.8	46.8	39.5	78.8	0.60	1.9	-0.58	
Richness	Total Richness	Richness	NEG	18.0	24.0	32.0	14.0	78.8	0.70	0.2	-0.48	X
BCGpct5	Percent Tolerant Individuals (BCG 5)	Tolerance	POS	13.5	33.7	64.0	50.5	75.8	0.53	0.6	0.41	
BCGpct3	Percent Sensitive Individuals (BCG 3)	Tolerance	NEG	14.8	29.9	67.6	52.8	72.7	0.53	0.9	-0.36	
BCGrich2	Richness (BCG 2)	Richness	NEG	1.0	2.5	6.0	5.0	72.7	0.60	0.5	-0.52	
FFGrichSC	Scraper Richness	Functional	NEG	5.0	8.0	11.0	6.0	72.7	0.51	0.3	-0.45	X
pctEPHMRLA	Percent Ephemerella	Composition	NEG	0.0	7.1	25.6	25.6	72.7	0.60	19.6	-0.43	
richEbcg13	Ephemeroptera Richness (BCG 1-3)	Richness	NEG	3.0	5.0	8.0	5.0	72.7	0.61	0.2	-0.54	
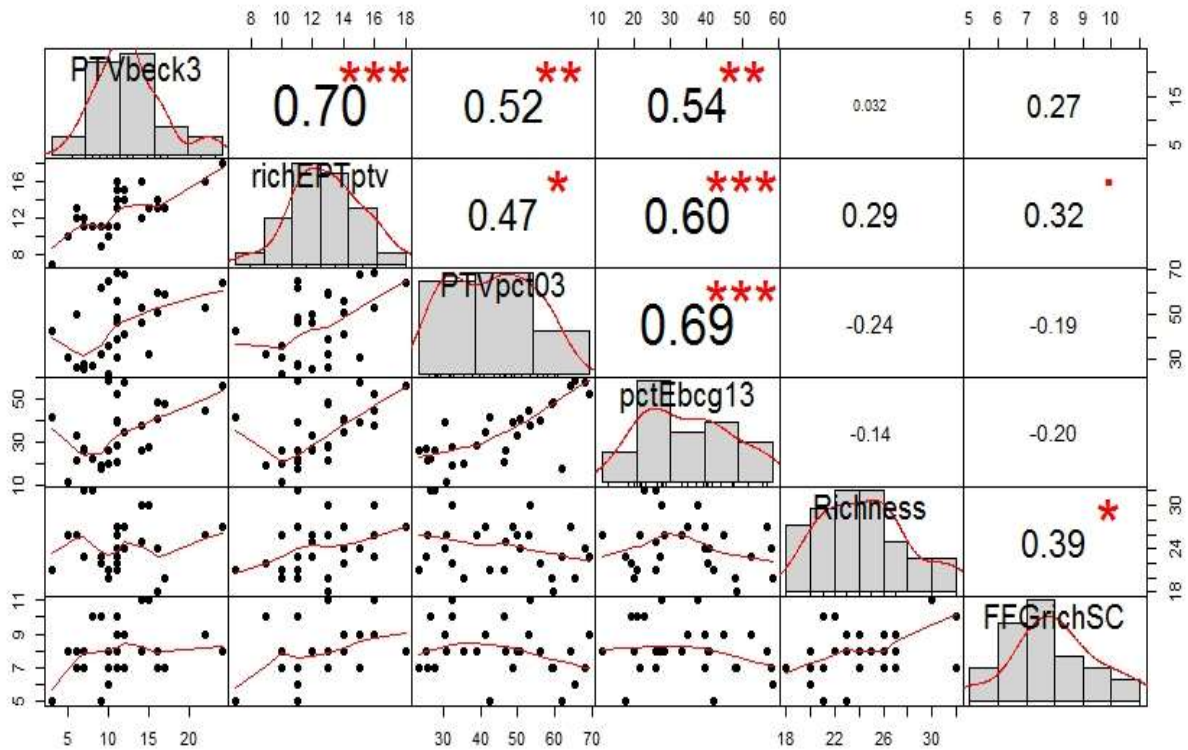
Figure 3-10. Fall Spearman rank correlation coefficients of final selected metrics within the LD condition. * indicates significance at p ≤ 0.05, ** indicates significance at p ≤ 0.01, and *** indicates significance at p ≤ 0.001.
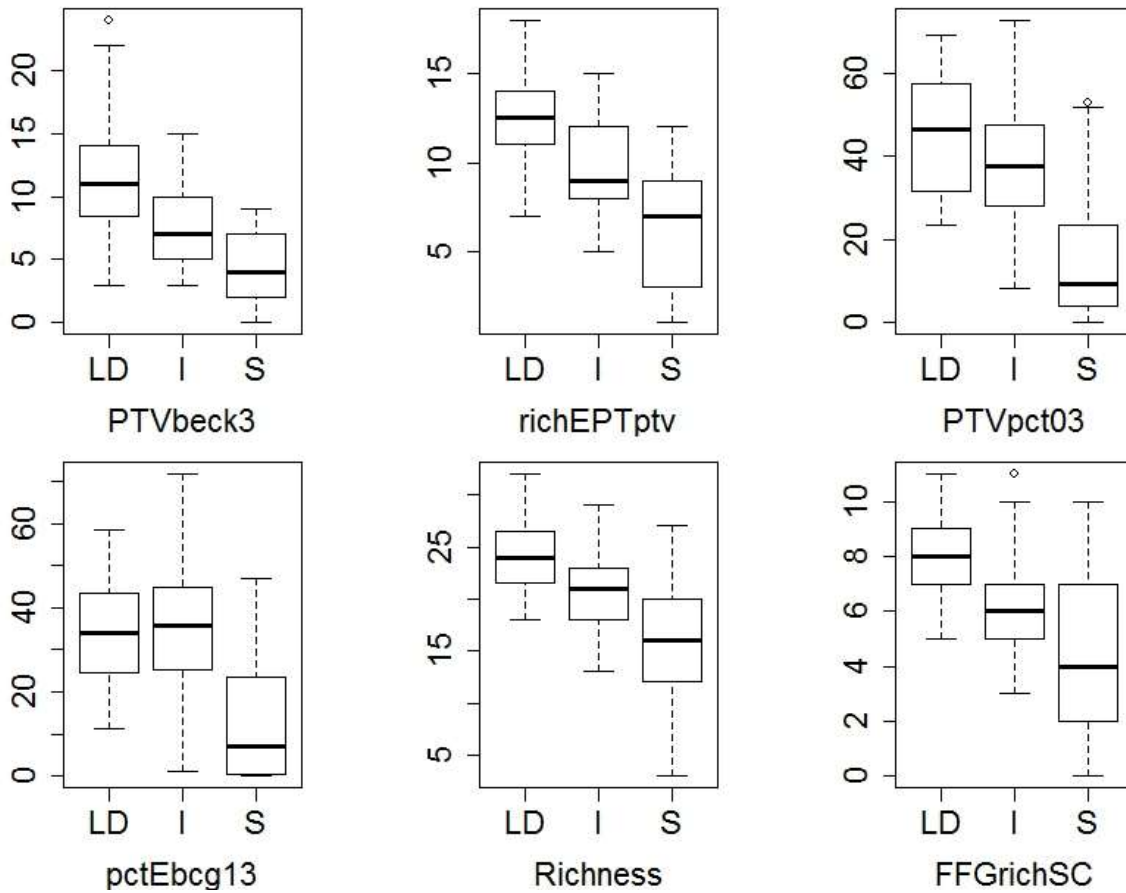
Figure 3-11. Metric boxplots for selected fall metrics. All selected metrics were negative response metrics.
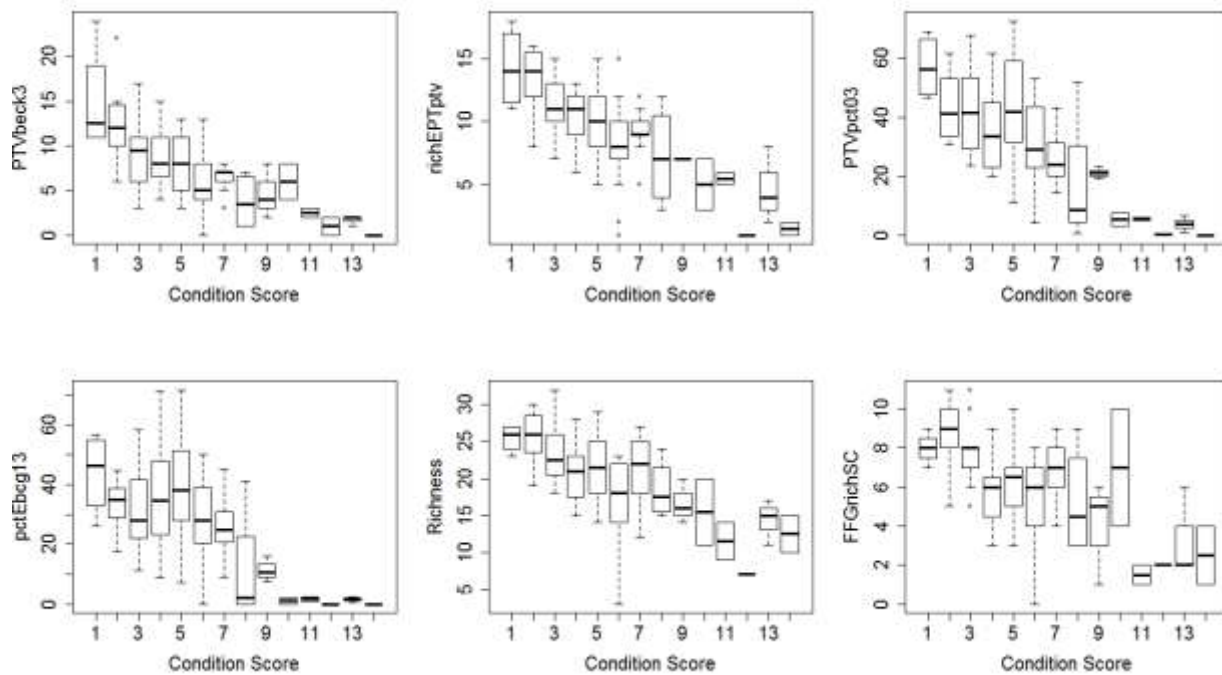
Figure 3-12. Fall metric responses to the stressor gradient composite scores developed in chapter 2, where 0 indicated the least amount of disturbance and 14 indicated the most amount of disturbance.

Metrics were then standardized and averaged together to obtain a fall SWMMI using the calibration dataset. Evaluation of DE between the LD and S conditions indicated high performance (DE = 88%, Figure 3-13). The fall SWMMI was significantly responsive to abiotic stressors (PCA; $R^2 = 0.45$, $p < 0.001$, Figure 3-14). The calibration samples within the LD condition did not show correlation values that suggested the fall SWMMI was being affected by natural variables (Table 3-9). All Spearman rank correlations were adequately low ($r \leq 0.50$) and most natural parameters were not significantly correlated ($p > 0.05$).

As with the summer SWMMI, four impairment thresholds were evaluated including the 5th, 10th, and 25th percentile of LD condition, along with the BSP described by Smith et al. (2007). Resulting impairment thresholds were 53, 56, 59, and 57, respectively. However, similar to the summer analysis, the high confidence 25th and low confidence 5th percentile thresholds were disregarded. Both the 10th percentile and the BSP thresholds closely agreed with each other (Figure 3-13), which implies that either may serve to adequately reduce the likelihood of making type I and type II error during future assessments. However, since the results between the 10th percentile and BPS were so closely matched, and because the 10th percentile of the LD condition was selected for

the summer SWMMI, the 10th percentile of the LD condition was selected as the impairment threshold for the fall SWMMI.
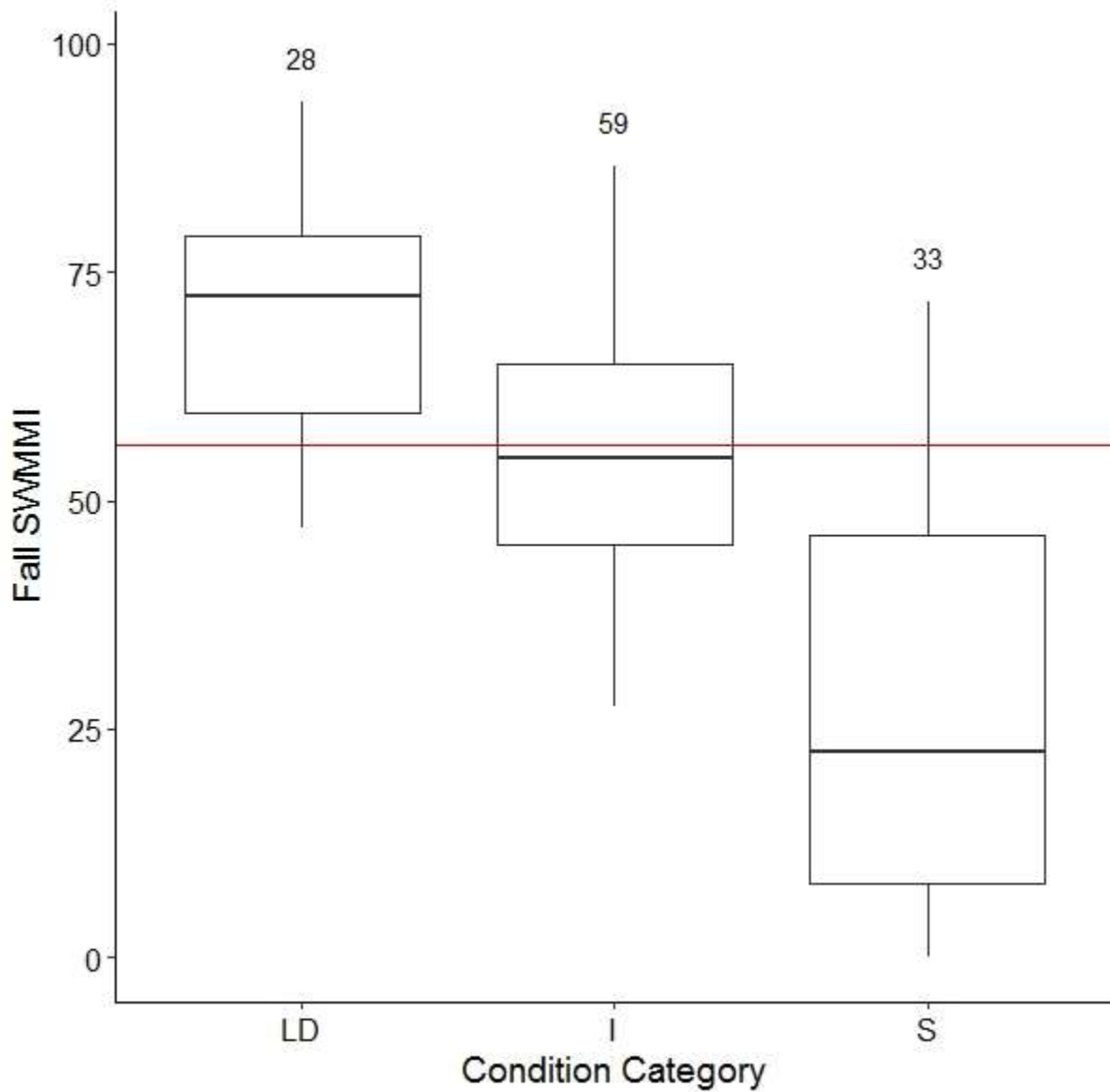
Figure 3-13. Fall SWMMI of the calibration dataset versus condition category. Numbers indicate sample size in each condition category. Red line indicates impairment threshold based on the 10th percentile of samples of the LD condition (56).
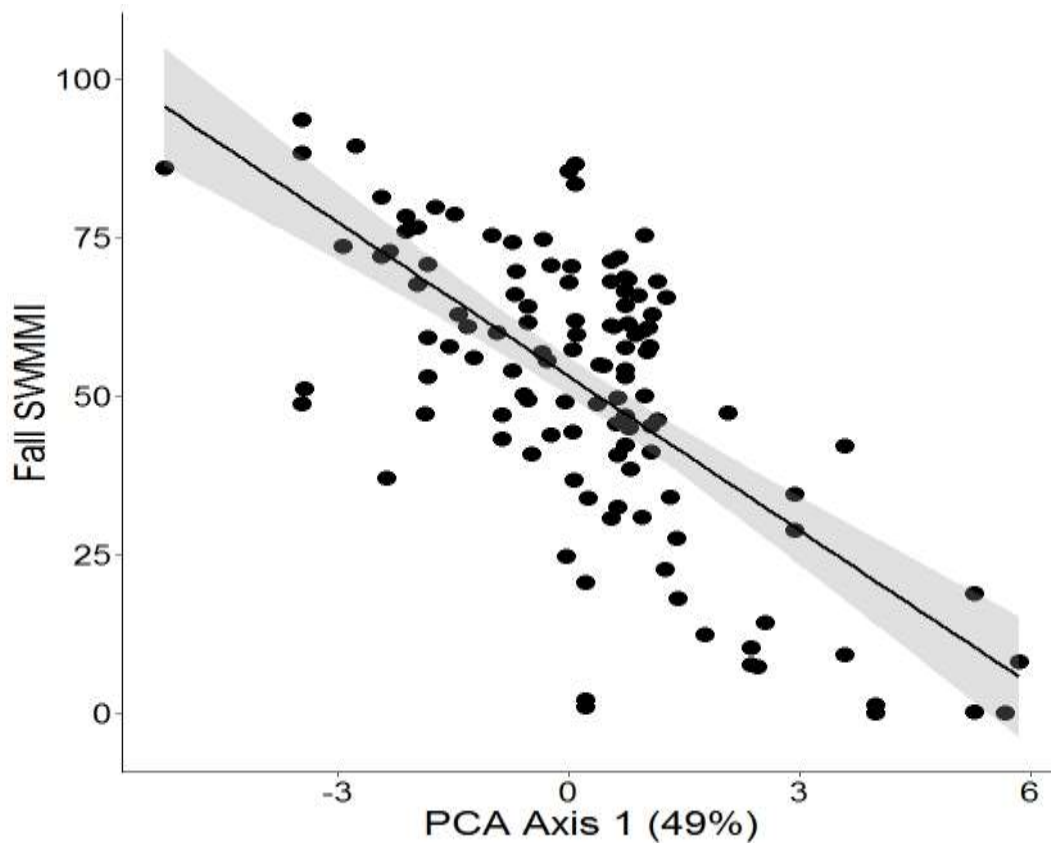
Figure 3-14. Fall SWMMI response to stress. The black line and shaded area indicate linear model fit and standard error, respectively. Based on PCA analysis, higher positive values indicated more stressful abiotic conditions and lower values indicated less stressful abiotic conditions.

Table 3-9. Spearman rank correlation coefficients between natural parameters and fall SWMMI scores within the LD condition of the calibration dataset. * indicates significance at $p \leq 0.05$, and ** indicates significance at $p \leq 0.01$.

Parameters	LD Calibration Dataset Correlations
Month	0.50**
Basin Size (km^2)	-0.26
Latitude	-0.20
Longitude	0.24
Slope	-0.01
Alkalinity	-0.44*

The fall validation metrics were standardized based on values from the calibration data. Standardized metrics were then averaged together to obtain a fall SWMMI score for each sample. Evaluation of CE using the selected impairment threshold between the LD

and S condition produced good results suggesting the calibration dataset was not over fitted, and the metrics are responsive to the stressor gradient (CE = 87%, Figure 3-15). The fall SWMMI was also checked against natural factors, which resulted in relatively low Spearman rank correlations and few significant correlations among the parameters measured (Table 3-10). Since no correlations exceeded thresholds (r < 0.50) or were significant (p > 0.05), further evaluation of natural factors and restandardization of the fall SWMMI was not necessary.
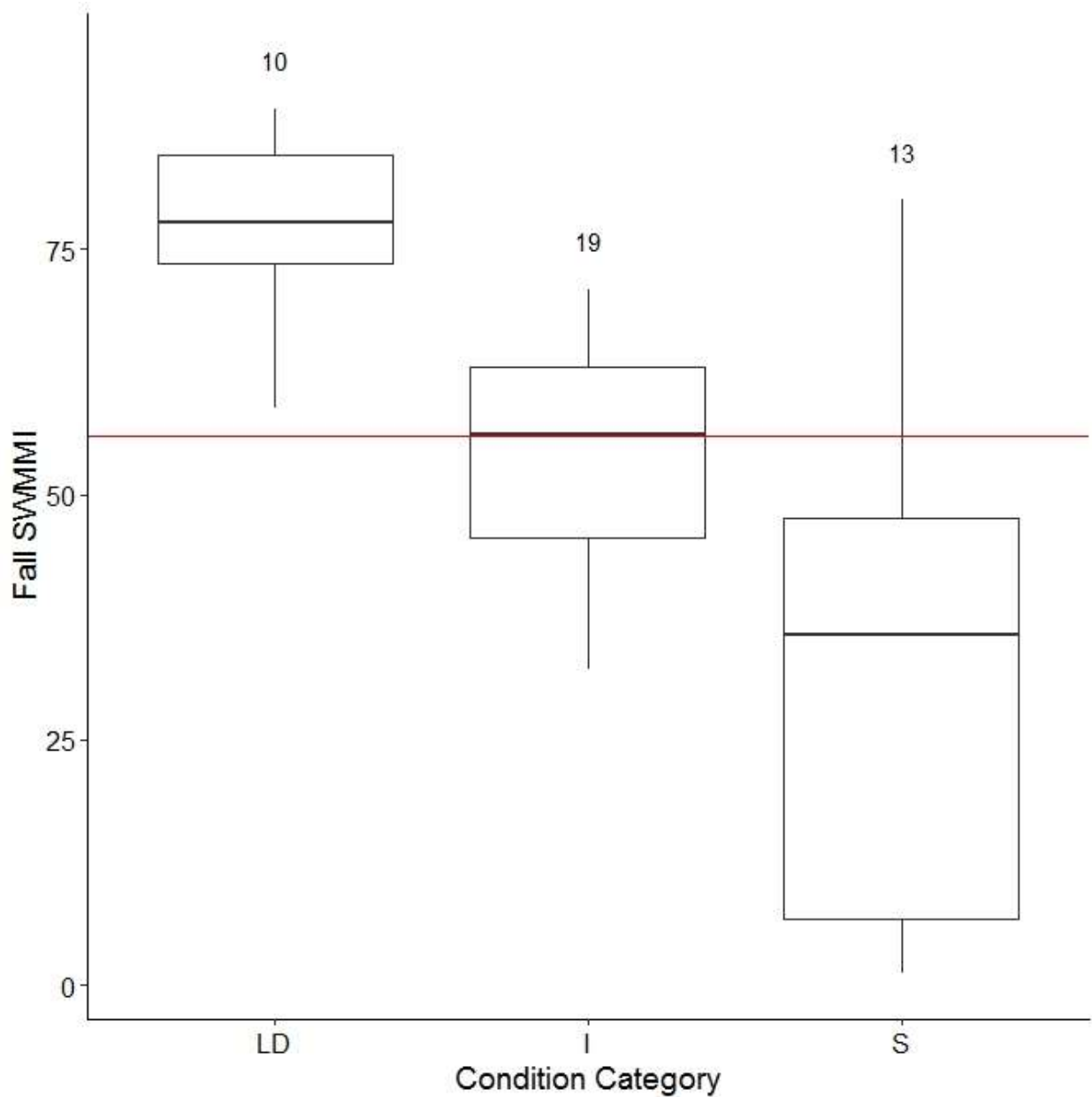
Figure 3-15. Fall SWMMI of the validation dataset versus condition category. Numbers indicate sample size in each condition category. Red line indicates impairment threshold

based on the 10th percentile of samples in the LD condition of the calibration dataset (56).

Table 3-10. Spearman rank correlation coefficients between natural parameters and fall SWMMI scores within the LD condition of the validation dataset. * indicates significance at $p \leq 0.05$.

Parameters	LD Validation Dataset Correlations
Month	0.10
Basin Size (km^2)	-0.41
Latitude	0.26
Longitude	0.39
Slope	0.16
Alkalinity	-0.43

Both datasets were then combined to obtain the final metric standardization values, final fall SWMMI impairment threshold, and to calculate intrasite and temporal precision. The final fall SWMMI had a DE of 89%, CE of 86%, and a final impairment threshold of 57 (Figure 3-16). There were 18 samples from 9 sites that were available for intrasite precision estimation. Method precision measured by the CV of replicate samples suggested that the fall SWMMI had a CV within the recommended limits (CV = 14.1%, Table 3-11). Temporal precision (CI90 of 1 sample) of all stressor gradients was 12.8 (Table 3-12) and the LD category was 11.4 (Table 3-13) points.
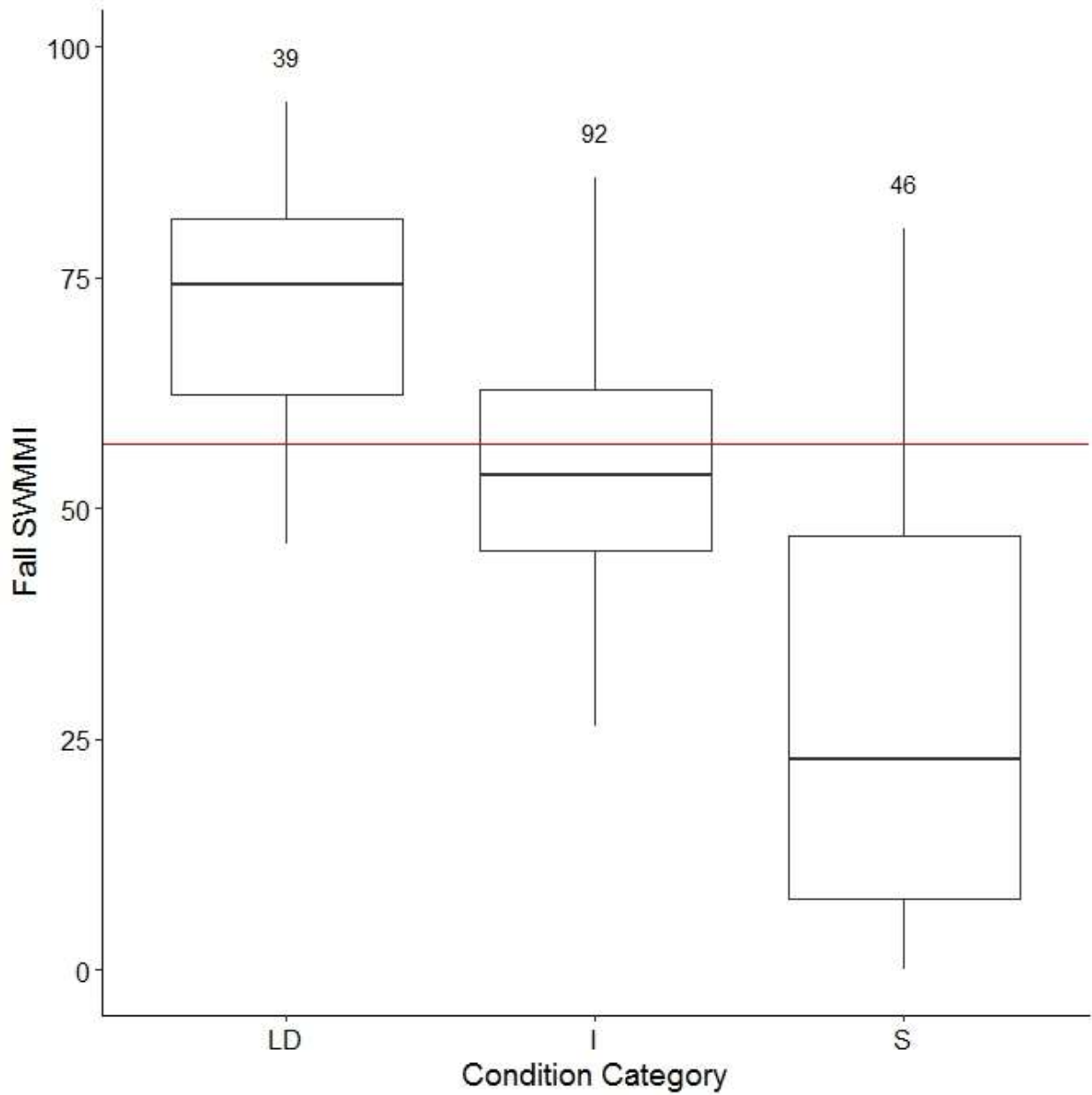
Figure 3-16. Fall SWMMI calculated from the final combined dataset versus condition category. Numbers above each boxplot indicate sample size in each condition category. Red line indicates impairment threshold based on the 10th percentile of samples in the LD condition (57).

Table 3-11. Fall SWMMI intrasite precision estimates using replicate samples (samples collected at the same site on the same day). MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. Fall SWMMI precision estimates are highlighted in gray.

Metric	MSE	9 Sites 18 Samples CI90 (1 sample)	Mean	sd	CV
SWMMI	83.9	11.7	46.2	7.4	14.1
PTVbeck3	358.3	24.2	32.0	12.7	39.4
richEPTptv	78.9	11.4	57.3	7.2	11.1
PTVpct03	130.5	14.6	41.7	9.0	26.4
pctEbcg13	100.0	12.8	47.8	7.6	30.7
Richness	368.9	24.6	48.6	13.7	23.3
FFGrichSC	208.3	18.5	50.0	11.8	22.2

Table 3-12. Fall SWMMI temporal precision estimates using samples collected over several years across all condition gradients. MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. Fall SWMMI precision estimates are highlighted in gray.

Metric	MSE	27 Sites 58 Samples CI90 (1 sample)	Mean	sd	CV
SWMMI	99.2	12.8	45.9	8.3	31.7
PTVbeck3	378.3	24.9	40.0	15.1	49.1
richEPTptv	151.7	15.8	52.1	9.9	29.5
PTVpct03	201.4	18.2	46.3	10.9	40.5
pctEbcg13	275.8	21.3	45.2	12.5	37.4
Richness	368.2	24.6	48.0	15.9	45.2
FFGrichSC	534.3	29.6	44.6	18.5	41.5

Table 3-13. Fall SWMMI temporal precision estimates using samples collected over several years, but only within the LD condition category. MSE indicates ANOVA mean square error, sd indicates standard deviation, and CV indicates coefficient of variation. Fall SWMMI precision estimates are highlighted in gray.

Metric	MSE	4 Sites 9 Samples CI90 (1 sample)	Mean	sd	CV
SWMMI	79.3	11.4	79.8	8.6	8.6
PTVbeck3	465.4	27.6	71.5	22.1	32.3
richEPTptv	232.7	19.5	79.5	14.9	18.8
PTVpct03	180.8	17.2	77.6	14.0	17.6
pctEbcg13	470.8	27.8	63.1	22.1	34.9
Richness	224.0	19.1	73.0	14.0	20.0
FFGrichSC	302.1	22.3	76.0	16.8	22.3

Discussion

To date, this index development process is the most exhaustive and rigorous that PADEP has completed. It is important to note that all the novel and principal MMI development processes completed here were first vetted through peer reviewed publications. Both the summer and fall SWMMI have proven effective at distinguishing between LD and S conditions, thereby providing an applicable and defensible tool for the Commonwealth, and the larger region. Results indicate that the method and resulting SWMMIs are precise, and can detect ecological changes due to anthropogenic stress or restoration efforts over time. This assessment tool can be used in other large rivers outside of the Commonwealth if sample methods are adhered to and ecological regions are comparable (e.g., Blocksom and Johnson 2009).

Differences in metric selection between the two SWWMIs indicate that the summer and fall SWMMI are responsive to stress in different ways due to the unique macroinvertebrate communities that exist between seasons. To a lesser extent, like in Chapter 2, there was also some evidence that natural factors such as alkalinity or specific conductance may be influencing the macroinvertebrate community. Natural relationships with the macroinvertebrate community were weakly significant during natural factor correlation analysis. However, with the current dataset, it is difficult to determine whether these relationships are truly natural or are the result of the inherent incorporation of anthropogenic influences into the LD condition. With more data, it is possible that even more accurate and precise assessment tools could be achieved by compensating for major basin or alkalinity. Future refinements of this protocol should further explore additional classification possibilities.

The metric selection results of the summer evaluation produced fewer acceptable metrics than the fall and specifically leaned toward tolerance metrics as being the most appropriate responders to anthropogenic stress. However, metrics types such as richness and functional groups were relatively lacking. This seems logical since many groups of macroinvertebrates are either in a diapause period of their lifecycle or are too young to identify, but it does not suggest that important groups of taxa were missing during the summer. Both sensitive Ephemeroptera using BCG attributes (pctEbcg13) and richness of sensitive Ephemeroptera, Plecoptera, and Trichoptera using BCG attributes (richEPTbcg) were included in the final summer metric set; suggesting that although there were fewer viable metrics, there were ample metric types to select from. Conversely, the fall SWMMI evaluation tended to select a greater number of acceptable metrics and leaned more toward richness metrics as being the dominant viable metric types and specifically included one functional feeding group metric, scraper richness (FFGrichSC). Again, this makes sense from a logical standpoint since fall collections have the potential to produce more identifiable organisms than summer collections at LD sites. This, coupled with lower taxa richness observed at S sites would increase the DE of richness metrics during the fall season, thereby bringing richness metrics to the top during metric selection. Thus, each tool used separately provides useful information that could not be elucidated if seasonal collections were composited into one assessment tool.

It is no surprise that several of the metrics in the wadeable MMI developed by Chalfant (2012) were also included in the SWMMIs. Both the wadeable MMI and at least one of the SWMMIs included total richness, sensitive EPT taxa richness, percent sensitive individuals, and the Beck's index. It is well known that metrics that include macroinvertebrate groups such as Ephemeroptera, Plecoptera, and Trichoptera are routinely found to be responsive across the greater Mid-Atlantic region (Klemm et al., 2003, Pond et al., 2013, Smith at al. 2017). This suggests that macroinvertebrate communities in semi-wadeable rivers, although somewhat different than wadeable communities, share many of the functional and compositional characteristics. This also helps to confirm that these macroinvertebrate groups are important taxonomic cohorts that consistently and predictably respond to degradation. This is particularly true when these taxonomic groups are coupled with sensitivity attributes. Both the wadeable MMI and the SWMMIs preferentially selected metrics with some sort of sensitivity attribute tied to the metric.

One of the major differences between these SWMMIs, and other MMIs, is the incorporation of BCG attribute metrics along with PTV attribute metrics. Including both sensitivity attributes in a single MMI may serve as a check and balance. Yet, it could be argued that using BCG metrics may be a concern when PADEP has not produced a full list of BCG attributes for every taxon. This concern was addressed through evaluation of

the calibration dataset. It was determined that subsample size of BCG metrics consistently came very close, and often matched the total subsample amount. In fact, there were only 6 out of the 275 samples (2%) used for development that had BCG subsample sizes out of recommended subsample ranges (200 ± 40, Chalfant 2012). This suggests that the BCG information was robust enough for incorporation into both SWMMIs. It also suggests that refinement of the BCG attributes for PADEP should continue. This is particularly important given the results of this work, where BCG metrics tended to outperform PTV metrics during metrics development.

It must also be noted that the summer metric analysis selected pctDOM as one of the best responders to anthropogenic stress, which can potentially be problematic if taken directly on face value without further evaluating the taxa list in the subsample. For instance, the pctDOM metric may show poor conditions, but the dominant taxon in the subsample is a sensitive mayfly. There were only two macroinvertebrate samples within the LD condition that exhibited a relatively mediocre pctDOM score (> 25% dominant) with a relatively sensitive taxon. In both cases *Isonychia* spp. (PTV of 3 and a BCG of 3) was the dominant taxon. Some of the more common taxa dominance issues in wadeable streams result from high abundances of sensitive taxa such as *Allocapnia* spp. and *Prosimulium* spp. These taxa tend to dominate the total number of individuals in winter and spring samples, respectively. One of the reasons that the summer metric analysis selected pctDOM may have been because these issues are less likely to occur in the summer within LD conditions. However, based on the possibility for this to still occur, it will be important to further evaluate the taxon that is driving the pctDOM metric during assessments.

The accuracy and precision of both SWMMIs was determined to be within acceptable limits. Yet the fall SWMMI tended to have more methodological variability (measured using CV) than the summer SWWMI. This result was expected after several biologists described difficulty with collecting samples during the generally higher flows in the fall. In some cases, there were also reports of difficulty finding the exact sampling location that was clearly observed during the summer. Higher SWMMI score variability during the fall may also be the result of having a significantly smaller sample size of replicate data to work with. Consequent to these observations, it is recommended that biologists attempting macroinvertebrate sample collection during the fall should only sample when flows are low enough to confidently reach and observe the targeted habitat. Median flows or lower are recommended. In addition, future refinements to this method should seek to include a greater number of replicate samples for both seasons, which will ultimately decrease the likelihood that outlier collections are affecting precision estimation. Regardless, both the summer and the fall SWMMIs showed good intrasite precision, which suggests that the method can be used for assessment purposes. For temporal precision, the summer and fall SWMMI CI90 within the LD condition was 14.2

and 11.4 points, respectively, which suggests that observed score changes at a site over time of 15 and 12 points for the summer and fall SWMMI, respectively, could be considered changes due to more than natural variability and sampling error. Chalfant (2012) found that wadeable streams in the Commonwealth typically have temporal variations between 6 and 13 points. This suggests that the large river SWMMIs experience slightly more variation on average than wadeable streams, which was expected given the complexity and variability observed in large rivers.

Overall, this work indicates that the summer and fall SWMMI are highly effective at distinguishing between good and poor environmental conditions in large semi-wadeable rivers. In addition, a high degree of precision shows that natural variability was acceptably accounted for and sampling error was as low as other recognized macroinvertebrate methods in wadeable streams. This greatly increases the confidence in using these indices as tools for evaluating the impacts to semi-wadeable large river water quality. However, one unique aspect to this sampling method and assessment tool is that it creates multiple sites across the width of a river if chemical and physical transect analysis dictates. One of the greatest challenges to using this method is determining how to assess a river holistically if results between sites at the same location suggest different outcomes. However, this situation is not particularly unique in large river assessments, and arriving at a final assessment conclusion is not insurmountable. For instance, the Ohio River Valley Water Sanitation Commission (ORSANCO) has also developed a biological assessment tool for large rivers (Emery et al. 2003) in which results among multiple sites are evaluated to make one assessment decision. Further discussion of how PADEP intends on using this tool to make ALU assessments was not within the scope of this chapter and will be discussed in more detail in the following chapter.

Literature Cited

Applegate, J. M., P. C. Baumann, E. B. Emery, and M. W. Wooten. 2007. First steps in developing a multimetric macroinvertebrate index for the Ohio River. River Research and Applications 23: 683–697.

Astin, L. E. 2007. Developing biological indicators from diverse data: The Potomac Basin-wide Index of Benthic Integrity (B-IBI). Ecological Indicators 7: 895–908.

Baptista, D. F., R. G. de Souza, C. A. Vieira, R. Mugnai, A. S. Souza, and R. S. de Oliveira. 2011. Multimetric index for assessing ecological condition of running waters in the upper reaches of the Piabanha-Paquequer-Preto Basin, Rio de Janeiro, Brazil. Zoologia. 28 (5): 619–628.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton,

Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency. Office of Water. Washington, D.C.

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 Biological Integrity (IBI) for "True" Limestone Streams. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/An%20Index%20of%20Biological%20Integrity-Limestone%20Streams.pdf

Chalfant, B. 2012. A benthic index of biotic integrity for wadeable freestone streams in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Technical%20Documentation/freestoneIBImarch2012.pdf

Charles, D.F., A. P. Tuccillo, and T.J. Belton. 2010. Diatoms and the biological condition gradient in New Jersey rivers and streams: A basis for developing nutrient guidance levels. New Jersey Department of Environmental Protection. Trenton, NJ.

Chutter, F. M. 1972. An empirical biotic index of the quality of water in South African streams and rivers. Water Research 6:19–30.

Davies, S.P. and S.K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251–1266.

Davis, W.S., and T.P. Simon. 1995. Introduction to Biological assessment and criteria: tools for water resource planning and decision making, W.S. Davis and T.P. Simon, eds. (pp. 3 – 6). CRC Press, Boca Raton.

Emery E.B., T.P. Simon, F.H. McCormick, P.L. Angermeier, J.E. Deshon, C.O. Yoder, R.E. Sanders, W.D. Pearson, G.D. Hickman, R.J. Reash, and J.A. Thomas. 2003. Development of a multimetric index for assessing the biological condition of the Ohio River. Transactions of the American Fisheries Society. 132 (4): 791–808.

Gerritsen, J., J. Burton, and M. T. Barbour. 2000. A stream condition index for West Virginia wadeable streams. Prepared for USEPA Office of Water and USEPA Region 3. EPA-822-B00-001. U.S. EPA, Office of Water, Washington, D.C.

Gerritsen, J., and B. Jessup. 2007. Identification of the biological condition gradient for freestone (noncalcareous) streams of Pennsylvania. Prepared for United States Environmental Protection Agency and Pennsylvania Department of Environmental Protection. Tetra Tech, Inc. Owings Mills, Maryland.

Hawkins, C.P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. Ecological Applications 16(4): 1277–1294.

Hilsenhoff, W.L. 1977. Use of arthropods to evaluate water quality of streams. Technical Bulletin Number 100. Wisconsin Department of Natural Resources. 15 pp. Madison, Wisconsin.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. The Great Lakes Entomologist 20(1): 31–39.

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. Journal of the North American Benthological Society 7(1): 65–68.

Homer, C.G., J.A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N.D. Herold, J.D. Wickham and K. Megown, 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing 81:345–354.

Karr, J.R. and D.R. Dudley. 1981. Ecological perspective on water quality goals. Environmental Management 5: 55–68.

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 & W.S. Davis, 2003. Development and evaluation of a macroinvertebrate Biotic Integrity Index (MIBI) for regionally assessing Mid-Atlantic Highland streams. Environmental Management 31: 656–669.

Le, S., J. Josse, and F. Husson. 2008. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 25(1): 1–18.

Merritt, R. W., K. W. Cummins, and M. B. Berg (editors). 2008. An Introduction to the aquatic insects of North America. 4th edition. Kendall/Hunt Publishing Company, Dubuque, Iowa.

Olivero Sheldon, A., A. Barnett, and M. G. Anderson. 2015. A Stream Classification for the Appalachian Region. The Nature Conservancy, Eastern Conservation Science, Eastern Regional Office. Boston, MA.

PADEP. 2007. Pennsylvania DEP Multihabitat Stream Assessment Protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/Multihabitat%20Protocol.pdf

Peterson, B.G., P. Carl, K. Boubt, R. Bennett, J. Ulrich, E. Zivot, M. Lestel, K. Balkissoon, and D. Wuertz. 2015. PerformanceAnalytics: Econometric tools for performance and risk analysis. https://cran.rproject.org/web/packages/PerformanceAnalytics

Pond, G. P., J. E. Bailey, B. M. Lowman, and M. J. Whitman, 2013. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. Environmental Monitoring and Assessment 185:1515–1540.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at http://www.R-project.org/.

Revelle, W. 2017. psych: Procedures for Psychological, Psychometric, and Personality Research. https://cran.r-project.org/web/packages/psych

Royer, T. V., C. T Robinson, and G. W. Minshall. 2001. Development of macroinvertebrate-based index for bioassessment of Idaho Rivers. Environmental Management. 27: 627–636.

Scopel, C., 2014. Create watersheds and trace downstream in your ArcGIS Online web map. Environmental Systems Research Institute. Redlands, CA. https://blogs.esri.com/esri/arcgis/2014/12/12/create-watersheds-and-trace-downstream-in-your-arcgis-online-web-map/

Smith, Z. M., C. Buchanan, and A. Nagel. 2017. Refinement of the Basin-Wide Benthic Index of Biotic Integrity for Non-Tidal Streams and Wadeable Rivers in the Chesapeake Bay Watershed. Interstate Commission on the Potomac River Basin (ICPRB). Rockville, MD.

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

Stribling, J. B., B. K. Jessup, and D. L. Feldman. 2008. Precision of benthic macroinvertebrate indicators of stream condition in Montana. Journal of the North American Benthological Society 27: 58–67.

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

USEPA. 2016b. National Rivers and Streams Assessment 2008-2009 Technical Report. EPA-841-R-16-008. U.S. Environmental Protection Agency, Washington, DC.

Weigel, B. M. and J. J. Dimick. 2011. Development, validation, and application of a macroinvertebrate based index of biotic integrity for nonwadeable rivers of Wisconsin. Journal of the North American Benthological Society 30(3):665–679.

Wessell, K. J., R. W. Merritt, J. G. O. Wilhelm, J. D. Allan, K. W. Cummins, and D. G. Uzarski. 2008. Biological evaluation of Michigan's non-wadeable rivers using macroinvertebrates. Aquatic Ecosystem Health and Management 11: 335–351

Wickham, H. 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. https://cran.r-project.org/web/packages/ggplot2

Wickham, H. 2011. plyr: The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software. 40(1): 1–29. https://cran.r-project.org/web/packages/plyr

Wickham, H., and R. Francois. 2016. dplyr: A Grammar of Data Manipulation. https://cran.r-project.org/web/packages/dplyr

Wright, J.F. 1994. Development of RIVPACS in the UK and the value of the underlying data-base. Limnética. 10 (1):15–31.

Chapter 4. Collecting Data and Making Aquatic Life Use Assessment Determinations

Introduction

Assessment of ALU in large semi-wadeable rivers can be a daunting and complex process. To appropriately assess biological communities in large rivers and to increase the efficiency of ALU assessments, PADEP will separate large rivers into two categories: semi-wadeable and non-wadeable. Semi-wadeable rivers are defined as predominantly free-flowing systems with drainage areas > 2,600 km^2, and have physical characteristics that allow for riffle and run sections to occur with relative frequency. These river systems tend to lack a well-defined and navigable U-shaped channel for any significant distance and frequently present difficulties for both wadeable and non-wadeable macroinvertebrate data collection methodologies. Well over half of the large rivers within the Commonwealth are considered semi-wadeable (Figure 4-1).
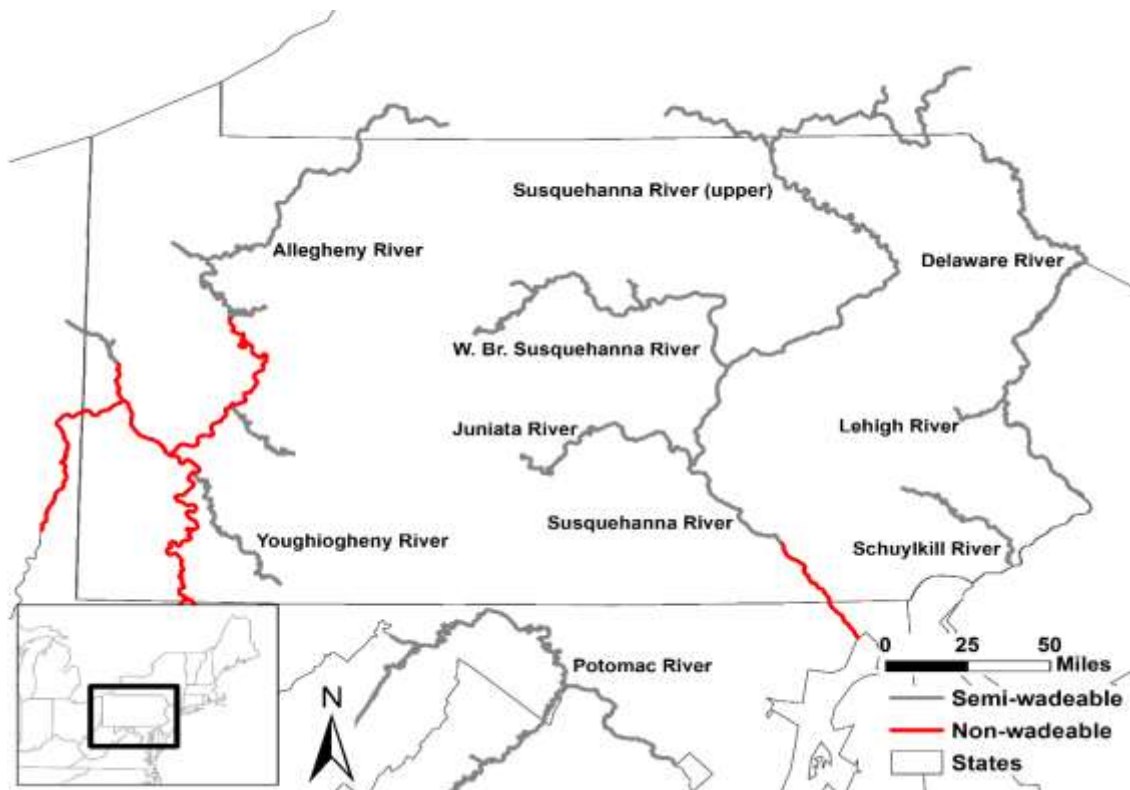
Figure 4-1. Large Rivers that are semi-wadeable and non-wadeable rivers throughout the Commonwealth. Assessment determinations will be made for semi-wadeable rivers using this assessment method. PADEP continues to develop assessment methods for non-wadeable large rivers.

Several studies have shown that semi-wadeable rivers can express substantial and reliable differences in water quality across their width for great distances, which can help drive variations observed in the macroinvertebrate communities that inhabit these regions (Guild et al. 2014, PADEP 2014, Shull and Pulket 2015). The water quality differences across the width of large semi-wadeable rivers are usually the result of major tributary inputs that do not mix. Additionally, each major tributary input is driven by both the natural and anthropogenic influences within the respective basin. Given these chemical and physical realities, it is critical that biological community complexity is also measured and imputed. It was not surprising that results in Chapter 1 suggested that macroinvertebrate communities were consistently different between major water quality zones across the width of some semi-wadeable rivers. Additionally, within each zone these communities were more similar than communities between different zones, even with repeated sampling visits across multiple years. Macroinvertebrate community results in Chapter 1 demonstrated that although differences in field parameters (e.g., specific conductance or turbidity), may only show a discrepancy of 10% at base flow conditions, these measurements were either reliably measuring a direct effect, or were serving as surrogates for macroinvertebrate community preferences. Based on these findings, differences of 10% or greater in any field parameter may indicate a change in macroinvertebrate communities. Furthermore, as discrepancies between water quality influences increase (i.e., due to increased anthropogenic stress in a major tributary), there is also an expected increase in biological community discrepancy between zones across the width of a river. Ultimately, these differences can lead to discrepancy in ALU assessment determinations, which requires additional discussion.

To date, no other large river biological assessment tool has sought to understand and deal with these chemical, physical, and biological differences within one location separately. On the contrary, many large river collection methods have been created to capture and composite these variables into one measure; thereby, accounting for, but not giving heed to these important differential aspects (Applegate et al. 2007, Wessell et al. 2008, Blocksom and Johnson 2009, Weigel and Dimick 2011). Final assessments using these tools average or generalize biological condition to provide valuable assessment information, but they do not consider potentially important details. This effectively obscures the ability to not only account for biological community degradation, but it also reduces the ability to track major sources of impacts driving degradation. Even more problematic are the large river biological collection methods that only collect data along the shoreline of a large semi-wadeable river (Merritt et al. 2005, Angradi 2006, USEPA 2013). These methods also attempt to capture and composite large river variability, but completely miss a large portion of biological information that resides in a river. Shoreline collection methods can be particularly questionable when making large scale inferences about water quality conditions, because shoreline habitats are more likely affected by minor tributary influences and point source discharges (PADEP 2014,

Shull and Pulket 2015). Ultimately, shoreline collection methods may result in mischaracterizing and misinforming the public about thousands of kilometers of large rivers as was done in past reports (e.g., USEPA 2016). Consequently, large river assessment tools and reports using shoreline collection methods should be limited to describing local scale impacts at discrete locations. Due to these concerns in large river methods, PADEP spent several years developing and refining the transect collection method, and because of this method, each SWMMI can not only be used to make one assessment determination (elucidated in this chapter) that is reflective of overall water quality, but also produce results that retain the unique aspects of water quality variations. This should greatly improve the validity of each assessment on large semi-wadeable rivers, as well as provide important source tracking information.

The goal of this chapter is to lay the framework for how PADEP intends on making ALU assessment determinations in large semi-wadeable rivers within the boundaries of the Commonwealth. The previous chapters of this report evaluate the complexity of large semi-wadeable rivers and provide assessment tools to compensate; however, the implementation of these tools is equally important and requires discussion. Working on, in, and around large semi-wadeable rivers can be challenging and dangerous. This discussion seeks to provide guidance on when and how to collect these data, as well as important safety considerations. Making accurate assessment decisions requires both a sufficient number of data types (e.g., physical, chemical and biological) and a specificity of those data within a particular water influence and season. Ultimately, ALU assessment determinations will be rather straightforward and similar to wadeable stream assessments if data can only be collected in one season and water influences are well mixed. However, ALU determinations when water influences are not well mixed and when data are collected during both the summer and fall will require additional evaluation and discussion.

Safety Procedures

Appropriate ALU assessment begins with data collection that follows proper guidelines and safety procedures. Safety is of utmost concern when working in and on large semi-wadeable rivers. Individuals that lack experience and education related to navigating these waterbodies put themselves and others at risk of injury or loss of life. In most, if not all cases, some form of watercraft is required for this sampling method. Whether a kayak, canoe, or powered boat is used, PADEP recommends – and in most cases Pennsylvania Fish and Boat Commission (PFBC) requires by law – that individuals using watercrafts take boating courses and obtain a boating safety certificate. However, whether laws apply to certain individuals or watercraft or not, PADEP strongly recommends that all individuals regardless of age, experience, or navigation method

obtain and keep a boating safety certification on them during semi-wadeable surveys. More information on boating courses and safety certifications are available here:

Large semi-wadeable rivers present difficulties for both wadeable and non-wadeable water navigation. Certain locations on rivers such as the Susquehanna River around Harrisburg, PA may be fully or partially wadeable during certain times of year, but there are also highly variable currents and drop-offs that are dangerous to wading collectors. Therefore, personal floatation devices (PFD) should always be worn. However, PFDs should not be used to facilitate wading or navigating the waterbody. Navigation using a watercraft can be equally if not more challenging than wading. Many boats that frequent the large semi-wadeable rivers such as the Susquehanna River are outfitted with modified jet propulsion to reduce the risk of serious damage associated with impacting the streambed with the lower unit. Therefore, consideration for modified equipment may be needed to conduct this method. In addition, riffles and bedrock substrate often dominate these rivers regardless of size. Consequently, PADEP recommends that at least two individuals work together during these surveys. This is critical for many safety reasons, and particularly important when a powered watercraft is underway. While in motion, PADEP recommends that the individual not operating the powered watercraft is actively looking for obstacles on and just under the surface. This individual should be actively and clearly communicating obstacles to the operator. This is best done with the use of hand signals that are agreed upon between collectors before venturing out on the water. All of this is not to say that macroinvertebrate sampling in large semi-wadeable rivers is overly risky. In fact, with the proper training and experience this sampling method can be conducted efficiently and safely using limited resources, and without restrictions based on the collector's physical stature.

Inherent in these recommendations is the understanding that collectors must be well informed of the unique characteristics of each river they intend to sample and be aware of current flow and weather conditions. Equipment considerations may be different between similar sized large semi-wadeable rivers, and certainly different when visiting multiple semi-wadeable rivers of different sizes. For instance, boat selection and safety equipment lists would be different depending on whether the collector is visiting a 2,600 km^2 river or 20,000 km^2 river. These differing characteristics require thorough planning before going into the field so resources are not wasted due to a lack of preparation. It is also imperative that collectors have access to and continually check river flow conditions. There will be situations when weather in one part of the Commonwealth may be optimal for sampling a site, but weather in another part of the Commonwealth – two or three days before the sampling period – created conditions that reduces the ability to collect. Analysis of flow conditions at all available points upstream of the intended

sampling locations are highly recommended. Commonly used discharge data are available here:

https://waterdata.usgs.gov/pa/nwis/current/?type=flow

Data Collection

What to Collect

Transect surveys and water chemistry samples should be collected before macroinvertebrate sampling. All transect surveys are conducted using a clean and calibrated field meter that collects water temperature, specific conductance, dissolved oxygen, pH, and – preferably – turbidity. Transects are taken at regular intervals (each point representing approximately 5-10% of the total wetted width) across the width of the river. This is the least amount of information required to determine if major water quality influences exist at the sampling location. A major water quality influence that shows a 10% difference or greater in any field meter parameter necessitates a unique water chemistry and macroinvertebrate sample be collected. It is often helpful to create a drawing or map of the site with transect points depicted to help delimit the locations of water influence transitions across the width of the river. When new major water influences are identified, additional transects and water chemistry samples will be collected on upstream reaches to identify and characterize the source of variable water quality across the width of the targeted reach of river. Transect and water chemistries will also be collected downstream, if possible, to determine the extent that the water influences do not mix. Multiple water chemistry samples over a period are recommended to provide additional information for assessments. Transect and water chemistry data collected through the season prior to macroinvertebrate collection provides the most utility for complete and accurate assessments, but routine year-round data collection has also proven useful for ALU assessments and method development efforts. Transect and water chemistry data collection prior to macroinvertebrate collections also provides opportunities to complete a thorough reconnaissance of the targeted semi-wadable river.

There are three water chemistry collection methods that are acceptable when making ALU assessments: fixed-width sampling, isokinetic sampling (USGS 2006), and discrete mid-channel/mid-depth sampling (Shull 2013). Fixed width and isokinetic sampling is useful for characterizing water chemistry for the entire width of the surface water, while discrete samples are useful for characterizing unique water quality influences. Water chemistry characterization is also driven by the need to identify sources and causes of impairment. Water chemistries should include a comprehensive list of constituents including total and dissolved nitrogen and phosphorus species, total metals, and ions. For water chemistry collection, PADEP routinely uses its Bureau of Laboratories

standard analysis code 612 (Table 4-1) for semi-wadeable assessments across the Commonwealth. In addition, constituents such as dissolved metals, pesticides, hormones, wastewater compounds, and pharmaceuticals will be added if necessary.

Table 4-1. Chemical parameters collected according to standard analysis code 612 during routine semi-wadeable river surveys.

Test Code	Test Description	Reporting Limit	Units	Reference
00095	Specific Conductance at 25C	1	UMH/CM	SM 2510B
00403	pH	0	pH UNITS	SM 4500H-B
00410	Alkalinity, pH 4.5	0	MG/L	SM 2320B
00530	Total Suspended Solids	2	MG/L	USGS-I-3765
00600A	Total Nitrogen	0.064	MG/L	SM 4500N-ORG
00602A	Dissolved Nitrogen	0.064	MG/L	SM 4500N-ORG
00608A	Dissolved Ammonia	0.02	MG/L	EPA 350.1
00610A	Total Ammonia	0.02	MG/L	EPA 350.1
00630A	Total Nitrite and Nitrate	0.04	MG/L	EPA 353.2
00631A	Dissolved Nitrite and Nitrate	0.04	MG/L	EPA 353.2
00665A	Total Phosphorous	0.01	MG/L	EPA 365.1
00666A	Dissolved Phosphorous	0.01	MG/L	EPA 365.1
00671A	Dissolved Orthophosphate	0.01	MG/L	EPA 365.1
00680	Total Organic Carbon	0.5	MG/L	SM 5310C
00900	Total Hardness	0.11	MG/L	SM 2340A+B and EPA 200.7 revision 4.4
00916A	Total Calcium	30	UG/L	EPA 200.7 Revision 4.4
00927A	Total Magnesium	10	UG/L	EPA 200.7 Revision 4.4
00929A	Total Sodium	200	UG/L	EPA 200.7 Revision 4.4
00937A	Total Potassium	1	MG/L	EPA 200.7 Revision 4.4
00940	Total Chloride	0.5	MG/L	EPA 300.0
00945	Total Sulfate	1	MG/L	EPA 300.0
01007A	Total Barium	10	UG/L	EPA 200.7 Revision 4.4
01022K	Total Boron	200	UG/L	EPA 200.7 Revision 4.4
01042H	Total Copper	4	UG/L	EPA 200.8 Revision 5.4
01045A	Total Iron	20	UG/L	EPA 200.7 Revision 4.4
01051H	Total Lead	1	UG/L	EPA 200.8 Revision 5.4
01055A	Total Manganese	10	UG/L	EPA 200.7 Revision 4.4
01067H	Total Nickle	4	UG/L	EPA 200.8 Revision 5.4
01082A	Total Strontium	10	UG/L	EPA 200.7 Revision 4.4
01092A	Total Zinc	10	UG/L	EPA 200.7 Revision 4.4
01105A	Total Aluminum	200	UG/L	EPA 200.7 Revision 4.4
01132A	Total Lithium	25	UG/L	EPA 200.7 Revision 4.4
01147H	Total Selenium	7	UG/L	EPA 200.8 Revision 5.4
70300	Total Dissolved Solids	2	MG/L	USGS-I-1750
70507A	Total Orthophosphate	0.01	MG/L	EPA 365.1
82550	Osmotic Pressure	1	MOSM/KG	DEP BOL
99020	Total Bromide	50	UG/L	EPA 300.1

A habitat assessment is also required for each semi-wadable macroinvertebrate sample. This habitat protocol has undergone several iterations from Plafkin et al. (1989), but PADEP uses a 12 parameter – 20-point scoring method (Chalfant 2012). Currently, it is recommended that all 12 parameters are recorded when conducting habitat assessments in a semi-wadeable river. Instream parameters such as instream cover, epifaunal substrate, and embeddedness are the most reliable habitat indicators for large semi-wadeable rivers. Instream cover evaluates the percent makeup of the substrate (boulders, cobble, other rock material) and submerged objects (logs, undercut banks) that provide refuge for fish. Epifaunal Substrate evaluates riffle quality, i.e. areal extent relative to stream width and dominant substrate materials that are present. Embeddedness estimates the percent (vertical depth) of the substrate interstitial spaces filled with fine sediments. These three instream habitat measurements can be summed to provide a possible range of 0 (indicating worst possible instream conditions) to 60 (indicating best possible instream conditions) points at each sampling site. Combined instream habitat scores < 30 indicate poor habitat conditions and will be used to inform a cause of ALU impairment. The other parameters in the habitat assessment are also useful, but tend to become difficult to measure as river size increases. Future refinements to this method will seek to improve rapid habitat assessments in large rivers.

Macroinvertebrate samples are collected using 6D-200 field collection and laboratory subsampling procedures detailed in PADEP protocols (Chalfant 2012). Briefly, this consists of 6 one-minute kicks evenly distributed across the entire width of the waterbody using a 500 µm mesh D-frame net. With the transect method modification, additional 6D-200 samples are collected in accordance with the number of major water quality influences discovered during water quality transect data collection. For instance, if three distinctly different water influences were discovered, then three 6D-200 samples are collected, with each sample composited across the entire width of the respective water influence. As with all macroinvertebrate sample collections, it is important not to collect water chemistry in the same location during or after the macroinvertebrate collection kicks have been conducted. Once field collections are complete, samples are taken to the laboratory for analysis. Macroinvertebrate processing and identification is conducted using standard PADEP methods, which are described in detail in Appendix A of Chalfant (2012). Metric calculation, standardization processes, and SWMMI scoring is described in detail in Appendix A of this report. The current taxa list with both PTV and BCG attributes used to develop this method is provided in Appendix B of this report.

Where and When to Collect

For ALU assessment purposes, each reach of river is assessed by the macroinvertebrate collection site immediately downstream. The length of each

assessed reach is then determined by where the next potential impact to water quality exists upstream (i.e., major tributary or major city). Therefore, the location of each upstream macroinvertebrate collection site should reflect this pattern. More explicitly, each macroinvertebrate collection site along the longitudinal gradient is determined by several factors including, where sufficient riffle-run habitat exists, where changes in physiographic and demographic characteristics occur, and where additional major tributaries enter the system. Ideally, macroinvertebrate collections sites will occur at every viable riffle-run habitat, but at the very least, it is necessary to bracket major potential impacts to each system such as a major tributary or change in land use. In the example provided below (Figure 4-2), two semi-wadeable rivers converge to form another semi-wadeable river. Below this confluence, multiple water quality transects show that water influences do not mix so the non-mixing water influences were mapped out. Hence, site 1 requires two unique 6D-200 samples composited completely within the delineated zone of each water influence. Additional macroinvertebrate collections sites (sites 2 – 5) are added upstream of the confluence to characterize each major water influence and to bracket the demographic characteristics across the drainage (e.g., communities shown as red areas, other land use transitions). It is important to note that water quality transect sites – used to delineate the area of specific water influences – can be collected at a higher frequency of locations than macroinvertebrate collections.
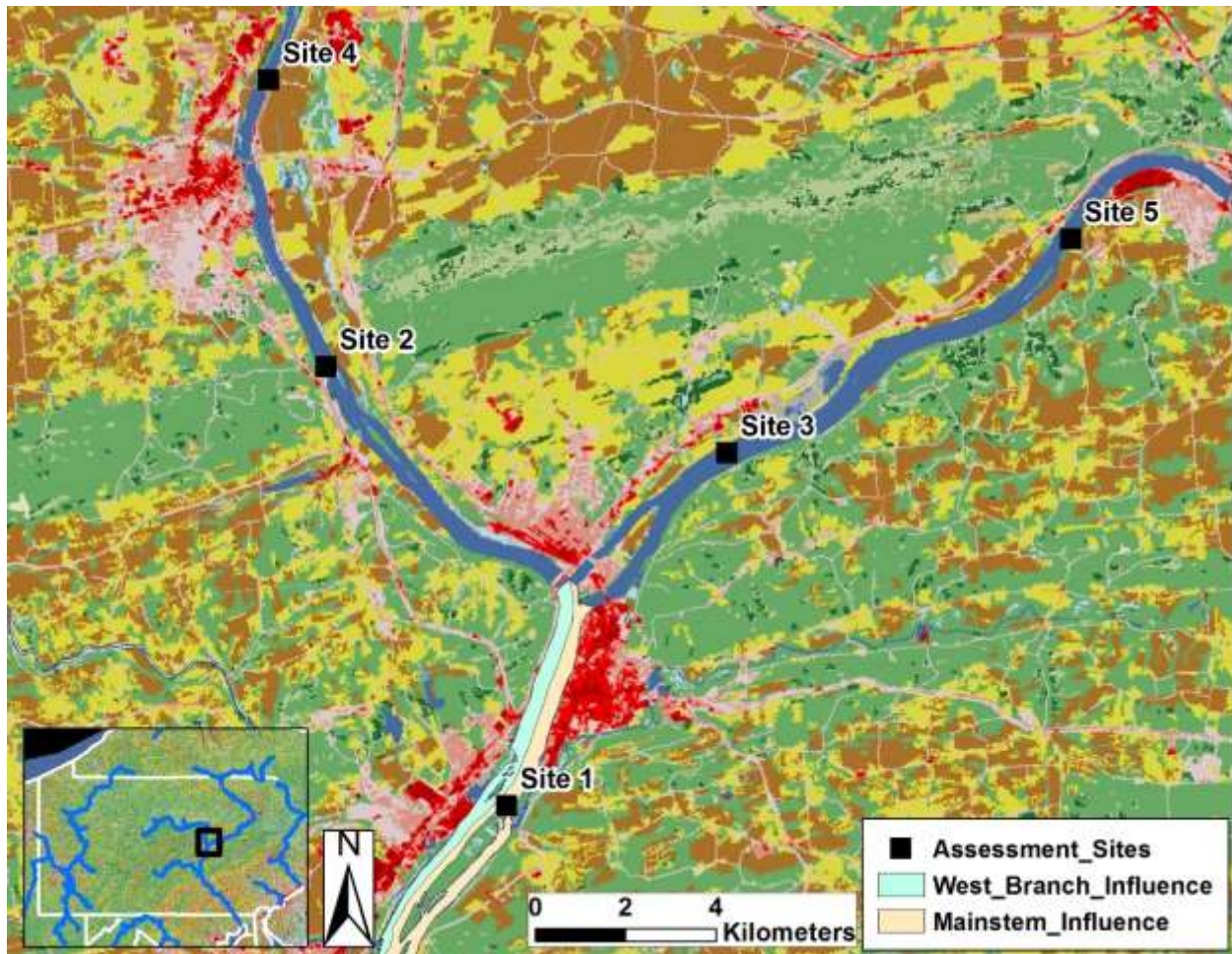
Figure 4-2. Macroinvertebrate collection sites on a large semi-wadeable river. Site locations were selected to bracket major land use changes and tributary inputs.

It must be noted that macroinvertebrate samples collected at or above median flow conditions can cause a lower confidence in results. Often, this situation is not realized until after the macroinvertebrate kicks and instream habitat have been conducted in the field. Although this situation was rare in over 400 samples collected for development of this index, it did occur. Therefore, certain rules should be established for determining when a sample is applicable for making assessment decisions. If collectors are not confident that the sampling was representative of the site (e.g., collector could not collect the 6 kicks relatively evenly across the entire width of the influence), then samples should not be used for assessments. Additionally, if collectors have not collected the sample in an appropriate riffle-run habitat type, then samples should not be used for assessments. The responsibility for this is decision is left to the most experienced individual in the field during the sampling event. Individuals that are experienced with working in large rivers quickly realize when an inappropriate sample

has been collected and will either discard the sample or annotate their concerns in the comments section of the field form.

Aquatic Life Use assessments using the SWMMI

Both SWMMIs are accurate and precise tools for making ALU assessment determinations. Ideally, assessment decisions stemming from SWMMI scores will understand and compensate for the complexity of the biological communities that exist in these rivers, as well as conform to the conventional reporting format within the integrated reports. Results from Chapter 2 showed two unique macroinvertebrate communities being collected across the summer and fall season. These results were further validated in Chapter 3 when different metrics responded to anthropogenic stress. Thus, each SWMMI is independently applicable when making ALU assessment determinations. This determination is based on EPA guidance, which says that all biological communities PADEP has assessment methods for must be evaluated on a stand-alone basis (USEPA 2002). Consequently, each SWMMI is functionally equivalent to having two completely different biological assessment tools (e.g., fish MMI and a macroinvertebrate MMI). Therefore, it is not appropriate to average both SWMMI scores to obtain an overall result. It is also not appropriate to favor the results of one SWMMI over the other. Impairment thresholds for each SWMMI are provided in the results section of Chapter 3. PADEP will always strive to collect as much information as possible to make the most accurate assessment decisions. However, based on independent applicability, it is also understood that only one SWMMI is required to make an ALU assessment determination for a semi-wadeable river.

The following situation provides an example of this biological assessment rule. Multiple summer and fall samples were collected at the same site (Figure 4-3). Based on transect analysis there was one major influence, so one macroinvertebrate sample was collected evenly across the width of the river during each visit. There was a total of five samples collected; two samples during the summer and three samples during the fall. The summer samples consistently showed reduced, but attaining SWMMI scores, yet the fall samples resulted in impaired scores. The fall biological community was not supporting the Aquatic Life Use; therefore, PADEP would determine that this section of the river is impaired. It may be concluded from this example that one SWMMI is more sensitive than the other; however, that is not the case. Examination of the entire development dataset showed no preference for one SWMMI consistently selecting for one assessment decision when biological communities were close to thresholds.
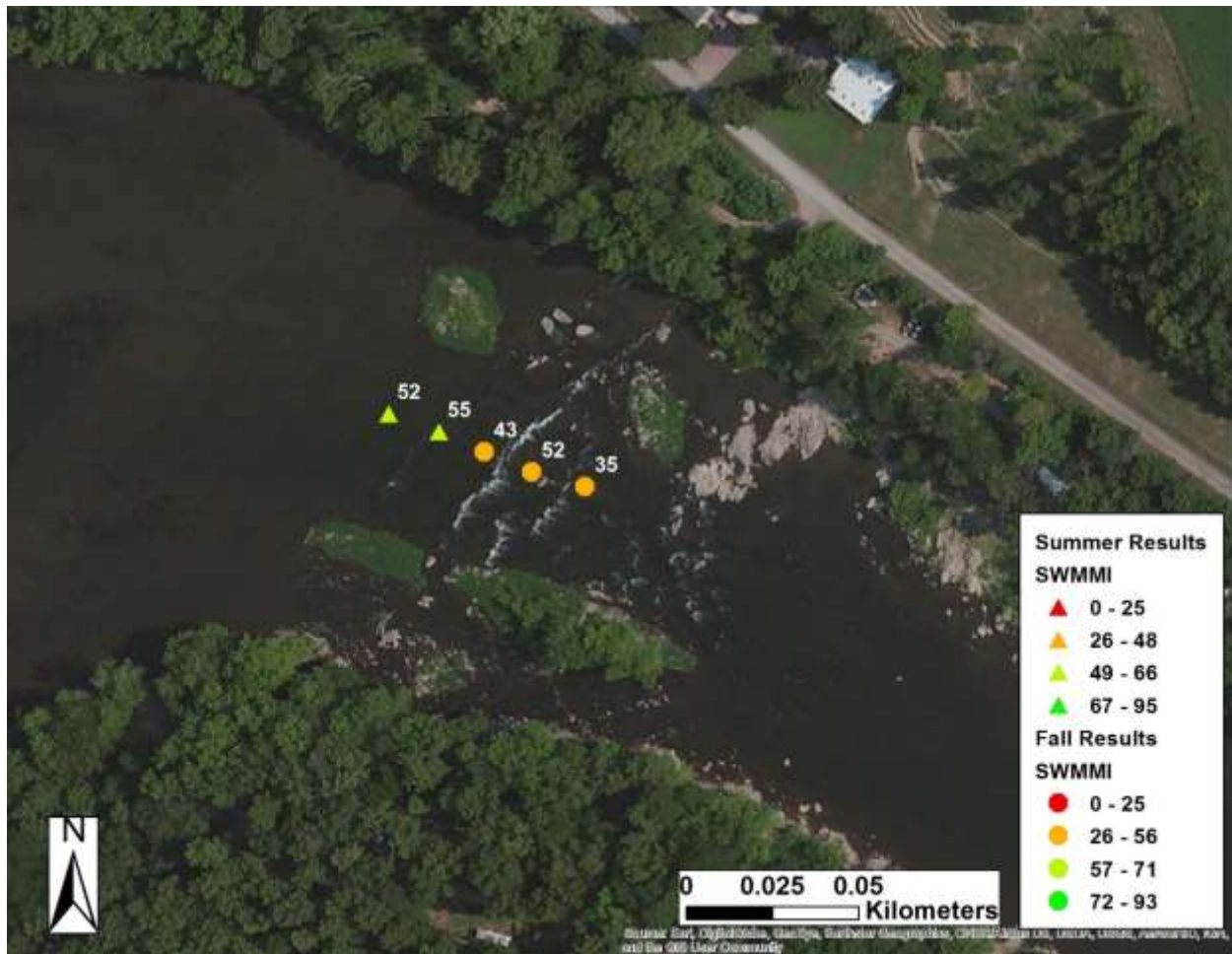
Figure 4-3. Multiple summer and fall SWMMI results over time at one location on a semi-wadeable river. Location of points on the map do not indicate exact sample location; points were moved slightly to illustrate the results of sampling. Points are labeled with the SWMMI score.

The transect method can also produce multiple SWMMI results at any given location based on the number of major water influences discovered during transect data collection. The transect method selects major water influences constituting more than 10% of the width of any river; hence, one impaired macroinvertebrate sample represents a major portion of that river. Therefore, one impaired SWMMI score is significant enough to warrant impairment of the river segment. Just over 10% of a river's width may seem like an insignificant amount of area to be concerned with, especially if remaining portion of the river suggests ALU attainment. However, the large scale of these rivers must be accounted for. For example, one impaired SWMMI score representing just over 10% of a 1.0 km wide river is approximately 100 m. Yet, due to the nature of these semi-wadeable rivers, that measured degradation in the biological community may persist for up to 105 km (as demonstrated in Chapter 1). Therefore, the

actual scale of impact on that river is 10.5 km^2, which is roughly equivalent to the area of a moderately large town in the Commonwealth. As stated before, assessment determinations that are currently in the Integrated Report are not broken into segments across the width of a river; therefore, PADEP must impair the entire reach of that river. Collection of samples across the entire width of the river will not only create a better representation of overall conditions in the river, but it will also more appropriately address ecological concerns in critical habitats for many important species. Understanding and accounting for major water influences in impairment decisions introduces the understanding of continuity between factors causing impairment (i.e., impaired major tributaries, and local scale impacts) and the greater riverine ecosystem at a regulatory level, which has been lacking in the past.

Assessment decisions based on averaging multiple sample results have been done for large rivers. For example, ORSANCO will typically analyze multiple biological results collected randomly in one pool and make assessment decisions based on averaging the scores. If PADEP decides to collected additional macroinvertebrate samples in large semi-wadeable rivers using a conditional (e.g., certain number of samples based on river size) or a random method, then it would also make sense to average results to create a more holistic assessment decision. However, the transect method specifically targets observed variations in water quality and measures biological conditions within those regions. Consequently, the additional information provided by the transect method is neither random or arbitrary, and therefore, SWMMI scores between different sites across the width of a river should not be averaged. It could also be argued that this method may neglect minor influences along each shoreline that affect critical habitats for many organisms. However, it must be understood that PADEP already has assessment methods in place to account for these concerns. For example, many minor water influences in large rivers are smaller wadeable streams. PADEP already has assessment methods for all the small wadeable streams within the Commonwealth; therefore, poor water quality conditions observed in minor river influences are addressed using wadeable methods (PADEP 2007, Botts 2009, Chalfant 2012). Minor river influences can also result from specific point and non-point source discharges. Again, PADEP has permitting and compliance practices in place to ensure these potential concerns are addressed.

Additional Application Considerations

There will be situations when data resulting from SWMMI scores are not used in making ALU assessment determinations. In fact, PADEP uses the wadeable method developed by Chalfant (2012) for several other purposes, including, but not limited to cause and effect surveys and incremental improvement reports. These surveys can collect biological information in areas that are not appropriate for making ALU assessment

determinations. For example, two macroinvertebrate samples were collected on a semi-wadeable river near a city in Pennsylvania, just downstream of a sewage treatment plant (Figure 4-4). In this example, the SWMMI results showed that a major portion of this semi-wadeable river (laterally) was being impacted by a facility, perhaps, not operating within permitted limits. Sampling locations specifically targeted one city's sewage treatment facility, but were not necessarily representative of river conditions in this area. Therefore, it would not be appropriate to use these results in making assessment decisions on this river. However, this example does illustrate the usefulness of the semi-wadeable biological collection method for other purposes. This example also illustrates the necessity to differentiate between ALU assessments and reports on local scale impacts. All ALU assessments on semi-wadeable rivers should examine the longitudinal scale that each macroinvertebrate sample represents. If a macroinvertebrate sample is determined to be more representative of a local scale impact, then compliance actions may be more appropriate than making an ALU impairment determination.
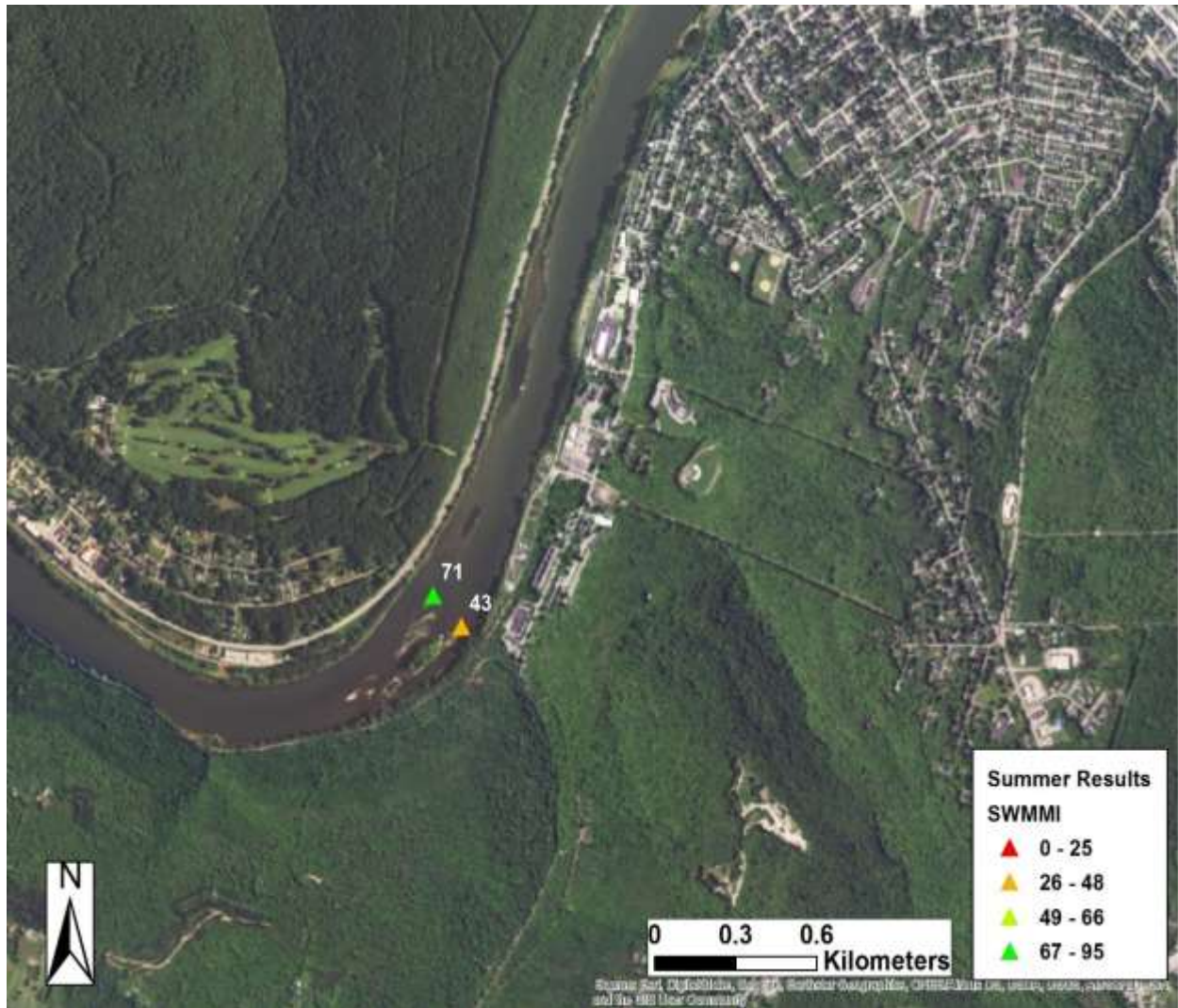
Figure 4-4. Summer SWMMI sample results on a semi-wadeable river showing impairment of biological condition in one major influence and attainment in the other major influence. Points are labeled with their respective SWMMI score.

The SWMMIs may also be used to evaluate whether conditions are degrading or improving at a given site (e.g., trend analysis). It is important to note that this type of analysis is different than making assessment determinations using an impairment threshold. Methodological error is already incorporated during the development of the impairment threshold, so using variability measurements as "gray areas" while making assessment determinations is not appropriate (Stribling et al. 2008). However, for analyses such as trend analysis, the temporal CI90 can be used to decide whether a macroinvertebrate community changes over time. When SWMMI scores at the same site change over time beyond the CI90, there is a high level of confidence that the biological community change was driven by human influences. The summer SWWMI CI90 for all sites (where repeat data were available) was 14.7 points, which suggests

that observed score changes at a site over time of 15 points or more can be considered a change in condition. The fall SWWMI CI90 for all sites (where repeat data were available) was 12.8 points, which suggests that observed score changes at a site over time of 13 points or more can be considered a change in condition. It could be argued that using the CI90 of sites within the LD condition is a more appropriate and conservative threshold for determining trends in the macroinvertebrate community over time. However, repeat data in the LD condition was lacking and additional data should be collected before using that threshold.

Literature Cited

Angradi, T. R. (editor) 2006. Environmental monitoring and assessment program: Great River ecosystems field operations manual. EPA/620/R-06/002. Office of Research and Development, US Environmental Protection Agency, Washington, DC

Applegate, J. M., P. C. Baumann, E. B. Emery, and M. W. Wooten. 2007. First steps in developing a multimetric macroinvertebrate index for the Ohio River. River Research and Applications 23: 683–697.

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 Biological Integrity (IBI) for "True" Limestone Streams. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/An%20Index%20of%20Biological%20Integrity-Limestone%20Streams.pdf

Chalfant, B. 2012. A benthic index of biotic integrity for wadeable freestone streams in Pennsylvania. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Technical%20Documentation/freestoneIBImarch2012.pdf

Guild, K., A. Anthony, M. Bilger, and J. Holt. 2014. Assessment of Passive and Active Macroinvertebrate collection methods in adjacent reaches on the upper main stem of the Susquehanna River. Journal of the Pennsylvania Academy of Science 88: 47–56.

Merritt, R. W, J. D. Allan, K. W. Cummins, K. J. Wessell, and J. O. Wilhelm. 2005. Qualitative biological and habitat protocols for Michigan's non-wadeable rivers. Michigan Department of Environmental Quality, Lansing, Michigan.

PADEP. 2007. Pennsylvania DEP Multihabitat Stream Assessment Protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/Multihabitat%20Protocol.pdf

PADEP. 2014. 2012-13 Susquehanna River Sampling and Assessment Report. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Water%20Quality%20Standards/Susquehanna_Report_2013_FinalDraft.pdf

Shull, D. R. 2013. Surface water collection protocol. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/Methodology/2015%20Methodology/Surface%20Water%20Collection%20Protocol.pdf

Shull, D. R., and M. Pulket. 2015. Causal Analysis of the Smallmouth Bass decline in the Susquehanna and Juniata Rivers. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania. http://files.dep.state.pa.us/Water/Drinking%20Water%20and%20Facility%20Regulation/WaterQualityPortalFiles/SusquehannaRiverStudyUpdates/SMB_CADDIS_Report.pdf

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.

USEPA. 2002. Consolidated assessment and listing methodology: Toward a compendium of best practices. U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds. Washington, DC

USEPA. 2013. National Rivers and Streams Assessment 2013-2014: Field Operations Manual – Non-Wadeable. EPA-841-B-12-009a. U.S. Environmental Protection Agency, Office of Water. Washington, DC.

USEPA. 2016. National Rivers and Streams Assessment 2008-2009: A collaborative study. EPA-841-R-16-007. U.S. Environmental Protection Agency, Washington, DC.

USGS. 2006. National Field Manual for the Collection of Water Quality Data. Chapter 4.1 - Surface Water Sampling: Collection methods at flowing-water and still-water sites. U.S. Geological Survey, Water Resources Office of Water Quality. Reston, VA. https://water.usgs.gov/owq/FieldManual/chapter4/pdf/Chap4_v2.pdf

Weigel, B. M. and J. J. Dimick. 2011. Development, validation, and application of a macroinvertebrate based index of biotic integrity for nonwadeable rivers of Wisconsin. Journal of the North American Benthological Society 30(3):665–679.

Wessell, K. J., R. W. Merritt, J. G. O. Wilhelm, J. D. Allan, K. W. Cummins, and D. G. Uzarski. 2008. Biological evaluation of Michigan's non-wadeable rivers using macroinvertebrates. Aquatic Ecosystem Health and Management 11: 335–351

Appendix A: Metric and Index Calculations

Two samples for each MMI (Summer and Fall) are provided to show the metric and index calculation process step-by-step. The summer and fall SWMMI calculations are separated into their respective sections for clarity.

Summer Metric and Index Calculation

The following summer samples were collected on the Delaware River near Callicoon, NY and the Mahoning River near Edinburg, PA.

Summer sample collected from the Delaware River near Callicoon, NY September 9th, 2016	
Taxa Name	Number of Individuals
Acroneuria	1
Agnetina	2
Baetisca	1
Brachycentrus	1
Cheumatopsyche	14
Chimarra	7
Chironomidae	10
Corbiculidae	8
Helicopsyche	14
Hydrobiidae	13
Hydropsyche	7
Isonychia	11
Lepidostoma	2
Leucrocuta	8
Maccaffertium	16
Micrasema	1
Oecetis	2
Oligochaeta	7
Optioservus	25
Physidae	1
Plauditus	7
Stenelmis	14
Teloganopsis	22
Tricorythodes	3

Summer sample collected from the Mahoning River near Edinburg, PA August 23rd, 2016	
Taxa Name	Number of Individuals
Ancylidae	1
Baetis	19
Cheumatopsyche	71
Chironomidae	20
Corbiculidae	1
Gammarus	39
Hirudinea	1
Hydropsyche	40
Orconectes	1
Stenelmis	3
Tricorythodes	1
Turbellaria	12

Percent Tolerant Individuals using BCG attribute 5 (BCGpct5)

$$= \left(\sum n_{indvBCG5}/N \right) *100$$

Where $n_{indvBCG5}$ is the number of individuals in the subsample with a BCG value of 5, and N is the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Acroneuria	1	3
Agnetina	2	3
Baetisca	1	2
Brachycentrus	1	3
Cheumatopsyche	14	5
Chimarra	7	4
Chironomidae	10	5
Corbiculidae	8	5
Helicopsyche	14	3
Hydrobiidae	13	4
Hydropsyche	7	5
Isonychia	11	3
Lepidostoma	2	2
Leucrocuta	8	3
Maccaffertium	16	3
Micrasema	1	3
Oecetis	2	3
Oligochaeta	7	5
Optioservus	25	4
Physidae	1	5
Plauditus	7	
Stenelmis	14	5
Teloganopsis	22	3
Tricorythodes	3	5

There are 64 individuals with a BCG of 5, and a total of 197 individuals in the subsample.

(64/197)*100 = 32.5%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Ancylidae	1	4
Baetis	19	5
Cheumatopsyche	71	5
Chironomidae	20	5
Corbiculidae	1	5
Gammarus	39	4
Hirudinea	1	5
Hydropsyche	40	5
Orconectes	1	4
Stenelmis	3	5
Tricorythodes	1	5
Turbellaria	12	5

There are 168 individuals with a BCG of 5, and a total of 209 individuals in the subsample.

(168/209)*100 = 80.4%

Percent Intolerant Individuals using PTV attributes 0-3 (PTVpct03)

$$= \left(\sum_{i=0}^{3} n_{indvPTVi} \Big/ N \right) * 100$$

Where $n_{indvPTVi}$ is the number of individuals in a sub-sample with PTV of i, and N = the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Pollution Tolerance Value
Acroneuria	1	0
Agnetina	2	2
Baetisca	1	4
Brachycentrus	1	1
Cheumatopsyche	14	6
Chimarra	7	4
Chironomidae	10	6
Corbiculidae	8	4
Helicopsyche	14	3
Hydrobiidae	13	8
Hydropsyche	7	5
Isonychia	11	3
Lepidostoma	2	1
Leucrocuta	8	1
Maccaffertium	16	3
Micrasema	1	2
Oecetis	2	8
Oligochaeta	7	10
Optioservus	25	4
Physidae	1	8
Plauditus	7	4
Stenelmis	14	5
Teloganopsis	22	2
Tricorythodes	3	4

There are 78 individuals with a PTV value of 0-3, and a total of 197 individuals in the subsample.

(78/197)*100 = 39.6%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Pollution Tolerance Value
Ancylidae	1	7
Baetis	19	6
Cheumatopsyche	71	6
Chironomidae	20	6
Corbiculidae	1	4
Gammarus	39	4
Hirudinea	1	8
Hydropsyche	40	5
Orconectes	1	6
Stenelmis	3	5
Tricorythodes	1	4
Turbellaria	12	9

There are 0 individuals with a PTV value of 0-3, and a total of 209 individuals in the subsample.

(0/209)*100 = 0%

Hilsenhoff Index using BCG attributes (BCGindex2)

$$= \sum_{i=1}^{6} [(i * n_{indvBCGi})] \Big/ N_{BCG}$$

Where $n_{indvBCGi}$ is the number of individuals in a sub-sample with a BCG of i, and N_{BCG} is the total number of individuals with BCG values in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Acroneuria	1	3
Agnetina	2	3
Baetisca	1	2
Brachycentrus	1	3
Cheumatopsyche	14	5
Chimarra	7	4
Chironomidae	10	5
Corbiculidae	8	5
Helicopsyche	14	3
Hydrobiidae	13	4
Hydropsyche	7	5
Isonychia	11	3
Lepidostoma	2	2
Leucrocuta	8	3
Maccaffertium	16	3
Micrasema	1	3
Oecetis	2	3
Oligochaeta	7	5
Optioservus	25	4
Physidae	1	5
Plauditus	7	
Stenelmis	14	5
Teloganopsis	22	3
Tricorythodes	3	5

There are 0 individuals with a BCG of 1, 3 with a BCG of 2, 78 with a BCG of 3, 45 with a BCG of 4, 64 with a BCG of 5, 0 with a BCG of 6, and a total of 190 BCG individuals in the subsample.

[(1 * 0) + (2 * 3) + (3 * 78) + (4 * 45) + (5 * 64) + (6 * 0)]/190 = 3.89

Mahoning River near Edinburg, PA

118

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Ancylidae	1	4
Baetis	19	5
Cheumatopsyche	71	5
Chironomidae	20	5
Corbiculidae	1	5
Gammarus	39	4
Hirudinea	1	5
Hydropsyche	40	5
Orconectes	1	4
Stenelmis	3	5
Tricorythodes	1	5
Turbellaria	12	5

There are 0 individuals with a BCG of 1, 3 with a BCG of 2, 78 with a BCG of 3, 45 with a BCG of 4, 64 with a BCG of 5, 0 with a BCG of 6, and a total of 209 BCG individuals in the subsample.

$$[(0 * 1) + (0 * 2) + (0 * 3) + (41 * 4) + (168 * 5) + (0 * 6)]/209 = 4.80$$

Percent Dominant Taxon (pctDOM)

$$= \left(\sum n_{indvDOM} / N \right) *100$$

Where $n_{indvDOM}$ is the number of individuals of the dominant taxon in the subsample, and N is the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals
Acroneuria	1
Agnetina	2
Baetisca	1
Brachycentrus	1
Cheumatopsyche	14
Chimarra	7
Chironomidae	10
Corbiculidae	8
Helicopsyche	14
Hydrobiidae	13
Hydropsyche	7
Isonychia	11
Lepidostoma	2
Leucrocuta	8
Maccaffertium	16
Micrasema	1
Oecetis	2
Oligochaeta	7
Optioservus	25
Physidae	1
Plauditus	7
Stenelmis	14
Teloganopsis	22
Tricorythodes	3

There are 25 individuals of the dominant taxon, *Optioservus* spp., and a total of 197 individuals in the subsample.

(25/197)*100 = 12.7%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals
Ancylidae	1
Baetis	19
Cheumatopsyche	71
Chironomidae	20
Corbiculidae	1
Gammarus	39
Hirudinea	1
Hydropsyche	40
Orconectes	1
Stenelmis	3
Tricorythodes	1
Turbellaria	12

There are 71 individuals of the dominant taxon, *Cheumatopsyche* spp., and a total of 209 individuals in the subsample.

(71/209)*100 = 34.0%

Percent Ephemeroptera using BCG attributes 1-3 (pctEbcg13)

$$= \left(\sum_{i=1}^{3} n_{EphemBCGi} \Big/ N \right) *100$$

Where $n_{EphemBCGi}$ is the number of Ephemeroptera individuals in a sub-sample with BCG of i, and N = the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Acroneuria	1	3
Agnetina	2	3
Baetisca	1	2
Brachycentrus	1	3
Cheumatopsyche	14	5
Chimarra	7	4
Chironomidae	10	5
Corbiculidae	8	5
Helicopsyche	14	3
Hydrobiidae	13	4
Hydropsyche	7	5
Isonychia	11	3
Lepidostoma	2	2
Leucrocuta	8	3
Maccaffertium	16	3
Micrasema	1	3
Oecetis	2	3
Oligochaeta	7	5
Optioservus	25	4
Physidae	1	5
Plauditus	7	
Stenelmis	14	5
Teloganopsis	22	3
Tricorythodes	3	5

There are 58 Ephemeroptera individuals with BCG values of 1-3, and a total of 197 individuals in the subsample.

(58/197)*100 = 29.4%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Ancylidae	1	4
Baetis	19	5
Cheumatopsyche	71	5
Chironomidae	20	5
Corbiculidae	1	5
Gammarus	39	4
Hirudinea	1	5
Hydropsyche	40	5
Orconectes	1	4
Stenelmis	3	5
Tricorythodes	1	5
Turbellaria	12	5

There are 0 Ephemeroptera individuals with BCG values of 1-3, and a total of 209 individuals in the subsample.

(0/209)*100 = 0%

Richness of Sensitive Ephemeroptera, Plecoptera, and Trichoptera taxa using BCG attributes 1-3 (richEPTbcg)

$$= n_{taxaEPTbcg}$$

Where $n_{taxaEPTbcg}$ is the number of taxa belonging to the orders Ephemeroptera, Plecoptera, and Trichoptera that have BCG attributes of 1-3.

Delaware River near Callicoon, NY

Taxa Name	Biological Condition Gradient Value
Acroneuria	3
Agnetina	3
Baetisca	2
Brachycentrus	3
Cheumatopsyche	5
Chimarra	4
Chironomidae	5
Corbiculidae	5
Helicopsyche	3
Hydrobiidae	4
Hydropsyche	5
Isonychia	3
Lepidostoma	2
Leucrocuta	3
Maccaffertium	3
Micrasema	3
Oecetis	3
Oligochaeta	5
Optioservus	4
Physidae	5
Plauditus	
Stenelmis	5
Teloganopsis	3
Tricorythodes	5

There are 5 Ephemeroptera taxa with BCG attributes of 1-3, 2 Plecoptera taxa with BCG attributes of 1-3, and 5 Trichoptera taxa with BCG attributes of 1-3.

5 + 2 + 5 = 12

Mahoning River near Edinburg, PA

Taxa Name	Biological Condition Gradient Value
Ancylidae	4
Baetis	5
Cheumatopsyche	5
Chironomidae	5
Corbiculidae	5
Gammarus	4
Hirudinea	5
Hydropsyche	5
Orconectes	4
Stenelmis	5
Tricorythodes	5
Turbellaria	5

There are 0 Ephemeroptera taxa with BCG attributes of 1-3, 0 Plecoptera taxa with BCG attributes of 1-3, and 0 Trichoptera taxa with BCG attributes of 1-3.

0 + 0 + 0 = 0

Metric Standardization and Index Calculation

Final ceiling and floor standardization values are needed to standardize each metric. All standardized metrics are then multiplied by 100 to get the metric standardized score, and the score must range between 0 and 100. Final adjusted metrics scores are then averaged to get a final Summer SWMMI score on a 0 to100 scale.

Summer Metric Standardization Values

Metric	Floor Standardization (5th percentile)	Ceiling Standardization (95th percentile)
BCGpct5	28.5	80.6
PTVpct03	2.3	50.6
BCGindex2	3.76	4.76
pctDOM	14.4	46.8
pctEbcg13	0.4	49.7
richEPTbcg	1	10

For metrics like PTVpct03, pctEbcg13, and richEPTbcg (negative-response metrics), standardizations are calculated using the following equation:

(observed value - floor) / (ceiling - floor) * 100.

For metrics like BCGpct5, BCGindex2, and pctDOM (positive-response metrics) standardizations are calculated using the following equation:

(ceiling - observed value) / (ceiling - floor) * 100.

It is important to note that if a metric standardization score is < 0 then the score is set to 0, and if the metric standardization score is > 100 then the score is set to 100. This process creates the adjusted standardized metric score.

Delaware River near Callicoon, NY

Metric / SWMMI	Observed Value	Standardized Metric Score	Adjusted Standardized Metric Score
BCGpct5	32.5	92.3	92.3
PTVpct03	39.6	77.7	77.7
BCGindex2	3.89	86.7	86.7
pctDOM	12.7	105.2	100
pctEbcg13	29.4	59.1	59.1
richEPTbcg	12	122.2	100
Summer SWMMI	--	--	**86.0**

Mahoning River near Edinburg, PA

Metric / SWMMI	Observed Value	Standardized Metric Score	Adjusted Standardized Metric Score
BCGpct5	80.4	0	0
PTVpct03	0	-4.8	0
BCGindex2	4.80	-4.8	0
pctDOM	34.0	38.9	38.9
pctEbcg13	0	-0.9	0
richEPTbcg	0	-11.1	0
Summer SWMMI	--	--	**6.5**

Fall Metric and Index Calculation

The following fall samples were collected on the Delaware River near Callicoon, NY and the Mahoning River near Edinburg, PA.

Fall sample collected from the Delaware River near Callicoon, NY on December 16th, 2015	
Taxa Name	Number of Individuals
Acroneuria	5
Cheumatopsyche	3
Chimarra	2
Chironomidae	65
Cultus	1
Epeorus	9
Ephemerella	53
Helopicus	2
Hydropsyche	15
Isonychia	3
Lepidostoma	5
Leucrocuta	5
Maccaffertium	28
Nematoda	1
Neophylax	1
Oligochaeta	4
Ophiogomphus	2
Optioservus	15
Oulimnius	1
Paraleptophlebia	2
Psephenus	1
Rhyacophila	2
Stenacron	1
Stenelmis	4
Taeniopteryx	1
Teloganopsis	6

Fall sample collected from the Mahoning River near Edinburg, PA on December 15th, 2015	
Taxa Name	Number of Individuals
Allocapnia	6
Baetis	6
Caecidotea	1
Cheumatopsyche	10
Chironomidae	41
Gammarus	18
Hirudinea	6
Hydropsyche	14
Optioservus	1
Simulium	71
Tipula	1
Tricorythodes	1

Beck's Index using PTV attributes 0-2 (PTVBeck3)

$$= 3 * (n_{taxaPTV0}) + 2 * (n_{taxaPTV1}) + 1 * (n_{taxaPTV2})$$

Where $n_{taxaPTV0}$ is the number of taxa with a PTV attribute of 0, $n_{taxaPTV1}$ is the number of taxa with a PTV attribute of 1, and $n_{taxaPTV2}$ is the number of taxa with a PTV attribute of 2.

Delaware River near Callicoon, NY

Taxa Name	Pollution Tolerance Value
Acroneuria	0
Cheumatopsyche	6
Chimarra	4
Chironomidae	6
Cultus	2
Epeorus	0
Ephemerella	1
Helopicus	2
Hydropsyche	5
Isonychia	3
Lepidostoma	1
Leucrocuta	1
Maccaffertium	3
Nematoda	9
Neophylax	3
Oligochaeta	10
Ophiogomphus	1
Optioservus	4
Oulimnius	5
Paraleptophlebia	1
Psephenus	4
Rhyacophila	1
Stenacron	4
Stenelmis	5
Taeniopteryx	2
Teloganopsis	2

There are 2 taxa with PTV attributes of 0, 6 taxa with PTV attributes of 1, and 4 taxa with PTV attributes of 2.

3 * (2) + 2 * (6) + 1 * (4) = 22

Mahoning River near Edinburg, PA

Taxa Name	Pollution Tolerance Value
Allocapnia	3
Baetis	6
Caecidotea	6
Cheumatopsyche	6
Chironomidae	6
Gammarus	4
Hirudinea	8
Hydropsyche	5
Optioservus	4
Simulium	6
Tipula	4
Tricorythodes	4

There are 0 taxa with PTV attributes of 0, 1 or 2.

3 * (0) + 2 * (0) + 1 * (0) = 0

<u>Richness of Sensitive Ephemeroptera, Plecoptera, and Trichoptera taxa using PTV attributes 0-4 (richEPTptv)</u>

$$= n_{taxaEPTptv}$$

Where $n_{taxaEPTptv}$ is the number of taxa belonging to the orders Ephemeroptera, Plecoptera, and Trichoptera that have PTV attributes of 0-4.

Delaware River near Callicoon, NY

Taxa Name	Pollution Tolerance Value
Acroneuria	0
Cheumatopsyche	6
Chimarra	4
Chironomidae	6
Cultus	2
Epeorus	0
Ephemerella	1
Helopicus	2
Hydropsyche	5
Isonychia	3
Lepidostoma	1
Leucrocuta	1
Maccaffertium	3
Nematoda	9
Neophylax	3
Oligochaeta	10
Ophiogomphus	1
Optioservus	4
Oulimnius	5
Paraleptophlebia	1
Psephenus	4
Rhyacophila	1
Stenacron	4
Stenelmis	5
Taeniopteryx	2
Teloganopsis	2

There are 8 Ephemeroptera taxa with PTV attributes of 0-4, 4 Plecoptera taxa with PTV attributes of 0-4, and 4 Trichoptera taxa with PTV attributes of 0-4.

8 + 4 + 4 = 16

Mahoning River near Edinburg, PA

Taxa Name	Pollution Tolerance Value
Allocapnia	3
Baetis	6
Caecidotea	6
Cheumatopsyche	6
Chironomidae	6
Gammarus	4
Hirudinea	8
Hydropsyche	5
Optioservus	4
Simulium	6
Tipula	4
Tricorythodes	4

There are 1 Ephemeroptera taxa with PTV attributes of 0-4, 1 Plecoptera taxa with PTV attributes of 0-4, and 0 Trichoptera taxa with PTV attributes of 0-4.

1 + 1 + 0 = 2

Percent Intolerant Individuals using PTV attributes 0-3 (PTVpct03)

$$= \left(\sum_{i=0}^{3} n_{indvPTVi} \Big/ N \right) *100$$

Where $n_{indvPTVi}$ is the number of individuals in a sub-sample with PTV of i, and N = the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Pollution Tolerance Value
Acroneuria	5	0
Cheumatopsyche	3	6
Chimarra	2	4
Chironomidae	65	6
Cultus	1	2
Epeorus	9	0
Ephemerella	53	1
Helopicus	2	2
Hydropsyche	15	5
Isonychia	3	3
Lepidostoma	5	1
Leucrocuta	5	1
Maccaffertium	28	3
Nematoda	1	9
Neophylax	1	3
Oligochaeta	4	10
Ophiogomphus	2	1
Optioservus	15	4
Oulimnius	1	5
Paraleptophlebia	2	1
Psephenus	1	4
Rhyacophila	2	1
Stenacron	1	4
Stenelmis	4	5
Taeniopteryx	1	2
Teloganopsis	6	2

There are 125 individuals with a PTV value of 0-3, and a total of 237 individuals in the subsample.

(125/237)*100 = 52.7%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Pollution Tolerance Value
Allocapnia	6	3
Baetis	6	6
Caecidotea	1	6
Cheumatopsyche	10	6
Chironomidae	41	6
Gammarus	18	4
Hirudinea	6	8
Hydropsyche	14	5
Optioservus	1	4
Simulium	71	6
Tipula	1	4
Tricorythodes	1	4

There are 6 individuals with a PTV value of 0-3, and a total of 176 individuals in the subsample.

(6/176)*100 = 3.4%

Percent Ephemeroptera using BCG attributes 1-3 (pctEbcg13)

$$= \left(\sum_{i=1}^{3} n_{EphemBCGi} \middle/ N \right) *100$$

Where $n_{EphemBCGi}$ is the number of Ephemeroptera individuals in a sub-sample with BCG of i, and N = the total number of individuals in the subsample.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Acroneuria	5	3
Cheumatopsyche	3	5
Chimarra	2	4
Chironomidae	65	5
Cultus	1	1
Epeorus	9	2
Ephemerella	53	2
Helopicus	2	3
Hydropsyche	15	5
Isonychia	3	3
Lepidostoma	5	2
Leucrocuta	5	3
Maccaffertium	28	3
Nematoda	1	
Neophylax	1	3
Oligochaeta	4	5
Ophiogomphus	2	3
Optioservus	15	4
Oulimnius	1	2
Paraleptophlebia	2	2
Psephenus	1	4
Rhyacophila	2	2
Stenacron	1	4
Stenelmis	4	5
Taeniopteryx	1	3
Teloganopsis	6	3

There are 106 Ephemeroptera individuals with BCG values of 1-3, and a total of 237 individuals in the subsample.

(106/237)*100 = 44.7%

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Biological Condition Gradient Value
Allocapnia	6	3
Baetis	6	5
Caecidotea	1	5
Cheumatopsyche	10	5
Chironomidae	41	5
Gammarus	18	4
Hirudinea	6	5
Hydropsyche	14	5
Optioservus	1	4
Simulium	71	5
Tipula	1	5
Tricorythodes	1	5

There are 0 Ephemeroptera individuals with BCG values of 1-3, and a total of 176 individuals in the subsample.

(0/176)*100 = 0%

Total Taxa Richness (Richness)

$$= n_{taxa}$$

Where n_{taxa} is the total number of taxa in the subsample.

Delaware River near Callicoon, NY

Taxa Name
Acroneuria
Cheumatopsyche
Chimarra
Chironomidae
Cultus
Epeorus
Ephemerella
Helopicus
Hydropsyche
Isonychia
Lepidostoma
Leucrocuta
Maccaffertium
Nematoda
Neophylax
Oligochaeta
Ophiogomphus
Optioservus
Oulimnius
Paraleptophlebia
Psephenus
Rhyacophila
Stenacron
Stenelmis
Taeniopteryx
Teloganopsis

There are 26 taxa in the subsample.

Mahoning River near Edinburg, PA

Taxa Name

Allocapnia
Baetis
Caecidotea
Cheumatopsyche
Chironomidae
Gammarus
Hirudinea
Hydropsyche
Optioservus
Simulium
Tipula
Tricorythodes

There are 12 taxa in the subsample.

Richness of taxa in the Functional Feeding Group Scrapers (FFGrichSC)

$$= n_{sctaxa}$$

Where n_{sctaxa} is the number of scraper taxa.

Delaware River near Callicoon, NY

Taxa Name	Number of Individuals	Functional Feeding Group
Acroneuria	5	PR
Cheumatopsyche	3	FC
Chimarra	2	FC
Chironomidae	65	CG
Cultus	1	PR
Epeorus	9	SC
Ephemerella	53	CG
Helopicus	2	PR
Hydropsyche	15	FC
Isonychia	3	CG
Lepidostoma	5	SH
Leucrocuta	5	SC
Maccaffertium	28	SC
Nematoda	1	CG
Neophylax	1	SC
Oligochaeta	4	CG
Ophiogomphus	2	PR
Optioservus	15	SC
Oulimnius	1	SC
Paraleptophlebia	2	CG
Psephenus	1	SC
Rhyacophila	2	PR
Stenacron	1	SC
Stenelmis	4	SC
Taeniopteryx	1	SH
Teloganopsis	6	CG

There are 9 scraper taxa in the subsample.

Mahoning River near Edinburg, PA

Taxa Name	Number of Individuals	Functional Feeding Group
Allocapnia	6	SH
Baetis	6	CG
Caecidotea	1	CG
Cheumatopsyche	10	FC
Chironomidae	41	CG
Gammarus	18	CG
Hirudinea	6	PR
Hydropsyche	14	FC
Optioservus	1	SC
Simulium	71	FC
Tipula	1	SH
Tricorythodes	1	CG

There is 1 scraper taxon in the subsample.

Metric Standardization and Index Calculation

Final ceiling and floor standardization values are needed to standardize each metric. All standardized metrics are then multiplied by 100 to get the metric standardized score, and the score must range between 0 and 100. Final adjusted metrics scores are then averaged to get a final fall SWMMI score on a 0 to100 scale.

Fall Metric Standardization Values

Metric	Floor Standardization (5th percentile)	Ceiling Standardization (95th percentile)
PTVBeck3	2	15
richEPTptv	2	15
PTVpct03	3.3	65.3
pctEbcg13	0	62.3
Richness	11	27
FFGrichSC	2	10

For all fall metrics (negative-response metrics), standardizations are calculated using the following equation:

$$\text{(observed value - floor) / (ceiling - floor) * 100.}$$

It is important to note that if a metric standardization score is < 0 then the score is set to 0, and if the metric standardization score is > 100 then the score is set to 100. This process creates the adjusted standardized metric score.

Delaware River near Callicoon, NY

Metric / SWMMI	Observed Value	Standardized Metric Score	Adjusted Standardized Metric Score
PTVBeck3	22	153.8	100
richEPTptv	16	107.7	100
PTVpct03	52.7	79.7	79.7
pctEbcg13	44.7	71.7	71.7
Richness	26	93.7	93.7
FFGrichSC	9	87.5	87.5
Fall SWMMI	--	--	**88.8**

Mahoning River near Edinburg, PA

Metric / SWMMI	Observed Value	Standardized Metric Score	Adjusted Standardized Metric Score
PTVBeck3	0	-15.4	0
richEPTptv	2	0	0
PTVpct03	3.4	0.1	0.1
pctEbcg13	0	0	0
Richness	12	6.25	6.25
FFGrichSC	1	-12.5	0
Fall SWMMI	--	--	**1.1**

1

Appendix B: Taxa Table

The following table lists the PTV, BCG, and FFG attributes of taxa used by PADEP for method development. The FFG abbreviations are as follows CG is collector-gatherer, FC is filter-collector, PI is piercer, PR is predator, SC is scraper, SH is shredder, and UK is unknown.

Taxa Name	PTV	BCG	FFG
Collembola			
Collembola	9		CG
Isotomidae	9		CG
Isotomurus	9		CG
Onychiuridae	9		CG
Onychiurus	9		CG
Podura	9		CG
Poduridae	9		CG
Ephemeroptera			
Acentrella	4	3	SC
Acerpenna	6	3	CG
Ameletidae	0		CG
Ameletus	0	2	CG
Anthopotamus	4	3	CG
Arthroplea	3		SC
Arthropleidae	3		SC
Attenella	2	2	SC
Baetidae	6	3	CG
Baetis	6	5	CG
Baetisca	4	2	CG
Baetiscidae	3		CG
Barbaetis	6		CG
Brachycercus	3		CG
Caenidae	7		CG
Caenis	7	5	CG
Callibaetis	9	4	CG
Centroptilum	2	3	CG
Choroterpes	2	2	CG
Cinygmula	1	1	CG
Cloeon	4	3	CG
Dannella	3	3	CG
Diphetor	6	2	CG
Drunella	1	2	SC
Epeorus	0	2	SC
Ephemera	2	2	CG
Ephemerella	1	2	CG
Ephemerellidae	2		CG
Ephemeridae	4		CG
Ephemeroptera			
Ephoron	2	3	CG
Eurylophella	4	2	SC
Fallceon	6		CG
Habrophlebia	4	3	CG
Habrophlebiodes	6	2	SC
Heptagenia	4	3	SC
Heptageniidae	3		SC
Heterocloeon	2	3	SC
Hexagenia	6	4	CG
Isonychia	3	3	CG
Isonychiidae	3		CG
Leptohyphes	4		CG
Leptophlebia	4	3	CG
Leptophlebiidae	4	2	CG
Leucrocuta	1	3	SC
Litobrancha	6	1	CG
Maccaffertium	3	3	SC
Metrotopus	2		CG
Neoephemera	3		CG
Neoephemeridae	3		CG
Nixe	2	1	SC
Paraleptophlebia	1	2	CG
Penelomax	2		CG
Pentagenia	4		CG
Plauditus	4		CG
Polymitarcyidae	2		CG
Procloeon	6	4	CG

Taxon			
Pseudocloeon	4	3	CG
Rhithrogena	0	2	CG
Serratella	2	3	CG
Siphlonurus	7		CG
Siphloplecton	2	2	CG
Stenacron	4	4	SC
Stenonema	4	4	SC
Stenonema(old genus)	3	3	SC
Teloganopsis	2	3	CG
Tricorythidae	4		CG
Tricorythodes	4	5	CG

Odonata

Taxon			
Aeshna	5	4	PR
Aeshnidae	3		PR
Amphiagrion	5		PR
Anax	5	4	PR
Aphylla	4		PR
Argia	6	4	PR
Arigomphus	4	4	PR
Basiaeschna	2	4	PR
Boyeria	2	3	PR
Calopterygidae	5	4	PR
Calopteryx	6	4	PR
Celithemis	2		PR
Chromagrion	4		PR
Coenagrionidae	8	4	PR
Cordulegaster	3	3	PR
Cordulegastridae	3		PR
Cordulia	4		PR
Corduliidae	5		PR
Didymops	4	4	PR
Dorocordulia	4		PR
Dromogomphus	4	4	PR
Enallagma	8	4	PR
Epiaeschna	2		PR
Epitheca	4		PR
Erythemis	5		PR
Erythrodiplax	5		PR
Gomphaeschna	2	4	PR
Gomphidae	4	3	PR
Gomphus	5	4	PR
Hagenius	3	3	PR
Helocordulia	2		PR
Hetaerina	6	4	PR
Ischnura	9	4	PR
Ladona	6		PR
Lanthus	5	2	PR
Leucorrhinia	6		PR
Libellula	8		PR
Libellulidae	9		PR
Macromia	2	4	PR
Nannothemis	6		PR
Nasiaeschna	2		PR
Nehalennia	7		PR
Neurocordulia	3		PR
Odonata			PR
Ophiogomphus	1	3	PR
Pachydiplax	8		PR
Pantala	7		PR
Perithemis	4		PR
Petaluridae	5		PR
Plathemis	3		PR
Progomphus	5	3	PR
Somatochlora	1	2	PR
Stylogomphus	4	4	PR
Stylurus	4	4	PR
Sympetrum	4		PR
Tachopteryx	5		PR
Tramea	4		PR
Williamsonia	4		PR

Plecoptera

Taxon			
Acroneuria	0	3	PR
Agnetina	2	3	PR
Allocapnia	3	3	SH
Alloperla	0	1	CG
Amphinemura	3	3	SH
Arcynopteryx	2		PR
Attaneuria	3	2	PR
Bolotoperla	2		SH
Capnia	1		SH
Capniidae	3	3	SH
Chloroperlidae	0	2	PR
Clioperla	2		PR
Cultus	2	1	PR
Diploperla	2	2	PR

Taxon				Taxon			
Diura	2		PR	Yugus	2	1	PR
Eccoptura	2	2	PR	Zapada	2		SH
Hansonoperla	3		PR	Zealeuctra	0		SH
Haploperla	0	3	PR	Capnura	1		SH
Helopicus	2	3	PR	Remenus	2	1	PR
Hydroperla	1		PR	Utacapnia	1		SH
Isogenoides	0	1	PR	Utaperla	0		PR
Isoperla	2	2	PR	Paranemoura	2		SH
Leuctra	0	2	SH	Hemiptera			
Malirekus	2	1	PR	Aquarius	9		PR
Megaleuctra	0		SH	Belostoma	9		PR
Nemocapnia	1		SH	Belostomatidae	9		PR
Nemoura	1	1	SH	Buenoa	8		PR
Nemouridae	2	3	SH	Ceratocombidae	9		PR
Neoperla	3	2	PR	Ceratocombus	9		PR
Oconoperla	2		PR	Corixidae	8	5	PR
Oemopteryx	3	2	SH	Gelastocoridae	8		PR
Ostrocerca	2	1	SH	Gelastocoris	8		PR
Paracapnia	1	2	SH	Gerridae	9		PR
Paragnetina	1	2	PR	Gerris	9		PR
Paraleuctra	0	1	SH	Halobates	9		PR
Peltoperla	2	1	SH	Hebridae	8		PR
Peltoperlidae	2	2	SH	Hebrus	8		PR
Perlesta	4	3	PR	Hemiptera			
Perlidae	3	3	PR	Hesperocorixa	5	5	PR
Perlinella	2	2	PR	Hydrometridae	9		PR
Perlodidae	2	2	PR	Lethocerus	9		PR
Plecoptera			PR	Limnoporus	9		PR
Podmosta	2		SH	Merragata	8		PR
Prostoia	2	3	SH	Mesovelia	9		PR
Pteronarcyidae	0		SH	Mesoveliidae	9		PR
Pteronarcys	0	2	SH	Metrobates	9		PR
Rasvena	0	1	PR	Micracanthia	8		PR
Shipsa	2	1	SH	Microvelia	9		PR
Soyedina	0	1	SH	Naucoridae	8		PR
Strophopteryx	3	3	SH	Neoplea	8		PR
Suwallia	0	1	CG	Nepa	8		PR
Sweltsa	0	3	PR	Nepidae	8		PR
Taenionema	3	1	SH	Notonecta	8		PR
Taeniopterygidae	2	3	SH	Notonectidae	8		PR
Taeniopteryx	2	3	SH	Ochteridae	8		PR
Tallaperla	0	1	SH	Ochterus	8		PR
Viehoperla	2		SH	Palmacorixa	8	4	PR

Taxon				Taxon			
Pelocoris	8		PR	Ceraclea	3	3	CG
Pentacora	8		PR	Ceratopsyche	5	4	FC
Pleidae	8		PR	Cernotina	6		PR
Ramphocorixa	8	4	PR	Cheumatopsyche	6	5	FC
Ranatra	8		PR	Chimarra	4	4	FC
Rhagovelia	9		PR	Clostoeca	5		SH
Rheumatobates	9		PR	Culoptila	1	3	SC
Salda	8		PR	Cyrnellus	8	5	FC
Saldidae	8		PR	Dibusa	4		SC
Saldula	8		PR	Diplectrona	0	2	FC
Sigara	8	4	PI	Dolophilodes	0	2	FC
Steinovelia	9		PR	Fabria	4		SH
Trepobates	9		PR	Frenesia	4		SH
Trichocorixa	8	5	PR	Glossosoma	0	3	SC
Veliidae	8		PR	Glossosomatidae	0	3	SC
Neuroptera				Glyphopsyche	3		SH
Climacia	1		PI	Goera	0	1	SC
Neuroptera	3		PR	Hagenella	5		SH
Sisyra	1		PI	Helicopsyche	3	3	SC
Sisyridae	1		PI	Helicopsychidae	3		SC
Megaloptera				Hesperophylax	4	3	CG
Chauliodes	4	4	PR	Heteroplectron	5	1	SH
Corydalidae	3		PR	Homoplectra	5		FC
Corydalus	4	4	PR	Hydatophylax	2	2	SH
Megaloptera	8		PR	Hydropsyche	5	5	FC
Nigronia	2	3	PR	Hydropsychidae	5		FC
Sialis	6	5	PR	Hydroptila	6	5	SC
Neohermes	2		PR	Hydroptilidae	4		PI
Trichoptera				Ironoquia	3		SH
Adicrophleps	2	1	SH	Ithytrichia	6		SC
Agapetus	0	3	SC	Lepidostoma	1	2	SH
Agarodes	3	2	SH	Lepidostomatidae	1	2	SH
Agraylea	8	4	CG	Leptoceridae	4		PR
Agrypnia	7		SH	Leptocerus	3		SH
Anabolia	5		SH	Leptophylax	2		SH
Apatania	3	2	SC	Leucotrichia	6	4	SC
Arctopora	5		SH	Limnephilidae	4	3	SH
Arctopsyche	1		FC	Limnephilus	3	3	SH
Banksiola	2		SH	Lype	2	2	CG
Beraeidae	3		SC	Macrostemum	3	4	FC
Brachycentridae	1	2	FC	Madeophylax	4		SH
Brachycentrus	1	3	FC	Mayatrichia	4		SC
Calamoceratidae	5		SH	Micrasema	2	3	SH

Taxon				Taxon			
Molanna	6	2	SC	Lepidoptera			
Molannidae	6		SC	Acentria	5		SH
Mystacides	4	3	CG	Acigona	5		SH
Nectopsyche	3	3	SH	Archanara	5		SH
Neophylax	3	3	SC	Archips	5		SH
Neotrichia	2		SC	Bellura	5		SH
Neureclipsis	7	3	FC	Chilo	5		SH
Nyctiophylax	5	4	PR	Coleophoridae	6		SH
Ochrotrichia	4		SC	Colephora	6		SH
Odontoceridae	0	1	SH	Eoparargyractis	5		SH
Oecetis	8	3	PR	Langessa	5		SH
Oligostomis	5		SH	Lepidoptera	5		SH
Onocosmoecus	3		SH	Munroessa	5		SH
Orthotrichia	6		SH	Neocataclysta	5		SH
Oxyethira	3	2	CG	Nepticulidae	5		SH
Palaeagapetus	1	1	SH	Noctuidae	5		SH
Parapsyche	0	1	FC	Nymphula	7		SH
Philarctus	3		SH	Nymphuliella	5		SH
Philopotamidae	3		FC	Ostrinia	5		SH
Phryganea	8		SH	Parapoynx	5		SH
Phryganeidae	4		SH	Petrophila	5	5	SC
Phylocentropus	5	4	FC	Pyralidae	5		SH
Platycentropus	4	3	SH	Schoenobius	5		SH
Polycentropodidae	6		FC	Simyra	5		SH
Polycentropus	6	4	FC	Stigmella	5		SH
Potamyia	5	3	FC	Synclita	5		FC
Protoptila	1	2	SC	Tortricidae	5		SH
Pseudostenophylax	0	3	SH	Cosmopterigidae	5		SH
Psilotreta	0	1	SC	Cosmopteryx	5		SH
Psychomyia	2	3	CG	Lymnaecia	5		SH
Psychomyiidae	2	3	CG	Coleoptera			
Ptilostomis	5	2	SH	Acilius	5	4	PR
Pycnopsyche	4	3	SH	Agabetes	5	4	PR
Rhyacophila	1	2	PR	Agabus	5	4	PR
Rhyacophilidae	1		SC	Anacaena	5		PR
Sericostomatidae	3		SH	Anchytarsus	5	2	SH
Setodes	2	2	CG	Ancyronyx	2	4	CG
Stactobiella	2		SC	Auleutes	6		SH
Triaenodes	6	3	SH	Bagous	6		SH
Trichoptera				Berosus	5	5	PR
Uenoidae	3		SC	Bidessonotus	5		PR
Wormaldia	0	1	FC	Bledius	5		PR
Beraea	3		SC	Brachybamus	6		SH

Brachyvatus	5		PR	Helophorus	5		SH
Carpelimus	5		PR	Histeridae	5		SH
Celina	5		PR	Hydaticus	5		PR
Chaetarthria	5		PR	Hydraena	6		CG
Chrysomelidae	5		SH	Hydraenidae	6		CG
Coleoptera				Hydrobius	5		PR
Colymbetes	5		PR	Hydrocanthus	5		PR
Copelatus	5	4	PR	Hydrochara	5		PR
Coptotomus	5		PR	Hydrochidae	5		SH
Crenitis	5		PR	Hydrochus	5		SH
Curculionidae	6		SH	Hydrophilidae	5		PR
Cybister	5	4	PR	Hydrophilus	5		PR
Cymbiodyta	5		PR	Hydroporus	5	4	PR
Cyphon	8		SC	Hydrovatus	5	4	PR
Derallus	5		PR	Hygrotus	5		PR
Desmopachria	5		PR	Ilybius	5	4	PR
Dibolocelus	5		PR	Laccobius	5		PR
Dicranopselaphus	4		SC	Laccophilus	5	4	PR
Dineutus	4	4	PR	Laccornis	5		PR
Disonycha	5		SH	Limnebius	6		CG
Donacia	5		SH	Liodessus	5		PR
Dryopidae	5		SC	Lioporius	5		PR
Dryops	5		SC	Lissorhoptrus	6		SH
Dubiraphia	6	4	SC	Listronotus	6		SH
Dytiscidae	5	4	PR	Lixellus	6		SH
Dytiscus	5		PR	Lixus	6		SH
Ectopria	5	3	SC	Lutrochidae	6		UK
Elmidae	5		CG	Lutrochus	6		UK
Elodes	8		SC	Macronychus	2	4	SC
Enochrus	5		PR	Matus	5		PR
Eubrianax	4		SC	Microcylloepus	2	4	SC
Euhrychiopsis	6		SH	Nebrioporus	5		PR
Flavohelodes	8		SC	Neohaemonia	5		SH
Gonielmis	5		SC	Noteridae	5		PR
Graphoderus	5		PR	Ochthebius	6		CG
Gyrinidae	4	4	PR	Optioservus	4	4	SC
Gyrinus	4	4	PR	Ordobrevia	5		SC
Haliplidae	5		SH	Oreodytes	5		PR
Haliplus	5		SH	Oulimnius	5	2	SC
Helichus	5	4	SC	Paracymus	5		PR
Helochares	5		PR	Pelenomus	6		SH
Helocombus	5		PR	Peltodytes	5		SH
Helophoridae	5		SH	Perenthis	6		SH

Phytobius	6		SH	Blera	10		CG
Prasocuris	5		SH	Brachypogon	6		PR
Promoresia	2	2	SC	Brachypremna	4		SH
Pronoterus	5		PR	Callicera	10		CG
Psephenidae	4		SC	Caloparyphus	8		CG
Psephenus	4	4	SC	Campsicnemus	4		CG
Psephidonus	5		PR	Caricea	6		PR
Pyrrhalta	5		SH	Ceratopogon	6	4	PR
Rhantus	5		PR	Ceratopogonidae	6	4	PR
Scirtes	8		SC	Ceriana	10		CG
Scirtidae	8		SC	Chalcosyrphus	10		CG
Spanglerogyrus	4		PR	Chaoboridae	8		PR
Sperchopsis	5		PR	Chaoborus	8		PR
Staphylinidae	5		PR	Chelifera	6	4	PR
Stenelmis	5	5	SC	Chelipoda	6		PR
Stenopelmus	6		SH	Chironomidae	6	5	CG
Stenus	5		PR	Chrysogaster	10		CG
Steremnius	6		SH	Chrysops	7	5	PI
Stictotarsus	5		PR	Clinocera	6	4	PR
Suphis	5		PR	Clinohelea	6		PR
Suphisellus	5		PR	Cnephia	4	3	FC
Tanysphyrus	6		SH	Cordilura	6		SH
Thinobius	5		PR	Cryptolabis	4	3	CG
Tropisternus	5		PR	Culex	8		FC
Uvarus	5	4	PR	Culicidae	8		FC
Diptera				Culicoides	10	4	PR
Acanthocnema	6		SH	Culiseta	8		FC
Aedes	8		FC	Dactylolabis	4		SH
Alluaudomyia	6	4	PR	Dasyhelea	6	4	CG
Anopheles	8		FC	Diachlorus	6		PR
Antocha	3	4	CG	Dicranota	3	3	PR
Arctoconopa	4		SH	Diptera			
Argyra	4		PR	Dixa	1	2	CG
Asyndetus	4		PR	Dixella	1		CG
Athericidae	2		PR	Dixidae	1	2	CG
Atherix	2	3	PR	Dohrniphora	6		CG
Atrichopogon	2	4	PR	Dolichocephala	5		PR
Atylotus	6		PI	Dolichopodidae	4		PR
Bezzia	6	4	PR	Dolichopus	4		PR
Bittacomorpha	8		CG	Ectemnia	1		FC
Bittacomorphella	8		CG	Elliptera	4		SH
Blepharicera	0	1	SC	Empididae	6	4	PR
Blephariceridae	0		SC	Ephydridae	6	5	PI

Taxon				Taxon			
Erioptera	7	4	CG	Neoplasta	6		PR
Eristalinus	10		CG	Nilobezzia	6		PR
Euparyphus	8		CG	Nymphomyiidae	6		SC
Forcipomyia	6	4	SC	Odontomyia	8		CG
Glutops	5		PR	Oreogeton	6		PR
Gonomyia	4		SH	Oreothalia	6		PR
Greniera	6		FC	Ormosia	6	3	CG
Haematopota	6		PR	Orthacheta	6		PR
Hedriodiscus	8		SC	Orthopodomyia	8		FC
Helius	4		SH	Oxycera	8		SC
Helophilus	10		CG	Palaeodipteron	6		SC
Hemerodromia	6	4	PR	Palpomyia	6	4	PR
Hercostomus	4		PR	Paradelphomyia	4		SH
Hexatoma	2	3	PR	Pedicia	6	3	PR
Hybomitra	6		PR	Pelastoneurus	4		PR
Hydromyza	6		SH	Pelecorhynchidae	5		PR
Hydrophorus	4		PR	Pericoma	4	5	CG
Hypocharassus	4		PR	Phalacrocera	4		SH
Johannsenomyia	6		PR	Phaonia	6		PR
Labostigmina	8		CG	Philosepedon	10		CG
Leptoconops	6		PR	Phoridae	6		CG
Leptopsilopa	6		CG	Pilaria	7	4	PR
Leptotarsus	4		SH	Prionocera	4		SH
Liancalus	4		PR	Probezzia	6	4	PR
Limnophila	3	4	PR	Proclinopyga	6		PR
Limnophora	6		PR	Prosimulium	2	3	FC
Limonia	6	4	SH	Protoplasa	6		CG
Lipsothrix	4	4	SH	Pseudolimnophila	2	4	PR
Lispe	6		PR	Psilopa	6		CG
Lispoides	6		PR	Psorophora	8		PR
Mallochohelea	6	4	PR	Psychoda	10	5	CG
Mallota	10		CG	Psychodidae	10	5	CG
Mansonia	8		FC	Ptychoptera	8		CG
Megaselia	6		CG	Ptychopteridae	8		CG
Merycomyia	6		PR	Rhabdomastix	4		SH
Metachela	6		PR	Rhamphomyia	6		PR
Mochlonyx	8		PR	Rhysophora	6		SH
Molophilus	4	3	SH	Roederiodes	6		PR
Monohelea	6		PR	Sargus	8		CG
Muscidae	6		PR	Scathophagidae	6		SH
Myolepta	10		CG	Sciomyzidae	10		PR
Nemotelus	8		CG	Sericomyia	10		CG
Neoascia	10		CG	Serromyia	6		PR

Taxon			
Simuliidae	6		FC
Simulium	6	5	FC
Spaziphora	6		SC
Sphaeromias	6		PR
Spilogona	6		PR
Spilomyia	10		CG
Stegopterna	6		FC
Stilobezzia	6	4	PR
Stilpon	6		PR
Stratiomyidae	8	6	CG
Stratiomys	5		CG
Sympycnus	4		PR
Syrphidae	10		CG
Tabanidae	6	5	PI
Tabanus	5	5	PR
Tachytrechus	4		PR
Telmatoscopus	10	5	CG
Telmaturgus	4		PR
Thaumalea	6		SC
Thinophilus	4		PR
Threticus	10		CG
Tipula	4	5	SH
Tipulidae	4	4	SH
Toxorhynchites			PR
Trichoclinocera	6		PR
Triogma	4		SH
Twinnia	6		FC
Ulomorpha	4		PR
Uranotaenia	8		FC
Wyeomyia	8		FC
Other			
Amphipoda	6	4	CG
Crangonyctidae	4		CG
Crangonyx	4	4	CG
Gammaridae	4		CG
Gammarus	4	4	CG
Haustoriidae	5		CG
Hyalella	8	4	CG
Monoporeia	5		CG
Stygonectes	4		CG
Hydridae	4		PR
Ampullaridae	7		SC
Viviparidae	7	4	CG

Taxon			
Ancylidae	7	4	SC
Lymnaeidae	7	5	SC
Physidae	8	5	SC
Planorbidae	6	5	SC
Branchiobdellida	6	4	CG
Cambaridae	6	4	CG
Cambarus	6	4	CG
Decapoda		4	UK
Fallicambarus	6		CG
Orconectes	6	4	CG
Procambarus	6		SH
Cladocera	5		FC
Spongillidae	4		FC
Tubificidae	10	5	CG
Valvatidae	2	4	SC
Asellidae	8	5	CG
Caecidotea	6	5	CG
Isopoda	8	5	CG
Lirceus	8	6	CG
Micromelaniidae	7		SC
Pomatiopsidae	8		SC
Bithyniidae	7		SC
Hydrobiidae	8	4	SC
Pleuroceridae	7	4	SC
Petasidae	4		PR
Hydracarina	7	4	PR
Margaritiferidae	5		FC
Unionidae	4		FC
Corbiculidae	4	5	FC
Dreissenidae	5		FC
Sphaeriidae	8		FC
Bryozoa	4		FC
Cavidae	4		PR
Gastropoda			
Hirudinea	8	5	PR
Nematoda	9		CG
Nematomorpha	9		CG
Nemertea	6	4	PR
Oligochaeta	10	5	CG
Ostracoda	8		CG
Polychaeta	10		FC
Turbellaria	9	5	PR