

2015

## Fish communities as indicators of ecological health in West Virginia rivers

Alison M. Anderson

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Anderson, Alison M., "Fish communities as indicators of ecological health in West Virginia rivers" (2015). *Graduate Theses, Dissertations, and Problem Reports*. 5101.  
<https://researchrepository.wvu.edu/etd/5101>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

**FISH COMMUNITIES AS INDICATORS OF ECOLOGICAL HEALTH IN  
WEST VIRGINIA RIVERS**

Alison M. Anderson

Dissertation submitted to the  
Davis College of Agriculture, Natural Resources, and Design  
At West Virginia University  
in partial fulfillment of the requirements of the degree of

Doctor of Philosophy  
in  
Forest Resources Management

Dr. J. Todd Petty, Committee Chair  
Dr. George Merovich  
Dr. Stuart A. Welsh  
Dr. Yong-Lak Park  
Dr. Nathaniel P. Hitt

School of Natural Resources  
Wildlife and Fisheries Resources Program

Morgantown, West Virginia  
2015

Keywords: bioassessment, community ecology, ecosystem health, ecological indicators, Index of Biotic Integrity, mountain-top mining, stream fish ecology, water quality, watershed management

Copyright 2015 Alison M. Anderson

## ABSTRACT

### **Fish communities as indicators of ecological health in West Virginia Rivers**

Alison M. Anderson

Anthropogenic altered landscapes from the extraction of natural resources, urbanization, agricultural development, dam construction, and general conversion from forested to non-forested landscape have resulted in the loss of biodiversity and changes in ecosystem function for aquatic environments. The hierarchical structure of stream ecosystems, in which sites are nested in streams, and streams are nested in watersheds (at multiple spatial scales), provides a unique opportunity for evaluating the influence of both natural and anthropogenic impacts on aquatic communities. In essence, landscape alterations at one scale of the hierarchical structure of stream ecosystems can ultimately impact the distribution of an organism or assembly of the stream community. Throughout this research, we have used detailed landscape data and a large fish community dataset to develop management tools and analyses in order to evaluate the condition of rivers and streams in West Virginia.

For Chapter 1, we developed a fish-based Index of Biotic Integrity (IBI), a common ecological tool for assessing the health of fish communities, in order to inform state and federal regulatory agencies of the impairment status of warm water, wadeable stream and rivers in West Virginia. Based on fish distributions and assemblage metrics within reference sites, we identified 5 distinct biomonitoring regions for which we constructed separate warm water IBIs. Final lists consisting of 7 – 9 metrics were retained within each region for the inclusion into a final IBI. Common metrics retained in the final IBIs included measures of benthic associated species, total species richness, clean-gravel spawning species, tolerance to stressors, feeding classification (e.g., invertivores, omnivore-herbivores) and taxonomic group membership (e.g., Family Cyprinidae, Family Cottidae, darters, and madtoms). In general, final IBI scores from each of the biomonitoring regions were sensitive to at least one anthropogenic stressor, such as surface mining, agriculture, and urbanization. The Mon CA-RV region had the highest correlation between IBI score, benthic macroinvertebrate indices (West Virginia Stream Condition Index and Genus-Level Indicator of Most Probable Stream Status), total fish abundance, Mid-Atlantic Highlands IBI scores, and specific conductance. The Ohio CA biomonitoring region had the highest correlation between IBI scores and % surface mining. Patterns of IBI and metric response in most regions were consistent with other studies showing negative impacts of anthropogenic land-use on stream condition. However, the differential response of IBIs to land-use patterns and other measures of biotic condition indicate that stressor responses are region and organism specific. Using both a fish and benthic macroinvertebrate measure together should enhance current biomonitoring and assessment criteria in addition to providing multiple avenues for evaluating current and future land-use practices.

In the remaining chapters, we evaluated alternative methods for portions of IBI development and relationships of the final IBI and metrics to both natural and anthropogenic landscape characteristics. Specifically for Chapter 2, we used a modeling framework (i.e., boosted regression tree modeling) in order to generate an anthropogenic condition gradient based on currently defined reference sites. We used the reference condition probabilities to locate additional reference sites that span a gradient of natural environmental conditions to be used in bioassessment development as well as define regions of high quality for future sampling or conservation efforts. We then used this larger sample of reference

sites to model the expected condition of trait-based fish community metrics using a variety of natural landscape variables (i.e., drainage area, elevation, and distance from a source river) in chapter 3. In order to generate predictive models, we used a boosted regression tree (BRT) framework in which we also analyzed the effectiveness of BRT models by developing thresholds of model use using Monte Carlo simulations. Overall, the metrics analyzed for chapter 3 showed distinct regional difference in their natural condition BRT models as well as their correlations with anthropogenic stressors within each biomonitoring region. Using predictive models in bioassessment development has allowed for the production of IBIs and metrics that are sensitive to landscape and fish community alterations.

Finally, for Chapter 4, we recognized the importance of the hierarchical nature of aquatic systems in determining local fish community structure. We evaluated different measures of fish community assembly (species richness, diversity, WV IBI scores, and proportion of tolerant individuals) and how local and neighborhood level landscape structure in addition to natural variables, can impact stream community health in Central Appalachian watersheds. In order to account for the nested structure of our data, we employed a mixed-effects modeling approach, which has the ability to incorporate non-independent sampling locations by establishing a fixed-effect grouping variable. For our analysis, we used neighborhood (HUC 12 watershed) as a grouping variable. Generally, each community variable, with exception of % tolerant individuals, was responsive to local landscape structure. These variables demonstrated decreases with increase in measures of surface mining intensity. However, they also demonstrated increases with increases in residential development indicating that residential development in the Central Appalachians may serve as a refuge from mining related stressors by providing increased levels of nutrients and release from degraded water quality. Interestingly, we observed no landscape control on % tolerant individuals within this region. The lack of anthropogenic and natural controls on tolerant individuals may point to increased homogenization of fish communities dominated by tolerant individuals in this degraded landscape.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b>	<b>VI</b>
-------------------------	-----------

<b>CHAPTER 1: FISH BASED INDEX OF BIOTIC INTEGRITY FOR WARM WATER WADEABLE STREAMS IN WEST VIRGINIA</b>	<b>1</b>
---	----------

<b>ABSTRACT</b>	<b>2</b>
<b>1.0 INTRODUCTION</b>	<b>4</b>
<b>2.0 METHODS</b>	<b>7</b>
2.1 SOURCES OF DATA	7
2.2 SITE CLASSIFICATION	9
2.3 DATA ANALYSIS	11
2.3.1 Community Classification into Biomonitoring Regions	11
2.3.2 Metric Selection and Index Construction	12
2.3.3 Metric scoring and aggregation into final indices	14
2.3.4 Index Performance	15
<b>3.0 RESULTS</b>	<b>15</b>
3.1 COMMUNITY CLASSIFICATION INTO BIOMONITORING REGIONS	15
3.2 METRIC EVALUATION AND SELECTION	16
3.3 INDEX PERFORMANCE	17
<b>4.0 DISCUSSION</b>	<b>18</b>
<b>LITERATURE CITED</b>	<b>24</b>
<b>TABLES</b>	<b>29</b>
<b>FIGURES</b>	<b>39</b>
<b>APPENDIX A</b>	<b>43</b>
<b>APPENDIX B</b>	<b>51</b>
<b>APPENDIX C</b>	<b>53</b>

<b>CHAPTER 2: LANDSCAPE BASED MODELS FOR IDENTIFYING LEAST IMPACTED STREAMS IN CENTRAL APPALACHIAN WATERSHEDS</b>	<b>68</b>
---	-----------

<b>ABSTRACT</b>	<b>69</b>
<b>1.0 INTRODUCTION</b>	<b>71</b>
<b>2.0 METHODS</b>	<b>75</b>
2.1 LANDUSE/LANDCOVER DATA	75
2.2 CURRENT REFERENCE POOL	76
2.3 BOOSTED REGRESSION TREE MODELS	77
2.4 SITE SELECTION	78
<b>3.0 RESULTS</b>	<b>79</b>
3.1 BOOSTED REGRESSION TREE MODELS	79
3.2 REFERENCE SITE SELECTION	80
3.3 SAMPLING PRIORITY SELECTION	81
3.4 HIERARCHICAL CLASSIFICATION OF REFERENCE CONDITION	82
<b>4.0 DISCUSSION</b>	<b>83</b>
<b>LITERATURE CITED</b>	<b>87</b>
<b>TABLES</b>	<b>90</b>

<b>FIGURES</b>	<b>94</b>
----------------	-----------

---

<b><u>CHAPTER 3: MODELING THE NATURAL VARIATION OF TRAIT-BASED FISH COMMUNITY STRUCTURE IN WEST VIRGINIA</u></b>	<b>99</b>
--	-----------

---

<b>ABSTRACT</b>	<b>100</b>
<b>1. INTRODUCTION</b>	<b>102</b>
<b>2. METHODS</b>	<b>104</b>
2.1 LANDSCAPE DATA	105
2.2 FISH COMMUNITY DATA & WV IBI METRICS	105
2.3 REFERENCE SITES	106
2.4 STATISTICAL ANALYSES	107
2.4.1 Modeling Expectations	107
2.4.2 Monte Carlo Analysis	108
2.4.3 Evaluation of Adjusted Metric Performance	108
<b>3. RESULTS</b>	<b>109</b>
3.1 DESCRIPTION OF REFERENCE SITES	109
3.2 BOOSTED REGRESSION TREE MODELS	110
3.2.1 Modeling Expectations	110
3.2.2 Monte Carlo Analysis	111
3.3 EVALUATION OF ADJUSTED WV IBI METRICS	113
<b>4.0 DISCUSSION</b>	<b>113</b>
<b>LITERATURE CITED</b>	<b>118</b>
<b>TABLES</b>	<b>122</b>
<b>FIGURES</b>	<b>129</b>
<b>APPENDIX A</b>	<b>131</b>

---

<b><u>CHAPTER 4: LOCAL AND NEIGHBORHOOD SCALE CONTROLS ON FISH COMMUNITY STRUCTURE IN CENTRAL APPALACHIAN WATERSHEDS</u></b>	<b>143</b>
--	------------

---

<b>ABSTRACT</b>	<b>144</b>
<b>1.0 INTRODUCTION</b>	<b>146</b>
<b>2.0 METHODS</b>	<b>149</b>
2.1 STUDY AREA	149
2.2 FISH COMMUNITY VARIABLES	149
2.3 LANDSCAPE ATTRIBUTES	150
2.4 STATISTICAL ANALYSES	151
<b>3.0 RESULTS</b>	<b>153</b>
3.1 LOCAL AND NEIGHBORHOOD LANDSCAPE PRINCIPAL COMPONENTS ANALYSES	153
3.2 SPATIAL AUTOCORRELATION OF FISH COMMUNITY VARIABLES	154
3.3 RESPONSES OF FISH COMMUNITY TO LOCAL AND NEIGHBORHOOD LAND-USE	154
<b>4.0 DISCUSSION</b>	<b>156</b>
<b>LITERATURE CITED</b>	<b>161</b>
<b>TABLES</b>	<b>164</b>
<b>FIGURES</b>	<b>168</b>

## ACKNOWLEDGEMENTS

I would first like to thank my family, especially my mother, Peggy Key, fathers, Sam Key and Brian Anderson, and numerous brothers and sisters for supporting me unconditionally throughout this process. I would like to extend my thank you to my mother for impressing upon me at a very early age the importance and beauty of nature and the sciences. I owe all that I am to my mother. All my current successes in life are due to her moral, intellectual, and physical education and words cannot describe the respect and admiration I have for her. I would not have developed a passion for water conservation and management and a love for aquatic organisms if it was not for her. I would also like to thank my sister-in-law, Dr. Christine Anderson. Without knowing, she has been an outstanding female role model for both my personal and professional development in the field of ecology.

I would also like to thank past and present graduate students, staff, and faculty of West Virginia University. I would especially like to thank my friends and colleagues, Eric R. Merriam and Eric M. Miller. I have worked very closely with both individuals since arriving on campus in 2009. Our travels to remote regions of West Virginia and Kentucky and months spent collecting field data together will be among my favorite memories of graduate school. I would also like to thank my friends and colleagues Donna Hartman, Dr. Patricia Mazik, Justine Whitaker, Andrew Watson, Dr. Brock Huntsman, Ross Andrew, Daniel Hanks, and Geriann Albers for their help with emotional, physical, and statistical support during this research process. I would also like to thank the multitude of undergraduate and graduate students who helped in the lab and the field for their hard work and comedic relief while traveling around the state.

In addition, I would like to thank John Wirts, Jason Morgan, and Jeff Bailey and the West Virginia Department of Environmental Protection for their support and allowing me to lead in the development of the West Virginia Index of Biotic Integrity (IBI; Chapter 1). It has surely been a one-of-a-kind experience. I would also like to thank Dan Cincotta and Dave Thorne and the West Virginia Department of Natural Resources for providing assistance with data collect and supply. Dan Cincotta also provided invaluable information during my teaching assistantship. His knowledge of fishes within West Virginia and the country are unparalleled and I am honored to have worked closely with him during my time at WVU. I would also like to thank Louis Reynolds from the US EPA for his analytical support during the IBI process.

Lastly, I would like to thank the members of my graduate committee – Dr. J. Todd Petty, Dr. George Merovich, Dr. Stuart Welsh, Dr. Yong-Lak Park, and Dr. Nathaniel P. Hitt – for their guidance and encouragement throughout my dissertation. I thank Dr. Park for introducing me to spatial data analysis. I thank Dr. Welsh for transferring some of his vast knowledge on stream fish ecology to me during the dissertation process. I thank Dr. Hitt for providing endless discussions on fish ecology and statistical modeling. Similarly, I would like to thank Dr. Merovich for never-ending statistical support and discussions throughout graduate school. In addition, he provided me with invaluable teaching experiences that will help further my professional and personal careers. Finally, I would like to thank my graduate advisor, Dr. Todd Petty. His endless trust, patience, support, and criticisms will continue to inspire my future in aquatic ecology.

Funding for this project was provided, in part, by West Virginia Department of Environmental Protection, Appalachian Research Initiative for Environmental Sciences (ARIES), and West Virginia University.

**CHAPTER 1: FISH BASED INDEX OF BIOTIC INTEGRITY FOR WARM WATER  
WADEABLE STREAMS IN WEST VIRGINIA**



## **Abstract**

Despite persistent efforts to protect stream water quality, approximately 41% of West Virginia (WV) streams and rivers have been classified as impaired based either on water quality or benthic macroinvertebrate-based biological criteria. However, WV lacks a formal fish based index of biotic integrity (IBI), and this limits our ability to unambiguously identify biologically impaired waterbodies. The over-riding objective of this research was to construct and validate a fish based IBI for warm water Wadeable West Virginia streams. Specifically, we: 1) compiled a comprehensive traits table for all fish species in the state; 2) identified fish biomonitoring regions; 3) identified reference/least disturbed sites; and 4) identified fish assemblage metrics that were responsive to anthropogenic stressors. Based on fish distributions and assemblage metrics within reference sites, we identified five distinct biomonitoring regions for which we constructed separate warm water IBIs. These regions included: Monongahela River Central Appalachian-Ridge and Valley (Mon CA-RV), Ohio and Monongahela River Western Allegheny Plateau (Ohio-Mon WAP), Ohio River Central Appalachians (Ohio CA), Upper Kanawha River drainage (UK), and Potomac River drainage (Potomac). All fish assemblage metrics were evaluated within each biomonitoring region for their overall range, correlation with drainage area, discrimination between reference and stressed conditions, correlations with land-use, and redundancy with other metrics. Final lists consisting of 7 – 9 metrics were retained within each region for the inclusion into a final IBI. Common metrics retained in the final IBIs included measures of benthic associated species, total species richness, clean-gravel spawning species, tolerance to stressors, feeding classification (i.e., invertivores, omnivore-herbivores) and taxonomic group membership (i.e., Family Cyprinidae, Family Cottidae, darters, and madtoms). In general, final IBI scores from each of the biomonitoring regions were sensitive to at least one anthropogenic stressor, such as surface mining, agriculture, and/or urbanization. The strongest correlations were observed in the Mon CA-RV and Ohio CA biomonitoring regions between IBI scores and specific conductance and % surface mining, respectively. Correlations between IBI scores in all regions and the density of point-source pollution discharges was low, however each region showed

variable responses to IBI scores and anthropogenic land-use variables. When compared to other measures of biotic integrity, the IBIs showed varying results among regions. The Mon CA-RV region had the highest correlation between IBI score, benthic macroinvertebrate indices (West Virginia Stream Condition Index and Genus-Level Indicator of Most Probable Stream Status), total fish abundance, and Mid-Atlantic Highlands IBI scores. The other regions also showed a positive correlation with Mid-Atlantic Highland IBI scores and total fish abundance, however their relationships with benthic macroinvertebrate index scores were lacking. Patterns of IBI and metric response in most regions were consistent with other studies showing negative impacts of anthropogenic land-use on stream condition. However, the differential response of IBIs to land-use patterns and other measures of biotic condition indicate that stressor responses are region and organism specific. Using both a fish and benthic macroinvertebrate measure together should enhance current biomonitoring and assessment criteria in addition to providing multiple avenues for evaluating current and future land-use practices.

## **1.0 Introduction**

Community assemblages have been widely used as aquatic bioindicators for a variety of state (see Lyons et al., 2001; Schleiger 2000), federal (e.g., U.S. EPA Environmental Monitoring and Assessment Program), and international (see Lyons et al., 1995; Bozzetti and Schulz 2004) biomonitoring programs. Measures of community composition (e.g., species richness and biotic indices) have been adopted to evaluate the response and condition of streams to environmental degradations. Those relationships, in combination with traditional water quality monitoring, then get used to inform management decisions such as impairment listings, establishment of conservation areas (Karr, 1990), or implementation of remediation efforts (Merovich et al., 2013).

The index of biotic integrity (IBI; Karr, 1981) was developed in order to monitor and evaluate the condition of streams in the United States using fish community assemblages. The IBI summarizes the composition of the fish community by incorporating measures of species richness or composition, trophic composition, life history strategies, and individual fish abundances or conditions. The evaluation of community assemblages leads to a more ecologically relevant analysis of environmental stressors due to a multispecies response, integrating the impacts from multiple ecosystems stressors over long periods of time (Attrill, 2002).

Currently multimetric indices are used to assess biologic response to anthropogenic effects have largely been developed using benthic macroinvertebrates. Both benthic macroinvertebrates and fish species are relatively easy to collect with standardized methods. However, their differences in life-history strategies and dispersal patterns lead to differences in their responses to environmental changes. Fishes, due to their relatively high mobility and long life, are thought to represent watershed scale and chronic stressors while benthic macroinvertebrates represent local degradation (Freund and Petty, 2007). The differences between assemblage groups allow biomonitoring programs to utilize groups concurrently or one assemblage over another in systems where diversity in one is lacking (Griffith et al., 2004).

Anthropogenic impacts in West Virginia have a long and diverse history from large scale surface and underground mining and timber harvest operations to rural and urban development and acid precipitation. These alterations on the landscape have profound impacts on the aquatic ecosystems. The West Virginia Stream Condition Index (WVSCI; Gerritsen et al., 2000) currently is the primary bioassessment tool that determines biological stream impairment for West Virginia. The response of WVSCI to anthropogenic stressors is well-documented with negative responses to increases in acid mine drainage (AMD) or acid precipitation (Freund and Petty, 2007; Merovich and Petty, 2010) as well as negative responses to residential development and mountain-top/valley-fill operations (Merriam et al., 2011; Merriam et al., 2013). A recently developed genus-level benthic macroinvertebrate index (GLIMPSS; Pond et al., 2012) responds to landscape stressors in a similar fashion as WVSCI, with decreases in index scores as residential development increases (Merriam et al., 2011; Merriam et al., 2013) and is currently being evaluated for use in bioassessment due to its increased sensitivity due to taxonomic refinement.

Acid mine drainage from legacy coal mining operations in conjunction with increased acid precipitation from elevated concentrations of sulfur and nitrogen dioxides has lead to increased dissolved solids (e.g.,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ) and acidity in stream systems (Skousen et al., 2000; Driscoll et al., 2001). Increases in stream acidity have resulted in reduced productivity and biodiversity within aquatic ecosystems (Driscoll et al., 2001) as well as decreases in a family level benthic macroinvertebrate multimetric index (WVSCI; Freund and Petty, 2007; Merovich and Petty, 2010) and fish-based index of biotic integrity scores (Freund and Petty, 2007). Within West Virginia the impacts from AMD and acid precipitation alone have resulted in approximately 12% (~1400 miles) of streams becoming impaired by low pH (<6.0; WVDEP, 2014).

Actively mined regions of West Virginia have differing water quality characteristics and impacts on aquatic ecosystems than the historic impacts associated with acid mine drainage. The southern coalfields (i.e., Mountain-top/valley-fill mining region) of West Virginia are characterized by an alkaline mine drainage which is high in sulfates and total dissolved solids, but neutral, or slightly alkaline, in pH (Hartman et al., 2005). Large scale surface mining conducted in this region has lead to significant

alterations in the benthic macroinvertebrate community through decreases in sensitive taxa and overall WVSCI scores (Merriam et al., 2011). Regardless of the mining type, approximately 17% of all stream miles in West Virginia are impacted by mine drainage which was indicated by elevated sulfate levels (>50 mg/L; WVDEP, 2014).

These mining activities can alter detrital processing, shifting food webs from allochthonous to autochthonous organic inputs as forested headwaters are removed and primary production is increased (Hill et al., 1995, US EPA, 2011). Degradation, or alteration, at the base of the food web may have effects on the trophic structure resulting in reduced biomass of fishes and other stream organisms (US EPA, 2011). For example, Daniel et al., (2014) detected low mining density threshold responses which have negative effects on fish assemblage diversity and evenness, and abundances of species with specific life history strategies, or habitat preferences. US EPA (2003) found similar findings in that Mid-Atlantic Highlands IBI (MAHIBI; McCormick et al., 2001) scores downstream of valley fills and surface mining operations, were an average of 10 points less than their unmined counterparts.

Within West Virginia, regulatory agencies have been using benthic macroinvertebrates, paired with water quality, to enforce environmental laws and regulations. The addition of a fish based bioassessment tool could provide a more ecologically holistic measure of stream impairment while helping preserve the integrity of some of the larger scenic rivers where benthic macroinvertebrate data are lacking. Currently, fisheries biologists within the state rely on the Mid-Atlantic Highlands IBI (McCormick et al., 2001), or its modification (Detenbeck and Cincotta, 2008), to assess the condition of fish assemblages. However, the diverse geology and large scale anthropogenic land use changes across the state may require a finer scale index of biotic integrity to accurately quantify these impacts. Consequently, due to the lack of a cohesive IBI at the state level, the primary objective of this research was to develop a fish based index of biotic integrity for warm water wadeable streams in West Virginia. In order to accomplish this objective we: 1) compiled a comprehensive traits table for fish species in the state; 2) identified reference conditions across a wide range of naturally occurring contexts (e.g., stream size, drainage basin); and 3)

integrated metrics that were responsive to anthropogenic into region-specific fish based indices of biotic integrity for wadeable West Virginia streams.

## **2.0 Methods**

### *2.1 Sources of Data*

Statewide fish community data were combined from various sampling sources (Table 1). Sampling sites were selected for years 1997, 1998, and 2000 – 2013. Only electrofishing (backpack, parallel wires, and barge) sampling types were used (N=1089). Fish community data consisted of identification of each fish captured to species and their abundances. Since IBI development relies on the classification of individual species, any individual not identified to species was removed from the sample. In addition, hybrid species were also removed because they could not be classified into trait groups. If additional environmental (habitat and/or water quality) or benthic macroinvertebrates samples were taken at the time of sampling (paired samples) those data were also included in the dataset. Additional benthic macroinvertebrate, habitat, and water quality data were added to the dataset if they matched sampling locations and were sampled within two years of the fish collection. Benthic macroinvertebrate data were in the form of stream condition indices developed for West Virginia based on family (WVSCI; Gerritsen et al., 2000; N=148) or genus-level (GLIMPSS; Pond et al., 2012; N=123) identification. Habitat data consisted of a total habitat score from the EPA's Rapid Bioassessment Protocol Visual-Based habitat assessment (RBP-VBHA; N=367). Water quality data, when available, primarily consisted of specific conductance ( $\mu\text{S}/\text{m}$ ; N=610), pH (N=548), fecal coliform density (colonies/L; N=222), and dissolved oxygen (mg/L; N=335).

Each fish species encountered within the state (N=171), encompassing all stream sizes, were classified based on several natural history based traits. The traits included life history aspects such as spawning, trophic guild, distribution, tolerance, and family classification. Traits for each individual species were collected from a variety of sources: Fish Traits Database (Frimpong and Angermeier, 2010), Freshwater

Fishes of Virginia (Jenkins and Burkhead, 1994), EPA's Rapid Bioassessment Protocols for Streams and Rivers (Barbour et al., 1999), with input from professionals from West Virginia Department of Environmental Protection, W.V. Division of Natural Resources, and U.S. Environmental Protection Agency (Region 3). Appendix A lists the classification of all species into trait categories used to calculate fish community metrics.

Sampling locations were then input to ArcGIS and joined with segment level watersheds (1:24,000 scale). Locations of sampling points were evaluated against the National Hydrography Dataset (NHD-24 K) to ensure site locations were attributed to the correct segment-level watershed. In order to reduce replication of the community data, sampling locations were further reduced by selecting the most recent sampling event within each segment level watersheds and by using only wadeable streams (7 – 400 km<sup>2</sup>). Each sampling location was assigned local and cumulative landscape attributes, major drainage basin (Monongahela, Ohio, Potomac, and Upper Kanawha), and Level III Ecoregion (Omerick, 1987).

Landscape characteristics for all 1:24,000 segment-level watersheds (SLWs) within the state of West Virginia were quantified using spatial analysis functions in ArcGIS ArcMap 10.0 (Environmental Systems Research Institute, Redlands, California). Segment-level watersheds are inter-confluence based watersheds. Measures of several landscape attributes for each segment-level watershed were quantified at the local (i.e., within individual SLWs) and cumulative (i.e., all SLWs upstream of a given sampling location) scale for each SLW (Strager et al., 2009). Land cover classifications were derived from the 2009 and 2010 National Agriculture Imagery Program (NAIP) orthophotography with a 1-meter pixel resolution at a scale of 1:10,000. Land cover types included open water, forest, grass and agricultural lands, and barren development. The mining-permit boundaries layer developed by the Technical Applications in GIS (TAGIS) office within WVDEP enabled further differentiation into mining-related open water (i.e., slurry impoundments), barren (i.e., active mine lands) and grasslands (i.e., reclaimed mine lands) from non-mining land cover. All mining-related cover classes were summed into a measure of total surface mining. The density (#/km<sup>2</sup>) of surface mining, underground mining, sewage, and septic

national pollution discharge elimination system (NPDES) permits were calculated from data obtained from WVDEP. The West Virginia Statewide Addressing and Mapping Board structures layer (WV SAMB) was used to calculate the density of residential and commercial structures ( $\#/km^2$ ). Natural landscape variables for each SLW were summarized including basin area ( $km^2$ ), mean elevation (m), and swim distance (km). Swim distance was defined as the minimum downstream distance (km) from the outflow of a SLW to the inflow of a SLW with a basin area  $\geq 200 km^2$  (Hitt and Angermeier, 2011).

## *2.2 Site Classification*

Stream ecological assessments rely on two major components: measurement of some ecological resource and a reference condition (Hawkins et al., 2010). A reference condition is considered a benchmark condition to which all other measurements are compared. Without a baseline condition, little can be inferred about the ecological condition due to natural variation among sites (Stoddard et al., 2006; Whittier et al., 2007; Hawkins et al., 2010). Identification of reference sites by the WVDEP were determined by a series of water quality and habitat characteristics along with identification of surrounding and upstream sources of pollution (Table 2; Pond et al., 2012). A short list of reference sites (N=55) using these criteria were determined prior to any analysis of the fish community datasets for IBI development.

Since most reference sites currently in WVDEP's database are on first and second order streams, a concerted effort was made to select some candidates on streams with larger watershed areas from the fish database being used for IBI development. In order to address large streams and areas where reference sites are difficult to identify, WVDEP established additional levels of reference condition (Level II & Level III). While Level I reference sites meet all reference site criteria (Table 2), Level II reference sites fail to meet one or more of them by a narrow margin. For example, Level II reference sites may be deficient in one RBP habitat parameter. Level III reference site designations are generally reserved for rivers and large streams ( $\geq 155.4 km^2$ ). Level III reference sites generally meet RBP habitat and water quality criteria at the assessment site, but because of their size generally have point source discharges



within their drainage or more land development and human disturbances than would be allowed for smaller streams designated as Level I or Level II. Level III reference sites were generally located in least disturbed segments of rivers and streams where local and upstream disturbances are minimized or distant to the site. It should be noted that best professional judgment by experienced personnel is an important part of the initial and final selection of Level I, Level II, and Level III reference sites. Additional reference sites were selected, after community classification into biomonitoring regions, from the remaining pool of sites that were previously sampled in order to establish a regional reference condition and to increase the reference sample sizes in each region. The final pool of reference sites would be characterized as minimally to least disturbed reference condition (Stoddard et al., 2006).

All fish sampling data were divided into reference, stressed, and non-reference sites. Stressed sites were defined as meeting at least one of the abiotic criteria (physical or chemical) shown in Table 2. These criteria are similar to the original WVSCI and GLIMPSS stress site criteria and cover a broad range of potential stressor variables across WV. Non-reference included all sites that were not classified as either reference or stressed due to either lack of abiotic data or did not meet the requirements to be classified as either site type. These non-reference sites were used in combination with reference and stressed sites for evaluating metric response to stressors, metric correlations, and standardizing metric values.

Sites were identified as being either warm water or cold water by an evaluation of the fish community data. Warm/cold water designation and criteria were determined by West Virginia Department of Environmental Protection. The threshold for the number of coldwater species present depended on overall richness. If species richness was less than or equal to five, then the presence of one cold indicator species resulted in the site being identified as a cold water site; if richness was between 5 and 10 then the presence of two cold water indicator species were required to be identified as cold water; if richness was greater than ten, then 3 cold water indicator species were required. Additionally, if sculpin species (*Cottus* spp.) were amongst the top three most numerous, the site was deemed to be cold water. All other comparable samples were identified as warm water. Cold water indicator species used for this exercise

were brook trout (*Salvelinus fontinalis*), mountain redbelly dace (*Chrosomus oreas*), longnose dace (*Rhinichthys cataractae*), and any sculpin species (*Cottus spp.*).

## 2.3 Data Analysis

### 2.3.1 Community Classification into Biomonitoring Regions

The variability in fish community distributions in West Virginia is relatively unknown. McCormick et al. (2001) found no regional differences; however Detenbeck and Cincotta (2008) detected ecoregional differences in fish IBI metrics and final scores. In order to evaluate the influences of ecoregion and drainage basin on community similarity, fish community assemblages were grouped into natural classes based on inferences generated from community similarity analyses. Combinations of geographical classification factors were evaluated in order to help explain the natural variability in fish communities found at the reference sites. Analysis of similarity (ANOSIM; Clarke, 1993) was conducted in order to evaluate the differences among the major classification groupings. The differing classifications were then evaluated with mean similarity analysis (MEANSIM; Van Sickle, 1997). The following combinations of strata were used: Level III ecoregion (Central Appalachian, Ridge & Valley, and Western Allegheny Plateau); major drainage basin (Ohio, Monongahela, Potomac, and Upper Kanawha); ecobasin (combined level III ecoregions X major drainage basin).

Classification strength was determined using a Bray-Curtis similarity matrix of the reference site fish communities and MEANSIM by comparing the average within-class similarity (W) to the average between-class similarity (B). A final classification strength (CS) was calculated (W-B). A dendrogram was generated based on MEANSIM to evaluate the clustering of similar sites based on fish community composition. Each analysis was performed using the original set of reference sites defined by WV DEP, commonly occurring fish species (>2.5% occurrence),  $\log(x+1)$  transformed fish abundances, and Bray-Curtis distances. Significance for analysis was based on 1,000 permutations. Final biomonitoring regions

were determined based on dendrogram grouping, available fish data, and to include distinct watershed boundaries (e.g., HUC8 or HUC12 boundaries).

### 2.3.2 Metric Selection and Index Construction

An extensive list of fish community trait and taxonomic based metrics (N=128) were compiled from the Mid-Atlantic Highland IBI (McCormick et al., 2001), its modification (Detenbeck and Cincotta, 2008), and selected traits from the Fish Traits Database (Frimpong and Angermeier, 2010). Modifications were made to several of the metrics to exclude tolerant or specific species (see Appendix B for complete list and description of all metrics). All metric calculations were conducted in program R version 3.1.2 (R Core Team, 2014) using matrix algebra and package vegan version 2.0-6 (used for richness calculations, Oksanen et al., 2013). Metrics for consideration in the IBI were then assigned an expected response to stressors: positive metrics decrease with increases in stressors, while negative metrics increase with increases in stressors.

Metric evaluation for consideration in a final index followed traditional techniques (e.g., Stoddard et al., 2008; Pond et al., 2012). After species classification and determination of biomonitoring regions, metrics were evaluated within each biomonitoring region using a step-wise selection process. The overall process of metric selection included: evaluating to insure metrics had sufficient range; relationship with drainage area; responsiveness (i.e., ability to distinguish between reference and stressed sites); response to human disturbance; and redundancy with other metrics.

The first step in this process, each metric was evaluated for their range. A metric was retained in the metric pool if it had a non-zero 25<sup>th</sup> percentile over all sites sampled. This evaluation was conducted in order to ensure rare metrics were not being evaluated for inclusion in a final IBI.

The second step in the selection process, evaluated the relationship of each metric with drainage area at reference sites only. Within West Virginia, other studies have demonstrated the importance of stream temperature, ecoregion, and distance to a source to fish community structure (Detenbeck and Cincotta,

2008; Hitt and Angermeier, 2011). Predictive models used in bioassessment programs allows for the comparison of observed fish community assemblages of a sampling location to what is expected in the absence of human disturbance (Observed:Expected; Flotemersch et al., 2006). The expected assemblage is generated using linear models based on regionally specific reference sites. Specifically, fish community metrics can be predicted for all wadeable streams in West Virginia using reference site based models generated using surrounding landscape characteristics. This approach is based on the concept that any significant departure from the baseline reference condition (i.e. expected value under natural landscape conditions only) is indicative of a disturbed system. Metrics were adjusted for natural variables after the range test to allow rare metrics (i.e. metrics with too many zeros) within each region to be excluded and to ensure metrics have high enough variability to discriminate among sites in different conditions (Stoddard et al., 2008). Fish community based metrics are commonly adjusted for watershed area during Index of Biotic Integrity construction (e.g. McCormick et al., 2001). Some fish community metrics were also transformed (e.g. arc-sine or  $\log_{10}(x+1)$ ) depending on its check for normality with a Shapiro-Wilks test. Metrics with significant ( $p$ -value  $<0.05$ ) relationships with drainage area were then predicted based on the linear model equation. Those metrics were then adjusted using an observed/expected formula.

Raw (i.e., metrics not adjusted using linear models) and adjusted (observed/expected) metrics were then evaluated for their discrimination efficiency (DE). Discrimination efficiency (i.e. responsiveness) was calculated as the number of stressed sites that fell below the 25<sup>th</sup> percentile (for positive metrics,) or fell above the 75<sup>th</sup> percentile (for negative metrics,) of the reference distribution in each biomonitoring region (Blocksom and Johnson, 2009). A metric had to exhibit discrimination efficiency above 60% prior to further evaluation with anthropogenic stressors.

Each metric was then evaluated for their relationship with environmental stressors using Spearman's correlation. Metrics were correlated with % cumulative surface mining, structure density, total agriculture, development, and total forest along with pH, and specific conductance. Redundancy of

metrics was evaluated with Spearman correlation. Any metric which was highly correlated ( $>|0.90|$ ) with another metric was considered for removal from IBI development. This procedure produced a pool of potential metrics that are either correlated with human disturbance, were highly discriminatory, or both. From this pool, a selection of metrics, or all metrics, could be scored and combined to produce a final IBI.

### 2.3.3 Metric scoring and aggregation into final indices

A final set of metrics within each biomonitoring region was selected to be aggregated into an index of biotic integrity. Metrics within each region were selected to represent diversity, spawning, and trophic diversity, or feeding guild of the fish communities. In addition, at least one “negative” metric (increases with increases in stress) was selected in most biomonitoring regions. Since raw proportional and richness metrics, in addition to adjusted metrics, were selected each metric was normalized (scored) in order to convert each metric to the same scale. Metrics were scored following Blocksom (2003) in which the 5<sup>th</sup> and 95<sup>th</sup> percentiles for all sites with fish present in each region were calculated and used as the floor and ceiling, respectively. For positive metrics scoring was based on the following equation:  $(\text{metric value} - \text{floor}) / (\text{ceiling} - \text{floor}) * 100$ . For negative metrics, scoring was based on the following equation:  $(\text{ceiling} - \text{metric value}) / (\text{ceiling} - \text{floor}) * 100$ . Again, these formulas normalized each metric to be dimensionless, and ranging from 0 – 100, so they could be combined into a final, unit-less, index value. For sites that exceed the 95<sup>th</sup> percentile of the full distribution of sites, resulting in normalized metrics to be greater than 100, those values were corrected to the maximum score of 100. Similarly, if a site had metric values below the 5<sup>th</sup> percentile, above 95<sup>th</sup> percentile for negative metrics, received a score of zero. Final index values for each region were calculated as the average of all the metrics for that region with final IBI scores ranging from 0 – 100.

Finally, four categories of impairment (excellent, good, fair, poor) were created based upon the reference distribution. For three of the biomonitoring regions (Ohio CA, Upper Kanawha, and Ohio-Mon WAP), the final impairment thresholds (between good and fair) were based upon the 25<sup>th</sup> percentile of the reference distribution. Due to the high number and high quality condition of the reference pool in the

Mon CA-RV region, the 10<sup>th</sup> percentile of the reference distribution was used to determine stream impairment.

#### 2.3.4 Index Performance

Differences in IBI scores between each site type classification (reference vs. stressed vs. other) was evaluated with an analysis of variance (ANOVA) within each biomonitoring region. It was expected that reference sites, on average, would score higher than stressed sites. However, the response of the non-reference sites, would span the full range of conditions seen at both the reference and stressed sites in each region since they could potentially contain reference and stressed quality streams. Therefore, no significant differences between other-reference sites and other-stressed sites were expected.

Final IBI score responses in each biomonitoring region were evaluated against anthropogenic landscape variables, water quality parameters, and other measures of biotic stream condition such as the West Virginia Stream Condition Index (WVSCI), Genus-Level Indicator of Most Probable Stream Status (GLIMPSS), and the Mid-Atlantic Highlands Index of Biotic Integrity (MAH IBI). In addition, low numbers of duplicate samples (sites visited again during the sampling time frame) were retained in some of the regions. Due to the low sample sizes, IBI scores were only visually compared to an expected 1:1 relationship line.

### 3.0 Results

#### 3.1 Community Classification into Biomonitoring Regions

Analysis of similarity (ANOSIM) indicated that classification using Major Basin, Ecoregion, and Ecoregion-Major Basin combinations all produced significant differences ( $p < 0.05$ ) between groups (Table 3). However, Ecoregion-Major Basin combinations produced the highest classification percentage (CS%), followed by Major Basin and then Ecoregion (13.5, 9.1, and 7.2%, respectively). This indicates that classification based solely on Basin or Ecoregion would be insufficient. Mean similarity analysis (MEANSIM) based on the Ecoregion-Major Basin classifications then indicated which groups were more

similar based on their fish community structure (Table 3). Upon evaluation of the MEANSIM dendrogram (Figure 1), general regions were selected for Index of Biotic Integrity development. Those regions include: Upper Kanawha (UK); Potomac; Ohio-Monongahela Western Allegheny Plateau (Ohio-MonWAP); Monongahela Ridge/Valley-Central Appalachian (Mon CA-RV); and Ohio Central Appalachians (Ohio CA). However, this initial classification resulted in regional boundaries that intersected watershed boundaries and were difficult to distinguish on the landscape. In order to make the regionalization more biologically relevant and amiable to interest groups that may utilize the index, either whole HUC8 watersheds were combined or distinct dividing lines were used, such as HUC12 outflows, based on which general region they intersected in order to form the biomonitoring regions (Table 4; Figure 2). The only HUC8 that was split based on a distinct dividing line was the Elk watershed. The dividing line for this watershed occurred at the outflow of Suttons Dam, a HUC 12 outflow. Due to the low sample size (N=105) and uncertainty of reference condition, the Potomac biomonitoring region was removed from further analysis and IBI development. Final sample sizes and site type classification are given in Table 5 – 8 with summaries of natural landscape variables.

### *3.2 Metric evaluation and selection*

The metric selection process resulted in final West Virginia warm water IBI models that incorporated 7 fish community metrics for the Mon CA-RV, Ohio CA, and Upper Kanawha biomonitoring regions. The Ohio-Mon WAP had a total of 8 metrics selected for inclusion in a final IBI. Table 9 summarizes species metric performance within each biomonitoring region. See Appendix C for biomonitoring region specific results from all metric testing and evaluations. Within each region, at least one measure of trophic structure, spawning preference, tolerance, and diversity were selected to evaluate stream conditions.

Thresholds for each scoring method for final metrics within the Mon CA-RV, Ohio CA, Ohio-Mon WAP, and Upper Kanawha are presented in Tables 10 – 13. For each region, final IBI scores were calculated based on the 5<sup>th</sup> (floor) and 95<sup>th</sup> (ceiling) of the full distribution of sites, excluding sites with zero individuals.

### *3.3 Index performance*

The final IBI scores within each region were compared between known reference, stressed, and non-reference sites to determine the ability of the IBI to discriminate between site types. The Upper Kanawha region had the highest range of IBI scores across all site types (0 – 100). Reduced maximum scores were exhibited in the Mon CA-RV (max = 97), Ohio CA (max = 92), and Ohio Mon-WAP (max = 82) across all site types. A post-hoc test (TukeyHSD) showed which site types were statistically different based on ANOVA results. The Mon CA-RV region has distinct separation in mean IBI scores between reference-stressed and reference-non-reference sites with no statistical distinction between stressed and non-reference sites (Figure 3). The Ohio CA and Ohio and Mon WAP biomonitoring regions found all three groups to be statistically different from one another (Figure 3). However, the Upper Kanawha biomonitoring region showed no significant difference between any of the site types due to the high variability of IBI scores within each site type (Figure 3).

Spearman correlations for final IBI scores within biomonitoring region against stream characteristics indicate that in some regions the IBI is responsive to anthropogenic land use patterns as well as to other measures of biotic conditions (Table 14). The Mon CA-RV region showed the strongest positive correlation, among all the regions, with biotic measures of stream conditions (WVSCI, GLIMPSS, Fish Abundance, and MAH IBI). All regions exhibited negative relationships, of varying strength, with specific conductance (SPC), as well as cumulative percentages of surface mining (C. Surface Mining) and development (C. Development). Strong relationships with drainage area and elevation were not detected in any of the regions, indicating that IBI scores are not biased towards large or low elevation streams. All regions exhibited positive relationships, of varying strength, with cumulative percent forest (C. Forest).

Duplicate samples (i.e., samples within the same segment level watershed) were retained within the Mon CA-RV (N=12), Ohio CA (N=4), and Ohio-Mon WAP (N=4) biomonitoring regions to evaluate the temporal variation in IBI scores between years. These duplicate samples were taken in different years and were not used to construct the final IBIs. The Mon CA-RV duplicate samples deviate strongly from



the 1-to-1 relationship that was expected (Figure 4). Three of the 12 samples in the Mon CA-RV had initial IBI scores of 0 due to no fish being captured during the original sampling. Duplicate samples were taken in 2013 following chemical stream restoration of acid mine drainage (AMD) in the Three Forks watershed. The resulting data captured an increase in IBI scores following restoration efforts indicating that the IBI for the Mon CA-RV region is sensitive to stream improvements. The duplicate samples in the Ohio CA and Ohio-Mon WAP show little deviation from the 1-to-1 relationship that was expected (Figure 4). No duplicate samples were located in the Upper Kanawha biomonitoring region. Due to the small sample size of replicate sites within each region a statistical test could not be constructed with any reliability.

IBI scores exceeding the 75<sup>th</sup> percentile of the reference distribution in the Ohio CA ( $IBI \geq 76.82$ ), Ohio and Mon WAP ( $IBI \geq 64.44$ ), and Upper Kanawha ( $IBI \geq 67.89$ ) were classified as having “Excellent” biotic integrity. Scores between the 75<sup>th</sup> and 25<sup>th</sup> percentiles for the Ohio CA (76.82 – 46.12), Ohio and Mon WAP (64.44 – 50.92), and Upper Kanawha (67.89 – 39.79) were identified as having “Good” biotic integrity. For the Ohio CA, Ohio and Mon WAP, and Upper Kanawha, any site exceeding the 25<sup>th</sup> percentile of the reference distribution was considered not impaired. Any IBI score below the 25<sup>th</sup> percentile, for each region, was considered impaired. Impaired sites were divided into two categories, “Degraded” and “Severely Degraded” based on the 5<sup>th</sup> percentile of the reference distribution within each biomonitoring region. The Mon CA-RV had a larger sample size of high quality reference sites than the other region, due to increased sampling efforts in 2013, therefore the 10<sup>th</sup> percentile of the reference distribution was used as the impairment threshold for this region ( $\leq 56.15$  is impaired).

#### **4.0 Discussion**

The IBIs developed within most of biomonitoring regions are robust and practical tools for evaluating the impacts to water quality and aquatic wildlife. Strong regional differences in fish community assemblages allowed us to account for some of the natural variation on the landscape which enabled the development

of IBIs sensitive to anthropogenic disturbances. Regional differences in metric and overall IBI response to anthropogenic stressors further demonstrated the importance and need for region specific definitions of reference and stressed conditions. Even though the IBIs developed span a wide range of stream sizes across the state, there are still a large number of streams the IBIs cannot assess. These stream types include headwater streams (<7 km<sup>2</sup>), large rivers (>400 km<sup>2</sup>), cold water streams, and Potomac River tributaries. Until more reliable methods of accounting for the natural variability associated with these different site types are incorporated into IBI development, then segregation into bioassessment groups will continue to be important.

The development of IBIs for West Virginia warm water, wadeable streams followed common standardized techniques for selecting fish community metrics (Stoddard et al., 2008). Attempts were made to select metrics from key ecological categories (i.e. trophic, reproduction, and tolerance) in order to generate IBIs that give an overall view of stream condition. Metrics were selected if they exhibited a predictable relationship with at least one anthropogenic landscape or water quality variable. In addition, metrics needed to demonstrate an ability to distinguish between reference and stressed streams. These criteria ensured that the final indexes in most of the biomonitoring regions were sensitive and responsive to the anthropogenic impacts evaluated. Several surrounding state water agencies have also incorporated IBIs into their routine of water quality monitoring as well as impairment delineation. Many of these programs have calibrated, or developed, regionally specific IBIs using their own data (e.g., Ohio EPA, 1987; Roth et al., 2000; Compton et al., 2003). Even though state-specific IBIs are becoming more common for regulatory purposes, region (Lyons et al., 1995; Daniels et al., 2002; Bozzetti and Schulz 2004; Pont et al., 2006), watershed (Kimmel and Argent, 2006), or habitat (Lyons, 2006; Mohamed, 2014) specific IBIs also exist. The metrics selected for this study, when compared to other state, regional, and international studies, demonstrate that similar fish assemblage groups (e.g., Family Cyprinidae and Darter-Madtom-Sculpin assemblages) and traits (e.g., gravel spawning and tolerance) are important in assessing integrity of aquatic systems.

Regionalization and stratification of sites used in bioassessment development has made fish-based multimetrics a practical tool for evaluating the impacts of water quality, habitat modification, and land-use changes on stream fish communities. Large scale, regional processes, such as catchment extent and physiography, has resulted in substantial variation in fish community structure (Angermeier et al., 2000). This variation in fish distribution and trait-based assemblage structure has led to a push for the development of regionally adjusted criteria for fish indices of biotic integrity (Smogor and Angermeier, 1998). In some cases, large scale process can have such a strong influence on fish assemblage structure that the effects of in-stream processes (i.e., local habitat and water quality) may not be detected (Angermier and Winston, 1999). Segregating sites based on major drainage basin and physiogeographic region (i.e., ecoregion) removed some of the confounding responses due to latitude, longitude, and elevation. The incorporation of drainage area in metric responses effectively removed any stream size dependence of the final IBIs. In addition, the removal of strictly cold water streams from our analysis relatively increased the homogenization of natural variables within each region while still maintaining differences between biomonitoring regions. By effectively removing portions of the natural variation known to influence fish community structure, the IBIs can now be used to assess the effects of land-use changes and remediation, or mitigation, efforts, on fish community health.

Segregation of stream types into temperature classes are common in IBIs that are developed in regions with high geographic variation and stream temperature regimes (see Lyons et al., 1996 and Lyons, 2012). Models predicting in-stream temperatures have found that there are several local and regional environmental factors controlling stream temperature. These variables can include elevation, watershed precipitation, slope, riparian cover, aspect, and air temperature (Segura et al., 2015). These measurements, in conjunction with continuous stream temperature data, could be used to generate predictions of maximum daily mean water temperature for all wadeable streams in West Virginia.

The immediate need for stream temperature modeling becomes important when evaluating the current classification of stream temperature classes using species assemblages. There were two duplicate stream

samples that shifted from cold water streams in one year to a warm water stream in the following years (Mon CA-RV biomonitoring region). One of these shifts has been attributed to the addition of a top-release dam upstream of the sampling location. However, the reason for the shift of the second site from a cold water species assemblage to a warm water assemblage is unknown. These shifts demonstrate the need to develop non-fish based classification criteria or *a priori* expectations of stream temperature regimes prior to sampling in order to evaluate temperature impaired cold water streams that can no longer support a cold water fish assemblage. In addition, the biological status and locations of current cold water streams are unknown. Monitoring these locations for economically valuable natural resources (i.e. trout fishing) in the face of increased anthropogenic and climate changes will be important as the development of a bioassessment program progresses.

Of all the streams assessed in West Virginia, increased fecal coliforms, increased iron, and degraded biological condition, based on the West Virginia Stream Condition Index (WVSCI) scores, are the top three leading causes of impairment. Despite the vast efforts in protecting designated stream uses, approximately 41% of streams and rivers in West Virginia are considered impaired and have been placed on West Virginia's 303(d) impaired streams list (WV DEP, 2014). Even though the benthic macroinvertebrate-based WVSCI scores have proven useful in determining stream impairment due to its high correlation with aquatic stressors and contaminants, it may not represent the entire stream ecosystem. Outside of the Mon CA-RV region, there was no correlation between WVSCI and WV IBI scores, indicating the stressors within each region have potential to influence assemblages differently. Benthic macroinvertebrate-based multimetric indices can be highly responsive to relatively low levels of acid mine drainage while fish IBI scores are much less responsive (Freund and Petty, 2007). Similarly, macroinvertebrates are able to better detect small increases in nutrient concentrations than fish communities (Justus et al., 2010). However, fish communities are more responsive to watershed level disturbances and flow regime changes.

Selecting which index to use to assess stream health can be difficult since assemblage groups respond differently to anthropogenic impacts. However, multimetric indices allow for a wide range of responses to ecosystem stressors since each metric may behave differently providing different information regarding the type and duration of stressors (Herman and Nejadhashemi, 2015). Metrics selected within each region for the WV IBIs each performed differently when compared to anthropogenic land-use and water quality measures allowing for IBIs to be responsive to a variety of stream characteristics. Of the sites evaluated for IBI development statewide, a portion had associated benthic macroinvertebrate data (N=148) in which 35.8% of the sites were impaired based on WVSCI scores (impairment < 68). However, using the thresholds for impairment based on IBI scores in each region, 54.7% of the sites with benthic macroinvertebrate data are considered impaired, with only 23.6% of those sites being impaired under both criteria.

As would be expected, the least-impacted, or reference, sites had the higher IBI scores within each region. However, when site types were compared within each region, some questions about the uncertainty of the IBI were formed. For example, the Mon CA-RV biomonitoring region did not exhibit a significant difference between stressed and non-reference sites. This lack of a significant difference did not come as a surprise due to the definition of the non-reference site type. Non-reference sites were placed there because there was either insufficient data to elevate the site to reference or to classify it as a stressed site, or the site was of intermediate quality and did not meet the standards of reference or stressed sites. Either outcome should produce a category of sites that span a wide range of environmental conditions and may not be statistically different from either reference or stressed sites, because it may contain both.

The Upper Kanawha region however, demonstrates no significant difference between any of the site types. The exact cause for the lack of difference is unknown. This region is known for its high quality streams and the majority of the streams in this region are historically cold water systems. The criteria used to classify sites into reference/stressed and cold/warm should be closely evaluated for this region in order to ensure streams are being stratified appropriately, which may not be the same criteria used for the

other regions. Additionally, the high landscape heterogeneity in this region may indicate that the recommended reference site sample size (N=34 – 40; Yoder and Rankin, 1995) be met for IBI development in this region. Based on this information, it was determined that until more sufficient classification criteria are developed for this region, the IBI will continue to be evaluated.

Among-year variability between IBI scores within each biomonitoring region should continue to be evaluated as duplicate samples are generated. Even though there were duplicate samples within each biomonitoring region, the temporal variability among years, or within a year, with these low numbers of duplicate samples, cannot be accurately determined. The Mon CA-RV biomonitoring region had the most duplicate samples due to increased sampling efforts by the West Virginia DEP in 2013. However, this region also demonstrated the highest variation between duplicate samples due to chemical stream restoration efforts. The increase in IBI scores in the Mon CA-RV biomonitoring region demonstrates the sensitivity of the IBI in detecting in-stream improvements. However, the impacts of chemical and physical restoration efforts provide an additional source of variability in IBI development since the influence and extent of these efforts are largely unknown in West Virginia.

Osbourne et al. (1992), Osbourne and Wiley (1992), and Hitt and Angermeier (2011) have suggested that the influence of stream order, or size, and stream location in the drainage network can all have impacts on species richness resulting in differences in IBI scores. The effects of drainage area on metric values were evaluated and adjusted appropriately resulting in final IBI scores that were uncorrelated with drainage area. However, the influence of stream position in the drainage network, in addition to the condition of neighboring streams, was not considered during IBI development for West Virginia. However, the influence of drainage position (i.e., swim distance) was evaluated in relation to final IBI scores within each region with no effect detected. The influence of the overall condition of surrounding streams and anthropogenic changes on the landscape on local IBI scores will continue to be evaluated.

## Literature Cited

- Angermeier, P.L., R.A. Smogor, and J.R. Stauffer. 2000. Regional frameworks and candidate metrics for assessing biotic integrity in Mid-Atlantic Highlands streams. *Transactions of the American Fisheries Society*. 129:962 – 981.
- Angermeier, P.L. and M.R. Winston. 1999. Characterizing fish community diversity across landscape: prerequisite for conservation. *Ecological Applications*. 9:335 – 349.
- Attrill, M.J. 2002. Community-level indicators of stress in aquatic ecosystems. Pages 473 – 508 in S.M. Adams, editor. *Biological indicators of aquatic ecosystem stress*. American Fisheries Society, Bethesda, Maryland.
- Barbour, M. T., J. Gerritsen, B. D. Snyder, J. B. Stribling, and C. Faulkner. 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. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. *Journal of Environmental Management*. 31:670 – 682.
- Blocksom, K. A. and B. R. Johnson. 2009. Development of a regional macroinvertebrate index for large river bioassessment. *Ecological Indicators*. 9:313 – 328.
- Bozzetti, M. and U.H. Schulz. 2004. An index of biotic integrity based on fish assemblages for subtropical streams in southern Brazil. *Hydrobiologia*. 529:133 – 144.
- Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*. 18:117 – 143.
- Compton, M. C., G. J. Pond, and J. F. Brumley. 2003. Development and application of the Kentucky index of biotic integrity (KIBI). Kentucky Department for Environmental Protection, Division of Water, Water Quality Branch. Frankfort, KY.
- Daniel, W. M., D. M. Infante, R. M. Hughes, Y. P. Tsang, P. C. Esselman, D. Wieferich, K. Herreman, A. R. Cooper, L. Wang, and W. W. Taylor. 2014. Characterizing coal and mineral mines as a regional source of stress to stream fish assemblages. *Ecological Indicators* 50:50 – 61.
- Daniels, R.A., K. Riva-Murray, D.B. Halliwell, D.L. Vana-Miller, and M.D. Bilger. 2002. An index of biotic integrity for Northern Mid-Atlantic Slope drainages. *Transactions of the American Fisheries Society*. 131: 1044 – 1060.
- Detenbeck, N. E. and D. A. Cincotta. 2008. Comparability of a regional and state survey: effects on fish IBI assessment for West Virginia, U.S.A. *Hydrobiologia*. 603:279 – 300.
- Driscoll C.T., G.B. Lawrence, A.J. Bulger, T.J. Butler, C.S. Cronan, C. Eagar, K.F. Lambert, G.E. Likens, J.L. Stoddard, K.C. Weathers. 2001. Acidic deposition in the Northeastern United States: Sources and inputs, ecosystem effects, and management strategies. *BioScience* 51:180–198

- Flotemersch, J. E., J. B. Stribling, and M. J. Paul. 2006. Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.
- Freund, J.G., and J.T. Petty. 2007. Response of fish and macroinvertebrate bioassessment indices to water chemistry in a mined Appalachian watershed. *Journal of Environmental Management*. 39:707 – 720. DOI 10.1007/s00267-005-0116-3.
- Frimpong E. A., and P. L. Angermeier. 2010. Trait-based approaches in the analysis of stream fish communities. *American Fisheries Society Symposium*. 73:109 – 136.
- Gerritsen, J., J. Burton, and M. T. Barbour. 2000. A stream condition index for West Virginia wadeable streams. Tetra Tech, Inc., Owings Mills, Maryland.
- Griffith, M.B., B.H. Hill, F.H. McCormick, P.R. Kaufmann, A.T. Herlihy, and A.R. Selle. 2004. Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. *Ecological Indicators*. doi:10.1016/j.ecolind.2004.11.001.
- Hartman, K. J., J. D. Kaller, J. W. Howell, and J. A. Sweka. 2005. How much do valley fills influence headwater streams? *Hydrobiologia*. 532:1 – 19.
- Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society*. 29:312 – 343.
- Herman, M.R., and A.P. Nejadhashemi. 2015. A review of macroinvertebrate- and fish-based stream health indices. *Ecohydrology & Hydrobiology*. In press. <http://dx.doi.org/10.1016/j.ecohyd.2015.04.001>.
- Hill, W. R., M. G. Ryon, and E. M. Schilling. 1995. Light limitation in a stream ecosystem: responses by primary producers and consumers. *Ecology*. 76:1297 – 1309.
- Hitt, N. P. and P. L. Angermeier. 2011. Fish community and bioassessment response to stream network position. *Journal of the North American Benthological Society*. 30:296 – 309.
- Jenkins, R.E., and N.M. Burkhead. 1994. *Freshwater fishes of Virginia*. American Fisheries Society, Bethesda, Maryland.
- Justus, B.G., J.C. Petersen, S.R. Femmer, J.V. Davis, J.E. Wallace. 2010. A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams. *Ecological Indicators*. 10:627 – 638.
- Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries*. 6: 21 – 27.
- Karr, J.R. 1990. Biological integrity and the goal of environmental legislation: lessons for conservation biology. *Conservation Biology*. 4:244 – 250.
- Kimmel, W.G. and D.G. Argent. 2006. Development and application of an index of biotic integrity for fish communities of wadeable Monongahela River tributaries. *Journal of Freshwater Ecology*. 21:183 – 190.



- Lyons, J. 2006. A fish-based index of biotic integrity to assess intermittent headwater streams in Wisconsin, USA. *Environmental Monitoring and Assessment*. 122:239 – 258. DOI: 10.1007/s10661-005-9178-1
- Lyons, J. 2012. Development and validation of two fish-based indices of biotic integrity for assessing perennial coolwater streams in Wisconsin, USA. *Ecological Indicators*. 23:402 – 412.
- Lyons, J., S. Navarro-Perez, P.A. Cochran, E. Santana C. and M. Guzman-Arroyo. 1995. Index of biotic integrity based on fish assemblages for the conservation of streams and rivers in West-Central Mexico. *9:569 – 584*.
- Lyons, J., R.R. Piette, and K.W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warmwater rivers. *Transactions of the American Fisheries Society*. 130:1077 – 1094.
- Lyons, J., L. Wang, and T. D. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management*. 16:241 – 256.
- McCormick, F. H., R. M. Hughes, P. R. Kaufman, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands Region. *Transactions of the American Fisheries Society*. 130:857 – 877.
- Merovich, G. T. and J. T. Petty. 2010. Continuous response of benthic macroinvertebrate assemblages to a discrete disturbance gradient: consequences for diagnosing stressors. *Journal of the North American Benthological Society*. 29:1241 – 1257.
- Merovich, G.T., J.T. Petty, M.P. Strager, and J.B. Fulton. 2013. Hierarchical classification of stream condition: a house-neighborhood framework for establishing conservation priorities in complex riverscapes. *Freshwater Science*. 32:874 – 891.
- Merriam, E. R., J. T. Petty, G. T. Merovich, J. B. Fulton, and M. P. Strager. 2011. Additive effects of mining and residential development on stream conditions in a central Appalachian watershed. *Journal of the North American Benthological Society*. 30:399 – 418.
- Merriam, E.R., J.T. Petty, M.P. Strager, A.E. Maxwell, and P.F. Ziemkiewicz. 2013. Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined central Appalachian watershed. *Freshwater Science*. 32:1246 – 1259.
- Mohamed, A.R.. 2014. A fish index of biotic integrity for evaluation of fish assemblage environment in restored Chybash Marsh, Iraq. *Global Journal of Biology, Agriculture, and Health Sciences*. 3:32 – 37.
- Ohio EPA. 1987. *Biological Criteria for the Protection of Aquatic Life. Volume II: Users Manual for Biological Field Assessment of Ohio Surface Waters*. Ohio EPA, Div. of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package. R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>

- Omernik, J. M. 1987. Ecoregions of the conterminous United States Map (scale 1, 7,500,000). *Annals of the Association of American Geographers*. 77:118 – 125.
- Osborne, L. L., and M. J. Wiley. 1992. Influence of tributary spatial position on the structure of warmwater fish communities. *Canadian Journal of Fisheries and Aquatic Sciences*. 49:671–681.
- Osborne, L. L, S. L. Kohler, P. B. Bayley, D. M. Day, W. A. Bertrand, M. J. Wiley, and R. Sauer. 1992. Influence of stream location in a drainage network on the index of biotic integrity. *Transactions of the American Fisheries Society*. 121:635–643.
- Pond, G. J., J. E. Bailey, B. M. Lowman, and M. J. Whitman. 2012. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. *Environmental Monitoring and Assessment*. DOI 10.1007/s10661-012-2648-3.
- Pont, D., B. Hugueny, U. Beier, D. Goffaux, A. Melcher, R. Noble, C. Rogers, N. Roset, and S. Schmutz. 2006. Assessing river biotic condition at a continental scale: A European approach using functional metrics and fish assemblages. *Journal of Applied Ecology*. 43:70 – 80.
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Roth, N. E., M. T. Southerland, J. C. Chaillou, P. F. Kazyak, S. A. Stranko. 2000. Refinement and validation of a fish index of biotic integrity for Maryland streams. Versar, Inc., Columbia, MD.
- Schleiger, S.L. 2000. Use of an index of biotic integrity to detect effects of land uses on stream fish communities in West-Central Georgia. *Transactions of the American Fisheries Society*. 129:1118 – 1133.
- Segura, C., P. Caldwell, G. Sun, S. McNulty, and Y. Zhang. 2015. A model to predict stream water temperature across the conterminous USA. *Hydrological Processes*. 29:2178 – 2195.
- Skousen J.G., A. Sexstone, and P.F. Ziemkiewicz. 2000. Acid mine drainage control and treatment. In Barnhisel RI, Darmody RG, Daniels WL, eds, *Reclamation of Drastically Disturbed Lands*, 2nd ed. Agronomy Monograph 41. American Society of Agronomy and American Society for Surface Mining and Reclamation, Madison, WI, USA, pp 131–168.
- Smogor, R. A., and P.L. Angermeier. 1999. Effects of drainage basin and anthropogenic disturbance on relations between stream size and IBI metrics in Virginia. Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida, 249-272.
- 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*. 16:1267 – 1276.
- Strager, M. P., J. T. Petty, J. M. Strager, and J. Barker-Fulton. 2009. A spatially explicit framework for quantifying downstream hydrologic conditions. *Journal of Environmental Management*. 90:1854 – 1861.

US EPA. 2003. Ecological assessment of streams in the coal mining region of West Virginia using data collected by the U.S. EPA and environmental consulting firms. U. S. Environmental Protection Agency, National Exposure Research Laboratory, Washington DC

US EPA. 2011. The effects of mountaintop mines and valley fills on aquatic ecosystems of the central Appalachian coalfields. U. S. Environmental Protection Agency, Office of Research and Development, Washington DC EPA/600/R-09/138F.

Van Sickle, J. 1997. Using Mean Similarity dendrograms to evaluate classifications. *Journal of Agricultural, Biological, and Environmental Statistics*. 2:370 – 388.

West Virginia Department of Environmental Protection (WVDEP). 2014. West Virginia Integrated Water Quality Monitoring and Assessment Report. Charleston, West Virginia.

Whittier, T.R., J.L. Stoddard, D.P. Larsen, and A.T. Herlihy. 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. *Journal of the North American Benthological Society*. 26:349 – 360.

Yoder, C.O., and E.T. Rankin. 1995. Biological response signatures and the area of degradation value: new tools for interpreting multimetric data. *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton, FL, 263-286.

## Tables

**Table 1:** A total of 1089 fish community samples were collected statewide from the sources listed.

<b>Data Sources</b>	<b>Number of Samples</b>
West Virginia University	128
West Virginia DEP	266
West Virginia DNR (Stream Classification Survey, REMAP)	525
Federal (MAHA, MAIA, NRSA, PEIS, EMAP)	38
Reports from Consulting Companies	135

**Table 2:** Water and habitat quality criteria used to identify reference and stressed site types in each region based on Pond et al. (2012). Level I reference met all criteria, while Level II could be deficient in 1 criterion. Level III reference site criteria could fall deficient in the numbers of point source discharge permits which is typical of large streams and rivers.

<b>Criteria</b>	<b>Reference</b>	<b>Stressed</b>
<b>NPDES Point Source</b>	None	
<b>Dissolved Oxygen (mg/L)</b>	$\geq 5.0$	
<b>pH (Std. Units)</b>	6.0 – 9.0	< 5.0
<b>Specific Conductance (<math>\mu\text{mhos/cm}</math>)</b>	< 500	
<b>Fecal Coliform Bacteria (colonies/100mL)</b>	< 800	
<b>State WQ Violations</b>	none	
<b>U. S. EPA-RBP VBHA metric scores:</b>		
<b>Epifaunal substrate, channel alteration, sediment deposition</b>	$\geq 11$	
<b>Bank vegetative protection, riparian vegetative zone width</b>	$\geq 6$	

**Total RBP habitat score**  $\geq 130$

**Table 3:** ANOSIM and MEANSIM results of the reference site analysis for region selection. The number of classification groups (n groups), within group similarity (W), and between group similarity (B) for each classification scenario. Classification strength (CS%) represents the difference of within group and between group similarities. The significant ( $p < 0.05$ ) model with the highest R statistic was selected as the grouping structure.

<b>Classification</b>	<b>n groups</b>	<b>W</b>	<b>B</b>	<b>W-B (CS%)</b>	<b>p</b>	<b>R</b>
<b>Major Basin</b>	4	0.35	0.26	9.10	0.001	0.26
<b>Ecoregion</b>	3	0.33	0.26	7.20	0.001	0.21
<b>Ecoregion-Basin</b>	9	0.40	0.27	13.50	0.001	0.38

**Table 4:** Final biomonitoring region classifications by HUC8 watershed name (HUC8 number) with total sample sizes (N=1089 statewide).

<b>Mon CA &amp; RV (N=253)</b>	<b>Ohio CA (N=254)</b>	<b>Ohio &amp; Mon WAP (N=279)</b>	<b>UK (N=198)</b>	<b>Potomac (N=105)</b>
Cheat (5020004)	Coal (5050009)	Big Sandy (5070204)	Gauley (5050005)	S. Branch Potomac (2070001)
Elk (5050007)	Elk (5050007)	Little Kanawha (5030203)	Greenbrier (5050003)	N. Branch Potomac (2070002)
Tygart Valley (5020001)	Tug (5070201)	Little Muskingum- Middle Island (5030201)	Lower New (5050004)	Cacapon (2070003)
Youghiogheny (5020006)	Upper Guyandotte (5070101)	Lower Guyandotte (5070102)	Middle New (5050002)	Potomac (2070004)
	Upper Kanawha (5050006)	Lower Kanawha (5050008)		Shenandoah Hardy (2070006)
		Lower Monongahela (5020005)		Shenandoah Jefferson (2070007)
		Raccoon-Symmes (5090101)		
		Twelvepole (5090102)		
		Upper Monongahela (5020003)		
		Upper Ohio (5030101)		
		Upper Ohio-Shade (5030202)		
		Upper Ohio-Wheeling (5030106)		
		West Fork (5020002)		

**Table 5:** Natural landscape summary statistics and sample sizes for all site types in Mon CA and RV region. Values are presented as mean (minimum-maximum).

Site Type	N	Drainage Area (km <sup>2</sup> )	Swim Distance (km)	Elevation (m)
Reference	23	82.76 (8.07 – 343.85)	51.91 (0 – 208.06)	628.58 (419 – 984)
Non-Reference	111	56.89 (7.95 – 357.60)	28.53 (0 – 120.08)	604.75 (320 – 1166)
Stressed	30	41.10 (7.47 – 234.63)	20.82 (0 – 81.11)	523.89 (354 – 1006)
Cold	77	29.31 (8.38 – 248.57)	41.37 (0 – 175.18)	725.37 (414 – 1232)

**Table 6:** Natural landscape summary statistics and sample sizes for all site types in Ohio CA region. Values are presented as mean (minimum-maximum).

Site Type	N	Drainage Area (km <sup>2</sup> )	Swim Distance (km)	Elevation (m)
Reference	18	53.77 (7.34 – 307.98)	31.35 (0 – 66.98)	440.00 (251 – 703)
Non-Reference	202	73.48 (7.29 – 392.99)	27.85 (0 – 112.59)	364.72 (190 – 790)
Stressed	24	34.46 (7.34 – 154.18)	41.62 (1.7 – 109.5)	365.58 (249 – 575)
Cold	2	13.85 (13.45 – 14.25)	77.11 (59.37 – 94.84)	581.06

**Table 7:** Natural landscape summary statistics and sample sizes for all site types in Ohio and Mon WAP region. Values are presented as mean (minimum-maximum).

Site Type	N	Drainage Area (km <sup>2</sup> )	Swim Distance (km)	Elevation (m)
Reference	21	79.20 (9.73 – 390.57)	32.10 (0 – 78.36)	310.97 (210 – 493)
Non-Reference	227	93.92 (7.38 – 384.85)	30.57 (0 – 137.51)	271.79 (175 – 506)
Stressed	27	98.68 (8.04 – 357.76)	24.89 (0 – 110.43)	302.56 (183 – 528)

**Table 8:** Natural landscape summary statistics and sample sizes for all site types in Upper Kanawha region. Values are presented as mean (minimum-maximum).

Site Type	N	Drainage Area (km <sup>2</sup> )	Swim Distance (km)	Elevation (m)
Reference	21	74.02 (9.55 – 335.09)	37.39 (0 – 135.93)	745.38 (399 – 1035)
Non-Reference	147	96.81 (8.09 – 392.90)	35.95 (0 – 134.68)	645.95 (223 – 1074)
Stressed	11	65.74 (7.01 – 351.70)	21.21 (0 – 49.78)	632.98 (395 – 954)
Cold	19	42.75 (7.63 – 133.71)	66.22 (2.52 – 118.57)	931.20 (518 – 1137)

**Table 9:** Final metrics selected for inclusion in IBIs for each biomonitoring region and their correlation coefficients with cumulative surface mining (%), development (%), grassland (%), agriculture (%), forest (%), and structure density (#/km<sup>2</sup>). Metrics were also correlated with specific conductance (SPC) and pH. Discrimination efficiency (%DE) was also calculated.

<b>Metric</b>	<b>SM</b>	<b>Develop</b>	<b>Grass.</b>	<b>Agri.</b>	<b>Forest</b>	<b>SD</b>	<b>SPC</b>	<b>pH</b>	<b>DE</b>
<b>Mon CA-RV</b>									
Adj.R_Cyprinid	-0.324	-0.185	-0.098	-0.195	0.190	0.034	-0.437	0.408	70.00
Adj.P_IN	-0.225	-0.239	-0.282	-0.149	0.249	0.123	-0.361	0.321	73.33
Adj.P_Fish2	-0.218	-0.282	-0.337	-0.265	0.350	-0.034	-0.407	0.197	76.67
Adj.R_Benthic	-0.323	-0.214	-0.106	-0.236	0.237	-0.001	-0.502	0.438	86.67
Adj.R_Fish2	-0.302	-0.197	-0.118	-0.211	0.225	0.032	-0.543	0.354	86.67
Adj.R_McC_CGS	-0.304	-0.303	-0.198	-0.300	0.306	-0.150	-0.601	0.320	90.00
P_Benthic2	-0.363	-0.208	-0.167	-0.202	0.253	-0.010	-0.575	0.314	93.33
<b>Ohio CA</b>									
Adj.R_Richness2	-0.482	-0.040	0.316	0.229	0.291	0.474	-0.325	-0.102	79.17
Adj.R_DMS	-0.516	-0.084	0.331	0.264	0.311	0.430	-0.379	-0.074	95.83
Adj.R_RGS	-0.446	-0.029	0.332	0.239	0.250	0.427	-0.352	-0.065	70.83
Adj.R_NGL2	-0.349	0.060	0.306	0.223	0.171	0.409	-0.240	-0.002	58.33
Adj.P_IN	-0.221	-0.235	0.001	0.128	0.276	0.234	-0.192	-0.114	58.33
Adj.P_Tol	0.022	0.192	0.177	-0.007	-0.209	-0.056	0.016	-0.018	50.00
SW Trophic	-0.227	-0.073	0.064	0.138	0.257	0.367	-0.015	-0.034	75.00
<b>Ohio-Mon WAP</b>									
Adj.R_DMS	-0.081	-0.388	-0.245	-0.193	0.297	-0.301	-0.176	-0.125	88.89
Adj.R_Cyp -bndseat	-0.122	-0.198	-0.072	-0.199	0.237	-0.104	-0.122	-0.120	81.48
Adj.R_RGS	0.021	-0.279	-0.171	-0.239	0.249	-0.174	-0.139	-0.048	77.78
Adj.R_Int	-0.136	-0.293	-0.161	-0.161	0.290	-0.191	-0.209	-0.107	70.37
Adj.R_Richness2	-0.212	-0.265	-0.137	-0.159	0.284	-0.195	-0.162	-0.167	70.37
P_OH_CAAN	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	62.96
Adj.P_Tol	-0.014	0.089	0.041	0.016	-0.040	-0.044	0.177	0.173	51.85
Adj.P_NGL2	0.120	-0.162	-0.115	-0.042	0.054	-0.196	-0.074	-0.109	51.85
<b>Upper Kanawha</b>									
Adj.R_Int	-0.354	-0.307	-0.082	-0.007	0.287	-0.133	-0.399	0.056	63.64
Adj.P_Tol	0.155	0.425	0.271	0.323	-0.418	0.400	0.318	0.091	63.64
R_Benthic	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	54.55
Adj.R_FISH	-0.225	-0.217	-0.219	0.094	0.032	-0.140	-0.199	0.172	45.45

---

Adj.P_IN	-0.059	-0.394	-0.280	-0.293	0.327	-0.406	-0.228	0.010	45.45
Adj.R_Cyprinid	-0.241	-0.164	-0.173	0.149	0.025	-0.065	-0.214	0.158	45.45
Adj.R_CGS_RGS	-0.214	-0.316	-0.328	-0.035	0.120	-0.288	-0.340	0.068	54.55

---



**Table 10:** Final metrics selected for the Mon CA-RV region with metric description and direction. Metrics direction is either positive (decreases with increases in stress) or negative (increases with increases in stress). The Ceiling (95<sup>th</sup> percentile) and Floor (5<sup>th</sup> percentile) were used for scoring criteria.

<b>Metric</b>	<b>Description</b>	<b>Direction</b>	<b>Ceiling</b>	<b>Floor</b>
P_Benthic2.DEP	Proportion of benthic individuals minus tolerant	Positive	0.683	0
Adj.R_Fish2.DEP	Adjusted species richness minus tolerant	Positive	1.215	0
Adj.R_McC_CGS	Adjusted clean gravel spawner richness	Positive	1.326	0
Adj.P_Fish2.DEP	Adjusted proportion of non-tolerant individuals	Positive	1.537	0
Adj.P_IN	Adjusted proportion of invertivore individuals	Positive	1.506	0
Adj.R_Benthic	Adjusted benthic species richness	Positive	1.370	0
Adj.R_Cyprinid	Adjusted Cyprinidae richness	Positive	1.326	0

**Table 11:** Final metrics selected for the Ohio CA region with metric description and direction. Metrics direction is either positive (decreases with increases in stress) or negative (increases with increases in stress). The Ceiling (95<sup>th</sup> percentile) and Floor (5<sup>th</sup> percentile) were used for scoring criteria.

<b>Metric</b>	<b>Description</b>	<b>Direction</b>	<b>Ceiling</b>	<b>Floor</b>
Adj.R_Fish2.DEP	Adjusted non-tolerant species richness	Positive	1.766	0
Adj.P_IN	Adjusted proportion invertivore individuals	Positive	1.751	0
Adj.R_DMS	Adjusted darter-madtom-sculpin richness	Positive	1.325	0
Adj.P_Tol	Adjusted proportion of tolerant individuals	Negative	1.702	0.413
Adj.R_RGS	Adjusted rock-gravel spawner richness	Positive	1.447	0
Adj.R_NGL2	Adjusted non-guarding lithophils minus tolerant richness	Positive	1.950	0
SW_TROPHIC	Shannon-Weaver Trophic diversity index	Positive	3.455	1.689

**Table 12:** Final metrics selected for the Ohio-Mon WAP region with metric description and direction. Metrics direction is either positive (decreases with increases in stress) or negative (increases with increases in stress). The Ceiling (95<sup>th</sup> percentile) and Floor (5<sup>th</sup> percentile) were used for scoring criteria.

<b>Metric</b>	<b>Description</b>	<b>Direction</b>	<b>Ceiling</b>	<b>Floor</b>
Adj.R_Fish2.DEP	Adjusted non-tolerant species richness	Positive	1.339	0.371
Adj.R_DMS	Adjusted darter-madtom-sculpin richness	Positive	1.158	0.265
Adj.R_Int	Adjusted intolerant species richness	Positive	1.640	0
Adj.P_Tol	Adjusted proportion tolerant individuals	Negative	1.675	0.726
Adj.R_RGS	Adjusted rock-gravel spawner richness	Positive	1.191	0.358
Adj.R_NGL2	Adjusted non-guarding lithophil richness minus tolerant species	Positive	1.454	0
P_OH_CAAN	Proportion of omnivore-herbivore minus Central Stoneroller	Negative	0.655	0.119
Adj.R_Cyprinid_BNDSEAT	Adjusted Cyprinidae richness minus Blacknose Dace and Creek Chub	Positive	1.412	0.276

**Table 13:** Final metrics selected for the Upper Kanawha region with metric description and direction. Metrics direction is either positive (decreases with increases in stress) or negative (increases with increases in stress). The Ceiling (95<sup>th</sup> percentile) and Floor (5<sup>th</sup> percentile) were used for scoring criteria.

<b>Metric</b>	<b>Description</b>	<b>Direction</b>	<b>Ceiling</b>	<b>Floor</b>
Adj.P_IN	Adjusted proportion of invertivore individuals	Positive	1.808	0.011
Adj.R_Int	Adjusted intolerant species richness	Positive	2.952	0
Adj.P_Tol	Adjust proportion of tolerant individuals	Negative	2.412	0.269
R_Benthic	Benthic species richness	Positive	1.850	10
Adj.Richness	Adjusted total species richness	Positive	2.001	0.405
Adj.R_CGS_RGS	Adjusted clean and rock-gravel spawning species richness	Positive	1.735	0.354
Adj.R_Cyprinid	Adjusted Cyprinidae species richness	Positive	2.067	0.426

**Table 14:** Spearman correlation coefficients for final IBI scores within each biomonitoring region against stream characteristics. Biotic variables include the West Virginia Stream Condition Index (WVSCI), Genus-Level Index of Most Probable Stream Status (GLIMPSS), total fish abundance, and the Mid-Atlantic Highlands Index of Biotic Integrity scores (MAH IBI). In stream characteristics compared were specific conductance (SPC) and pH, along with cumulative percentages of surface mining (C. Surface Mining), development (C. Development), agriculture (C. Agriculture), and total forest (C. Forest). Relationships with cumulative densities of structures (C. Structure Density) and National Pollution Discharge Elimination System Permits (C. NPDES Permit Density) were also evaluated.

<b>Variable</b>	<b>Mon CA-RV</b>	<b>Ohio CA</b>	<b>Ohio-Mon WAP</b>	<b>Upper Kanawha</b>
<b>WVSCI</b>	0.7922	-0.0793	0.0721	0.3980
<b>GLIMPSS</b>	0.7746	-0.1078	-0.053	0.3186
<b>Fish Abundance</b>	0.7707	0.5269	0.3174	0.5155
<b>MAH IBI</b>	0.7859	0.6866	0.6434	0.6256
<b>SPC</b>	-0.6640	-0.2736	-0.1983	-0.2949
<b>pH</b>	0.1110	-0.1065	-0.1842	0.1355
<b>Drainage Area</b>	0.2617	0.2372	-0.0995	0.0968
<b>Elevation</b>	0.1111	-0.3009	0.1342	0.1976
<b>C. Surface Mining</b>	-0.2026	-0.4148	-0.1348	-0.2585
<b>C. Development</b>	-0.2656	-0.1187	-0.3558	-0.3463
<b>C. Agriculture</b>	-0.3110	0.2083	-0.2003	0.0007
<b>C. Structure Density</b>	-0.0007	0.4265	-0.2103	-0.2628
<b>C. Forest</b>	0.3374	0.3326	0.3149	0.1816
<b>C. NPDES Permit Density</b>	-0.0412	0.1394	-0.1968	-0.2687

**Table 15:** Impairment category thresholds for each biomonitoring region based on the distribution of the reference sites. The 25th percentile within the Ohio CA, Ohio-Mon WAP, and Upper Kanawha and the 10<sup>th</sup> percentile in the Mon CA-RV were used to make the distinction between impaired (i.e. Degraded and Severely Degraded) and non-impaired (Good and Excellent) streams.

<b>Condition</b>	<b>Mon CA-RV</b>	<b>Ohio CA</b>	<b>Ohio-Mon WAP</b>	<b>Upper Kanawha</b>
<b>Excellent</b>	$\geq 62.36$	$\geq 77.92$	$\geq 66.63$	$\geq 69.30$
<b>Good</b>	62.36 – 56.15	77.92– 48.13	66.63 – 53.16	69.30 – 47.48
<b>Degraded</b>	56.15 – 45.38	48.13 – 23.64	53.16 – 43.21	47.48– 39.34
<b>Severely Degraded</b>	$< 45.38$	$< 23.64$	$< 43.21$	$< 39.34$

## Figures

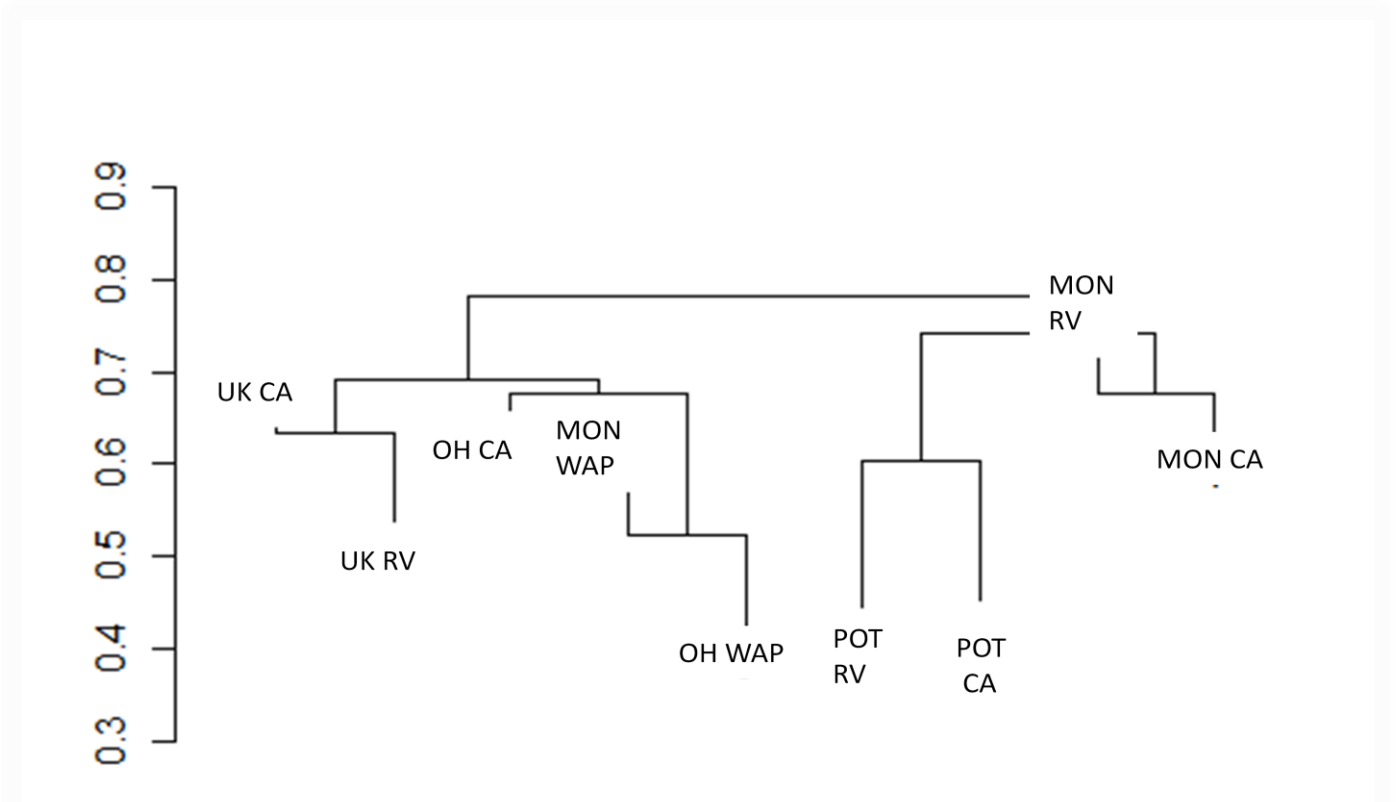


Figure 1: MEANSIM dendrogram groups based on fish species abundances at all reference sites. Based on the dendrogram grouping and available samples biomonitoring regions were determined. We grouped the nine classifications listed into five biomonitoring regions. These regions include: 1) Upper Kanawha Central Appalachians (UK CA) and Upper Kanawha Ridge/Valley (UK RV); 2) Monongahela Ridge/Valley (MON RV) and Monongahela Central Appalachians (MON CA); 3) Potomac Ridge/Valley (POT RV) and Potomac Central Appalachians (POT CA); 4) Monongahela Western Allegheny Plateau (MON WAP) and Ohio Western Allegheny Plateau (OH WAP); and 5) Ohio Central Appalachians (OH CA).

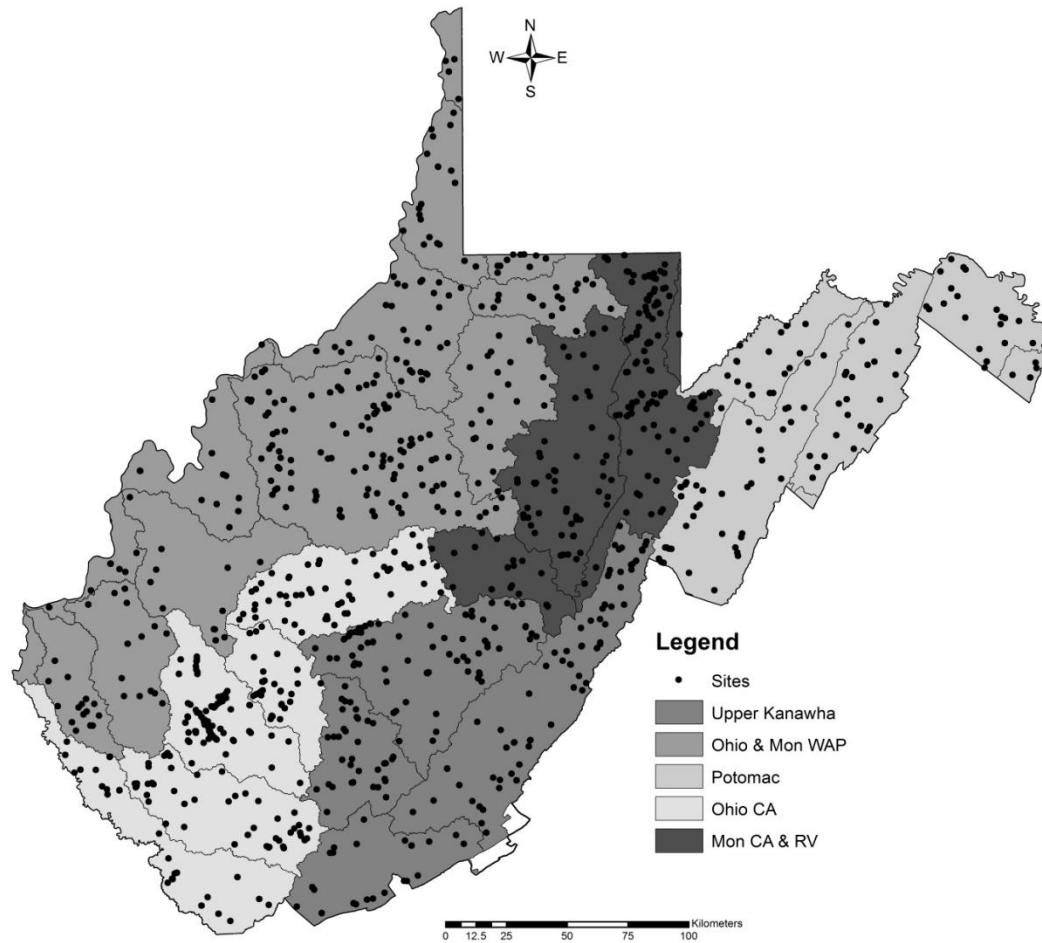


Figure 2: Final fish IBI biomonitoring regions based on similarity analysis and arrangement to include distinct watershed boundaries from whole HUC8 or HUC12 (Elk watershed) watersheds. The final biomonitoring regions include the Upper Kanawha, Ohio and Monongahela Western Allegheny Plateau (Ohio & Mon WAP), Potomac, Ohio Central Appalachians (Ohio CA), and Monongahela Central Appalachians and Ridge and Valley (Mon CA & RV).

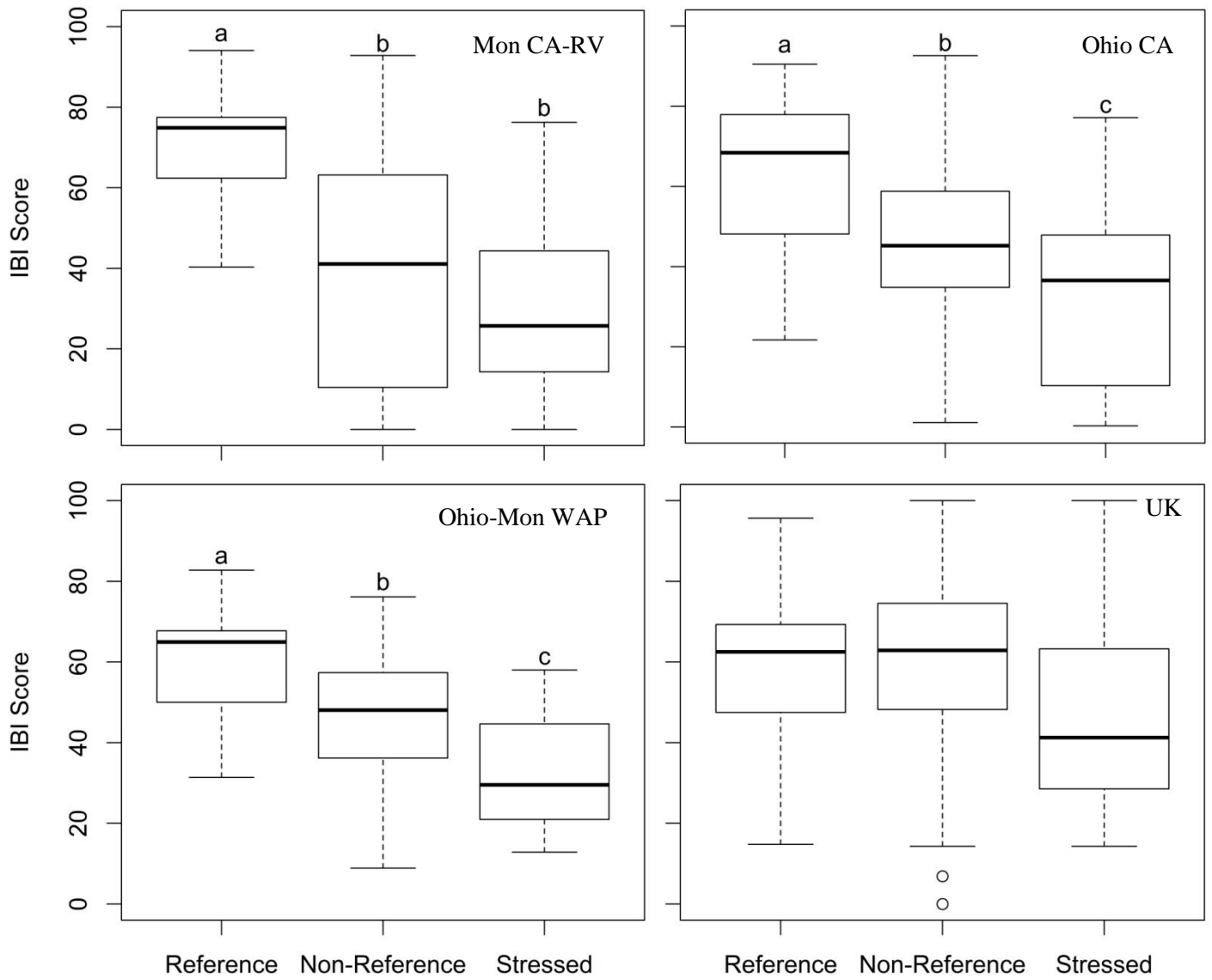


Figure 3: Final IBI scores for each biomonitoring region. Different lowercase letters indicate a significant difference between groups (Reference, Non-Reference, or Stressed) based on ANOVA results for Mon CA-RV, Ohio CA, Ohio-Mon WAP, and Upper Kanawha (UK).



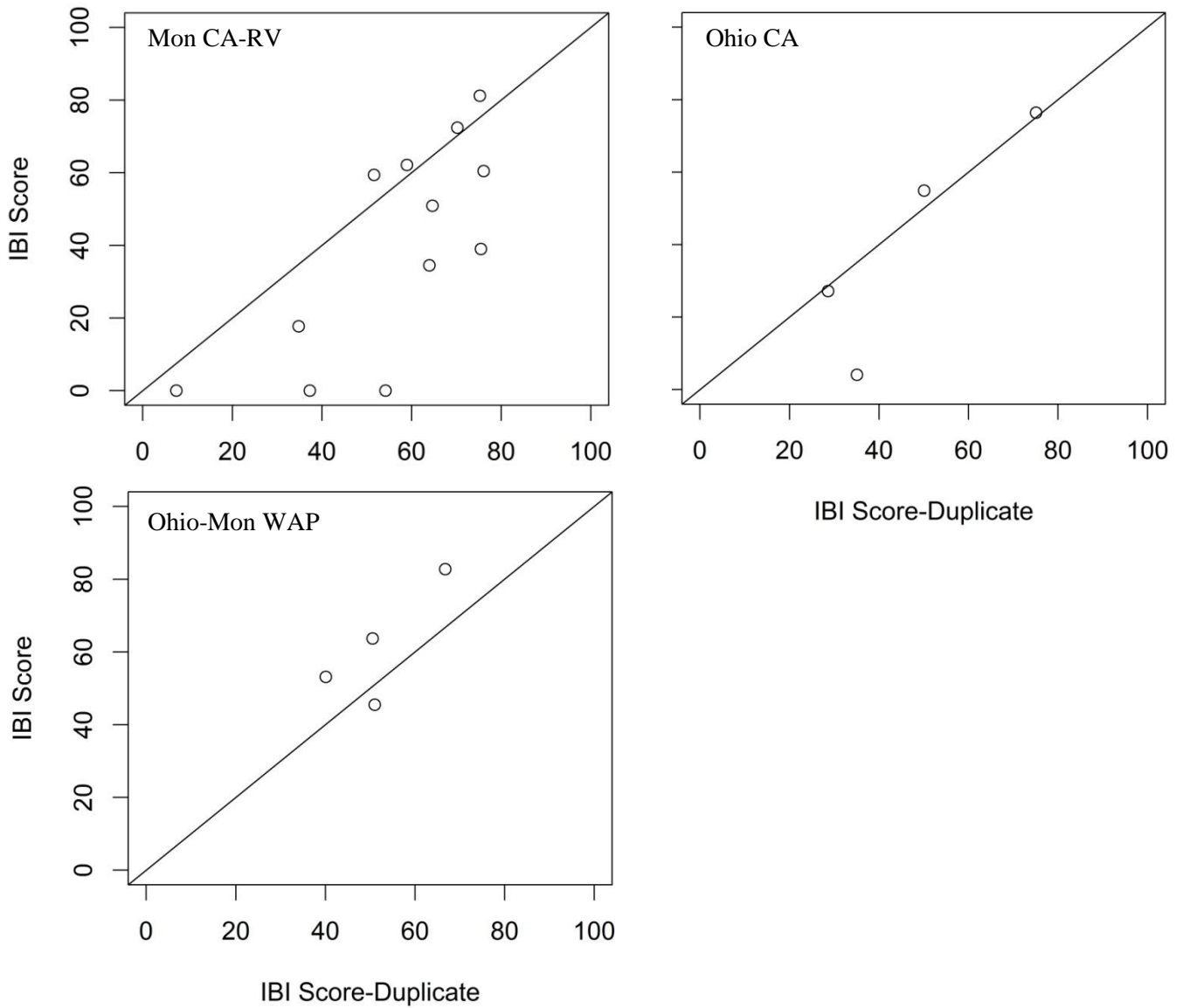


Figure 4: Final IBI scores for original (y-axis) and duplicate (x-axis) samples within the Mon CA-RV, Ohio CA, and Ohio-Mon WAP biomonitoring regions. The solid line represents a 1:1 relationship.

1 **Appendix A**

2 All 171 species occurring in the dataset were classified based on native status, spawning, feeding, tolerance values, and other data. Native status  
 3 consists of only species native to the Ohio, Monongahela (Mon), Potomac, and Upper Kanawha (UK) drainages or all drainages (WV). Spawning  
 4 consisted of rock-gravel spawners (RG), gravel-sand spawners (GS), non-guarding lithophils (NGL), cavity spawners (CAV), and clean gravel  
 5 spawners (CGS). The feeding (trophic) category consisted of invertivore-piscivore (IP), invertivore (IN), macro-omnivore (MO), and omnivore-  
 6 herbivore (OH). Tolerance values ranged from intolerant (I), moderate tolerance (M), and tolerant (T). Other classifications included benthic  
 7 species (B), game species (G), and cold water species (C). Lithophilic spawners in sand to rock (LSR) is consisted of any species that was  
 8 classified as either RG or GS, or showing no substrate preference. Species are listed in alphabetical order according to their scientific name.

<b>Common</b>	<b>Scientific</b>	<b>Code</b>	<b>Family</b>	<b>Native</b>	<b>Spawning</b>	<b>Trophic</b>	<b>Tolerance</b>	<b>Other</b>
<b>Skipjack herring</b>	<i>Alosa chrysochloris</i>	ALCH	Clupeidae	Ohio		IP	M	
<b>Rock bass</b>	<i>Ambloplites rupestris</i>	AMRU	Centrarchidae	Mon, Ohio		IP	M	G, B
<b>Black bullhead</b>	<i>Ameiurus melas</i>	AMME	Ictaluridae	Ohio		MO, OH	M	G, B
<b>Yellow bullhead</b>	<i>Ameiurus natalis</i>	AMNA	Ictaluridae	WV		MO, OH	T	G, B
<b>Brown bullhead</b>	<i>Ameiurus nebulosus</i>	AMNE	Ictaluridae	WV		MO, OH	T	G, B
<b>Western sand darter</b>	<i>Ammocrypta clara</i>	AMCL	Percidae	Ohio	LSR	IN, IP	I	B
<b>Eastern sand darter</b>	<i>Ammocrypta pellucida</i>	AMPE	Percidae	Ohio	LSR	IN, IP	I	B
<b>American eel</b>	<i>Anguilla rostrata</i>	ANRO	Anguillidae	Potomac, Ohio, UK		IP	T	
<b>Freshwater drum</b>	<i>Aplodinotus grunniens</i>	APGR	Sciaenidae	Ohio		IN, IP	M	G, B
<b>Central Stoneroller</b>	<i>Campostoma anomalum</i>	CAAN	Cyprinidae	WV	RG, CGS	MO, OH	T	B
<b>Goldfish</b>	<i>Carassius auratus</i>	CAAU	Cyprinidae	None		MO, OH	T	
<b>River carpsucker</b>	<i>Carpiodes carpio</i>	CACA	Catostomidae	Ohio, Mon		MO, OH	M	B
<b>Quillback</b>	<i>Carpiodes cyprinus</i>	CACY	Catostomidae	Mon, Ohio		MO, OH	T	B
<b>Highfin carpsucker</b>	<i>Carpiodes velifer</i>	CAVE	Catostomidae	Ohio		MO, OH	I	B
<b>Longnose sucker</b>	<i>Catostomus catostomus</i>	CACT	Catostomidae	Mon	RG, NGL	MO, OH	I	B, C
<b>White sucker</b>	<i>Catostomus commersonii</i>	CACO	Catostomidae	WV	GS, NGL	MO, OH	T	B

<b>Southern redbelly dace</b>	Chrosomus erythrogaster	PHER	Cyprinidae	Ohio		MO, OH	M	
<b>Mountain redbelly dace</b>	Chrosomus oreas	PHOR	Cyprinidae	UK	RG	MO, OH	I	C
<b>Redside dace</b>	Clinostomus elongatus	CLEL	Cyprinidae	Mon, Ohio	RG	IN, IP	I	
<b>Rosyside dace</b>	Clinostomus funduloides	CLFU	Cyprinidae	WV	RG	IN, IP	I	
<b>Mottled sculpin</b>	Cottus bairdii	COBA	Cottidae	Mon, Ohio, UK	CAV	IN, IP	M	B
<b>Blue Ridge sculpin</b>	Cottus caeruleomentum	COCA	Cottidae	Potomac		IN, IP	M	B
<b>Banded sculpin</b>	Cottus carolinae	COCR	Cottidae	UK	CAV	IN, IP	M	B
<b>Slimy sculpin</b>	Cottus cognatus	COCO	Cottidae	Potomac	CAV	IN, IP	I	B, C
<b>Potomac sculpin</b>	Cottus girardi	COGI	Cottidae	Potomac	CAV	IN, IP	M	B
<b>Kanawha sculpin</b>	Cottus kanawhae	COKA	Cottidae	UK		IN, IP	I	B, C
<b>Checkered Sculpin</b>	Cottus cf. cognatus	CORO	Cottidae	Potomac		IN, IP	M	B, C
<b>Bluestone sculpin</b>	Cottus sp.	COBL	Cottidae	UK		IN, IP	M	B
<b>Diamond darter</b>	Crystallaria cincotta	CRCI	Percidae	Ohio		IN, IP	I	B
<b>Grass carp</b>	Ctenopharyngodon idella	CTID	Cyprinidae	None		MO, OH	M	B
<b>Satinfin shiner</b>	Cyprinella analostana	CYAN	Cyprinidae	Potomac	CAV	OH	T	
<b>Whitetail shiner</b>	Cyprinella galactura	CYGA	Cyprinidae	UK	CAV	IN, IP	M	
<b>Spotfin shiner</b>	Cyprinella spiloptera	CYSP	Cyprinidae	WV	CAV	IN, IP	T	
<b>Steelcolor shiner</b>	Cyprinella whipplei	CYWH	Cyprinidae	Mon, Ohio	CAV	IN, IP	M	
<b>Common carp</b>	Cyprinus carpio	CYCA	Cyprinidae	None		MO, OH	T	G
<b>Gizzard shad</b>	Dorosoma cepedianum	DOCE	Clupeidae	Mon, Ohio		MO, OH	T	
<b>Threadfin shad</b>	Dorosoma petenense	DOPE	Clupeidae	Ohio		MO, OH	M	
<b>Appalachia darter</b>	Percina gymnocephala	PEGY	Percidae	UK	GS, CGS	IN, IP	I	B
<b>Streamline chub</b>	Erimystax dissimilis	ERDI	Cyprinidae	Ohio, UK	RG, NGL	OH	I	B
<b>Creek chubsucker</b>	Erimyzon oblongus	EROB	Catostomidae	Potomac	GS, NGL	MO, OH	I	B
<b>Grass pickerel</b>	Esox americanus	ESAM	Esocidae	Potomac, Ohio		IP	M	

<b>Northern Pike</b>	<i>Esox lucius</i>	ESLU	Esocidae	None	NGL	IP	I	G
<b>Muskellunge</b>	<i>Esox masquinongy</i>	ESMA	Esocidae	Mon, Ohio		IP	I	G
<b>Chain pickerel</b>	<i>Esox niger</i>	ESNI	Esocidae	Potomac		IP	M	G
<b>Greenside darter</b>	<i>Etheostoma blennioides</i>	ETBL	Percidae	Mon, Ohio, UK	RG, NGL	IN, IP	I	B
<b>Rainbow darter</b>	<i>Etheostoma caeruleum</i>	ETCA	Percidae	Mon, Ohio, UK	RG, CGS	IN, IP	M	B
<b>Bluebreast darter</b>	<i>Etheostoma camurum</i>	ETCM	Percidae	Mon, Ohio, UK	GS	IN, IP	I	B
<b>Fantail darter</b>	<i>Etheostoma flabellare</i>	ETFL	Percidae	WV	RG, CAV	IN, IP	M	B
<b>Longfin darter</b>	<i>Etheostoma longimanum</i>	ETLO	Percidae	None		IN, IP	I	B
<b>Spotted darter</b>	<i>Etheostoma maculatum</i>	ETMA	Percidae	Ohio	CAV	IN, IP	I	B
<b>Johnny darter</b>	<i>Etheostoma nigrum</i>	ETNI	Percidae	Mon, Ohio, UK	RG, CAV	IN, IP	M	B
<b>Tessellated darter</b>	<i>Etheostoma olmstedii</i>	ETOL	Percidae	Potomac	CAV	IN, IP	M	B
<b>Candy darter</b>	<i>Etheostoma osburni</i>	ETOS	Percidae	UK	GS	IN, IP	I	B
<b>Snubnose darter</b>	<i>Etheostoma simoterum</i>	ETSI	Percidae	None	RG	IN, IP	M	B
<b>Tippecanoe darter</b>	<i>Etheostoma tippecanoe</i>	ETTI	Percidae	Ohio	RG	IN, IP	I	B
<b>Variagate darter</b>	<i>Etheostoma variatum</i>	ETVA	Percidae	Mon, Ohio	GS, NGL	IN, IP	M	B
<b>Banded darter</b>	<i>Etheostoma zonale</i>	ETZO	Percidae	Mon, Ohio	NGL	IN, IP	I	B
<b>Tonguetied minnow</b>	<i>Exoglossum laurae</i>	EXLA	Cyprinidae	Mon, UK	RG, CGS	IN, IP	M	
<b>Cutlips minnow</b>	<i>Exoglossum maxillingua</i>	EXMA	Cyprinidae	Potomac	RG, CGS	IN, IP	I	
<b>Northern studfish</b>	<i>Fundulus catenatus</i>	FUCA	Fundulidae	None		IN, IP	I	
<b>Banded killifish</b>	<i>Fundulus diaphanus</i>	FUDI	Fundulidae	Potomac, Mon, Ohio		IN, IP	T	
<b>Mosquitofish</b>	<i>Gambusia affinis</i>	GAAF	Poeciliidae	None		IN, IP	T	
<b>Eastern mosquitofish</b>	<i>Gambusia holbrooki</i>	GAHO	Poeciliidae	None	CGS	IN, IP	T	
<b>Goldeye</b>	<i>Hiodon alosoides</i>	HIAL	Hiodontidae	Ohio	NGL	IN, IP	I	B
<b>Mooneye</b>	<i>Hiodon tergisus</i>	HITE	Hiodontidae	Mon, Ohio		IN, IP	I	
<b>Eastern silvery minnow</b>	<i>Hybognathus regius</i>	HYRE	Cyprinidae	Potomac		MO, OH	I	B

<b>Bigeye chub</b>	<i>Hybopsis amblops</i>	HYAM	Cyprinidae	Mon, Ohio	GS, NGL	IN, IP	M	
<b>Northern hogsucker</b>	<i>Hypentelium nigricans</i>	HYNI	Catostomidae	WV	RG, CGS, NGL	IN, IP	M	B
<b>Ohio lamprey</b>	<i>Ichthyomyzon bdellium</i>	ICBD	Petromyzontidae	Ohio	RG	MO, OH	M	B
<b>Northern Brook lamprey</b>	<i>Ichthyomyzon fossor</i>	ICFO	Petromyzontidae	Ohio	RG	MO, OH	I	B
<b>Mountain brook lamprey</b>	<i>Ichthyomyzon greeleyi</i>	ICGR	Petromyzontidae	Ohio	GS, CGS	MO, OH	I	B
<b>Silver lamprey</b>	<i>Ichthyomyzon unicuspis</i>	ICUN	Petromyzontidae	Ohio	GS	MO, OH	M	B
<b>Channel catfish</b>	<i>Ictalurus punctatus</i>	ICPU	Ictaluridae	Mon, Ohio, UK		MO, OH	T	B, G
<b>Smallmouth Buffalo</b>	<i>Ictiobus bubalus</i>	ICBU	Catostomidae	Ohio		OH	M	B
<b>Bigmouth buffalo</b>	<i>Ictiobus cyprinellus</i>	ICCY	Catostomidae	Ohio		OH	M	B
<b>Black Buffalo</b>	<i>Ictiobus niger</i>	ICNI	Catostomidae	Ohio		MO, OH	M	B
<b>Brook silverside</b>	<i>Labidesthes sicculus</i>	LASI	Atherinopsidae	Mon, Ohio		IN, IP	I	
<b>Least brook lamprey</b>	<i>Lampetra aepyptera</i>	LAAE	Petromyzontidae	Mon, Ohio	GS, CGS	MO, OH	I	B
<b>Longnose gar</b>	<i>Lepisosteus osseus</i>	LEOS	Lepisosteidae	Mon, Ohio		IP	M	G
<b>Redbreast sunfish</b>	<i>Lepomis auritus</i>	LEAU	Centrarchidae	Potomac	GS	IP	M	G
<b>Green sunfish</b>	<i>Lepomis cyanellus</i>	LECY	Centrarchidae	Mon, Ohio		IP	T	G
<b>Pumpkinseed</b>	<i>Lepomis gibbosus</i>	LEGI	Centrarchidae	Potomac, Mon, Ohio		IN, IP	M	
<b>Warmouth</b>	<i>Lepomis gulosus</i>	LEGU	Centrarchidae	Ohio		IP	M	
<b>Orangespotted Sunfish</b>	<i>Lepomis humilis</i>	LEHU	Centrarchidae	Ohio	GS	IP	M	
<b>Bluegill</b>	<i>Lepomis macrochirus</i>	LEMA	Centrarchidae	Potomac, Mon, Ohio		IN, IP	T	G
<b>Longear sunfish</b>	<i>Lepomis megalotis</i>	LEME	Centrarchidae	Potomac, Mon, Ohio		IN, IP	M	G
<b>Redear sunfish</b>	<i>Lepomis microlophus</i>	LEMI	Centrarchidae	None		IN, IP	M	
<b>American brook lamprey</b>	<i>Lethenteron appendix</i>	LAAP	Petromyzontidae	Ohio	GS, CGS	MO, OH	I	B
<b>White shiner</b>	<i>Luxilus albeolus</i>	LUAL	Cyprinidae	UK	RG	IN, IP	M	

<b>Striped shiner</b>	<i>Luxilus chrysocephalus</i>	LUCH	Cyprinidae	Mon, Ohio	RG	OH	T	
<b>Common shiner</b>	<i>Luxilus cornutus</i>	LUCO	Cyprinidae	Potomac	GS	OH	M	
<b>Rosefin shiner</b>	<i>Lythrurus ardens</i>	LYAR	Cyprinidae	UK	RG	IN, IP	M	
<b>Redfin shiner</b>	<i>Lythrurus umbratilis</i>	LYUM	Cyprinidae	Mon, Ohio		IN, IP	T	
<b>Speckled chub</b>	<i>Macrhybopsis aestivalis</i>	MAAE	Cyprinidae	Ohio		IP	I	B
<b>Shoal chub</b>	<i>Macrhybopsis hyostoma</i>	MAHY	Cyprinidae	Ohio		IN, IP	I	
<b>Silver chub</b>	<i>Macrhybopsis storeriana</i>	MAST	Cyprinidae	Mon, Ohio		IN, IP	I	B
<b>Pearl dace</b>	<i>Margariscus margarita</i>	MAMA	Cyprinidae	Potomac, Mon	GS, NGL	IN, IP	M	
<b>Smallmouth bass</b>	<i>Micropterus dolomieu</i>	MIDO	Centrarchidae	Mon, Ohio		IP	M	G
<b>Spotted bass</b>	<i>Micropterus punctulatus</i>	MIPU	Centrarchidae	Mon, Ohio, UK		IP	M	G
<b>Largemouth bass</b>	<i>Micropterus salmoides</i>	MISA	Centrarchidae	Potomac, Mon, Ohio		IP	M	G
<b>Spotted sucker</b>	<i>Minytrema melanops</i>	MIME	Catostomidae	Ohio	RG, NGL	OH	M	B
<b>White Perch</b>	<i>Morone americana</i>	MOAM	Moronidae	None		IP	M	G
<b>White bass</b>	<i>Morone chrysops</i>	MOCH	Moronidae	Ohio		IP	T	G
<b>Striped bass</b>	<i>Morone saxatilis</i>	MOSA	Moronidae	None		IP	I	G
<b>Silver redhorse</b>	<i>Moxostoma anisurum</i>	MOAN	Catostomidae	Mon, Ohio	RG, NGL	IN, IP	M	B
<b>Smallmouth redhorse</b>	<i>Moxostoma breviceps</i>	MOBR	Catostomidae	Ohio		IN, IP	M	B
<b>River redhorse</b>	<i>Moxostoma carinatum</i>	MOCA	Catostomidae	Ohio	RG, NGL	IN, IP	I	B
<b>Black redhorse</b>	<i>Moxostoma duquesni</i>	MODU	Catostomidae	Ohio	RG, NGL	IN, IP	I	B
<b>Golden redhorse</b>	<i>Moxostoma erythrurum</i>	MOER	Catostomidae	Potomac, Mon, Ohio	GS, CGS, NGL	IN, IP	I	B
<b>Shorthead redhorse</b>	<i>Moxostoma macrolepidotum</i>	MOMA	Catostomidae	Potomac	RG, NGL	IN, IP	M	B
<b>Bluehead chub</b>	<i>Nocomis leptocephalus</i>	NOLE	Cyprinidae	UK	RG, CGS	MO, OH	M	
<b>River chub</b>	<i>Nocomis micropogon</i>	NOMI	Cyprinidae	Potomac, Mon, Ohio	RG, CGS	IN, IP	M	
<b>Bigmouth chub</b>	<i>Nocomis platyrhynchus</i>	NOPL	Cyprinidae	UK	RG	IN, IP	M	
<b>Golden shiner</b>	<i>Notemigonus crysoleucas</i>	NOCY	Cyprinidae	Potomac, Mon, Ohio		MO, OH	T	
<b>Comely shiner</b>	<i>Notropis amoenus</i>	NOAM	Cyprinidae	Potomac	RG	IN, IP	T	
<b>Popeye shiner</b>	<i>Notropis ariommus</i>	NOAR	Cyprinidae	Mon, Ohio	RG, NGL	IN, IP	I	

<b>Emerald shiner</b>	<i>Notropis atherinoides</i>	NOAT	Cyprinidae	Mon, Ohio		MO, OH	M	
<b>River shiner</b>	<i>Notropis blennius</i>	NOBL	Cyprinidae	Ohio	GS, NGL	IN, IP	M	
<b>Bigeye shiner</b>	<i>Notropis boops</i>	NOBO	Cyprinidae	Ohio	GS, NGL	IN, IP	I	
<b>Silverjaw minnow</b>	<i>Notropis buccatus</i>	NOBU	Cyprinidae	WV	GS, NGL	IN, IP	T	
<b>Ghost shiner</b>	<i>Notropis buchanani</i>	NOBC	Cyprinidae	Mon, Ohio	GS, NGL	IN, IP	M	
<b>Spottail shiner</b>	<i>Notropis hudsonius</i>	NOHU	Cyprinidae	Potomac	GS, NGL	OH	M	
<b>Silver shiner</b>	<i>Notropis photogenis</i>	NOPH	Cyprinidae	Mon, Ohio, UK		IN, IP	T	
<b>Swallowtail shiner</b>	<i>Notropis procne</i>	NOPR	Cyprinidae	Potomac	GS, NGL	IN, IP	M	
<b>Rosyface shiner</b>	<i>Notropis rubellus</i>	NORU	Cyprinidae	WV	RG, NGL	IN, IP	I	
<b>New River shiner</b>	<i>Notropis scabriceps</i>	NOSC	Cyprinidae	UK	GS, NGL	IN, IP	I	
<b>Sand shiner</b>	<i>Notropis stramineus</i>	NOST	Cyprinidae	Ohio, UK	LSR	OH	M	
<b>Telescope shiner</b>	<i>Notropis telescopus</i>	NOTE	Cyprinidae	None	GS, NGL	IN, IP	M	
<b>Mimic shiner</b>	<i>Notropis volucellus</i>	NOVO	Cyprinidae	Mon, Ohio, UK		IN, IP	M	
<b>Channel shiner</b>	<i>Notropis wickliffi</i>	NOWI	Cyprinidae	Ohio		IN, IP	M	
<b>Mountain madtom</b>	<i>Noturus eleutherus</i>	NOEL	Ictaluridae	Ohio	CAV	IN, IP	I	B
<b>Stonecat</b>	<i>Noturus flavus</i>	NOFU	Ictaluridae	Mon, Ohio, UK	CAV	IN, IP	M	B
<b>Margined madtom</b>	<i>Noturus insignis</i>	NOIN	Ictaluridae	Potomac, UK	CAV	IN, IP	M	B
<b>Brindled madtom</b>	<i>Noturus miurus</i>	NOMU	Ictaluridae	Ohio, UK	CAV	IN, IP	M	B
<b>Northern madtom</b>	<i>Noturus stigmosus</i>	NOSG	Ictaluridae	Ohio	CAV	IN, IP	I	B
<b>Rainbow trout</b>	<i>Oncorhynchus mykiss</i>	ONMY	Salmonidae	None		IP	I	G, C
<b>Cheat minnow</b>	<i>Parahinichthys bowersi</i>	PABO	Cyprinidae	Mon		IN, IP	M	B
<b>Yellow perch</b>	<i>Perca flavescens</i>	PEFL	Percidae	None		IP	M	G
<b>Logperch</b>	<i>Percina caprodes</i>	PECA	Percidae	Mon, Ohio	GS, CGS	IN, IP	M	B
<b>Channel darter</b>	<i>Percina copelandi</i>	PECO	Percidae	Ohio	RG	IN, IP	I	B
<b>Gilt darter</b>	<i>Percina evides</i>	PEEV	Percidae	Ohio	GS, CGS	IN, IP	I	B
<b>Longhead darter</b>	<i>Percina macrocephala</i>	PEMA	Percidae	Ohio	RG, NGL	IN, IP	I	B

<b>Blackside darter</b>	<i>Percina maculata</i>	PEMC	Percidae	Mon, Ohio	GS, CGS	IN, IP	M	B
<b>Stripeback darter</b>	<i>Percina notogramma</i>	PENO	Percidae	None	CGS	IN, IP	I	B
<b>Sharpnose darter</b>	<i>Percina oxyrhynchus</i>	PEOX	Percidae	Mon, Ohio, UK	GS	IN, IP	I	B
<b>Slenderhead darter</b>	<i>Percina phoxocephala</i>	PEPH	Percidae	Ohio	GS	IN, IP	I	B
<b>Roanoke darter</b>	<i>Percina roanoka</i>	PERO	Percidae	None	GS, CGS	IN, IP	M	B
<b>Dusky darter</b>	<i>Percina sciera</i>	PESC	Percidae	Ohio	GS	IN, IP	M	B
<b>River darter</b>	<i>Percina shumardi</i>	PESH	Percidae	Ohio	GS, CGS	IN, IP	M	B
<b>Trout-perch</b>	<i>Percopsis omiscomaycus</i>	PEOM	Percopsidae	Ohio		IN, IP	M	B
<b>Suckermouth minnow</b>	<i>Phenacobius mirabilis</i>	PHMI	Cyprinidae	Ohio	GS, NGL	OH	M	B
<b>Kanawha minnow</b>	<i>Phenacobius teretulus</i>	PHTE	Cyprinidae	UK	RG, NGL	OH	I	B
<b>Bluntnose minnow</b>	<i>Pimephales notatus</i>	PINO	Cyprinidae	WV	CAV	MO, OH	T	
<b>Fathead minnow</b>	<i>Pimephales promelas</i>	PIPR	Cyprinidae	Ohio	CAV	MO, OH	T	
<b>Bullhead minnow</b>	<i>Pimephales vigilax</i>	PIVI	Cyprinidae	Ohio	CAV	MO, OH	M	
<b>Paddlefish</b>	<i>Polyodon spathula</i>	POSP	Polydontidae	Ohio	NGL	MO, OH	I	G
<b>White crappie</b>	<i>Pomoxis annularis</i>	POAN	Centrarchidae	Mon, Ohio		IP	T	G
<b>Black crappie</b>	<i>Pomoxis nigromaculatus</i>	PONI	Centrarchidae	Mon, Ohio		IP	M	G
<b>Flathead catfish</b>	<i>Pylodictis olivaris</i>	PYOL	Ictaluridae	Mon, Ohio, UK	CAV	IP	M	G, B
<b>Eastern Blacknose dace</b>	<i>Rhinichthys atratulus</i>	RHAT	Cyprinidae	WV	GS, CGS	MO, OH	T	B
<b>Longnose dace</b>	<i>Rhinichthys cataractae</i>	RHCA	Cyprinidae	WV	CGS	IN, IP	M	B, C
<b>Western blacknose dace</b>	<i>Rhinichthys obtusus</i>	RHOB	Cyprinidae	Ohio	CGS	MO, OH	T	B, C
<b>Brown trout</b>	<i>Salmo trutta</i>	SATR	Salmonidae	None	CGS	IP	I	G, C
<b>Brook trout</b>	<i>Salvelinus fontinalis</i>	SAFO	Salmonidae	WV	CGS	IP	I	G, C
<b>Sauger</b>	<i>Sander canadensis</i>	SACA	Percidae	Mon, Ohio		IP	M	G, B
<b>Walleye</b>	<i>Sander vitreus</i>	SAVI	Percidae	Mon, Ohio		IP	M	G, B
<b>Creek chub</b>	<i>Semotilus atromaculatus</i>	SEAT	Cyprinidae	WV	GS	IP	T	
<b>Fallfish</b>	<i>Semotilus corporalis</i>	SECO	Cyprinidae	Potomac	RG	IP	M	G



<b>Torrent sucker</b>	<i>Thoburnia rathoeca</i>	THRH	Catostomidae	Potomac, UK	RG, CGS, NGL	MO, OH	I	B
-----------------------	---------------------------	------	--------------	----------------	-----------------	--------	---	---

9

## Appendix B

Metrics evaluated for fish IBI development for biomonitoring regions in West Virginia and their brief descriptions. Each metric had a richness component (R\_) and a proportion (P\_) that was evaluated separately. The expected response of each metric was determined using expert knowledge and consisted of positive (+) metrics that decreased with increases in stress and negative (-) metrics that increased with increases in stress.

<b>Metric</b>	<b>Expected Response</b>	<b>Description</b>
Richness	+	Richness
Native	+	Native Status
Game	+	Classified Game fish from WV DNR
RGS	+	Rock and gravel spawning
GSS	+	Gravel and sand spawning
LSR	+	Lithophilic spawning
NGL	+	Non-guarding lithophilic spawning
MO	+	Macro-omnivore
IN	+	Invertivore
IP	+	Invertivore-Piscivore
ISEAT	+	Invertivore-Piscivore minus creek chub (SEAT)
Benthic	+	Benthic
Benthic_CACO	+	Benthic minus white sucker (CACO)
Cottid	+	Cottidae
Cyprinid	+	Cyprinidae
Cyprinid_BNDSEAT	+	Cyprinidae Family minus blacknose dace (RHOB & RHAT) and creek chub (SEAT)
Cyprinid_N	+	Native Cyprinidae
Cyprinid_NBNDSEAT	+	Native Cyprinidae Family minus blacknose dace (RHOB & RHAT) and creek chub (SEAT)
BND_CACO_SEAT	-	Blacknose Dace (RHOB & RHAT), white sucker (CACO), and creek chub (SEAT)
OH	-	Omnivore-Herbivore
OH_CAAN	-	Omnivore-Herbivore minus central stoneroller (CAAN)
OH_CAAN_CACO	-	Omnivore-Herbivore minus central stoneroller (CAAN) and white sucker (CACO)
Cold	+	Cold water specialists
Cold_SATR_ONMY	+	Cold water specialists minus brown (SATR) and rainbow (ONMY) trout
GameC	+	Reduced list of game fish
OH_NG	-	Non-game omnivore-herbivore
IBenthicNG	+	Benthic and non-game invertivore-piscivore
INonGameNB	+	Non-game and non-benthic invertivore-piscivore
DMS	+	Darter-madtom-sculpins
Percidae	+	Family Percidae
Sunfish	+	Sunfish (Family Centrarchidae)
Catfish	+	Family Ictaluridae

<b>Metric</b>	<b>Expected Response</b>	<b>Description</b>
Catostomidae	+	Family Catostomidae
McC_CGS	+	Clean Gravel Spawning (Mc Cormick)
CGS_RGS	+	Clean Gravel & Rock-gravel Spawning
CavitySpawn	+	Cavity Spawning
Fish2.DEP	+	Fish minus tolerant
Native2.DEP	+	Native Status minus tolerant
RGS2.DEP	+	Rock and gravel spawning minus tolerant
GSS2.DEP	+	Gravel and sand spawning minus tolerant
LSR2.DEP	+	Lithophilic spawning minus tolerant
NGL2.DEP	+	Non-guarding lithophilic spawning minus tolerant
IP2.DEP	+	Invertivore-Piscivore minus tolerant
Benthic2.DEP	+	Benthic minus tolerant
Cyprinid2.DEP	+	Family Cyprinidae minus tolerant
Cyprinid_N2.DEP	+	Native Family Cyprinidae minus tolerant
OH2.DEP	-	Omnivore-Herbivore minus tolerant
Cold2.DEP	+	Cold water specialist minus tolerant
Game2.DEP	+	Game fish minus tolerant
DMS2.DEP	+	Darter-madtom-sculpins minus tolerant
Tol.DEP	-	Tolerant
Mod.DEP	+	Moderate Tolerance
Int.DEP	+	Intolerant
Tol_Benthic.DEP	-	Tolerant Benthics
Int_Benthic.DEP	+	Intolerant Benthics
Tol_Cyprinid.DEP	-	Tolerant Family Cyprinidae
Int_Cyprinid.DEP	+	Intolerant Family Cyprinidae
Int_RGS.DEP	+	Intolerant Rock-gravel spawning
Int_GSS.DEP	+	Intolerant gravel-sand spawning
Int_LSR.DEP	+	Intolerant lithophilic spawning
Int_NGL.DEP	+	Intolerant non-guarding lithophilic spawning
McC_CGS2.DEP	+	Clean gravel spawning minus tolerant (Mc Cormick)
CGS_RGS2.DEP	+	Clean gravel and rock-gravel spawning minus tolerant
CavitySpawn2.DEP	+	Cavity Spawning minus tolerant
SW-Trophic	+	Shannon-Weaver Trophic Diversity Index

## Appendix C

Metric specific responses and results from the step-wise selection process. Metrics listed also passed a range test. Each set of metrics evaluated for each region are presented in separate tables.

**Table 1:** Spearman correlation coefficients for all metrics in the Mon CA-RV region with surface mining (%), development (%), grassland (%), agriculture (%), forest (%), structure density (#/km<sup>2</sup>; SD), specific conductance (SPC), and pH. Discrimination efficiency (%; DE) and 25<sup>th</sup> and 75<sup>th</sup> percentiles (reference distribution) were also calculated for each metric. Table is sorted by descending discrimination efficiency.

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
P_IP_BenthicNG	-0.352	-0.199	-0.158	-0.195	0.244	-0.011	-0.591	0.350	93.33	0.34	0.47
P_Benthic2.DEP	-0.363	-0.208	-0.167	-0.202	0.253	-0.010	-0.575	0.314	93.33	0.35	0.47
Adj.P_Benthic	-0.324	-0.194	-0.106	-0.183	0.182	-0.142	-0.597	0.442	90.00	0.74	1.20
Adj.R_McC_CGS	-0.304	-0.303	-0.198	-0.300	0.306	-0.150	-0.601	0.320	90.00	0.84	1.18
P_DMS	-0.337	-0.161	-0.151	-0.160	0.210	0.035	-0.544	0.381	90.00	0.21	0.38
P_DMS2.DEP	-0.337	-0.161	-0.151	-0.160	0.210	0.035	-0.544	0.381	90.00	0.21	0.38
Adj.P_Benthic_CACO	-0.351	-0.234	-0.139	-0.213	0.222	-0.149	-0.609	0.417	86.67	0.73	1.20
Adj.P_CavitySpawn2.DEP	-0.321	-0.181	-0.142	-0.134	0.199	-0.005	-0.564	0.359	86.67	0.62	1.38
Adj.R_Fish2	-0.310	-0.235	-0.201	-0.254	0.300	-0.015	-0.494	0.335	86.67	0.84	1.12
Adj.R_IP2	-0.312	-0.232	-0.199	-0.251	0.298	-0.014	-0.493	0.337	86.67	0.85	1.13
Adj.R_Benthic	-0.323	-0.214	-0.106	-0.236	0.237	-0.001	-0.502	0.438	86.67	0.85	1.14
Adj.R_Fish2.DEP	-0.302	-0.197	-0.118	-0.211	0.225	0.032	-0.543	0.354	86.67	0.89	1.11
P_McC_CGS	-0.223	-0.327	-0.189	-0.247	0.250	-0.176	-0.516	0.390	86.67	0.35	0.51
Adj.R_Native2	-0.277	-0.193	-0.169	-0.209	0.263	0.076	-0.512	0.338	83.33	0.74	1.21
Adj.P_McC_CGS2	-0.195	-0.256	-0.185	-0.204	0.252	-0.063	-0.460	0.224	83.33	0.61	1.36
Adj.R_IP_SEAT	-0.287	-0.122	-0.080	-0.146	0.162	0.106	-0.476	0.364	83.33	0.87	1.11
Adj.R_IP2.DEP	-0.307	-0.198	-0.123	-0.208	0.228	0.037	-0.548	0.354	83.33	0.89	1.13
P_Benthic2	-0.364	-0.220	-0.193	-0.200	0.265	-0.044	-0.578	0.312	83.33	0.23	0.46
Adj.R_CGS_RGS	-0.320	-0.250	-0.145	-0.238	0.260	0.006	-0.518	0.365	80.00	0.91	1.18
Adj.P_McC_CGS2.DEP	-0.184	-0.252	-0.185	-0.220	0.260	-0.050	-0.451	0.224	80.00	0.66	1.43
Adj.R_IP	-0.306	-0.156	-0.102	-0.182	0.193	0.064	-0.450	0.370	80.00	0.84	1.12
Adj.R_Benthic_CACO	-0.325	-0.170	-0.077	-0.209	0.207	0.070	-0.545	0.428	80.00	0.85	1.11

Adj.R_Mod	-0.291	-0.192	-0.153	-0.160	0.237	0.061	-0.424	0.322	80.00	0.78	1.20
Adj.R_Native2.DEP	-0.298	-0.186	-0.117	-0.197	0.222	0.090	-0.529	0.351	76.67	0.78	1.20
Adj.P_Native2	-0.201	-0.295	-0.352	-0.274	0.363	-0.028	-0.441	0.221	76.67	0.83	1.17
Adj.P_Native2.DEP	-0.202	-0.282	-0.321	-0.266	0.343	-0.002	-0.430	0.215	76.67	0.84	1.13
Adj.P_Fish2	-0.220	-0.295	-0.366	-0.271	0.368	-0.064	-0.418	0.207	76.67	0.84	1.17
Adj.P_IP2	-0.220	-0.295	-0.366	-0.271	0.368	-0.064	-0.418	0.207	76.67	0.84	1.17
Adj.P_Tol	0.074	0.071	0.145	0.125	-0.156	0.196	-0.047	0.386	76.67	0.65	1.31
Adj.P_CGS_RGS	-0.252	-0.352	-0.230	-0.284	0.306	-0.076	-0.448	0.471	76.67	0.85	1.12
Adj.P_CGS_RGS2	-0.199	-0.294	-0.227	-0.256	0.304	-0.035	-0.437	0.269	76.67	0.59	1.30
Adj.P_Fish2.DEP	-0.218	-0.282	-0.337	-0.265	0.350	-0.034	-0.407	0.197	76.67	0.88	1.13
Adj.P_IP2.DEP	-0.219	-0.283	-0.338	-0.265	0.350	-0.037	-0.408	0.198	76.67	0.86	1.13
Adj.P_IN	-0.225	-0.239	-0.282	-0.149	0.249	0.123	-0.361	0.321	73.33	0.86	1.16
Adj.P_Tol.DEP	0.055	0.053	0.102	0.110	-0.127	0.151	-0.066	0.404	73.33	0.70	1.26
Adj.R_IN	-0.298	-0.141	-0.110	-0.149	0.181	0.158	-0.440	0.404	73.33	0.75	1.14
Adj.R_Mod.DEP	-0.252	-0.176	-0.104	-0.158	0.206	0.180	-0.477	0.359	73.33	0.81	1.13
P_Mod.DEP	-0.165	-0.262	-0.288	-0.223	0.309	0.000	-0.464	0.180	73.33	0.39	0.67
Adj.P_CavitySpawn	-0.327	-0.118	-0.056	-0.019	0.101	0.125	-0.471	0.327	70.00	0.69	1.29
Adj.R_FISH	-0.318	-0.127	-0.076	-0.156	0.162	0.090	-0.438	0.388	70.00	0.82	1.16
Adj.R_LSR	-0.297	-0.157	-0.113	-0.189	0.203	0.053	-0.431	0.373	70.00	0.77	1.17
Adj.R_Cyprinid	-0.324	-0.185	-0.098	-0.195	0.190	0.034	-0.437	0.408	70.00	0.68	1.13
P_Mod	-0.185	-0.230	-0.303	-0.185	0.291	-0.011	-0.408	0.192	70.00	0.32	0.54
Adj.P_CGS_RGS2.DEP	-0.213	-0.257	-0.151	-0.214	0.230	0.066	-0.403	0.324	66.67	0.58	1.28
Adj.R_CyprinidN	-0.330	-0.180	-0.089	-0.177	0.168	0.029	-0.427	0.384	63.33	0.68	1.20
Adj.P_RGS2.DEP	-0.114	-0.103	-0.198	-0.207	0.296	0.160	-0.351	0.180	63.33	0.40	1.19
Adj.R_Native	-0.313	-0.116	-0.061	-0.137	0.141	0.119	-0.419	0.379	60.00	0.78	1.18
R_IP_NonGameNB	-0.237	-0.153	-0.142	-0.187	0.225	0.059	-0.416	0.324	60.00	2.00	5.00
SW_TROPHIC	-0.281	-0.108	-0.018	-0.073	0.098	0.109	-0.443	0.432	60.00	2.76	3.24
Adj.P_MO	-0.232	-0.071	0.074	0.010	-0.018	0.013	-0.421	0.452	56.67	0.47	1.56
P_Catostomidae	-0.114	0.027	0.050	-0.001	0.039	0.110	-0.329	0.363	56.67	0.02	0.08
Adj.P_Cyprinid_NBNDSEAT	-0.192	-0.181	-0.143	-0.129	0.203	0.241	-0.315	0.360	53.33	0.74	1.26
Adj.P_RGS	-0.182	-0.198	-0.179	-0.180	0.256	0.265	-0.314	0.383	53.33	0.57	1.25

Adj.P_Cyprinid_BNDSEAT	-0.191	-0.179	-0.141	-0.128	0.202	0.243	-0.314	0.358	53.33	0.74	1.27
Adj.P_NGL	-0.217	0.021	0.084	-0.032	0.012	0.100	-0.276	0.437	50.00	0.57	1.35
P_Cyprinid	0.005	-0.200	-0.022	-0.081	0.048	-0.008	-0.215	0.446	50.00	0.47	0.70
Adj.P_IP_SEAT	-0.121	-0.215	-0.325	-0.192	0.279	0.068	-0.245	0.124	43.33	0.86	1.15
R_Game	-0.201	-0.023	-0.024	-0.046	0.057	0.095	-0.295	0.285	43.33	2.00	5.00
R_GSS	-0.176	-0.059	0.107	-0.024	-0.036	-0.008	-0.396	0.317	43.33	2.00	3.00
R_Game2	-0.217	-0.112	-0.085	-0.123	0.142	-0.013	-0.358	0.263	43.33	2.00	4.00
Adj.P_CyprinidN	-0.006	-0.181	0.016	-0.064	0.010	-0.011	-0.176	0.409	40.00	0.82	1.13
Adj.P_LSR	0.003	-0.144	-0.022	-0.088	0.029	0.006	-0.179	0.445	40.00	0.78	1.21
Adj.P_Tol_Cyprinid	-0.029	-0.033	0.075	0.023	-0.048	0.071	-0.226	0.499	40.00	0.60	1.43
Adj.P_Tol_Cyprinid.DEP	-0.050	-0.033	0.080	0.024	-0.050	0.073	-0.232	0.499	40.00	0.68	1.31
Adj.P_GSS	-0.021	-0.043	0.120	0.012	-0.095	-0.068	-0.265	0.317	40.00	0.58	1.32
P_GameC	-0.064	-0.110	-0.193	-0.043	0.141	0.026	-0.167	0.206	40.00	0.01	0.04
Adj.P_IP_NonGameNB	-0.004	-0.106	0.032	0.020	-0.043	0.112	-0.183	0.390	30.00	0.50	1.29
Adj.P_IP	0.024	-0.136	-0.217	-0.118	0.150	0.040	-0.138	0.086	30.00	0.90	1.12
P_Game2.DEP	-0.114	-0.049	-0.098	-0.053	0.112	0.116	-0.151	0.257	30.00	0.02	0.04
R_Tol.DEP	-0.241	0.044	0.066	0.024	-0.029	0.195	-0.293	0.380	30.00	3.50	6.00
Adj.P_BND_CACO_SEAT	-0.060	-0.127	0.061	-0.025	-0.018	-0.122	-0.330	0.308	26.67	0.31	1.39
P_Game2	-0.031	-0.008	-0.089	-0.078	0.108	0.106	-0.080	0.157	23.33	0.02	0.05
P_Sunfish	-0.024	0.037	0.002	0.031	-0.044	0.207	0.024	0.172	23.33	0.00	0.02
Adj.P_OH_CAAAN	-0.180	-0.129	0.021	0.039	0.001	0.072	-0.334	0.496	20.00	0.60	1.36
P_Game	0.011	-0.008	-0.109	-0.056	0.076	0.141	-0.030	0.144	20.00	0.02	0.05
R_Tol_Cyprinid.DEP	-0.309	-0.057	0.002	-0.052	0.058	0.159	-0.373	0.436	20.00	3.00	4.00
Adj.R_Tol_Benthic	-0.288	-0.118	0.017	-0.135	0.116	0.065	-0.460	0.380	16.67	0.86	1.18
R_Tol	-0.246	-0.010	0.032	-0.042	0.029	0.171	-0.352	0.412	16.67	5.00	8.00
R_Tol_Cyprinid	-0.282	-0.075	-0.017	-0.069	0.069	0.149	-0.379	0.463	16.67	3.00	4.00
Adj.P_OH	-0.206	-0.092	0.040	0.012	0.007	0.075	-0.371	0.495	13.33	0.55	1.47
Adj.P_OH_NG	-0.205	-0.099	0.036	0.008	0.012	0.070	-0.370	0.497	13.33	0.55	1.47
Adj.P_OH_CAAAN_CACO	-0.193	-0.174	0.003	0.011	0.035	0.066	-0.346	0.480	10.00	0.50	1.44
Adj.P_Tol_Benthic.DEP	-0.231	-0.099	-0.005	-0.077	0.058	-0.097	-0.465	0.468	10.00	0.45	1.72
P_Native	0.019	-0.097	-0.058	-0.024	-0.003	0.084	-0.177	0.227	10.00	0.94	1.00

---

R_OH	-0.272	-0.017	0.036	-0.029	0.020	0.145	-0.357	0.400	10.00	2.00	4.00
R_OH_NG	-0.270	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	10.00	2.00	4.00
P_FISH	-0.121	-0.199	-0.174	-0.122	0.158	0.082	-0.409	0.534	0.00	1.00	1.00
P_Tol_Benthic	-0.252	-0.089	0.037	-0.084	0.045	-0.026	-0.458	0.505	0.00	0.16	0.34

---

**Table 2:** Spearman correlation coefficients for all metrics in the Ohio CA region with surface mining (%), development (%), grassland (%), agriculture (%), forest (%), structure density (#/km<sup>2</sup>; SD), specific conductance (SPC), and pH. Discrimination efficiency (%; DE) and 25<sup>th</sup> and 75<sup>th</sup> percentiles (reference distribution) were also calculated for each metric. Table is sorted by descending discrimination efficiency.

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.R_DMS	<b>-0.516</b>	-0.084	<b>0.331</b>	0.264	<b>0.311</b>	<b>0.430</b>	<b>-0.379</b>	-0.074	<b>95.83</b>	0.61	1.30
Adj.R_Benthic2.DEF	<b>-0.508</b>	-0.055	<b>0.315</b>	0.242	<b>0.300</b>	<b>0.462</b>	<b>-0.343</b>	-0.052	<b>95.83</b>	0.66	1.32
Adj.R_DMS2.DEF	<b>-0.516</b>	-0.084	<b>0.331</b>	0.264	<b>0.311</b>	<b>0.430</b>	<b>-0.379</b>	-0.074	<b>95.83</b>	0.61	1.30
Adj.R_IP_BenthicNG	<b>-0.496</b>	-0.029	<b>0.336</b>	0.277	0.271	<b>0.470</b>	<b>-0.334</b>	-0.016	<b>91.67</b>	0.66	1.28
R_CavitySpawn	<b>-0.528</b>	0.315	<b>0.315</b>	<b>0.315</b>	<b>0.315</b>	<b>0.315</b>	<b>0.315</b>	<b>0.315</b>	<b>91.67</b>	2.00	5.00
Adj.P_CavitySpawn	<b>-0.394</b>	-0.079	<b>0.301</b>	0.296	0.215	<b>0.325</b>	<b>-0.331</b>	-0.096	87.50	0.76	1.18
Adj.R_GSS	<b>-0.473</b>	-0.133	<b>0.342</b>	0.163	0.263	0.291	<b>-0.344</b>	-0.166	83.33	0.62	1.27
Adj.R_Benthic_CACO	<b>-0.482</b>	-0.025	<b>0.369</b>	0.241	0.248	<b>0.433</b>	<b>-0.348</b>	-0.047	83.33	0.74	1.26
Adj.R_IP2.DEF	<b>-0.474</b>	-0.046	0.293	0.225	0.296	<b>0.461</b>	<b>-0.316</b>	-0.076	83.33	0.60	1.33
Adj.R_Mod.DEF	<b>-0.478</b>	-0.043	0.276	0.219	0.283	<b>0.430</b>	<b>-0.306</b>	-0.114	83.33	0.71	1.24
Adj.R_IN	<b>-0.462</b>	-0.054	<b>0.323</b>	0.226	0.259	<b>0.450</b>	<b>-0.349</b>	-0.120	79.17	0.62	1.39
Adj.R_IP	<b>-0.471</b>	-0.057	0.292	0.204	0.298	<b>0.431</b>	-0.296	-0.093	79.17	0.61	1.31
Adj.R_IP_SEAT	<b>-0.473</b>	-0.048	0.287	0.206	<b>0.304</b>	<b>0.448</b>	-0.290	-0.095	79.17	0.56	1.32
Adj.R_Percidae	<b>-0.491</b>	-0.062	<b>0.340</b>	0.242	0.293	<b>0.438</b>	-0.349	-0.087	79.17	0.53	1.36
Adj.R_Richness2.DEF	<b>-0.482</b>	-0.040	<b>0.316</b>	0.229	0.291	<b>0.474</b>	-0.325	-0.102	79.17	0.58	1.30
Adj.R_Native2.DEF	<b>-0.469</b>	-0.042	<b>0.312</b>	0.210	0.277	<b>0.453</b>	-0.331	-0.114	79.17	0.60	1.33
Adj.R_RGS	<b>-0.446</b>	-0.029	<b>0.332</b>	0.239	0.250	<b>0.427</b>	-0.352	-0.065	70.83	0.65	1.33
Adj.R_Benthic	<b>-0.484</b>	-0.010	<b>0.406</b>	0.268	0.230	<b>0.461</b>	-0.327	-0.020	70.83	0.68	1.24
R_Tol_Benthic.DEF	<b>-0.233</b>	0.123	<b>0.413</b>	0.152	0.009	<b>0.345</b>	-0.151	0.023	70.83	1.50	2.00
Adj.Richness	<b>-0.466</b>	-0.035	<b>0.352</b>	0.212	0.251	<b>0.422</b>	-0.307	-0.103	66.67	0.65	1.29
Adj.R_Native	<b>-0.457</b>	-0.042	<b>0.348</b>	0.196	0.245	<b>0.402</b>	-0.312	-0.110	66.67	0.66	1.29
R_CGS_RGS	<b>-0.402</b>	0.024	<b>0.321</b>	0.290	0.218	<b>0.577</b>	-0.223	0.023	66.67	4.00	11.00
P_Percidae	-0.157	-0.054	0.029	0.182	0.180	<b>0.358</b>	0.035	0.124	66.67	0.09	0.19
Adj.R_NGL	<b>-0.397</b>	0.031	<b>0.363</b>	0.215	0.189	<b>0.459</b>	-0.289	-0.035	62.50	0.58	1.32
P_DMS	-0.125	-0.099	-0.017	0.166	0.191	0.280	0.044	0.154	62.50	0.18	0.27
Adj.R_NGL2.DEF	<b>-0.349</b>	0.060	<b>0.306</b>	0.223	0.171	<b>0.409</b>	-0.240	-0.002	58.33	0.65	1.70
Adj.P_IN	-0.221	-0.235	0.001	0.128	0.276	0.234	-0.192	-0.114	58.33	0.81	1.29



Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.P_IP_SEAT	-0.070	-0.185	-0.171	0.045	0.238	0.146	0.068	0.035	58.33	0.80	1.23
Adj.Proportion2.DEP	-0.074	-0.141	-0.099	0.119	0.207	0.241	0.006	0.055	58.33	0.80	1.21
Adj.P_Native2.DEP	-0.073	-0.139	-0.101	0.111	0.195	0.234	-0.039	0.040	58.33	0.82	1.22
P_IP	0.084	-0.151	<b>-0.340</b>	-0.096	0.198	-0.023	0.118	-0.026	58.33	0.58	0.80
P_IP2.DEP	-0.043	-0.068	-0.125	0.153	0.194	<b>0.340</b>	0.097	0.159	58.33	0.27	0.64
Adj.P_NGL	-0.257	-0.020	0.186	0.079	0.206	<b>0.369</b>	-0.092	0.020	54.17	0.58	1.43
Adj.P_RGS2.DEP	-0.136	-0.128	-0.007	0.095	0.187	0.212	-0.012	0.050	54.17	0.71	1.35
Adj.P_CGS_RGS2.DEP	-0.139	-0.148	-0.007	0.082	0.199	0.196	0.003	0.061	54.17	0.71	1.35
R_McC_CGS	-0.287	0.079	0.271	0.262	0.139	<b>0.543</b>	-0.064	0.135	54.17	2.50	5.00
R_RGS2.DEP	<b>-0.425</b>	-0.005	0.297	<b>0.308</b>	0.255	<b>0.572</b>	-0.256	-0.007	54.17	2.50	8.00
P_OH	-0.084	-0.084	-0.084	-0.084	-0.084	-0.084	-0.084	-0.084	54.17	0.20	0.42
P_OH_NG	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	54.17	0.20	0.42
P_IP_BenthicNG	-0.030	-0.040	-0.091	0.135	0.140	0.276	0.131	0.211	54.17	0.21	0.32
Adj.P_NGL2.DEP	-0.191	0.038	0.074	0.091	0.169	0.297	-0.065	0.018	50.00	0.32	2.24
Adj.P_Tol.DEP	0.022	0.192	0.177	-0.007	-0.209	-0.056	0.016	-0.018	50.00	0.81	1.17
Adj.P_Tol_Cyprinid.DEP	0.014	0.177	0.187	0.010	-0.205	-0.054	-0.042	-0.037	50.00	0.84	1.18
P_Catostomidae	-0.088	0.075	0.009	0.093	0.130	<b>0.386</b>	0.162	0.194	50.00	0.02	0.06
Adj.R_MO	<b>-0.340</b>	0.068	<b>0.446</b>	0.169	0.071	0.294	-0.266	-0.077	45.83	0.85	1.35
P_IP_NonGameNB	-0.020	-0.154	-0.154	-0.154	-0.154	-0.154	-0.154	-0.154	45.83	0.26	0.50
Adj.R_LSR	<b>-0.378</b>	-0.036	<b>0.335</b>	0.194	0.181	<b>0.344</b>	-0.244	-0.085	41.67	0.60	1.33
Adj.R_LSR2.DEP	<b>-0.341</b>	-0.022	<b>0.300</b>	0.183	0.183	<b>0.378</b>	-0.182	-0.058	41.67	0.48	1.44
Adj.P_RGS	-0.113	0.023	0.072	0.021	0.083	0.100	-0.017	0.028	41.67	0.64	1.38
Adj.P_McC_CGS2.DEP	-0.081	-0.016	-0.017	0.060	0.144	0.221	0.265	0.220	41.67	0.31	1.74
Adj.P_CGS_RGS	0.069	-0.042	0.062	0.001	-0.127	-0.083	-0.013	0.091	37.50	0.85	1.13
Adj.P_LSR2.DEP	-0.100	-0.011	0.009	0.182	0.130	0.282	0.034	0.183	37.50	0.03	1.46
R_CGS_RGS2.DEP	<b>-0.409</b>	-0.003	0.285	0.298	0.242	<b>0.577</b>	-0.216	0.026	37.50	2.50	8.50
Adj.R_Cyprinid	-0.299	-0.032	<b>0.330</b>	0.083	0.116	0.191	-0.237	-0.134	33.33	0.72	1.37
Adj.R_Cyprinid_BNDSEAT	-0.250	0.032	0.261	0.130	0.095	0.222	-0.149	-0.047	33.33	0.28	1.93
Adj.R_CyprinidN	<b>-0.301</b>	-0.047	<b>0.325</b>	0.075	0.118	0.178	-0.257	-0.141	33.33	0.72	1.37

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.R_Cyprinid_NBNDSEAT	-0.253	0.129	0.129	0.129	0.129	0.129	0.129	0.129	33.33	0.28	1.93
Adj.R_IP_NonGameNB	-0.215	0.002	0.002	0.002	0.002	0.002	0.002	0.002	33.33	0.53	1.26
Adj.P_GSS	-0.113	-0.079	0.071	-0.090	0.084	-0.136	-0.073	-0.239	33.33	0.54	1.25
Adj.P_Benthic	0.153	0.158	-0.054	0.068	-0.113	0.186	0.254	<b>0.301</b>	33.33	0.79	1.12
Adj.P_Benthic_CACO	0.169	0.159	-0.089	0.051	-0.108	0.184	0.251	0.298	33.33	0.80	1.14
Adj.P_Cyprinid_BNDSEAT	-0.188	0.054	0.219	0.077	0.053	0.096	-0.136	0.006	33.33	0.31	1.43
Adj.P_Cyprinid_NBNDSEAT	-0.190	0.080	0.080	0.080	0.080	0.080	0.080	0.080	33.33	0.31	1.43
R_Catostomidae	-0.261	0.133	0.278	0.220	0.082	<b>0.553</b>	-0.087	0.127	33.33	1.00	2.00
P_Cyprinid	-0.004	0.016	0.153	-0.068	-0.138	-0.213	-0.211	-0.240	33.33	0.63	0.75
P_CyprinidN	-0.012	0.011	0.159	-0.057	-0.139	-0.223	-0.247	-0.219	33.33	0.63	0.75
Adj.P_BND_CACO_SEAT	0.075	0.052	0.133	-0.030	-0.187	-0.170	-0.040	-0.061	29.17	0.38	1.19
P_OH_CAAAN	-0.162	-0.162	-0.162	-0.162	-0.162	-0.162	-0.162	-0.162	29.17	0.15	0.40
P_OH_CAAAN_CACO	-0.154	-0.154	-0.154	-0.154	-0.154	-0.154	-0.154	-0.154	29.17	0.15	0.32
P_LSR	0.006	0.006	-0.104	0.006	0.065	0.074	0.128	-0.011	20.83	0.38	0.71
P_MO	0.017	0.131	0.247	0.008	-0.227	-0.104	0.018	0.105	20.83	0.09	0.39
Adj.R_OH	<b>-0.344</b>	0.050	0.431	0.174	0.076	<b>0.311</b>	-0.263	-0.095	16.67	0.71	1.28
Adj.P_McC_CGS	0.241	0.153	-0.061	-0.025	-0.214	0.002	0.327	0.290	16.67	0.64	1.37
Adj.R_OH_CAAAN	<b>-0.376</b>	-0.044	<b>0.406</b>	0.161	0.117	<b>0.306</b>	-0.294	-0.133	12.50	0.74	1.25
Adj.R_OH_CAAAN_CACO	<b>-0.351</b>	-0.053	<b>0.358</b>	0.124	0.112	0.220	-0.289	-0.148	12.50	0.65	1.30
R_OH_NG	<b>-0.362</b>	0.100	<b>0.414</b>	0.252	0.123	<b>0.521</b>	-0.179	-0.016	12.50	2.00	4.50
R_McC_CGS2.DEP	-0.281	0.038	0.178	0.258	0.176	<b>0.541</b>	-0.042	0.132	12.50	1.00	4.00
Adj.R_BND_CACO_SEAT	-0.196	0.089	<b>0.349</b>	0.175	0.009	<b>0.336</b>	-0.039	0.079	8.33	0.74	1.37
Adj.R_Tol.DEP	<b>-0.330</b>	0.033	<b>0.375</b>	0.117	0.112	0.283	-0.233	-0.079	8.33	0.79	1.35
Adj.R_Tol_Cyprinid.DEP	-0.289	0.010	<b>0.360</b>	0.074	0.081	0.176	-0.227	-0.095	8.33	0.71	1.38
P_Native	0.024	-0.041	0.041	-0.120	-0.063	-0.158	-0.111	-0.129	8.33	1.00	1.00

**Table 3:** Spearman correlation coefficients for metrics in the Ohio and Mon WAP region with surface mining (%), development (%), grassland (%), agriculture (%), forest (%), structure density (#/km<sup>2</sup>; SD), conductivity (SPC), and pH. Discrimination efficiency (%; DE) and 25<sup>th</sup> and 75<sup>th</sup> percentiles indicated.

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.R_CyprinidN	-0.196	-0.198	-0.111	-0.239	0.271	-0.118	-0.153	-0.090	<b>92.59</b>	0.930	1.169
Adj.R_CavitySpawn2.DEP	-0.201	<b>-0.317</b>	-0.259	-0.235	<b>0.342</b>	-0.218	-0.234	-0.146	88.89	0.906	1.183
Adj.R_CyprinidN2.DEP	-0.127	-0.223	-0.132	-0.263	0.267	-0.119	-0.129	-0.040	88.89	0.822	1.560
Adj.R_DMS	-0.081	<b>-0.388</b>	-0.245	-0.193	0.297	<b>-0.301</b>	-0.176	-0.125	88.89	0.903	1.147
Adj.R_DMS2.DEP	-0.082	<b>-0.390</b>	-0.244	-0.194	0.297	<b>-0.302</b>	-0.176	-0.125	88.89	0.903	1.147
Adj.R_IP_BenthicNG	-0.195	<b>-0.380</b>	-0.215	-0.145	0.280	<b>-0.317</b>	-0.198	-0.141	88.89	0.928	1.142
Adj.R_NGL2.DEP	-0.140	<b>-0.301</b>	-0.152	-0.098	0.227	-0.232	-0.158	-0.150	88.89	0.758	1.305
Adj.P_CyprinidN2.DEP	-0.166	-0.258	-0.188	-0.204	0.265	-0.105	-0.191	0.087	85.19	0.343	1.789
Adj.P_Int_LSR.DEP	-0.152	-0.290	-0.250	-0.250	0.291	-0.214	-0.217	-0.137	85.19	0.405	1.805
Adj.R_CavitySpawn	-0.241	-0.250	-0.208	-0.197	0.295	-0.157	-0.235	-0.091	85.19	0.928	1.071
Adj.R_Int_Benthic.DEP	-0.184	<b>-0.332</b>	-0.223	-0.129	<b>0.317</b>	-0.254	-0.201	-0.195	85.19	0.812	1.162
Adj.R_Int_RGS.DEP	-0.064	<b>-0.360</b>	-0.220	-0.216	<b>0.317</b>	-0.263	-0.196	-0.058	85.19	0.791	1.159
Adj.R_RGS2.DEP	0.019	-0.292	-0.176	-0.221	0.248	-0.179	-0.157	-0.022	85.19	0.786	1.192
Adj.R_Cyprinid_BNDSEAT	-0.122	-0.198	-0.072	-0.199	0.237	-0.104	-0.122	-0.120	81.48	0.918	1.144
Adj.R_Cyprinid2.DEP	-0.018	-0.196	-0.072	-0.234	0.219	-0.083	-0.098	0.009	81.48	0.859	1.145
Adj.R_Int_LSR.DEP	-0.181	<b>-0.331</b>	-0.267	-0.254	<b>0.370</b>	-0.274	-0.194	-0.134	81.48	0.728	1.141
Adj.R_Benthic	-0.169	<b>-0.307</b>	-0.196	-0.158	0.275	-0.254	-0.209	-0.133	77.78	0.905	1.155
Adj.R_Benthic2.DEP	-0.202	<b>-0.359</b>	-0.222	-0.147	<b>0.307</b>	-0.299	-0.192	-0.223	77.78	0.812	1.188
Adj.R_RGS	0.021	-0.279	-0.171	-0.239	0.249	-0.174	-0.139	-0.048	77.78	0.815	1.181
Adj.P_Int_RGS.DEP	-0.014	-0.235	-0.163	-0.081	0.180	-0.097	-0.120	-0.178	74.07	0.626	1.432
Adj.R_Benthic_CACO	-0.190	-0.287	-0.160	-0.101	0.242	-0.225	-0.178	-0.176	74.07	0.874	1.188
Adj.R_CGS_RGS2.DEP	-0.117	<b>-0.350</b>	-0.215	-0.176	0.288	-0.281	-0.209	-0.112	74.07	0.840	1.150
Adj.R_Cyprinid	-0.105	-0.144	-0.036	-0.184	0.197	-0.062	-0.146	-0.064	74.07	0.895	1.115
Adj.R_IP_NonGameNB	-0.137	-0.216	-0.085	-0.220	0.257	-0.113	-0.171	-0.082	74.07	0.824	1.252
Adj.R_IP2.DEP	-0.213	-0.284	-0.146	-0.174	0.284	-0.209	-0.154	-0.150	74.07	0.814	1.188
Adj.R_LSR2.DEP	-0.103	<b>-0.370</b>	-0.247	-0.253	<b>0.352</b>	-0.293	-0.170	-0.156	74.07	0.691	1.147
Adj.R_NGL	-0.152	-0.274	-0.125	-0.124	0.216	-0.214	-0.153	-0.132	74.07	0.771	1.299

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.P_CavitySpawn2.DEP	-0.145	-0.109	-0.120	0.014	0.104	-0.122	-0.111	-0.178	70.37	0.570	1.105
Adj.R_CGS_RGS	-0.100	-0.295	-0.196	-0.195	0.280	-0.232	-0.225	-0.112	70.37	0.877	1.181
Adj.R_Int.DEP	-0.136	-0.293	-0.161	-0.161	0.290	-0.191	-0.209	-0.107	70.37	0.626	1.270
Adj.R_Mod.DEP	-0.208	-0.181	-0.059	-0.109	0.201	-0.155	-0.113	-0.147	70.37	0.827	1.275
Adj.R_Native2.DEP	-0.251	-0.294	-0.171	-0.170	<b>0.309</b>	-0.236	-0.171	-0.176	70.37	0.769	1.306
Adj.R_Percidae	-0.067	<b>-0.335</b>	-0.214	-0.129	0.248	<b>-0.303</b>	-0.150	-0.110	70.37	0.811	1.119
Adj.R_Richness2.DEP	-0.212	-0.265	-0.137	-0.159	0.284	-0.195	-0.162	-0.167	70.37	0.768	1.291
Adj.P_Cyprinid_NBNDSEAT	-0.140	-0.121	-0.049	-0.198	0.212	-0.050	-0.128	-0.013	66.67	0.927	1.161
Adj.P_Cyprinid2.DEP	-0.008	-0.184	-0.111	-0.096	0.157	-0.023	-0.169	0.166	66.67	0.301	1.747
Adj.P_Native2.DEP	-0.053	-0.152	-0.131	-0.091	0.108	-0.069	-0.177	-0.197	66.67	0.747	1.171
Adj.R_Cyprinid_NBNDSEAT	-0.189	-0.240	-0.133	-0.243	0.299	-0.148	-0.148	-0.148	66.67	0.641	1.225
Adj.R_GSS	-0.230	-0.242	-0.199	-0.176	0.295	-0.257	-0.163	-0.202	66.67	0.780	1.129
Adj.R_GSS2.DEP	-0.275	<b>-0.325</b>	<b>-0.310</b>	-0.179	<b>0.377</b>	<b>-0.327</b>	-0.223	-0.276	66.67	0.660	1.305
Adj.R_IN	-0.207	<b>-0.304</b>	-0.122	-0.144	0.266	-0.223	-0.176	-0.167	66.67	0.795	1.152
Adj.R_LSR	-0.083	<b>-0.304</b>	-0.181	-0.267	<b>0.301</b>	-0.242	-0.117	-0.121	66.67	0.802	1.174
P_Native	<b>-0.358</b>	-0.236	-0.245	-0.273	<b>0.319</b>	-0.276	-0.129	-0.157	66.67	1.000	1.000
Adj.P_CGS_RGS2.DEP	0.104	-0.179	-0.144	-0.143	0.139	-0.116	-0.140	-0.220	62.96	0.765	1.291
Adj.P_Int.DEP	0.086	-0.245	-0.214	-0.207	0.210	-0.209	-0.134	-0.074	62.96	0.549	1.738
Adj.P_IP2	-0.184	-0.213	-0.130	-0.190	0.160	-0.118	-0.099	-0.091	62.96	0.468	1.434
Adj.R_Int_NGL.DEP	-0.157	-0.268	-0.098	-0.039	0.196	-0.211	-0.165	-0.117	62.96	0.708	1.662
Adj.R_IP	-0.255	-0.247	-0.114	-0.108	0.245	-0.199	-0.156	-0.173	62.96	0.780	1.181
Adj.R_IP_SEAT	-0.267	-0.252	-0.117	-0.106	0.249	-0.199	-0.158	-0.176	62.96	0.758	1.199
P_CyprinidN	-0.146	-0.131	-0.110	-0.217	0.234	-0.070	-0.070	0.162	62.96	0.674	0.762
P_OH_CAAN	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	-0.035	62.96	0.142	0.301
Adj.P_Cyprinid_BNDSEAT	-0.058	-0.078	0.005	-0.113	0.139	0.000	-0.107	0.048	59.26	0.928	1.128
Adj.P_GSS2.DEP	-0.130	-0.220	-0.254	-0.145	0.247	<b>-0.321</b>	-0.115	-0.255	59.26	0.627	1.348
Adj.P_RGS	0.042	-0.146	-0.045	-0.166	0.149	-0.039	0.016	-0.022	59.26	0.731	1.220
Adj.R_Native	-0.276	-0.203	-0.094	-0.099	0.220	-0.175	-0.134	-0.168	59.26	0.839	1.163
Adj.R_Tol_Benthic.DEP	0.080	0.183	0.163	0.157	-0.201	0.177	0.019	0.188	59.26	0.830	1.298

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
P_Int_Benthic.DEP	0.019	-0.111	-0.140	0.017	0.066	-0.154	-0.074	-0.198	59.26	0.083	0.162
P_LSR	0.037	-0.141	-0.125	-0.194	0.191	-0.014	-0.030	0.099	59.26	0.623	0.708
Adj.P_Int_NGL.DEP	0.095	-0.153	-0.055	0.008	0.022	-0.193	-0.125	-0.051	55.56	0.553	1.338
Adj.P_IP_SEAT	-0.020	-0.117	-0.069	-0.016	0.041	-0.041	-0.076	-0.235	55.56	0.718	1.132
Adj.P_LSR2.DEP	0.146	-0.239	-0.208	-0.231	0.231	-0.102	-0.162	-0.018	55.56	0.654	1.535
Adj.P_RGS2.DEP	0.109	-0.143	-0.071	-0.106	0.103	-0.035	-0.100	-0.161	55.56	0.645	1.332
Adj.R_McC_CGS2.DEP	-0.053	-0.203	-0.182	-0.136	0.195	-0.213	-0.165	-0.137	55.56	0.752	1.188
Adj.Richness	-0.234	-0.166	-0.045	-0.068	0.173	-0.129	-0.124	-0.151	55.56	0.823	1.199
P_IP	0.025	-0.054	-0.088	-0.005	-0.003	-0.002	-0.030	-0.200	55.56	0.498	0.692
P_IP_NonGameNB	-0.024	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	55.56	0.247	0.471
P_OH	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	55.56	0.308	0.502
P_OH_CAAN_CACO	-0.062	-0.062	-0.062	-0.062	-0.062	-0.062	-0.062	-0.062	55.56	0.103	0.301
P_OH_NG	-0.024	0.018	0.018	0.018	0.018	0.018	0.018	0.018	55.56	0.308	0.502
Adj.P_BND_CACO_SEAT	0.034	0.084	0.023	0.052	-0.122	0.012	0.107	0.056	51.85	0.561	1.286
Adj.P_NGL2.DEP	0.120	-0.162	-0.115	-0.042	0.054	-0.196	-0.074	-0.109	51.85	0.665	1.306
Adj.P_Tol.DEP	-0.014	0.089	0.041	0.016	-0.040	-0.044	0.177	0.173	51.85	0.788	1.135
Adj.P_Tol_Cyprinid.DEP	0.021	0.009	-0.028	-0.074	0.043	-0.085	0.093	0.192	51.85	0.765	1.162
Adj.Proportion2.DEP	0.070	-0.085	-0.058	0.009	0.006	0.010	-0.166	-0.125	51.85	0.746	1.161
Adj.R_BND_CACO_SEAT	0.132	0.172	0.088	0.055	-0.145	0.095	-0.065	0.173	48.15	0.811	1.386
P_CGS_RGS	0.096	-0.143	-0.085	-0.149	0.107	-0.057	0.106	0.070	48.15	0.415	0.671
P_Cyprinid	-0.053	-0.075	-0.057	-0.126	0.157	-0.007	-0.059	0.190	48.15	0.674	0.787
Adj.P_GSS	-0.013	-0.032	-0.146	-0.049	0.050	-0.121	0.022	-0.026	44.44	0.732	1.076
Adj.P_IP2.DEP	0.043	-0.131	-0.077	-0.080	0.049	-0.047	-0.082	-0.196	44.44	0.648	1.192
P_DMS	0.171	-0.017	-0.048	-0.012	-0.034	-0.044	-0.063	-0.166	44.44	0.120	0.244
P_DMS2.DEP	0.171	-0.017	-0.048	-0.012	-0.034	-0.044	-0.063	-0.166	44.44	0.120	0.244
P_IP_BenthicNG	0.153	-0.036	-0.050	-0.002	-0.060	-0.071	-0.066	-0.132	44.44	0.173	0.287
SW_TROPHIC.cor	-0.097	-0.072	-0.025	0.041	0.009	-0.037	0.004	-0.225	44.44	2.720	2.982
Adj.P_IN	-0.014	-0.174	-0.049	-0.019	0.103	-0.049	-0.160	-0.141	40.74	0.810	1.137
Adj.P_McC_CGS2.DEP	0.196	-0.069	-0.096	-0.154	0.049	-0.070	-0.055	-0.053	40.74	0.538	1.336

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
P_Benthic2.DEP	0.158	-0.048	-0.068	-0.018	-0.043	-0.098	-0.063	-0.179	40.74	0.180	0.293
P_McC_CGS	0.116	-0.063	-0.033	-0.131	0.017	-0.025	0.125	0.092	40.74	0.204	0.334
P_Percidae	0.201	-0.014	-0.038	0.038	-0.012	-0.057	-0.057	-0.166	40.74	0.100	0.216
R_McC_CGS	0.007	-0.097	-0.125	-0.082	0.115	-0.112	-0.157	0.009	40.74	4.000	6.000
Adj.P_Benthic	0.115	-0.048	-0.041	-0.051	-0.044	-0.041	0.125	0.024	37.04	0.715	1.182
Adj.P_NGL	0.026	-0.028	0.041	-0.025	-0.056	-0.023	0.073	-0.030	37.04	0.704	1.120
Adj.P_Tol_Benthic.DEP	-0.014	-0.020	0.018	-0.012	-0.024	0.015	0.177	0.161	37.04	0.687	1.206
P_Benthic_CACO	0.074	-0.105	-0.063	-0.101	0.030	-0.068	0.097	-0.041	37.04	0.301	0.542
P_CavitySpawn	-0.074	0.058	0.054	-0.003	-0.049	-0.050	-0.127	-0.090	37.04	0.124	0.233
R_Catostomidae	-0.205	-0.160	-0.051	0.002	0.096	-0.160	-0.144	0.016	37.04	2.000	3.000
Adj.P_Catostomidae	-0.136	-0.039	0.022	-0.013	-0.040	-0.026	-0.014	-0.143	33.33	0.657	1.429
Adj.R_MO	-0.016	0.148	0.100	0.065	-0.098	0.092	0.008	0.098	33.33	0.822	1.150
P_Mod.DEP	0.055	0.059	0.029	0.155	-0.129	0.087	-0.136	-0.072	33.33	0.199	0.342
P_OH2.DEP	-0.005	-0.098	-0.063	0.049	0.089	-0.094	-0.124	0.009	25.93	0.000	0.089
R_Tol.DEP	-0.162	0.098	0.117	0.148	-0.109	0.032	-0.015	0.021	25.93	6.000	9.000
P_MO	-0.044	0.107	0.130	-0.030	-0.064	0.048	0.131	0.154	22.22	0.177	0.411
R_Tol_Cyprinid.DEP	-0.134	-0.014	0.026	0.032	0.024	-0.022	-0.166	0.038	22.22	5.000	6.000
P_Game	-0.141	0.075	0.072	0.194	-0.079	0.020	0.124	-0.148	11.11	0.001	0.030
R_OH	-0.074	0.066	0.088	0.009	-0.003	0.050	-0.050	-0.001	11.11	4.000	7.000
R_OH_CAAN	-0.087	0.066	0.093	0.021	-0.006	0.050	-0.061	-0.007	11.11	3.000	6.000
R_OH_CAAN_CACO	-0.076	0.056	0.064	0.031	-0.002	0.040	-0.044	-0.011	11.11	2.000	5.000
R_OH_NG	-0.080	-0.003	-0.003	-0.108	0.122	-0.020	-0.134	-0.040	11.11	4.000	6.000
R_Game	-0.171	0.090	0.081	0.198	-0.094	0.010	0.048	-0.046	3.70	1.000	4.000
Adj.P_GameC	-0.230	0.074	0.116	0.137	-0.036	0.067	0.129	-0.164	0.00	0.000	2.232
Adj.R_Game2.DEP	-0.284	0.002	0.049	0.004	0.095	0.015	-0.058	-0.194	0.00	0.000	1.658
P_Game2.DEP	-0.198	0.035	0.050	0.106	-0.009	0.020	0.071	-0.163	0.00	0.000	0.020
P_Sunfish	-0.195	-0.027	-0.021	0.121	0.044	-0.111	0.035	-0.188	0.00	0.000	0.010
R_GameC	-0.123	0.148	0.084	0.136	-0.104	0.035	0.047	0.021	0.00	0.000	2.000
R_OH2.DEP	-0.058	-0.100	-0.054	-0.037	0.138	-0.082	-0.085	-0.030	0.00	0.000	2.000

**Table 4:** Spearman correlation coefficients for metrics in the Upper Kanawha region with surface mining (%), development (%), grassland (%), agriculture (%), forest (%), structure density (#/km<sup>2</sup>; SD), conductivity (SPC), and pH. Discrimination efficiency (%; DE) and 25<sup>th</sup> and 75<sup>th</sup> percentiles indicated. Table is sorted by descending discrimination efficiency.

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.P_Cold	-0.010	-0.082	-0.021	-0.022	0.137	0.034	-0.123	-0.165	81.82	0.638	1.404
P_Int_RGS.DEP	-0.188	<b>-0.359</b>	-0.182	-0.147	<b>0.351</b>	<b>-0.363</b>	-0.255	-0.173	81.82	0.128	0.324
Adj.P_Cold_SATR_ONMY	-0.022	-0.107	-0.043	-0.030	0.128	0.000	-0.139	-0.193	72.73	0.370	1.494
P_Native2	-0.176	-0.258	-0.067	-0.041	0.274	-0.146	-0.102	0.192	72.73	0.216	0.368
P_IP_BenthicNG	-0.048	<b>-0.391</b>	<b>-0.316</b>	-0.168	0.189	<b>-0.390</b>	-0.100	0.049	72.73	0.192	0.396
Adj.R_IP_BenthicNG	-0.249	<b>-0.404</b>	<b>-0.376</b>	0.000	0.167	-0.294	-0.218	0.083	63.64	0.594	1.357
Adj.R_Int	<b>-0.340</b>	<b>-0.370</b>	-0.185	0.058	0.251	-0.123	<b>-0.336</b>	0.006	63.64	0.441	1.645
Adj.R_Int.DEP	<b>-0.354</b>	<b>-0.307</b>	-0.082	-0.007	0.287	-0.133	<b>-0.399</b>	0.056	63.64	0.555	1.464
Adj.R_Int_Benthic.DEP	<b>-0.319</b>	<b>-0.490</b>	<b>-0.333</b>	0.006	0.223	<b>-0.396</b>	<b>-0.326</b>	-0.058	63.64	0.708	1.210
Adj.P_Tol.DEP	0.155	<b>0.425</b>	0.271	<b>0.323</b>	<b>-0.418</b>	<b>0.400</b>	<b>0.318</b>	0.091	63.64	0.577	1.166
P_Cyprinid2	-0.127	-0.164	-0.026	-0.065	0.260	-0.121	-0.142	0.138	63.64	0.124	0.317
P_CyprinidN2	-0.119	-0.154	-0.014	-0.059	0.246	-0.113	-0.137	0.137	63.64	0.124	0.317
P_DMS	-0.128	<b>-0.439</b>	<b>-0.377</b>	-0.146	0.193	<b>-0.421</b>	-0.165	-0.006	63.64	0.108	0.370
P_RGS2.DEP	-0.043	<b>-0.340</b>	-0.279	<b>-0.314</b>	0.295	<b>-0.467</b>	-0.283	-0.028	63.64	0.233	0.484
P_Benthic2.DEP	-0.026	<b>-0.354</b>	-0.284	-0.167	0.159	<b>-0.378</b>	-0.084	0.061	63.64	0.192	0.396
P_CyprinidN2.DEP	-0.144	-0.167	-0.005	-0.045	0.249	-0.104	-0.142	0.159	63.64	0.124	0.317
P_DMS2.DEP	-0.128	<b>-0.439</b>	<b>-0.377</b>	-0.146	0.193	<b>-0.421</b>	-0.165	-0.006	63.64	0.108	0.370
P_Int_Benthic.DEP	-0.246	<b>-0.449</b>	-0.297	-0.118	0.292	<b>-0.491</b>	<b>-0.309</b>	-0.176	63.64	0.066	0.179
Adj.R_Native	-0.262	<b>-0.332</b>	<b>-0.323</b>	0.066	0.117	-0.236	-0.267	0.101	54.55	0.755	1.154
Adj.R_IN	-0.245	<b>-0.308</b>	<b>-0.300</b>	0.019	0.111	-0.234	-0.218	0.079	54.55	0.668	1.298
R_Benthic	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	-0.220	54.55	4.000	6.000
Adj.R_Benthic_CACO	-0.236	<b>-0.340</b>	<b>-0.343</b>	0.026	0.118	-0.254	-0.226	0.134	54.55	0.626	1.278
Adj.R_CGS_RGS	-0.214	<b>-0.316</b>	<b>-0.328</b>	-0.035	0.120	-0.288	<b>-0.340</b>	0.068	54.55	0.703	1.214
Adj.R_CGS_RGS2.DEP	-0.248	<b>-0.354</b>	<b>-0.302</b>	-0.060	0.170	<b>-0.303</b>	<b>-0.368</b>	0.002	54.55	0.732	1.224
Adj.P_Cyprinid2.DEP	-0.252	-0.204	-0.055	-0.063	<b>0.307</b>	-0.086	-0.163	0.052	54.55	0.744	1.536
R_CyprinidN	-0.221	-0.132	-0.124	0.156	0.043	-0.038	-0.205	0.226	54.55	3.000	5.000
P_Fish2	-0.165	-0.191	-0.014	-0.049	0.254	-0.081	-0.072	0.197	54.55	0.217	0.444

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
P_Tol	0.132	0.185	0.029	0.074	-0.232	0.101	0.072	-0.197	54.55	0.556	0.783
P_Percidae	-0.086	<b>-0.397</b>	<b>-0.388</b>	-0.293	0.205	-0.539	-0.281	-0.144	54.55	0.078	0.233
P_CGS_RGS2	-0.098	-0.119	0.017	-0.115	0.277	-0.099	-0.201	0.116	54.55	0.131	0.289
P_Native2.DEP	-0.114	<b>-0.396</b>	-0.230	-0.206	<b>0.307</b>	<b>-0.389</b>	-0.223	0.059	54.55	0.321	0.589
P_Int.DEP	-0.196	-0.196	0.024	-0.083	<b>0.319</b>	-0.081	-0.292	0.068	54.55	0.015	0.270
P_CGS_RGS2.DEP	-0.013	-0.313	-0.219	-0.298	<b>0.307</b>	<b>-0.411</b>	-0.265	-0.040	54.55	0.269	0.545
P_CavitySpawn2.DEP	-0.224	<b>-0.443</b>	-0.266	0.017	0.264	<b>-0.330</b>	-0.149	-0.039	54.55	0.072	0.243
Adj.R_FISH	-0.225	-0.217	-0.219	0.094	0.032	-0.140	-0.199	0.172	45.45	0.764	1.212
Adj.R_Benthic2	-0.211	-0.253	-0.206	0.098	0.071	-0.088	-0.043	0.204	45.45	0.645	1.619
Adj.R_Cyprinid	-0.241	-0.164	-0.173	0.149	0.025	-0.065	-0.214	0.158	45.45	0.791	1.136
Adj.R_Cyprinid_BNDSEAT	-0.240	-0.173	-0.152	0.151	-0.003	-0.052	-0.213	0.126	45.45	0.747	1.496
Adj.R_McC_CGS	-0.131	-0.248	<b>-0.306</b>	-0.032	0.079	-0.224	-0.283	0.067	45.45	0.646	1.455
Adj.R_Native2.DEP	<b>-0.347</b>	<b>-0.398</b>	<b>-0.320</b>	0.055	0.189	-0.270	<b>-0.311</b>	0.067	45.45	0.580	1.255
Adj.R_IP2.DEP	-0.276	-0.286	-0.233	0.015	0.119	-0.189	-0.257	0.059	45.45	0.721	1.294
Adj.R_Benthic2.DEP	-0.228	<b>-0.356</b>	<b>-0.349</b>	-0.008	0.122	-0.272	-0.205	0.088	45.45	0.541	1.281
Adj.R_McC_CGS2.DEP	-0.086	-0.263	-0.295	-0.111	0.077	-0.274	<b>-0.300</b>	-0.064	45.45	0.649	2.176
Adj.P_IN	-0.059	<b>-0.394</b>	-0.280	-0.293	<b>0.327</b>	-0.406	-0.228	0.010	45.45	0.671	1.372
Adj.P_Benthic_CACO	-0.016	0.079	0.019	0.142	-0.142	0.126	0.126	0.195	45.45	0.799	1.186
Adj.P_Catostomidae	<b>0.324</b>	-0.006	-0.149	-0.179	-0.022	-0.122	0.121	0.232	45.45	0.148	1.925
Adj.P_Fish2.DEP	-0.201	<b>-0.397</b>	-0.225	-0.229	<b>0.382</b>	<b>-0.317</b>	-0.270	-0.028	45.45	0.771	1.309
Adj.P_Mod.DEP	-0.114	<b>-0.348</b>	-0.257	-0.200	0.231	<b>-0.300</b>	-0.164	0.032	45.45	0.587	1.482
Adj.P_Tol_Cyprinid.DEP	0.135	<b>0.391</b>	0.243	<b>0.330</b>	<b>-0.408</b>	<b>0.369</b>	0.284	0.125	45.45	0.618	1.199
R_RGS	-0.237	-0.229	-0.274	0.065	0.072	-0.192	-0.235	0.203	45.45	3.000	6.000
R_RGS2.DEP	-0.271	-0.273	-0.273	0.053	0.117	-0.217	-0.258	0.169	45.45	2.000	5.000
P_Benthic2	-0.053	-0.157	-0.064	0.004	0.096	-0.079	0.083	0.290	45.45	0.042	0.177
P_Cyprinid_NBNDSEAT	-0.062	0.140	0.166	0.210	-0.173	0.198	0.132	<b>0.397</b>	45.45	0.129	0.362
P_Tol_Cyprinid	0.062	0.297	0.174	0.247	-0.262	0.328	0.150	-0.029	45.45	0.385	0.613
P_CGS_RGS	0.018	0.018	-0.031	-0.103	0.046	-0.040	-0.119	-0.019	45.45	0.598	0.780
P_CavitySpawn	-0.209	<b>-0.330</b>	-0.152	0.155	0.088	-0.222	-0.071	0.048	45.45	0.072	0.243



Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.R_Native2	<b>-0.328</b>	<b>-0.363</b>	-0.267	0.070	0.194	-0.186	-0.262	0.087	36.36	0.533	1.223
Adj.R_Game	-0.018	0.126	0.161	0.187	-0.177	0.094	0.000	0.259	36.36	0.000	1.445
Adj.R_NGL	-0.028	-0.183	-0.280	-0.007	0.028	-0.121	-0.115	0.101	36.36	0.616	1.393
Adj.R_IP	-0.224	-0.243	-0.201	0.011	0.084	-0.184	-0.215	0.092	36.36	0.673	1.263
Adj.R_Cyprinid2	<b>-0.308</b>	-0.223	-0.146	0.107	0.138	-0.088	-0.233	0.074	36.36	0.509	1.210
Adj.R_CyprinidN2	-0.297	-0.217	-0.141	0.106	0.136	-0.086	-0.231	0.073	36.36	0.468	1.233
Adj.R_Mod	-0.097	-0.132	-0.171	0.052	-0.018	-0.057	-0.021	0.198	36.36	0.614	1.702
Adj.R_Fish2.DEP	<b>-0.336</b>	<b>-0.304</b>	-0.237	0.070	0.121	-0.173	-0.265	0.094	36.36	0.590	1.248
Adj.R_Cyprinid2.DEP	<b>-0.361</b>	-0.212	-0.106	0.159	0.142	-0.036	-0.241	0.091	36.36	0.626	1.513
Adj.R_CyprinidN2.DEP	<b>-0.326</b>	-0.196	-0.083	0.159	0.120	-0.033	-0.218	0.094	36.36	0.615	1.614
Adj.P_NGL	0.000	-0.225	-0.288	-0.139	0.135	-0.166	-0.040	0.074	36.36	0.141	1.838
Adj.P_Cyprinid_BNDSEAT	-0.106	0.017	0.040	0.148	-0.067	0.133	0.054	0.248	36.36	0.716	1.365
Adj.P_OH	-0.083	0.269	0.162	<b>0.415</b>	<b>-0.301</b>	<b>0.385</b>	0.162	0.207	36.36	0.611	1.318
Adj.P_OH_NG	-0.084	0.268	0.161	<b>0.415</b>	<b>-0.300</b>	<b>0.385</b>	0.160	0.208	36.36	0.611	1.318
Adj.P_IP2.DEP	-0.084	-0.345	-0.194	-0.297	<b>0.338</b>	<b>-0.329</b>	-0.207	-0.030	36.36	0.686	1.422
Adj.P_Tol_Benthic.DEP	-0.005	0.245	0.120	0.239	-0.208	<b>0.320</b>	0.156	0.126	36.36	0.685	1.473
R_LSR	-0.180	-0.168	-0.253	0.049	0.067	-0.129	-0.195	0.250	36.36	3.000	6.000
R_MO	-0.192	-0.081	-0.140	0.252	-0.118	0.033	-0.069	0.186	36.36	2.000	4.000
R_Percidae	-0.237	<b>-0.330</b>	<b>-0.305</b>	-0.033	0.150	<b>-0.300</b>	-0.275	0.046	36.36	1.000	2.000
R_Catostomidae	0.134	-0.024	-0.197	0.003	-0.092	-0.067	0.000	<b>0.302</b>	36.36	1.000	1.000
R_CavitySpawn2.DEP	<b>-0.358</b>	-0.464	<b>-0.348</b>	0.096	0.213	<b>-0.319</b>	-0.241	0.024	36.36	1.000	2.000
P_Native	0.079	0.051	0.062	-0.024	-0.136	-0.079	0.190	0.166	36.36	0.586	0.795
P_RGS	-0.028	-0.035	-0.093	-0.054	-0.041	-0.101	-0.057	0.230	36.36	0.233	0.536
P_IP2	-0.028	-0.107	0.026	-0.116	0.205	-0.080	-0.012	0.240	36.36	0.131	0.343
P_Cyprinid	0.021	0.252	0.233	0.182	-0.097	<b>0.319</b>	0.074	-0.013	36.36	0.573	0.796
P_Mod	-0.080	-0.079	-0.048	0.048	0.052	0.011	0.066	0.278	36.36	0.007	0.154
P_McC_CGS2.DEP	0.099	-0.019	-0.110	-0.133	-0.017	-0.136	-0.065	0.164	36.36	0.007	0.199
SW_TROPHIC	0.086	-0.093	-0.097	-0.050	0.022	-0.176	-0.101	0.072	36.36	2.360	3.068
Adj.R_Fish2	-0.310	-0.263	-0.180	0.064	0.146	-0.090	-0.211	0.117	27.27	0.514	1.331

Metric	Surface Mining	Development	Grassland	Agriculture	Forest	SD	SPC	pH	DE	25th	75th
Adj.R_IP2	-0.162	-0.224	-0.173	-0.029	0.118	-0.160	-0.211	0.103	27.27	0.788	1.500
Adj.R_IP_SEAT	-0.237	-0.242	-0.197	0.011	0.092	-0.167	-0.215	0.041	27.27	0.708	1.289
Adj.R_GameC	-0.036	0.100	0.133	0.171	-0.112	0.074	-0.032	0.269	27.27	0.000	1.632
Adj.R_Game2	0.022	0.058	0.077	0.101	-0.104	0.028	-0.060	0.261	27.27	0.000	1.501
Adj.R_IP_NonGameNB	-0.161	-0.165	-0.145	-0.036	0.111	-0.157	-0.266	0.008	27.27	0.644	1.423
Adj.R_CGS_RGS2	-0.228	-0.232	-0.118	0.025	0.131	-0.114	-0.290	0.068	27.27	0.691	1.549
Adj.R_Game2.DEP	0.009	0.036	0.055	0.077	-0.086	-0.008	-0.099	0.277	27.27	0.000	1.545
Adj.R_Mod.DEP	-0.240	-0.178	-0.196	0.114	-0.017	-0.104	-0.131	0.104	27.27	0.501	1.318
Adj.P_MO	-0.096	0.258	0.174	<b>0.407</b>	-0.283	<b>0.390</b>	0.166	0.191	27.27	0.616	1.321
Adj.P_IP_SEAT	-0.075	<b>-0.326</b>	-0.191	-0.298	<b>0.337</b>	<b>-0.322</b>	-0.251	-0.086	27.27	0.682	1.412
Adj.P_Benthic	-0.014	0.082	0.022	0.133	-0.134	0.130	0.161	0.183	27.27	0.752	1.253
Adj.P_BND_CACO_SEAT	0.166	0.035	0.019	-0.119	0.023	-0.046	0.046	-0.294	27.27	0.436	1.262
R_RGS2	-0.264	-0.161	-0.145	0.129	0.078	-0.052	-0.187	0.226	27.27	1.000	3.000
R_LSR2	-0.245	-0.213	-0.185	0.079	0.122	-0.103	-0.213	0.250	27.27	1.000	3.000
R_Cold_SATR_ONMY	-0.155	-0.199	-0.113	0.020	0.161	-0.072	-0.237	0.014	27.27	1.000	2.000
R_DMS	<b>-0.332</b>	<b>-0.370</b>	<b>-0.326</b>	0.066	0.163	-0.265	-0.214	0.132	27.27	1.000	3.000
R_Tol_Cyprinid	-0.094	-0.021	-0.126	0.145	-0.099	-0.002	-0.075	<b>0.311</b>	27.27	2.000	3.000
R_CavitySpawn	<b>-0.338</b>	<b>-0.365</b>	-0.261	0.176	0.105	-0.193	-0.163	0.134	27.27	1.000	2.000
R_LSR2.DEP	-0.260	-0.211	-0.225	0.068	0.121	-0.113	-0.199	0.223	27.27	1.000	4.000
R_DMS2.DEP	<b>-0.332</b>	<b>-0.370</b>	<b>-0.326</b>	0.066	0.163	-0.265	-0.214	0.132	27.27	1.000	3.000
P_GSS	0.041	-0.081	-0.096	-0.093	0.137	-0.121	0.032	-0.110	27.27	0.073	0.312
P_RGS2	-0.163	-0.163	-0.085	-0.120	0.244	-0.158	-0.195	0.177	27.27	0.022	0.270

**CHAPTER 2: LANDSCAPE BASED MODELS FOR IDENTIFYING LEAST  
IMPACTED STREAMS IN CENTRAL APPALACHIAN WATERSHEDS**

## Abstract

In heavily impacted regions, such as the central Appalachians, finding marginally impacted or reference streams to sample for bioassessment programs can be difficult. As a consequence, reference sites often are biased towards small headwater catchments on federally or state protected lands, and extrapolation to the broader region may be unwarranted. Our objective in this study was to use land-cover data to identify streams across a broad range of natural contexts (drainage area, elevation, distance to large mainstem) that have a high probability of serving as a reference site for bioassessment. Boosted Regression Tree models were developed using an existing sample of previously defined reference sites and anthropogenic landscape variables in order to predict the probability of being in reference condition for all wadeable stream segments in the study region. Combined with in-stream water quality measures, the probability was then used as a flexible selection criterion for the evaluation of potential reference sites. The analysis of current reference sites produced a strong Boosted Regression Tree model (ROC=0.998; CV ROC=0.978) and contained a variety of cumulative and local landscape variables. Cumulative percentage of forest, surface mining, and development accounted for the majority of the variance explained. Other important variables included measures of local percentages of forest and development and structure density in addition to cumulative densities of structures and surface mining permits and percentage of grassland. Depending on the fish biomonitoring region, we were able to include an additional 16 – 57 reference sites spanning a wide range of drainage areas using a probability threshold of 90%. Within the Ohio CA region, there is a large presence of surface mining activities, resulting in the majority of larger streams having elevated levels of mining above what would be expected of a traditional reference stream. In order to select larger streams representative of this region, a least disturbed reference condition approach was utilized by reducing the probability criterion for larger streams in this region. The addition of these reference sites increased the maximum drainage area from 85 to 357 km<sup>2</sup> in the Mon CA-RV, 60 to 250 km<sup>2</sup> in the Ohio CA, 143 to 295 km<sup>2</sup> in the Ohio-Mon WAP, and 150 to 383 km<sup>2</sup> in the Upper Kanawha biomonitoring regions. These methods allow increased levels of human disturbance in larger

streams and rivers in order to generate accurate models of expected condition as a function of stream size for bioassessment development. Using flexible criteria for larger streams encouraged the selection of these streams in the reference pool, not only increasing the range of drainage areas for bioassessment but also generating a more holistic picture of least disturbed reference condition in remote or highly disturbed landscapes. The hierarchical classification of reference condition probabilities demonstrates geographical clustering of high quality segment-, HUC 12-, and HUC 8-scale watersheds within and between biomonitoring regions. The nested structure of watershed conditions provides a unique framework for evaluating reference condition and the influence of anthropogenic stressors on aquatic communities at multiple spatial scales.

**Keywords: Boosted Regression Tree models, Reference Condition, Bioassessment**

## **1.0 Introduction**

Aquatic monitoring programs collect and analyze data in order to evaluate the health and resource status of the nation's waterways. However, in order to assess the resource conditions, benchmark values are needed for comparison. Numeric criteria are typically generated based on the distribution of abiotic (i.e., physical or chemical criteria) or biotic (i.e., indicator species or multimetric indices) measurements. These numeric criteria, or thresholds, are generally developed using a reference condition approach (Karr and Chu, 1998). A reference condition is developed by evaluating the condition gradient of every stream or river ranging from pristine and free from human influence to completely degraded and potentially devoid of life. Ideally, the reference condition is based on pristine sites. However, these stream types constitute a very small proportion of all the streams in the United States. Since pristine streams are relatively rare outside of headwater systems, a "minimally" or "least" disturbed definition of the reference condition is used. A "minimally disturbed" condition is typically reserved for streams or rivers where there is very little evidence of human disturbance (Stoddard et al., 2006). However, a "least disturbed" condition is applied to streams and rivers that have the best condition given the current status of the human-dominated landscape (Stoddard et al., 2006).

Several types of bioassessment programs use the reference condition approach in order to quantify anthropogenic effects on freshwater ecosystems. Index of Biotic Integrity (IBI;Karr, 1986), Biological Condition Gradient (BCG; Davies and Jackson, 2006), and River InVertebrate Prediction and Classification Systems (RIVPACS; Wright et al., 1984) modeling are all methods of assessing freshwater ecosystems. Regardless of the underlying statistical techniques used to determine stream health (e.g., narrative criteria, index values, or species presence predictions), they all use a reference condition as the foundation of their programs. The reference condition approach allows sites to serve as replicates instead of comparing to the pre-impact condition of a stream, which can be unknown in most regions and is not important in assessing stressors or current ecosystem condition (Reynoldson et al., 1997).

The reference condition approach takes a pool of sites minimally exposed to anthropogenic stressors in order to account for the natural variation exhibited in the model organism (Reynoldson et al., 1997; Bailey et al., 1998; Bowman and Somers, 2005). Generally, test (non-reference) sites are exposed to some degree of anthropogenic stressors and the deviation away from the reference condition can be a quantifiable indicator of ecological stress under this approach. Regional reference conditions have been used to evaluate both indicators of biological condition and for predictive modeling. These reference sites are used to quantitatively predict the expected conditions at non-reference sites across ranges of natural conditions. For example, the influence of drainage area on diversity of stream fishes is widely known (Minshall et al., 1985), and reference sites can be used to account for this natural variation in order to get an accurate representation of how the fish community is responding to stressors regardless of stream size (Fausch et al., 1984). Large sample sizes of reference sites covering large regional extents are important for modeling expected conditions because they contain wide variations in fish communities as well as a wide range in natural conditions allowing for increased sensitivity in detecting responses to stressors (Riseng et al., 2006).

During reference site selection at the onset of a bioassessment program, it is important that the sites cover a wide range of natural variation as well as be exposed to the lowest levels of anthropogenic impacts for that region. Stringent reference site selection criteria are utilized in order to increase detection of anthropogenic impacts. However, the selection of sites can be difficult when the reference condition, or least disturbed condition, has not been defined or is extremely rare in heavily disturbed regions. In order to combat some of the ambiguity surrounding reference site selection, bioassessment programs generate a list of acceptable and unacceptable measurements of ecosystem stress in order for a site to be added to the reference pool. For example, the Mid-Atlantic Highlands Index of Biotic Integrity (MAH-IBI; McCormick et al., 2001) uses three different reference definitions (least, moderately, and most restrictive) to determine reference sites used in setting metric scoring thresholds and expectations. The least restrictive was based on chemical criteria and Rapid Bioassessment Protocol habitat measures with

watershed land-use, road density, and watershed condition criteria added as you progress to more restrictive reference sites. Similarly, the benthic macroinvertebrate genus-level index of most probable stream status (GLIMPSS; Pond et al., 2012) uses a strict set of water quality, local habitat measures (Rapid Bioassessment Protocol), and upstream point and nonpoint source discharges to select high quality reference sites (Pond et al., 2012). However, in regions where very few sites meet all criteria best professional judgment was used to determine the inclusion of a site into the reference pool, even though it failed one criterion, generating a reference pool of least and minimally disturbed sites.

Reliability in bioassessment programs assumes that the expected conditions based on the reference pool can accurately predict conditions and are unbiased (Hawkins et al., 2010b). However, the restrictive nature of screening criteria and strict focus on local habitat measures for reference site selection coupled with a probabilistic survey design, may unintentionally bias reference sites towards small streams. The high cost of ecological monitoring has led to the need for highly efficient sampling designs that are scientifically defensible, can characterize the regions ecological condition, meet project goals, and affect policy decisions. Even though probabilistic based sample design programs are able to make direct inferences about the target population and eliminate bias associated with site based criteria sampling high amounts of variability between samples can occur. This high variability can be influenced by geographic heterogeneity such as land use, stream size, and ecoregion.

Landscape and land-use data are frequently being used to screen sites by quantifying the relative human influence at the watershed scale (Whittier et al., 2007; Wang et al., 2008; Angradi et al., 2009; Yates and Bailey, 2010). A landscape-scale approach allows for the combination of multiple stressors in order to generate a stressor gradient. During reference site selection, focus has typically been placed on the number and types of point-source pollution discharges such as industrial and municipal wastewater systems. However, the importance of non-point source pollution as a result of complex interactions from industrial, agricultural, commercial, and residential land use practices is becoming more evident as major contributors of stream health (Merriam et al., 2011, 2013). Measurements of water quality can be used as



surrogates for overall anthropogenic land-use practices, however sampling can be costly, time consuming, and may require previous knowledge of potential stress. The use of remotely sensed landscape data would help inform, or reduce the need for, extensive water quality monitoring in reference site selection.

In most regions of West Virginia, historic data were not collected prior to anthropogenic impacts. Consequently, determining the health of stream is dependent on the reference condition. Candidate reference sites are evaluated based on past and present land-use disturbances followed by a field assessment examining all chemical, habitat, biological information as well as professional judgment from trained personnel. Like other bioassessment programs, West Virginia uses different levels of reference condition in which Level I sites meet all criteria, while Level II sites fail to meet one or more criteria by a narrow margin. Finally, Level III sites are reserved for larger rivers and streams ( $> \sim 150 \text{ km}^2$ ) which due to their size, may have some levels of point source discharges but generally meet RBP habitat and water quality criteria. Like other bioassessment programs that use probabilistic sampling designs and restrictive selection criteria, West Virginia reference sites may not capture a wide range of natural variation occurring on the landscape and may under represent larger rivers and streams. For example, the current reference sites used for fish bioassessment has an average drainage area approximately half ( $32.73 \text{ km}^2$ ) the size of the average wadeable stream in the state ( $61.15 \text{ km}^2$ ) and does not cover the full range of drainage areas ( $7 - 150 \text{ km}^2$ ) that the bioassessment program covers ( $7 - 400 \text{ km}^2$ ). This bias towards smaller streams in the reference pool may have important impacts on the expectations extrapolated to larger streams and rivers throughout the state.

In this study we demonstrate how we derived landscape based stressor gradient for wadeable streams of West Virginia and how we used those gradients to select region specific reference sites for bioassessment development. The goal of this study was to evaluate landcover/land-use of the current reference sites used in West Virginia in order to generate a list of candidate reference sites, representing the full range of natural variation seen on the landscape, and describe the distribution of stream conditions using spatial modeling techniques across the study region. Specifically, we are using known, pre-defined, reference

sites to identify the reference condition in remote, or under sampled, regions of West Virginia. This model will be used to address our 3 research objectives: 1) to identify potential reference sites within a pool of previously sampled sites; 2) to identify targeted sampling priorities along a continuum of predicted stream condition; and 3) to evaluate the distribution of stream conditions at multiple spatial scales.

## **2.0 Methods**

In order to identify candidate reference sites, we used the landscape condition observed at current reference sites to model the reference condition probability using Boosted Regression Tree modeling. The model was then used to predict reference condition probability statewide. Modeled probabilities, in conjunction with water quality data (i.e., specific conductance and pH), were used to determine the final reference pool for West Virginia streams. Finally, the hierarchical spatial structure of stream condition based on the reference condition probability was mapped to evaluate the geographical arrangement of reference conditions throughout the state.

### *2.1 Landuse/landcover data*

Landscape characteristics for all 1:24,000 segment-level watersheds (SLWs) within the state of West Virginia were quantified using spatial analysis functions in ArcGIS ArcMap 10.0 (Environmental Systems Research Institute, Redlands, California). In conjunction with flow tables, cumulative measures of several landscape attributes for each segment-level watershed (Strager et al., 2009) were quantified at the local (i.e., within individual SLWs) and cumulative (i.e., all SLWs upstream of a given sampling location) scale for each SLW. Land cover classifications were derived from the 2009 and 2010 National Agriculture Imagery Program (NAIP) orthophotography with a 1-meter pixel resolution at a scale of 1:10,000. Land cover types included open water, forest, grass and agricultural lands, and barren development. The mining-permit boundaries layer developed by the Technical Applications in GIS (TAGIS) office within WVDEP enabled further differentiation into mining-related open water (i.e., slurry

impoundments), barren (i.e., active mine lands) and grasslands (i.e., reclaimed mine lands) from non-mining land cover. All mining-related cover classes were summed into a measure of total surface mining. The density ( $\#/km^2$ ) of surface mining, underground mining, sewage, and septic national pollution discharge elimination system (NPDES) permits were calculated from data obtained from WVDEP. The West Virginia Statewide Addressing and Mapping Board structures layer (WV SAMB) was used to calculate the density of residential and commercial structures ( $\#/km^2$ ).

Natural landscape variables, such as drainage area ( $km^2$ ), elevation (m), and swim distance (km), were also generated in a GIS framework. Cumulative drainage area was generated in a similar fashion to cumulative landscape variables listed above; however we used segment level watershed area in conjunction with the stream flow tables in the stream accumulation tool. Digital Elevation models were generated for West Virginia in 2003 by the United States Geological Survey (USGS) and the Statewide Addressing and Mapping Board (SAMB) at a 3-meter ( $1/9^{th}$  arc-second) resolution. We used this data to summarize the maximum elevation for each segment level watershed used in this study. Swim distance was defined as the in-stream distance from the outflow of any segment level watershed to the inflow of a downstream segment level watershed of  $200 km^2$  or greater (Hitt and Angermeier, 2011). Measures of swim distance were generated for every segment level watershed in West Virginia using accumulated flow tables and the swim distance calculator located in the stream accumulation tool provided by NRAC and West Virginia University. Because, some river systems originate outside of West Virginia, the analysis was reduced to wadeable streams and rivers that have less than 1% of their cumulative drainage area outside of the state boundary.

## *2.2 Current reference pool*

A set of 848 stream segments were sampled using electrofishing (e.g., backpack, boat, parallel wires) from 2000 – 2013 for fish biomonitoring from a variety of data sources (e.g. West Virginia Department of Environmental protection, West Virginia Department of Natural Resources, and West Virginia University). A portion of the sites were evaluated after sampling, by West Virginia Department of

Environmental Protection (WVDEP) to be considered a reference site based on strict water quality standards used for environmental regulation along with habitat assessment (Table 2). Each site passing the water quality and habitat standards was evaluated in terms of local landscape disturbances and known point-source discharges upstream and final inclusion into the reference pool was determined by best professional judgment. The reference pool (N=54) used for model building spans across the Ohio, Monongahela, and Upper Kanawha drainages as well as over 3 different ecoregions (i.e. Central Appalachian, Western Allegheny Plateau, and the Ridge and Valley).

The spatial positioning of the reference sites leads to great variability in the natural and anthropogenic landscapes surrounding these streams making it difficult to set a statewide reference condition. Based on previous research done in this region (Chapter 1), the state has been divided into five fish biomonitoring regions based on fish community composition and natural landscape variation in order to get a more accurate representation of the reference condition at a smaller spatial extent. The five regions include: Ohio Central Appalachians (Ohio CA), Ohio-Monongahela Western Allegheny Plateau (Ohio-Mon WAP), Monongahela Ridge and Valley (Mon CA-RV), Upper Kanawha (UK) and the Potomac (not included in analysis due to low sample sizes) (Figure 1).

### *2.3 Boosted Regression Tree Models*

Boosted regression tree (BRT) modeling is an additive regression model technique which combines iteratively fit simple regression trees using a boosting algorithm (Elith et al., 2008). The incorporation of a boosting algorithm when fitting successive models increases predictive performance allowing modeling technique to be well suited for modeling ecological and landscape datasets with a complex covariance structure (Elith et al., 2008). Statewide boosted regression tree models were constructed relating presence of a reference site and both cumulative and local (i.e. segment level watershed scale) anthropogenic land cover and use data. Initial models were used to predict probability of the presence of a reference site to each wadeable (7 – 400 km<sup>2</sup>) segment level watershed (N=16,118) in the study region. Initial BRT models were constructed statewide based on the anthropogenic landscape differences between reference

(N=54) and stressed sites (N=80). The BRT model was built using a bag fraction of 0.50 with a learning rate of 0.001. A tree complexity of two was used to allow for one-way interactions between variables used in the model in order to more accurately predict in a complex landscape structure. Model predictive performance was assessed by calculating the mean model deviance and cross-validation predictive deviance from ten folds of the data (Elith et al., 2008). Predictor variables were removed if the removal of the variable did not increase the cross-validation error rate. Finally, relative importance of each variable retained in the simplified model was generated using an out-of-bag procedure outlined in Elith et al. (2008). Only anthropogenic variables were used in order to reduce any bias towards specific stream sizes or elevations that may have occurred during the best professional selection process of the initial reference sites. The model was then used to predict to all wadeable segment level watersheds statewide. Landscape modeled reference probabilities were then mapped to show the pattern of condition across the region hierarchically at the SLW, HUC 12, and HUC 8 scales. Conditions were established for both the HUC 12 and HUC 8 catchment scales by averaging the probability of reference for all wadeable segment-level watersheds within each scale. We recognize the importance of region specific reference sites however; due to sample size limitations within each region we were not able to generate region specific models. BRT models were constructed in the R environment (version 2.15.0; R Development Core Team, Vienna, Austria) with functions in package *gbm* and those provided by Elith et al. (2008).

#### *2.4 Site Selection*

In order to reduce the cost associated with large scale field collections, candidate reference sites were selected from all the sites that have been previously sampled in the past 10 years. A pool of candidate reference sites were generated based on probabilities ( $\geq 0.90$ ) determined by the BRT model. These sites were then elevated to a reference site status if they met the following water quality criteria: 1) conductivity  $< 500$  and 2)  $6 < \text{pH} < 9$  in order to ensure the highest quality sites were being selected. If no large ( $> 150 \text{ km}^2$ ) sites were selected using the methods listed previously, then the probability of reference criteria was relaxed in a stepwise fashion, in 0.10 increments for every  $100 \text{ km}^2$  increase in drainage area,

for each region until representatives across all stream sizes were selected. Reducing the probability of reference criteria across drainage areas allowed for the inclusion of least disturbed sites given stream size in all regions so the reference pool is not biased towards small streams. Additional sites were identified as field collection priorities and were selected along a drainage area and stream condition (i.e. probability of reference) continuum in order to guarantee a gradient of all anthropogenic stressors and all stream sizes occur within the fish biomonitoring database. The reference pool within each region were compared based on drainage area, elevation, and swim distance before and after the inclusion of additional reference sites as well as to the average wadeable stream for that region.

### **3.0 Results**

#### *3.1 Boosted Regression Tree Models*

BRT fit 5,150 trees for the statewide model predicting the occurrence of reference sites and made use of six cumulative and three local land-use predictor variables. The model provided an adequate fit to the data with training receiver operating characteristic (ROC) curve score of 0.998 and a cross-validation ROC curve score of 0.978 (standard error 0.011). The most influential variable in the model was cumulative % forest (relative influence of 45.19) with cumulative % surface mining having the second highest relative influence (27.10) followed by cumulative % development (9.31). The remaining six variables: local % forest, cumulative % grassland, cumulative surface mine permit density, local % development, cumulative structure density, and local structure density, all had similar relative influences ranging from 3.22 to 2.75. Upon examination of the partial dependence plots for the final set of variables included in the model (Figure 2), the probability of a segment level watershed being a reference site increases with both local and cumulative % forest and decreases sharply with slight increases in the other anthropogenic variables. For example, probability of a reference site does not increase until local % forest reaches approximately 85%. The final model was then used to predict reference condition probabilities to all wadeable segment level watersheds (Figure 3) based on their anthropogenic landscape structure. All segment level watersheds (N=16,118) were evaluated for their probability of a reference site. Only

29.79% (N=4,822) of those segment level watersheds met the initial reference site criteria of having a predicted reference probability  $\geq 0.90$ , which spans approximately 759 different streams and rivers throughout the study area.

### *3.2 Reference Site Selection*

From sites previously sampled, 187 (22.05%) had probabilities of reference  $\geq 0.90$  in addition to measures of specific conductance and pH. These potential reference sites were further evaluated on their water quality measurements, in which 116 (13.68%) sites that had not previously been identified as a reference met all selection criteria. The addition of these sites increased the number of sites in the reference pool from 54 to 170 statewide. The addition of sites across drainage areas (Figure 4) has increased the drainage size range of sites that can be evaluated (7 – 393 km<sup>2</sup>) and has increased the average drainage area (61.00 km<sup>2</sup>) to more closely resemble the average wadeable stream in West Virginia (61.15 km<sup>2</sup>).

Upon analysis of reference sites within each region (Table 3), the number of reference sites in the Mon CARV region increased from 17 to 74 sites, which was the greatest increase in any of the regions. The addition of 57 sites resulted in a small increase in the average drainage area from 30.94 km<sup>2</sup> to 49.96 km<sup>2</sup>. However, the range of the drainage areas increased (8.07 – 357.6 km<sup>2</sup>) to more closely represent the range of stream segments in this region (7.01 – 399.6 km<sup>2</sup>). In this region, there is also high variation in stream temperatures due to rapid elevation changes. Adding these reference sites also increased the range of elevation of the reference pool (555 – 1432 m) as well as including sites that are close proximity to larger streams (i.e. swim distance of 0.00 km) as well as more remote systems (i.e. swim distance of 208 km).

The Ohio Central Appalachian biomonitoring region exhibited an increase of 16 additional reference sites selected for inclusion in the original reference pool (10 sites). We were unable to select any previously sampled large sites using the same criteria used in other regions. By reducing the selection criteria to a probability of being a reference site to 0.80, we were able to select two additional large reference sites

while still meeting strict water quality criteria. The addition of these sites increased the average drainage area of the reference pool from 25.70 km<sup>2</sup> to 52.41 km<sup>2</sup>, which more closely resembles the average stream segment (55.05 km<sup>2</sup>) in this region.

In the remaining two biomonitoring regions, Ohio Mon WAP and Upper Kanawha, we were able to add 23 and 20 additional reference sites to the original reference pools, respectively. In the Ohio Mon WAP region, we were able to dramatically increase the average drainage area (29.37 km<sup>2</sup>) of the small reference pool (N=9) to an average drainage area (69.84 km<sup>2</sup>) and range (7.38 – 295.6 km<sup>2</sup>) to more closely resemble stream segments in this region (drainage area average= 54.45 km<sup>2</sup>; range= 7.0 – 399.45 km<sup>2</sup>). The addition of reference sites also increased the variety of network positions to include sites adjacent to larger streams. The Upper Kanawha region exhibited similar patterns as the Ohio Mon WAP with increases in drainage areas from 44.92 to 81.41 km<sup>2</sup>.

### *3.3 Sampling Priority Selection*

Boosted regression tree model probability of reference site occurrence generated a gradient on which sampling sites can be selected. Upon examining the distribution of sites across drainage areas, gaps in the data distribution have been located that should be filled in order to get an accurate representation of the effects stressors have on stream fish communities (Figure 4). For example, there is a need for large reference sites in the Ohio Central Appalachian biomonitoring region. Using the results generated from the BRT modeling, we were able to select large stream segments that have a high probability of reference ( $\geq 0.90$ ; Table 4). Using these criteria, we were able to locate 16 segment level watersheds in this region with a high probability of reference and were between 150 and 250 km<sup>2</sup>, however these segments only occurred on two different streams. For streams larger than 250 km<sup>2</sup>, we had to reduce the probability criteria to 0.60 in order to pick up three additional segment level watersheds, which all occur on the same stream, for targeted sampling. Additional large streams with high probability of reference were also identified in the other regions (Table 4) in order to aid future sampling efforts. A total of 593 stream



segments in the other three regions, covering approximately 44 different streams, were selected as future sampling priorities for monitoring efforts (Table 4).

In addition to the need of large reference sites, all regions were lacking large (>150 km<sup>2</sup>) streams with an intermediate (0.20—0.60) probability of reference. Using the results generated from BRT for all segment level watersheds in the state, we are able to pinpoint 45 stream segments, on approximately seven streams, for future targeted sampling in the Ohio Central Appalachian region (Table 5). A total of 205 stream segments, covering approximately 29 streams, in the other 3 regions were identified as large streams of intermediate quality (Table 5).

### *3.4 Hierarchical classification of reference condition*

We mapped reference condition probability at the SLW-scale (Figure 3) to identify potential reference sites and to evaluate environmental quality across the study region. There were distinct patterns and differences in SLW-scale probabilities at multiple spatial scales. Outside of the Ohio CA region, most regions contain clusters of probability sites within one or two HUC 8 watersheds, typically occurring in the upper reaches of catchments. However, there are isolated stream reaches within each biomonitoring region that are of both high and poor quality.

When we averaged reference condition probabilities from the SLW scale to the HUC 12 scale (Figure 5), the spatially clustering of high and low quality catchments across the study area become more apparent. For example, the upper catchments of the Elk, Cheat, and Tygart HUC 8 watersheds have high numbers of HUC 12 watersheds representing the best stream conditions in the study region. Consequently, these three HUC 8 watersheds make up the entirety of the Mon CA-RV biomonitoring region. Similarly, the Upper Kanawha biomonitoring region has clustering of HUC 12 watersheds with high reference condition probabilities occurring in the Greenbrier and Gauley HUC 8 watersheds. The Ohio-Mon WAP region also exhibits concentrations of HUC 12 watersheds with high reference condition probabilities in the Little Kanawha and Middle Ohio North HUC 8 watersheds. Outside of the lower portion of the Elk

watershed, the Ohio CA region displays many isolated HUC 12 watersheds of high reference condition probability.

When we averaged the SLW-scale reference condition probabilities to the HUC 8-scale (Figure 5), similar patterns of geographic clustering of high quality watersheds are apparent. At this scale, HUC 8 watersheds with clustering of high quality HUC 12 watersheds in their upper reaches typically have higher average reference condition probabilities. All biomonitoring regions have at least one, or a portion of one, HUC 8 watershed with a high average reference condition probability. Regardless of biomonitoring region, there are high-quality segment-level watersheds found in even the poorest quality HUC 12 watershed, nested in a range of HUC 8 watershed conditions.

#### **4.0 Discussion**

Anthropogenic landscape disturbances within a watershed can influence instream habitat and water quality conditions, ultimately impacting biological assemblages and overall stream health (Allan, 2004; Wang et al., 2008; Petty et al., 2010; Merriam et al., 2013). Most stream health assessments focus on instream habitat, water quality, and biological condition and their relationships with point-source pollution or specific land-use practices. However, with the expanded use and precision of remotely sensed land-use data, the use of the complex landscape structure in determining stream health and identifying reference streams and evaluating overall stream condition is becoming more feasible (Wang et al., 2008; Yates and Bailey, 2010; Merovich et al., 2013; Hughes et al., 2015).

We found that a large number of wadeable segment level watersheds within each region were in good condition (29.92%) based on their reference condition probabilities. Once larger streams were evaluated for their reference condition, the difference in the stressor gradients between regions became more apparent. The negative influence that small levels of surface mining had on the reference condition probability made it very rare that a large stream in the Ohio Central Appalachian region could be evaluated as a reference site since most have some measure of surface mining in their watershed. By

reducing the reference probability criteria, we are recognizing the need for regionally appropriate reference sites and that elevated levels of surface mining may need to be acceptable for generating a least disturbed reference condition in this region, especially for larger streams.

Our approach to identifying reference sites and establishing a stressor gradient has advantages relative to other methods. Most aquatic bioassessment programs summarize results based on site-specific data (i.e., point data). Our analysis attributes each point to a segment-level watershed, linking the cumulative landscape to monitoring data. Analyzing the data in a spatial framework then allows predictions to be made and evaluated for unsampled portions of the study regions. Probability based survey designs are useful since they improve accuracy and precision of regional assessments and are able to make generalized estimates of stream health throughout a region. However, predictions or generalizations about unsampled areas cannot be made using a randomized sampling design. In addition, our approach allowed us to generate *a priori* conditions for all wadeable stream reaches across each region. We were able to use the landscape condition at all wadeable segment level watersheds to establish a least-disturbed reference condition in each region rather than quantifying the effects of one or two dominant disturbances. Merovich et al. (2013) developed landscape-based BRT models to evaluate stream condition in watersheds heavily influenced by coal mining. Their evaluation of conditions at broad spatial scales provided insight on prioritizing sites for restoration based on hierarchical stream conditions. Evaluating our reference condition probabilities at multiple spatial scales allows us to make informed decisions about reference site selection and impacts broader spatial scales may have on aquatic community structure.

Much of the current discussion of reference site selection has been put in the context of human disturbance gradients to insure that least-disturbed sites are evaluated. However, the natural variability of the reference condition is one of the main determining factors in the success of bioassessment programs (Hawkins et al., 2010a). The availability of natural landscape variables, such as drainage area, remove the relative ambiguity of the 1<sup>st</sup> – 3<sup>rd</sup> order stream criteria of most bioassessment plans by placing distinct

thresholds on the size of streams that the program can accurately evaluate. For example, the Mid-Atlantic Highlands Index of Biotic Integrity (McCormick et al., 2001) indicate that their bioassessment tool is appropriate for 1<sup>st</sup> – 3<sup>rd</sup> order streams, which only ranges on average from 2.2 km<sup>2</sup> to 67 km<sup>2</sup> (Leopold et al., 1964), making it inappropriate for any larger streams in that region. Including elevation and stream network position to the natural variables that are considered during reference sites selection, more accurate bioassessment predictions under the reference condition can be generated rather than using solely drainage area. The exclusion of natural variables from our boosted regression tree model allowed us to generate a human disturbance gradient giving us the capacity to select reference sites across a range of natural conditions without incorporating any bias that might be associated with the original reference pool. Remotely sensed landscape data, like the ones used in this study, not only give us a clear picture of the anthropogenic and natural landscape, but we are also able to use this data in a statistical framework to quantify the effect on stream systems.

There are also several potential shortcomings associated with our approach to reference site selection. Even though the land-use dataset used in our analysis was extensive and the best available to us, there are some human disturbances that could not be accurately mapped, or their stressor relationship with stream health may be unknown. For example, the extent of acid mine drainage or acid precipitation was not evaluated at the landscape scale. Both stressors have been shown to reduce aquatic community health in central Appalachia (Freund and Petty 2007; Petty et al., 2010), but their impacts were not used in our models. However, the importance of these variables in structuring stream communities are well known, so we used water quality measures (i.e., pH and specific conductance) to evaluate “unknown” impacts that might be occurring on the landscape. In addition, the influences of legacy land-use practices can still have an impact on stream quality today, especially in regions with a diverse history of natural resource extraction (Foster et al., 2003).

The primary focus of this study has been on landscape-scale processes and factors controlling stream health. However, localized human activities can also have a dramatic impact on stream health. For

example, channelization, dredging, or bank erosion can have significant impacts on fish and benthic macroinvertebrate communities, ultimately impacting overall stream health (Wang et al., 2002). On the other hand, localized improvements to stream health were also not considered. The implementation of best management practices (BMPs), which are meant to reduce or minimize human impacts, cannot be taken in consideration so all measures of agricultural, surface mining, and urban development are treated equally. Similarly, instream mitigation efforts from habitat modifications and chemical restoration cannot be taken into consideration either. The inability to incorporate localized impacts demonstrates the need to evaluate, or field validate, the local condition of reference sites prior to bioassessment development in order to help account for some of the smaller scale processes.

## Literature Cited

- Allan, J.D. 2004. Landscapes and Riverscape: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*. 35:257 – 284.
- Angradi, T.R., M.S. Pearson, T.M. Jicha, D.L. Taylor, D.W. Bolgrien, M.F. Moffett, K.A. Blocksom, and B.H. Hill. 2009. Using stressor gradients to determine reference expectations for great river fish assemblages. *Ecological Indicators*. 9:748 – 764.
- Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwater Biology*. 39:765 – 774.
- Bowman, M. F. and K. M. Somers. 2005. Considerations when using the reference condition approach for bioassessment of freshwater ecosystems. *Water Qual. Res. J. Canada*. 40:347 – 360.
- Davies, S.P., and S.K. Jackson, 2006. The Biological Condition Gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16:1251 – 1266.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*. 77:802 – 813.
- Fausch, K.D., J. R. Karr, and P.R. Yant. 1984. Regional application of an Index of Biotic Integrity based on stream fish communities. *Transactions of the American Fisheries Society*. 113:39 – 55.
- Foster, D., F. Swanson, J. Aber, I. Burke, N. Brokaw, D. Tilman, and A. Knapp. 2003. The Importance of land-use legacies to ecology and conservation. *BioScience*. 53:77 – 88.
- Freund, J.G. and J.T. Petty. 2007. Response of fish and macroinvertebrate bioassessment indices to water chemistry in a mined Appalachian watershed. *Journal of Environmental Management*. 39:707 – 720. DOI 10.1007/s00267-005-0116-3.
- Hawkins, C.P., Y. Cao, and B. Roper. 2010a. Method of predicting reference condition biota affects performance and interpretation of ecological indices. *Freshwater Biology*. 55:1066 – 1085.
- Hawkins, C.P., J. R. Olson, and R.A. Hill. 2010b. The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society*. 29:312 – 343.
- Hughes, R.M., A.T. Herlihy, and J.C. Sifneos. 2015. Predicting aquatic vertebrate assemblages from environmental variables at three multistate geographic extents of the western USA. *Ecological Indicators*. 57:546 – 556.
- Karr, J.R., and E.W. Chu. 1998. *Restoring life in running waters: better biological monitoring*. Island Press. Washington, D.C.

- Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey, Champagne, Illinois, Special Publication
- Leopold, L. B., M. G. Wolman, and J. P. Miller. 1964. Fluvial processes in Geomorphology. W. H. Freeman, San Francisco, CA.
- McCormick, F.H., R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, and A.T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands region. Transactions of the American Fisheries Society. 130:857 – 877.
- Merovich, G.T., J.T. Petty, M.P. Strager, and J.B. Fulton. Hierarchical classification of stream condition: a house-neighborhood framework for establishing conservation priorities in complex riverscapes. Freshwater Science. 32:874 – 891.
- Merriam, E. R., Petty, J.T., Merovich, G.T., Fulton, J.B., and Strager, M.P., 2011. Additive effects of mining and residential development on stream conditions in a central Appalachian watershed. Journal of North American Benthological Society. 30:399 – 418.
- Merriam, E.R., J.T. Petty, M.P. Strager, A.E. Maxwell, and P.F. Ziemkiewz. 2013. Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined central Appalachian watershed. Freshwater Science. 32:1246 – 1259.
- Minshall, G.W., R.C. Petersen, and C.F. Nimz. 1985. Species richness in streams of different size from the same drainage basin. The American Naturalist. 125:16 – 38.
- Petty, J.T., J.B. Fulton, M.P. Strager, G.T. Merovich, J.M. Stiles, and P.F. Ziemkiewicz. 2010. Landscape indicators and thresholds of stream ecological impairment in an intensively mined Appalachian watershed. Journal of the North American Benthological Society. 29:1291 – 1309.
- Pond, G.J., J.E. Bailey, B.M. Lowman, and M.J. Whitman. 2012. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. Environmental Monitoring and Assessment. DOI 10.1007/s10661-012-2648-3.
- Riseng, C. M., M. J. Wiley, R. J. Stevenson, T. G. Zorn, and P. W. Seelbach. 2006. Comparison of coarse versus fine scale sampling on statistical modeling of landscape effects and assessment of fish assemblages of the Mukegon River, Michigan. American Fisheries Society Symposium. 48:555 – 575.
- Reynoldson, T.B., R.H. Norris, V.H. Resh, K.E. Day, and D.M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society. 16:833 – 852.
- 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. 16:1267 – 1276.
- Strager, M.P., Petty, J.T., Strager, J.M. and Barker-Fulton, J., 2009. A spatially explicit framework for quantifying downstream hydrologic conditions. Journal of Environmental Management. 90:1854 – 1861.

Wang, L., T. Brenden, P. Seelbach, A. Cooper, D. Allan, R. Clark Jr., M. Wiley. 2008. Landscape based identification of human disturbance gradients and reference conditions for Michigan streams. *Environmental Monitoring and Assessment*. 141:1 – 17.

Wang, L., J. Lyons, and P. Kanehl. 2002. Effects of watershed best management practices on habitat and fish in Wisconsin streams. *Journal of the American Water Resources Association*. 38:663 – 680.

Whittier, T.R., J.L. Stoddard, D.P. Larsen, A.T. Herlihy. 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. *Journal of the North American Benthological Society*. 26: 349 – 360.

Wright, J. F., D. Moss, P. D. Armitage, and M. T. Furse. 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data.

Yates, A. G. and R. C. Bailey. 2010. Selecting objectively defined reference sites for stream bioassessment programs. *Environmental Monitoring and Assessment*. 170:129 – 140.



## Tables

**Table 1:** Mean (Standard Deviation) for all Wadeable Segment Level Watersheds for Local and Cumulative Variables Evaluated in Initial Boosted Regression Tree Models.

<b>Variable</b>	<b>Local (SD)</b>	<b>Cumulative (SD)</b>
<b>Structure Density (#/km<sup>2</sup>)</b>	30.74 (63.55)	13.79 (19.76)
<b>NPDES Permit Density (#/km<sup>2</sup>)</b>	0.8811 (8.119)	0.3422 (0.6064)
<b>Surface Mine Permit Density (#/km<sup>2</sup>)</b>	0.0345 (0.6401)	0.0502 (0.1104)
<b>Underground Mine Permit Density (#/km<sup>2</sup>)</b>	0.0447 (0.6829)	0.0502 (0.1304)
<b>Sewage &amp; Septic Service Density (#/km<sup>2</sup>)</b>	0.3878 (3.456)	0.2122 (0.4213)
<b>Development (%)</b>	2.251 (6.403)	0.8913 (2.059)
<b>Surface Mining (%)</b>	0.5909 (3.587)	1.664 (5.392)
<b>Forest (%)</b>	75.13 (23.32)	84.11 (12.59)
<b>Grassland (%)</b>	10.55 (10.79)	6.576 (4.832)
<b>Agriculture (%)</b>	9.278 (15.59)	6.158 (8.130)

**Table 2:** Criteria used to identify reference (Bailey 2009) and stressed sites for generating a Boosted Regression Tree model.

<b>Criteria</b>	<b>Reference</b>	<b>Stressed</b>
<b>NPDES Permits</b>	None	--
<b>Non-point source pollution</b>	None	--
<b>Dissolved Oxygen (mg/L)</b>	$\geq 5.0$	--
<b>pH</b>	Between 6.0 & 9.0	$\leq 5.0$
<b>Conductivity (<math>\mu</math>mhos/cm)</b>	$< 500$	--
<b>Fecal Coliform bacteria</b>	$< 800$ colonies/100 mL	--
<b>U.S. EPA-RBP Habitat scores</b>	$\geq 11$ for epifaunal substrate, channel alteration, and sediment deposition	--
	$\geq 6$ for bank vegetation protection and riparian vegetation zone width	--
	$\geq 130$ for Total RBP habitat score	--

**Table 3:** Reference site summary statistics for original, additionally selected, and combined (original + additional) reference sites compared to all segment level watersheds (All SLW) in each region. Data are presented as minimum – maximum (mean).

<b>Reference</b>	<b># Sites</b>	<b>Drainage Area (km<sup>2</sup>)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>
<b>MonCARV</b>				
Original	17	12.14 – 84.91 (30.94)	555 – 1120 (858.56)	7.19 – 104.53 (43.97)
Additional	57	8.07 – 357.6 (55.63)	662 – 1432 (839.11)	0 – 208.06 (39.73)
Combined	74	8.07 – 357.6 (49.96)	555 – 1432 (843.69)	0 – 208.06 (40.70)
All SLW	2425	7.01 – 399.66 (77.44)	291 – 1457 (784.55)	0 – 217.55 (40.04)
<b>Ohio CA</b>				
Original	10	7.34 – 60.97 (25.70)	488 – 802 (660.80)	22.26 – 66.98 (42.66)
Additional	16	9.64 – 250.26 (67.24)	347 – 899 (471.83)	2.7 – 88.75 (30.99)
Combined	26	7.34 – 250.26 (38.99)	348 – 899 (553.21)	2.7 – 88.75 (37.87)
All SLW	3486	7.01 – 385.68 (55.07)	180 – 1050 (538.89)	0 – 129.52 (35.89)
<b>Ohio Mon WAP</b>				
Original	9	9.73 – 143.74 (29.37)	268 – 448 (381.37)	10.01 – 78.36 (45.52)
Additional	23	7.38 – 295.6 (85.67)	272 – 454 (367.89)	0 – 75.88 (27.52)
Combined	32	7.38 – 295.6 (69.84)	268 – 454 (371.55)	0 – 78.36 (32.58)
All SLW	6707	7.0 – 399.45 (54.45)	163 – 754 (348.08)	0 – 188.4 (40.24)
<b>Upper Kanawha</b>				
Original	18	8.93 – 150.62 (44.92)	448 – 1334 (914.00)	3.56 – 135.93 (48.6)
Additional	20	9.5 – 383.0 (114.26)	600 – 1268 (885.44)	0 – 132.37 (26.27)
Combined	38	8.93 – 383.0 (81.41)	451 – 1334 (898.86)	0 – 135.93 (36.85)
All SLW	3500	7.0 – 395.69 (56.97)	267 – 1476 (803.15)	0 – 158.39 (38.24)

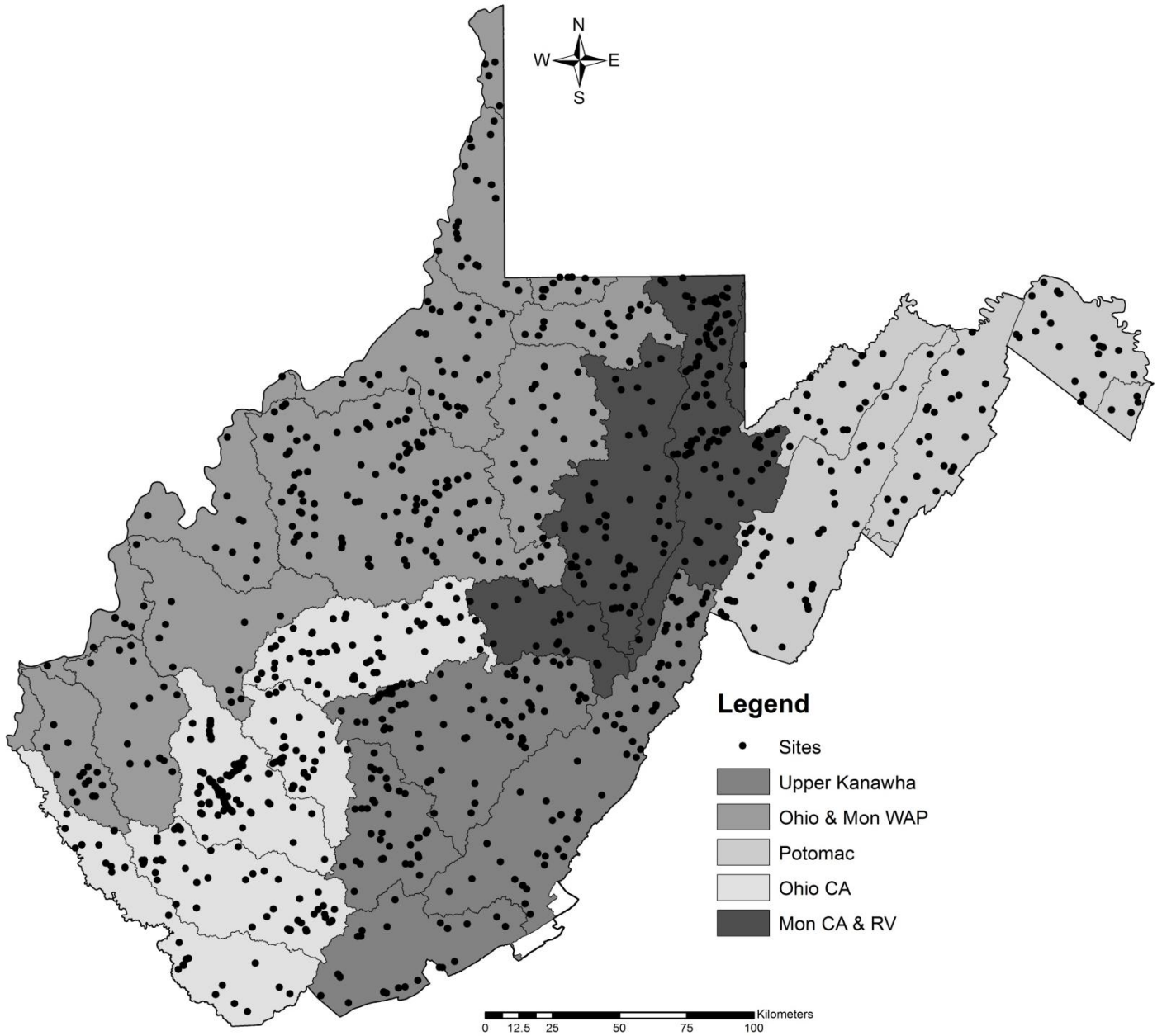
**Table 4:** Summary statistics for segment level watersheds (SLW) identified as large (>150 km<sup>2</sup>) and of reference condition within each region. Data are presented as minimum – maximum (mean).

<b>Reference</b>	<b># SLW</b>	<b>Drainage Area (km<sup>2</sup>)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>
<b>MonCARV</b>	264	150.06 – 399.65 (270.81)	443 – 1322 (831.04)	0 – 51.18 (4.62)
<b>Ohio CA</b>	16	171.39 – 212.99 (191.94)	356 – 440 (400.01)	0 – 35.45 (11.56)
<b>Ohio Mon WAP</b>	146	152.26 – 354.57 (228.71)	200 – 524 (357.77)	0 – 47.71 (5.71)
<b>Upper Kanawha</b>	183	150.48 – 390.97 (226.88)	583 – 1176 (893.68)	0 – 59.27 (15.30)

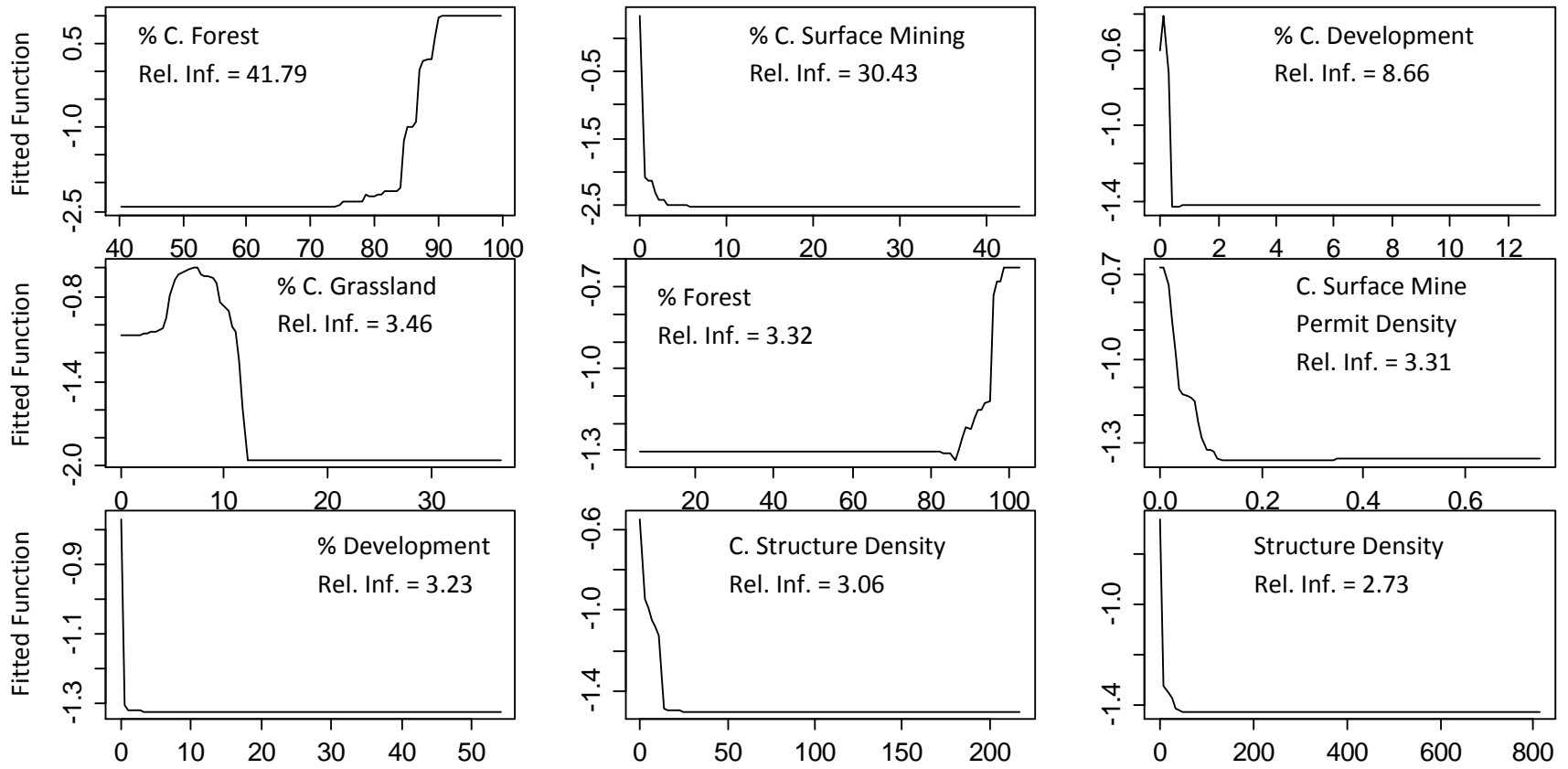
**Table 5:** Summary statistics for segment level watersheds (SLW) identified as large (>150 km<sup>2</sup>) and of intermediate condition within each region. Data are presented as minimum – maximum (mean).

<b>Intermediate</b>	<b># SLW</b>	<b>Drainage Area (km<sup>2</sup>)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>
<b>MonCARV</b>	57	152.56 – 395.76 (271.96)	512 – 1193 (785.30)	0 – 36.35 (3.73)
<b>Ohio CA</b>	45	153.67 – 357.74 (260.09)	277 – 661 (460.43)	0 – 35.61 (6.07)
<b>Ohio Mon WAP</b>	113	151.64 – 396.63 (227.03)	173 – 477 (308.71)	0 – 72.54 (13.81)
<b>Upper Kanawha</b>	35	150.55 – 395.69 (235.39)	467 – 978 (747.09)	0 – 30.23 (2.72)

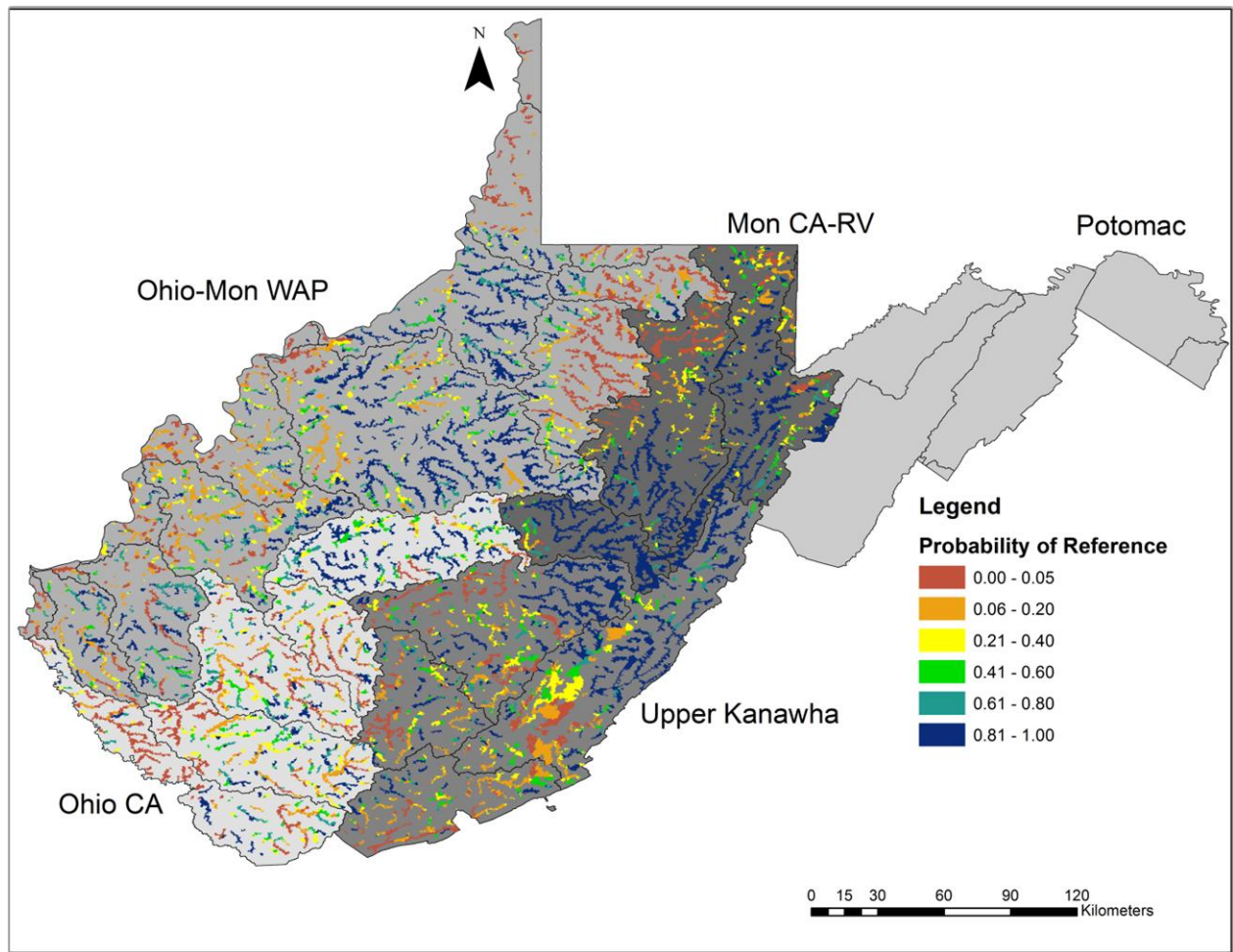
Figures



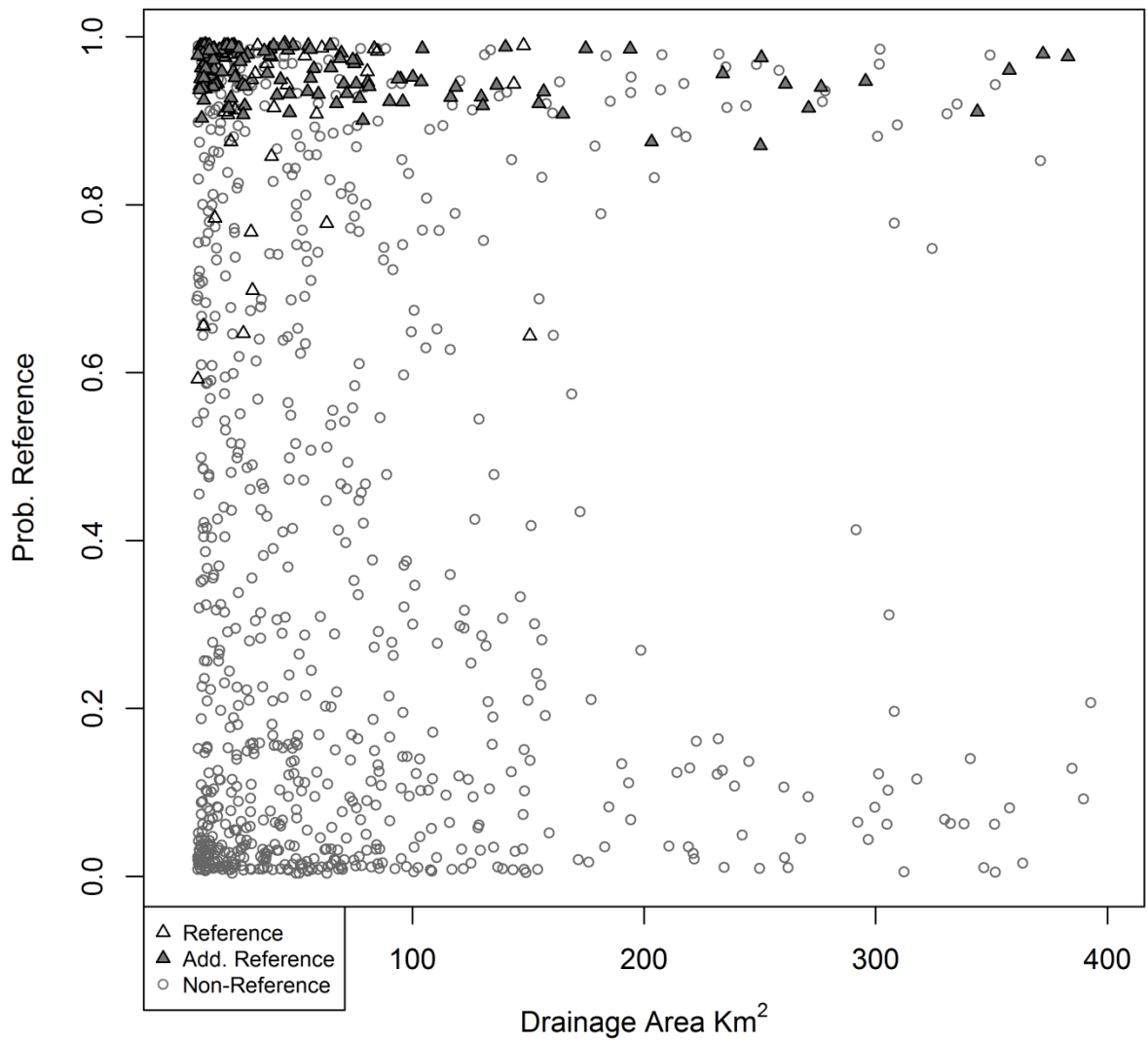
**Figure 1:** Fish biomonitoring regions for West Virginia Index of Biotic Integrity and site locations available for analysis. The regions consist of the Upper Kanawha, Ohio and Monongahela Western Allegheny Plateau (Ohio & Mon WAP), Potomac River, Ohio Central Appalachians (Ohio CA), and Monongahela Central Appalachians and Ridge-Valley (Mon CA & RV).



**Figure 2:** Relative function curves for the final variables selected to determine reference condition in order of highest relative influence (Rel. Inf.) from left to right.

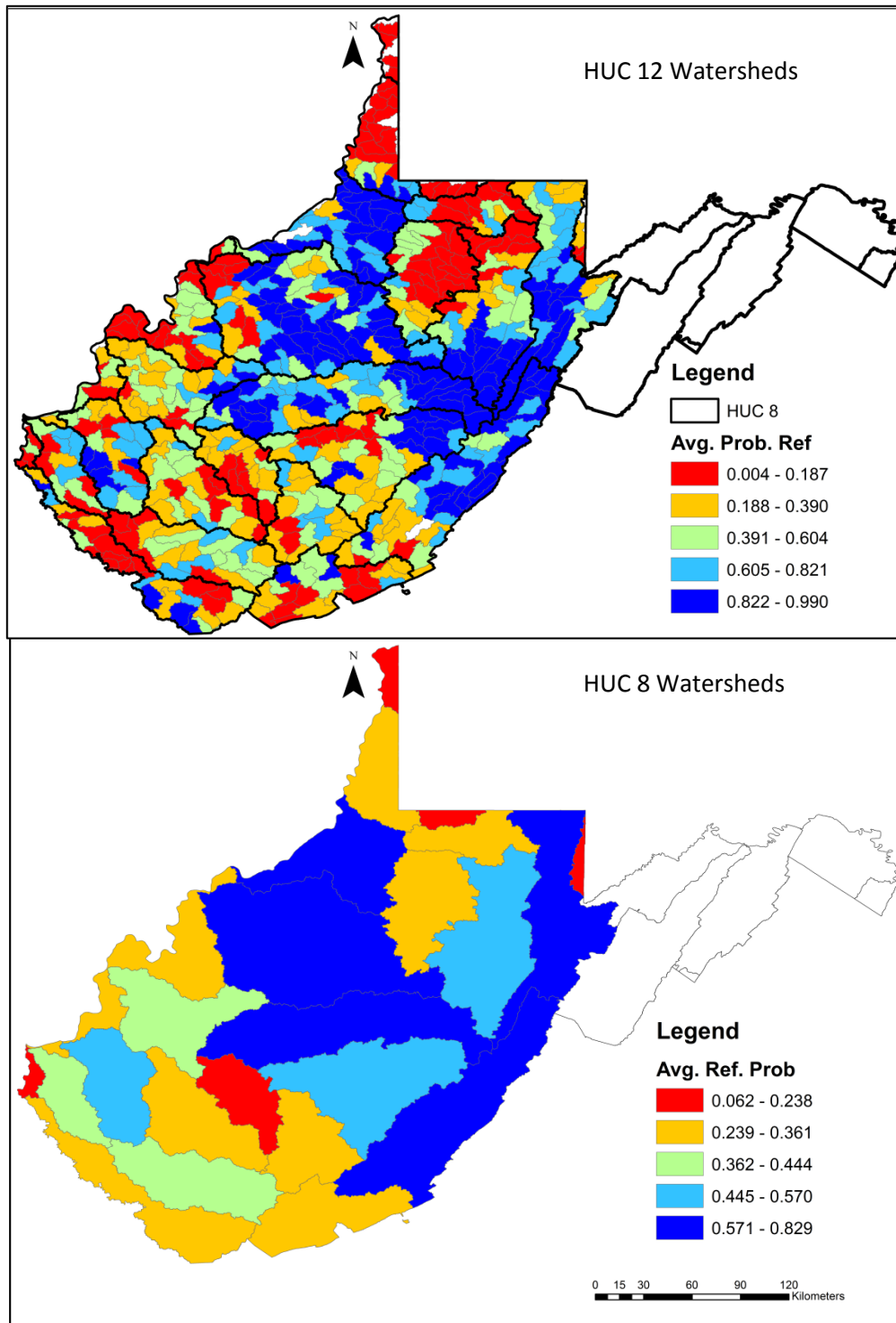


**Figure 3:** Probability of reference for all wadeable segment level watersheds in the study region generated from the final BRT model.



**Figure 4:** Gradient of probability of reference and drainage area for all sampled sites in the study area.





**Figure 5:** Average probability of reference at the HUC12 and HUC 8 watershed scale. Averages were generated from all wadeable segment level watersheds within each HUC 12 or HUC 8 watershed to generate a overall condition.

**CHAPTER 3: MODELING THE NATURAL VARIATION OF TRAIT-BASED FISH  
COMMUNITY STRUCTURE IN WEST VIRGINIA**

## **Abstract**

The trait-based structure of fish assemblages varies with the spatial extent and the heterogeneity of landscapes. Understanding the natural variation between fish assemblages will provide new insights on how community traits are controlled on the landscape as well as evaluate the response of these communities to environmental alterations. Accounting for complex natural gradients during bioassessment development can increase sensitivity of the final multimetric index to fish community alterations. We developed spatial Boosted Regression Tree (BRT) models to account for the natural variation observed in trait-based fish community assemblage structure. We used regionally defined reference sites to predict fish community metrics as a function of elevation, drainage area, and the distance from a stream site to potential riverine source populations (i.e. swim distance). We then evaluate the need to use BRT models by developing a criteria threshold using a randomization technique (Monte Carlo analysis) for selected community metrics based on 10,000 model iterations. The majority of fish community metrics evaluated with BRT produced a wide range of deviance explained using at least two of the natural landscape variables, most commonly drainage area and elevation. We evaluated the individual model results for metrics selected for use in the region-specific West Virginia Index of Biotic Integrity and adjusted metric values (observed:expected values) based on BRT predictions. Most metrics demonstrated thresholds with elevation and drainage area in which decreases in metric values were observed between 400 – 600 m and increases were observed from 25 – 100 km<sup>2</sup>, respectively. Adjusted metric values were correlated with known landscape and water quality stressors associated with surface mining and residential development. Even though, there were generally negative responses of all metrics in the Mon CARV and Ohio CA to mining related stressors (% surface mining, pH, specific conductance), the remaining regions exhibited no correlations with any of the anthropogenic variables analyzed. Evaluating the response and use of predictive modeling on fish community assemblage structure using boosted regression tree modeling sets the stage for developing bioassessment programs that are more sensitive to anthropogenic stressors when the appropriate community metrics are selected.

In addition, the predictive models allow for a wider variety of natural stream conditions to be evaluated under fewer assessment criteria. However, care should be taken in order to select trait-based indices that are responsive to major stressors of a region and to establish predictive model thresholds. Accounting for larger environmental processes and considering the influences of natural variation on fish trait assemblage structure, the deviations from the expected natural condition can now be linked directly to anthropogenic degradation.

## 1. Introduction

Local community assembly is a product of several biotic and abiotic hierarchical filters operating at multiple spatial and temporal scales (Poff, 1997). There are studies designed to investigate which types of factors (local vs. regional) are most important when evaluating community assembly (Resetarits, 2005; Hoeinghaus et al., 2007; Blanchet et al., 2014; Pease et al., 2015). Local factors, such as stream flow and morphology, and regional factors, such as ecoregion and climate, can impact the structure of fish communities (Wang et al., 2003; Pease et al., 2015). The debate over local versus regional processes has demonstrated that the main driving factors of community assembly can depend on the organisms being studied, the extent of anthropogenic disturbances, and natural landscape heterogeneity. For example, in undisturbed streams and watersheds, fish assemblages are influenced by reach-scale variables (i.e., channel morphology, fish cover, and substrate types) and as disturbance increases, the watershed-scale variables become increasingly important (Wang et al., 2003; Oliveira et al., 2012). However, in urban and agricultural dominated systems, land-use practices at the watershed scale are the main factors influencing stream fish communities over local-scale habitat features (Wang et al., 2001, 2002). Knowing the relative importance of natural and anthropogenic environmental variables occurring at multiple spatial scales is important moving forward in stream conservation and restoration efforts. However, little attention is given to studies evaluating the impacts of these landscape factors on the functional trait organization of fish assemblages.

Trait-based approaches in ecology are commonly used to evaluate patterns of species invasions, evaluate community response to habitat alterations, and explore the mechanisms driving community assembly. The first mention of a trait-based approach culminated with the guild concept in which species were combined if they exploited the same resource type (Root, 1967). Species are assigned functional groups, or guilds, based on different life-history traits such as spawning preference, feeding behavior, morphology, and behavioral adaptations to their environment. Aggregating individual species into measurable trait-based groups can help reduce the dimensionality of community level data providing

increased explanatory power over taxonomic data when evaluating ecosystem functions (Gagic et al., 2015). In addition, species traits can be used to examine complex relationships between the ecological functions of a stream and the environmental variables impacting them, both natural and anthropogenic. Commonly these classifications have been reserved for plants (Johnson, 1981), birds, and more recently fish species with the introduction of biological monitoring and assessment (Austen et al., 1994). Trait-based summaries of the fish community are the foundation for the development of multi-metric indices, like the Index of Biotic Integrity (IBI; Karr, 1981).

Fish community based biomonitoring programs aim to combine both the importance of functional diversity while attempting to capture changes in ecosystem function and species diversity. Overall, trait-based approaches perform better than traditional taxonomic approaches (e.g., number and abundances of species) in predicting ecosystem functions (Gagic et al., 2015). Traditionally, it was thought that trait-based groups are less geographically constrained when compared to individual species, allowing them to be easily compared across large geographic scales (Ibanez et al., 2009). However, more recent studies have demonstrated that trait-based functional diversity can differ along longitudinal stream gradients (Pease et al., 2012) as well as across physiogeographic regions (Pease et al., 2015). Biomonitoring and assessments of stream ecosystems using fish as a model organism cannot evaluate intra- and inter-specific variation in functional traits across large geographic extents. Instead these programs generate expectations of species assemblage as a function of natural variables (e.g. drainage area/volume, distance from source populations) in order to account for a portion of that variation. Understanding the baseline functional diversity and trait-based structure of fish communities along natural environmental gradients would enable bioassessment programs to effectively measure the community level response to anthropogenic disturbances across broad geographical areas.

Within West Virginia, there are extensive anthropogenic landscape alterations occurring due to natural resource extraction as well as low intensity residential development and agriculture that can have impacts on the fish communities. These large scale anthropogenic impacts have potential to shift community

assemblage structure; however the high amount of natural variation across the landscape makes it difficult to evaluate the magnitude of these impacts. Overall, correcting for natural gradients using predictive modeling prior to multimetric development or analysis of community structure decreases the among-site variability allowing final measures of trait-based community composition and resulting multimetric indices to be sensitive to alterations (Hawkins et al., 2010; Esselman et al., 2013). The main objective of this study was to construct and validate spatial models predicting trait-based metrics of fish community composition as a function of natural landscape variables using a reference condition approach for streams and rivers in West Virginia. In order to accomplish this objective we developed region specific models for a variety of trait-based community metrics using natural landscape variables and compared deviations of selected trait-based metrics from expected natural conditions to anthropogenic landscape and water quality stressors.

## **2. Methods**

This study focused on the development of spatial Boosted Regression Tree (BRT) models to generate a baseline condition and assess impacts from anthropogenic landscape stressors on trait-based measures stream fish communities in West Virginia. Regionally defined reference sites were used to predict fish community metrics as a function of elevation, drainage area, and the distance from a stream site to potential riverine source populations (i.e., swim distance). In order to evaluate the effectiveness or the need to use BRT models to predict community composition metrics, we generated null models of BRT performance. A null hypothesis was generated by producing a Monte Carlo estimate of p-values by randomizing the fish community metric prior to running a boosted regression tree model (Gotelli and Ellison, 2004). Thresholds of % variance explained were generated from significant (p-value <0.05) simulated Monte Carlo models. Fish community metrics were predicted if they met the required threshold for their biomonitoring region and adjusted (observed:expected). Finally, the deviation of selected adjusted metrics, based on their natural expected condition, was evaluated by relating them with known anthropogenic stressors.

## *2.1 Landscape Data*

Natural landscape variables, such as cumulative drainage area (km<sup>2</sup>), elevation (m), and swim distance (km), were generated in a GIS framework on a segment level watershed (National Hydrology Dataset 1:24K) basis using accumulated stream flow tables and the stream accumulation tool for ArcGIS provided by NRAC at West Virginia University (Strager et al., 2009). Digital Elevation Models were generated for West Virginia in 2003 by the United States Geological Survey (USGS) and the Statewide Addressing and Mapping Board (SAMB) at a 3-meter (1/9<sup>th</sup> arc-second) resolution. We used this data to summarize the maximum elevation for each segment level watershed used in this study. Swim distance was defined as the in-stream distance from the outflow of any segment level watershed to the inflow of a downstream segment level watershed of 200 km<sup>2</sup> or greater (Hitt and Angermeier, 2011). Measures of swim distance were generated for every segment level watershed in West Virginia using accumulated flow tables and the swim distance calculator located in the stream accumulation tool provided by NRAC and West Virginia University. Since some river systems originate outside of West Virginia, we reduced our analysis to wadeable streams and rivers (7 – 400 km<sup>2</sup>) that have less than 1% of their cumulative drainage area outside of the state boundary.

## *2.2 Fish Community Data & WV IBI Metrics*

A set of 937 stream segments were sampled using either backpack or boat electrofishing, or parallel wires from 2000 – 2013 for fish biomonitoring from a variety of institutions (e.g. West Virginia Department of Environmental protection, West Virginia Department of Natural Resources, and West Virginia University). Sampling locations were attributed to segment level watersheds in ArcGIS. In cases in which multiple sampling events occurred in a single segment level watershed, the most recent event was selected to characterize that stream segment. For all sampling events, individual fishes were identified to species either in the laboratory or in the field. Fish community data for each site was summarized according to feeding guild, spawning characteristics, habitat preference, tolerance to stressors, and species composition according to both their richness and proportions. Classifications of species into traits were



based on traits outlined in the West Virginia Index of Biotic Integrity (Chapter 1). Additional water quality data collected during fish sampling and remotely sensed and accumulated landscape data (Strager et al., 2011) were retained for further analysis. Only a portion of the fish sampling locations had additional water quality measures (specific conductance and pH) associated with them. Approximately 140 proportion and richness based community metrics were analyzed for their relationship with natural variables. The metrics utilized in West Virginia IBIs were used for additional analyses within their respective biomonitoring region.

### *2.3 Reference sites*

The spatial positioning of the reference sites leads to great variability in the natural and anthropogenic landscapes surrounding these streams making it difficult to set a statewide reference condition. Based on previous research done in this region (Anderson and Petty 2015, Chapter 1), the state has been split into five fish biomonitoring regions based on fish community composition and natural landscape variation in order to get a more accurate representation of the reference condition at a smaller spatial extent. The five regions include: Ohio Central Appalachians (Ohio CA), Ohio-Monongahela Western Allegheny Plateau (Ohio Mon WAP), Monongahela Ridge and Valley (Mon RV), Upper Kanawha (UK) and the Potomac (not included in analysis due to low sample sizes) (Figure 1).

A portion of the sampled sites were evaluated after sampling to be considered a reference site based on strict water quality standards used for environmental regulation along with habitat assessment (Table 1). However, each site passing the water quality and habitat standards were evaluated in terms of local landscape disturbances and known point-source discharges upstream and final inclusion into the reference pool was determined by best professional judgment (N=54). Additional reference sites (N=153) were identified using the methods outlined in Anderson and Petty (2015, Chapter 2) in order to ensure the full range of drainage areas of wadeable streams in each biomonitoring region was represented (Table 2). Since a reference condition approach is being used, summary statistics of natural landscape variables for reference (N=207) and non-reference (N=730) sites in each biomonitoring region are provided (Table 2).

## 2.4 Statistical Analyses

Boosted regression tree (BRT) models are additive regression models which combines iteratively fit simple regression trees using a boosting algorithm in order to increase predictive performance (Elith et al., 2008). This technique has proven to be well suited for modeling ecological and landscape datasets with a complex covariance structure (Merovich et al., 2013; Fleishman et al., 2014). Prior to BRT modeling of community metrics, each metric was evaluated on its range (maximum value – minimum value) to ensure proper variation in the data to warrant further analysis. Any metric with a range of at least 5 species (richness metric) or 5% (proportion metric) was further evaluated with a BRT model. In order to prevent over-application of BRT models, we also developed a guideline for model acceptance by calculating Monte Carlo estimates of p-values and variance explained (%) for each biomonitoring region. Using these thresholds, predictions of selected metrics were generated and the deviation of each metric from expected condition was evaluated against known landscape stressors.

### 2.4.1 Modeling Expectations

All fish community composition metrics at reference sites were analyzed in a boosted regression tree framework for their relationship with natural landscape features such as cumulative drainage area (km<sup>2</sup>), elevation (m), and swim distance (km) within each fish biomonitoring region. For all models, tree complexity and learning rate were held constant at 2 and 0.001, respectively. A tree complexity of two was used to allow for one-way interactions between variables used in the model in order to more accurately predict in a complex landscape structure. The learning rate, or shrinkage parameter, determines the contribution of each tree to the model and was held constant in order to allow for comparisons among community metrics within a region. However, bag fraction varied between regions depending on sample size. The Mon CARV region was assigned the lowest bag fraction (0.50) while the other three regions used a higher bag fraction (0.75) due to a relatively lower sample size (~40 reference sites). Initial BRT models were generated using all 3 landscape variables. Model predictive performance was assessed by calculating the mean model deviance and cross-validation predictive deviance from ten folds of the data

(Elith et al., 2008). Predictor variables were removed if the removal of the variable did not increase the cross-validation error rate. Finally, relative importance of each variable retained in the simplified model was generated using an out-of-bag procedure outlined in Elith et al., (2008).

#### 2.4.2 Monte Carlo Analysis

In order to evaluate the effectiveness or the need to use BRT models to predict community composition metrics, we generated null models of BRT performance. The model with the highest total deviance explained for richness and proportion based metrics in each region was used to generate region specific thresholds for model use. A null hypothesis was generated by producing a Monte Carlo estimate of a p-value by randomizing the fish community metric prior to running a BRT model (Gotelli and Ellison, 2004). Each randomized boosted regression tree model was performed using the same predictive variables and parameters as the initial predictive model. A final histogram of the total deviance explained (%) from each randomized BRT model was produced and the 95<sup>th</sup> percentiles of the distributions were calculated based on 10,000 iterations of the model. The 95<sup>th</sup> percentile of the total deviance explained for the simulated models served as conservative thresholds to determine whether an individual model would be used to predict community composition metrics within that region for that metric type. Only thresholds from significant (simulated p-value < 0.05) Monte Carlo simulations were evaluated.

#### 2.4.3 Evaluation of Adjusted Metric Performance

Metrics utilized in regionally developed Indices of Biotic Integrity for wadeable warm water West Virginia streams (WV IBI; Chapter 1) were evaluated against known landscape and water quality stressors. The metrics evaluated in the Monongahela Central Appalachians-Ridge and Valley (Mon CA-RV) included proportion benthic individuals minus tolerant (% Benthic-Tol), species richness minus tolerant (Richness-Tol), richness of clean gravel spawners (CGS), proportion of non-tolerant individuals (% Fish-Tol), proportion invertivores (% IN), benthic species richness (Benthic), and Cyprinidae species richness (Cyprinid). The metrics evaluated in the Ohio Central Appalachians included species richness minus tolerant (Richness-Tol), proportion invertivores (% IN), Shannon-Weaver Trophic Diversity (SW-

Trophic), darter-madtom-sculpin species richness (DMS), proportion of tolerant individuals (% Tol), rock-gravel spawner richness (RGS), and non-guarding lithophilic spawner richness minus tolerant (NGL-Tol). The Ohio and Monongahela Western Allegheny Plateau IBI consisted of species richness minus tolerant (Richness-Tol), darter-madtom-sculpin species richness (DMS), proportion of tolerant individuals (% Tol), rock-gravel spawner richness (RGS), and non-guarding lithophilic spawner richness minus tolerant (NGL-Tol), intolerant species richness (Int), and proportion of omnivore-herbivores minus Central Stonerollers (*Campostoma anomalum*; %OH-CAAN). Finally, the Upper Kanawha region IBI consisted of proportion invertivores (% IN), benthic species richness (Benthic), Cyprinidae species richness (Cyprinid), total species richness (Richness), and clean-gravel plus rock-gravel spawner richness (CGS+RGS). Within each regional WV IBI, portions of the metrics selected were adjusted for drainage area (log base10 transformed) using the linear relationship within the reference sites. For this analysis, we used the BRT predictions to adjust metric values using the same observed/expected approach outlined in the WV IBI. Only metrics with significant (i.e., % variance explained exceeds Monte Carlo simulation threshold) BRT models were adjusted. Trait-based metrics within each region were correlated using Pearson's correlation against specific conductance, pH, and cumulative measures of anthropogenic land-use (i.e., structure density, % surface mining) (Bernhardt and Palmer 2011; Merriam et al., 2011).

### **3. Results**

#### *3.1 Description of Reference Sites*

Prior to statistical analysis, reference sites were evaluated in each region to ensure reference sites spanned the range of the natural variables observed at all the sites (Table 2). Each biomonitoring region had at least 40 reference sites used to model fish community structure given natural variables. The average drainage area represented in each biomonitoring region ranged from a minimum of 55.22 km<sup>2</sup> in the Ohio CA to a maximum of 87.72 km<sup>2</sup> in the Upper Kanawha with the Mon CA-RV (56.21 km<sup>2</sup>) and Ohio-Mon WAP (74.35 km<sup>2</sup>) falling in between (Table 2). The non-reference sites followed a similar pattern with Mon CA-RV sites having the lowest average drainage area (43.69 km<sup>2</sup>) followed by the Ohio CA (55.22

km<sup>2</sup>), Upper Kanawha (84.72 km<sup>2</sup>), and Ohio-Mon WAP (91.24 km<sup>2</sup>). Even though the ranges of drainage areas were similar among biomonitoring regions, there were distinct differences in average elevations between regions. The Mon CA-RV and Upper Kanawha biomonitoring regions demonstrated a larger range and higher average elevations than the Ohio CA and Ohio-Mon WAP regions (Table 2). Swim distance values between reference and non-reference sites within and among biomonitoring regions was highly variable with large ranges, however the average swim distances did not exhibit any distinct patterns.

### *3.2 Boosted Regression Tree Models*

#### 3.2.1 Modeling Expectations

The relative influence of natural variables on trait-based community metrics within West Virginia varied between regions. For the IBI metrics in the Mon CA-RV region, drainage area was the only variable used in all seven models (Table 3). However, it was also found to be the least important variables in predicting two (%Benthic-Tol and %Fish-Tol) of the response variables with swim distance being the most important. Within the Ohio CA region, six of the seven metrics within the WV IBI were evaluated using BRT (Table 3). The Shannon-Weaver Trophic Diversity index was not evaluated since it is already an abstraction of the assemblage structure. In this region, drainage area was found to be the most important variable in all but one of the models (RGS). In that model, maximum elevation was the most important variable, followed by drainage area. The two proportional metrics evaluated used all three landscape variables. Models used to evaluate the influence of natural landscape variables on selected metrics in the Ohio-Mon WAP found drainage area to be important to each of the response variables (Table 3). However, one model (%OH-CAAN) had maximum elevation as the most important variable followed by drainage area. The Upper Kanawha region had similar patterns with the selected metrics evaluated (Table 3) in which drainage area was the most important variable in all models. In this region, maximum elevation was also found to be influential in all the models generated for the selected metrics. Model results for all of the other fish community metrics evaluated are presented in Appendix A. The Ohio CA

IBI region had two of the IBI metrics evaluated perform the best out of all the metrics in that region and were used for the Monte Carlo Analysis. Fitted function curves for selected metrics were evaluated for patterns in metric response to drainage area, elevation, and swim distance. Drainage area and maximum elevation were the two most common variables used in predicting trait-based community metrics regardless of the biomonitoring region. Generally, most metrics exhibited a similar pattern as rock-gravel spawner richness in the Ohio CA biomonitoring region (Figure 2) in which metrics increased sharply in value from ~25 – 75 km<sup>2</sup> in drainage area and decreased sharply at ~ 400 – 600 m in elevation.

### 3.2.2 Monte Carlo Analysis

Initial models of trait-based metrics were evaluated regionally and by metric type (i.e. richness vs. proportion) in order to select two metrics with the highest total deviance explained. These metrics were then used in a Monte Carlo randomization analysis (i.e., community metric randomized and BRT model generated) using 10,000 iterations. After each Monte Carlo iteration, the total deviance explained was retained and the 95<sup>th</sup> percentile of the simulated distribution was used as a conservative threshold. We determined if BRT models should be used to predict trait-based community composition metrics based on these thresholds. Overall, models exceeding thresholds were determined to exhibit a relationship with natural variables that was better than random and were further used to predict community structure given natural variables for all wadeable streams in West Virginia.

For the Mon CA-RV, %Lithophilic spawning individuals minus toternat (%LSR-Tol) and invertivore-piscivore species richness minus Creek Chub (*Semotilus atromaculatus*; IP-SEAT) had initial total deviances explained of 65.85% and 65.71%, respectively (Table 4). The proportional metric utilized drainage area and swim distance while the richness metric used drainage area and elevation. The average Monte Carlo simulated total deviance explained for the Mon CA-RV was less than 5 % for both metrics resulting in p-values < 0.0001. The 95<sup>th</sup> percentiles of the metrics based on the full distribution of total deviation explained over the 10,000 simulations were approximately 21% for both metrics (Table 4).

For the Ohio CA biomonitoring region, % Invertivore (% IN) and species richness minus tolerant (Richness-Tol) had initial total deviances explained of 81.25% and 78.88%, respectively (Table 4). The proportional metric utilized all 3 landscape variables while the richness metric used drainage area and elevation. The average Monte Carlo simulated total deviance explained for the Ohio CA was 6.3% and 4.8% for the proportion and richness metrics, respectively, resulting in p-values  $\leq 0.00005$  for both metrics. The 95<sup>th</sup> percentiles of the metrics based on the full distribution of total deviation explained over the 10,000 simulations were approximately 32.5% and 25.9% (Table 4).

Metrics measuring the proportion of native Cyprinidae individuals minus Blacknose Dace (*Rhinichthys atratulus*) and Creek Chub (%Cyprinid-BNDSEAT) and invertivore-piscivore richness minus tolerant (IP-Tol) had initial total deviances explained of 86.36% and 80.77%, respectively, in the Ohio-Mon WAP region (Table 4). All three natural landscape variables were found to be important in predicting both metrics. The average Monte Carlo simulated total deviance explained for the Ohio-Mon WAP was 6.2% and 6.1% for the proportion and richness metrics, respectively, resulting in p-values  $\leq 0.00005$  for both metrics. The 95<sup>th</sup> percentiles of the metrics based on the full distribution of total deviation explained over the 10,000 simulations were approximately 32.1% and 30.8% (Table 4).

For the Upper Kanawha biomonitoring region, % non-guarding lithophils minus tolerant (%NGL-Tol) and invertivores-piscivore richness (IP) had initial total deviances explained of 72.97% and 76.00%, respectively (Table 4). Both BRT models utilized drainage area. In addition to drainage area, the richness model used elevation while the proportion model used swim distance. The average Monte Carlo simulated total deviance explained for the Upper Kanawha was 5.8% and 5.2% for the proportion and richness metrics, respectively, resulting in p-values 0.0000 for both metrics. The 95<sup>th</sup> percentiles of the metrics based on the full distribution of total deviation explained over the 10,000 simulations were approximately 30.3% and 27.6% (Table 4).

### *3.3 Evaluation of Adjusted WV IBI Metrics*

The WV IBI metrics evaluated within each region were adjusted if they exceeded the 95<sup>th</sup> percentiles for their metric type (Richness vs. Proportion). One WV IBI metric in the Mon CA-RV, % non-tolerant individuals (%Fish-Tol), and the Ohio-Mon WAP, rock-gravel spawner richness (RGS), regions did not meet the required thresholds for their region to be adjusted using BRT models. These metrics were still evaluated for their correlation with known stressors, but they were not adjusted. The remaining WV IBI metrics were adjusted and correlated with anthropogenic landscape stressors and water quality measures (Table 5).

The correlations of WV IBI metrics to known landscape and water quality stressors (Table 5) in West Virginia showed strong regional differences. Overall, the Mon CA-RV region metrics demonstrated a strong negative correlation with specific conductance and strong positive correlations with pH. However, few patterns were evident with structure density and cumulative surface mining (%). Darter-madtom-sculpin richness and rock-gravel spawner richness metrics in the Ohio CA were the only metrics that showed a negative response to specific conductance. Overall, trait-based metrics in this region showed a stronger negative correlation with surface mining (%) but did not show any correlation with pH. We also observed a positive correlation of all metrics in the Ohio CA with structure density. The Ohio-Mon WAP and Upper Kanawha regions showed little response of their WV IBI metrics to any of the measures of anthropogenic stress.

## **4.0 Discussion**

A wide range of natural landscape conditions lead to high variation in natural fish community structure within and between biomonitoring regions. A full understanding of the patterns and processes of fish community structure in a natural or reference condition is an important step in successful ecosystem management. Bioassessment programs depend on the reference condition to set baseline expectations so the impacts of anthropogenic alterations can be accurately evaluated. However, unbiased assessments of



stream quality are difficult to develop due to confounding landscape variables. To help distinguish between the effects of natural and anthropogenic variables, highly stratified bioassessment programs are developed. Multimetric indices have been developed for different states (Roth et al., 1998), regions (McCormick et al., 2001; Daniels et al., 2002), stream sizes (Lyons et al., 2001; Lyons 2006), stream temperature classes (Lyons et al., 1996; Lyons et al., 2012), and individual rivers (Emery et al., 2003) or watersheds (Kimmel and Argent, 2006). Predictive models could be used to set baseline reference conditions using natural landscape variables that are known to drive the characterization of the strata used to distinguish groups for IBI development (Pont et al., 2009).

The separation of West Virginia into fish biomonitoring regions was established based on differences in fish community structure due to large scale environmental processes such as ecoregion and major drainage basin (Chapter 1). For example, the differences in drainage network patterns and underlying geology associated with ecoregions can ultimately lead to differences in trait-based structures of fish communities by favoring gravel associated spawners in the Ridge and Valley and Central Appalachian systems (Detenbeck and Cincotta, 2008). Differences in fish assemblage structure due to large scale ecoregional processes have been observed in other studies (Pease et al., 2011; Pond et al., 2012; Pease et al., 2015).

Early multimetric indices were developed at relatively small spatial scales and relied on best professional judgment to determine reference condition expectations. As the spatial extent of bioassessment programs have expanded, the need to control for natural abiotic variation has increased in order to detect differences between disturbed and undisturbed conditions (Esselman et al., 2013). Evaluating the expected trait-based community structure within each biomonitoring region removed some of the environmental variation associated with large scale processes. Our results from this study demonstrate general threshold responses of selected metric values to both elevation and drainage area. Between 400 – 600 m in elevation, we observed decreases in metric values within the Mon CA-RV, Ohio CA, and Ohio-Mon WAP biomonitoring regions. Even though elevation was an important predictor variable for metrics in

the Upper Kanawha region, highly variable relationships were observed. Drainage area on the other hand, exhibited similar relationships regardless of the metric or the biomonitoring region. Most metrics evaluated demonstrated a strong increase in metric value between drainage areas of 25 – 100 km<sup>2</sup>. These results correspond with other studies that have indicated that large, lower elevation streams typically harbor higher numbers of species, and in turn, a higher trait-diversity (Pease et al., 2012). Migration of fish species can occur from mainstem habitats into smaller adjacent streams increasing the overall species, and trait, diversity (Hitt and Angermeier, 2011). Even though no differences in multimetric index scores were observed between mainstem and headwater tributaries, metric values did differ between the two especially for intolerant, benthic, and cyprinid species richness (Hitt and Angermeier, 2011). For our study, swim distance was important in predicting benthic associated species, gravel spawning individuals, invertivores, and tolerant species across all biomonitoring regions.

Our study adds to the growing literature using BRT modeling with ecological datasets. BRT modeling, and other machine learning techniques, has proven beneficial in evaluating ecological data due to its ability to evaluate non-linear responses to predictors and incorporate multiple interactions and data types (De'ath, 2007). Cao et al., (2007), Esselman et al., (2013), and Daniel et al., (2014) used classification and regression, or boosted regression, trees to account for the natural environmental variation of trait-based community metrics. These studies found that adjusted metrics estimated the natural environmental gradient with higher precision compared to studies using traditional linear regression. Furthermore, these studies were then able to also evaluate the influence of anthropogenic processes on the remaining variance. Since natural differences in elevation, temperature, network position, and stream size are important factors in driving fish assemblages, they should all be taken into consideration when evaluating fish community response to anthropogenic stress. During IBI development, rather than using solely drainage area adjustments, more programs are using a predictive modeling approach similar to the methods used in this study which are based on several natural gradients. In fact, predictive based IBIs

have been developed for the western (Pont et al., 2009) and conterminous (Esselmand et al., 2013) United States, France (Oberdorff et al., 2002) and parts of Europe (Pont et al., 2007).

With the increased use of BRT modeling, we took additional precautions in our statistical analysis and evaluations of model results. The “significance” of each BRT model was evaluated by developing thresholds of the deviance explained using Monte Carlo simulations for each metric type within each region. Based on the results of the Monte Carlo analysis, care must be taken when making inferences using BRT models with less than 30 – 35% deviance explained since it may not be better than random. In addition, thresholds for model use may be region, or data, specific so thresholds should be developed based on the distribution and relationships in the data being analyzed. To our knowledge, this is one of the first studies to evaluate the statistical significance, or effectiveness, of boosted regression tree models in predicting ecological data by comparing it to a randomized data set.

Metrics adjusted based on expectations given their natural conditions showed region specific responses to anthropogenic stressors. The metrics evaluated in the Mon CA-RV and Ohio CA demonstrated overall negative responses to specific conductance. The causes of elevated specific conductance in these two regions can be linked to current (Hartman et al., 2005; Bernhardt and Palmer, 2011) and legacy coal mining (i.e., abandoned mine lands; Herlihy et al., 1990). Decreases in fish community metrics in acid mine drainage (AMD) impacted regions can be linked to decreased survival of juvenile and larval fish populations (Hafs et al., 2010). Even though the impacts of elevated mountain top/valley fill mining on population dynamics is unknown, temporal changes in both taxonomic and functional diversity of fish communities have been observed (Hitt and Chambers, 2014). Specifically, streams exposed to surface mining effluents had lower diversity than reference streams in addition to lower abundances and overall biomass, despite similarities in physical habitat conditions.

Unlike the other regions, the adjusted metrics evaluated in the Ohio CA exhibited an increase with increases in structure density. The narrow floodplains and steep topology of the region restricts the

expansion of development resulting in low overall percentages of land-use while densities of residential and commercial structures are high. The increased nutrients associated with residential development maybe augmenting aquatic systems, ultimately restructuring the food quality and quantity available to stream fishes (Wang et al., 2007). In a region dominated by current and legacy surface mining activities it is possible that the shift to a primary production dominated system with influxes of excess nutrients can lead to an inflation of fish production (Taylor et al., 2014) and potentially provide refuge from other surface mining related stressors.

The use of trait-based indices of community structure is a promising approach to help establish relationships between landscape stressors and fish community response. Predictive modeling of fish community assemblage structure, using boosted regression tree modeling, can be used to establish a more accurate reference condition for bioassessment development. In addition, the predictive models allow for a wider variety of natural stream conditions to be evaluated under fewer assessment criteria. However, care should be taken in order to select trait-based indices that are responsive to major stressors of a region and to establish predictive model thresholds. Accounting for larger environmental processes and considering the influences of natural variation on fish trait assemblage structure, the deviations from the expected natural condition can now be linked directly to anthropogenic degradation.

## Literature Cited

- Austen, D.J., P.B. Bayley, and B.W. Menzel. 1994. Importance of the guild concept to fisheries research and management. *Fisheries*. 19:12 – 20.
- Bernhardt, E. S. and M.A. Palmer, 2011. The environmental costs of mountaintop mining and valley fill operations for aquatic ecosystems of the Central Appalachians. *Annals of the New York Academy of Sciences*. 1223:39 – 57.
- Blanchet, S., M.R. Helmus, S. Brosse, and G. Grenouillet. 2014. Regional vs. local drivers of Phylogenetic species diversity in stream fish communities. *Freshwater Biology*. 59:450 – 462.
- Cao, Y., C.P. Hawkins, J. Olson, and M.A. Kosterman. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *Journal of North American Benthological Society*. 26:566 – 585.
- Daniel, W.M., D.M. Infante, R.M. Hughes, Y.P. Tsang, P.C. Esselman, D. Wieferich, K. Herreman, A.R. Cooper, L. Wang, W.W Taylor. 2014. Characterizing coal and mineral mines as a regional source of stress on stream fish assemblages. *Ecological Indicators*. 50:50 – 61.
- Daniels, R.A., K. Riva-Murray, D.B. Halliwell, D.L. Vana-Miller, and M.D. Bilger. 2002. An index of biotic integrity for Northern Mid-Atlantic Slope drainages. *Transactions of the American Fisheries Society*. 131:1044 – 1060.
- De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology*. 88:243 – 251.
- Detenbeck, N. E., and D. A. Cincotta. 2008. Comparability of a region and state survey: effects on fish IBI assessment for West Virginia, U.S.A. *Hydrobiologia*. 603:279 – 300.
- Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*. 77:802 – 813.
- 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*. 132:791 – 808.
- Esselman, P.C., D.M. Infante, L. Wang, A.R. Cooper, D. Wieferich, Y.P. Tsang, D.J. Thornbrugh, and W.W. Taylor. 2013. Regional fish community indicators of landscape disturbance to catchments of the conterminous United States. *Ecological Indicators*. 26:163 – 173.
- Fleishman, E., J.R. Thomson, E.L. Kalies, B.G. Dickson, D.S. Dobkin, and M. Leu. 2014. Projecting current and future location, quality, and connectivity of habitat for breeding birds in the Great Basin. *Ecosphere*. 5:1 – 29.
- Gagic, V., I. Bartomeus, T. Jonsson, A. Taylor, C. Winqvist, C. Fischer, E.M. Slade, et al., 2015. Functional identity and diversity of animals predict ecosystem functioning better than species-based indices. *Proceeding of the Royal Society Biological Sciences*. 282:20142620.

- Gotelli, N.J. and A.M. Ellison. 2004. *A Primer of Ecological Statistics*. Sinauer Associates, Inc., Sunderland, Massachusetts, U.S.A.
- Hafs, A.W., C.D. Horn, P.M. Mazik, and K.J. Hartman. 2010. Influences of acid mine drainage and thermal enrichment on stream fish reproduction and larval survival. *Northeastern Naturalist*. 17:575 – 592.
- Hartman, K.J., J.D. Kaller, J.W. Howell, and J.A. Sweka. 2005. How much do valley fills influence headwater streams? *Hydrobiologia*. 532:1 – 19.
- Hawkins, C.P., Y. Cao, and B. Roper. 2010. Method of predicting reference condition biota affects performance and interpretation of ecological indices. *Freshwater Biology*. 55:1066 – 1085.
- Herlihy, A.T., P.R. Kaufmann, M.E. Mitch, and D.D. Brown. 1990. Regional estimates of acid mine drainage impact on streams in the Mid-Atlantic and Southeastern United States. *Water, Air, and Soil Pollution*. 50:91 – 107.
- Hitt, N.P., and P.L. Angermeier. 2011. Fish community and bioassessment responses to stream network position. *Journal of the North American Benthological Society*. 30:296 – 309.
- Hitt, N.P., and D.B. Chambers. 2014. Temporal changes in taxonomic and functional diversity of fish assemblages downstream from mountaintop mining. *Freshwater Science*. 33:000 – 000. DOI: 10.1086/676997.
- Hoeinghaus, D.J., K.O. Winemiller, and J.S. Birnbaum. 2007. Local and regional determinants of stream fish assemblage structure: inferences based on taxonomic vs. functional groups. *Journal of Biogeography*. 34:324 – 338.
- Ibanez, C., J. Belliard, R.M. Hughes, P. Irz, A. Kamdem-Toham, N. Lamouroux, P.A. Tedesco, and T. Oberdorff. 2009. Convergence of temperate and tropical stream fish assemblages. *Ecography*. 32:658 – 670.
- Johnson, R. A. 1981. Application of the guild concept to environmental impact analysis of terrestrial vegetation. *Journal of Environmental Management*. 13:205 – 222.
- Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries*. 6:21 – 27.
- Kimmel, W.G. and D.G. Argent. 2006. Development and application of an index of biotic integrity for fish communities of wadeable Monongahela River tributaries. *Journal of Freshwater Ecology*. 21:183 – 190.
- Lyons, J. 2006. A fish-based index of biotic integrity to assess intermittent headwater streams in Wisconsin, USA. *Environmental Monitoring and Assessment*. 122:239 – 258.
- Lyons, J. 2012. Development and validation of two fish-based indices of biotic integrity for assessing perennial coolwater streams in Wisconsin, USA. *Ecological Indicators*. 23:402 – 412.

- Lyons, J., R.R. Piette, and K.W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warmwater rivers. *Transactions of the American Fisheries Society*. 130:1077 – 1094.
- Lyons, J., L. Wang, and T.D. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management*. 16:241 – 256.
- McCormick, F.H., R.M. Hughes, P.R. Kaufman, D.V. Peck, J.L. Stoddard, and A.T. Herlihy. 2001. Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands Region. *Transactions of the American Fisheries Society*. 130:857 – 877.
- Merovich, G.T., J.T. Petty, M.P. Strager, and J.B. Fulton. 2013. Hierarchical classification of stream condition: a house-neighborhood framework for establishing conservation priorities in complex riverscapes. *Freshwater Science*. 32:874 – 891.
- Merriam, E.R., J.T. Petty, G.T. Merovich, J.B. Fulton, and M.P. Strager. 2011. Additive effects of mining and residential development on stream conditions in a central Appalachian watershed. *Journal of North American Benthological Society*. 30: 399 – 418.
- Oberdorff, T., D. Pont, B. Hugueny, and J.P. Porchers. 2002. Development and validation of a fish-based index for the assessment of 'river health' in France. *Freshwater Biology*. 47:1720 – 1734.
- Oliveria, J.M., P. Segurado, J.M. Santos, A. Teixeira, M.T. Ferreira, and R.V. Cortes. 2012. Modelling stream-fish functional traits in reference conditions: regional and local environmental correlates. *Plos One*. 7:1 – 11.
- Pease, A.A., A.A. Gonzalez-Diaz, R. Rodiles-Hernandez, and K.O. Winemiller. 2012. Functional diversity and trait-environmental relationships stream fish assemblages in a large tropical catchment. *Freshwater Biology*. 57:1060 – 1075.
- Pease, A.A., J.M. Taylor, K.O. Winemiller, and R.S. King. 2011. Multiscale environmental influences on fish assemblage structure in Central Texas streams. *Transactions of the American Fisheries Society*. 140:1409 – 1427.
- Pease, A.A., J.M. Taylor, K.O. Winemiller, and R.S. King. 2015. Ecoregional, catchment, and reach-scale environmental factors shape functional-trait structure of stream fish assemblages. *Hydrobiologia*. 753:265 – 283.
- Poff, N.L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*. 16:391 – 409.
- Pond, G.J., J.E. Bailey, B.M. Lowman, and M.J. Whitman. 2012. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. *Environmental Monitoring and Assessment*. DOI 10.1007/s10661-012-2648-3.
- Pont, D., R.M. Hughes, T.R. Whittier, and S. Schmutz. 2009. A predictive index of biotic integrity model for aquatic-vertebrate assemblages of Western U.S. streams. *Transactions of the American Fisheries Society*. 138:292 – 305.

- Pont, D., B. Hugueny, and C. Rogers. 2007. Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. *Fisheries Management and Ecology*. 14:427 – 439.
- Resetarits, W.J. 2005. Habitat selection behavior links local and regional scales in aquatic systems. *Ecology Letters*. 8:480 – 486.
- Roth, N.E., J.D. Allan, and D.L. Erickson. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology*. 11:141 – 156.
- Root, R. B. 1967. The niche exploitation pattern of the blue-gray gnatcatcher. *Ecol. Monog.* 37: 317-350.
- Strager, M.P., J.T. Petty, J.M. Strager and J. Barker-Fulton. 2009. A spatially explicit framework for quantifying downstream hydrologic conditions. *Journal of Environmental Management*. 90:1854–1861.
- Taylor, J.M., R.S. King, A.A. Pease, and K.O. Winemiller. 2014. Nonlinear response of stream ecosystem structure to low-level phosphorous enrichment. *Freshwater Biology*. 59:969 – 984.
- Wang, L., J. Lyons, P. Kanehl, and R. Bannerman. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. *Environmental Management*. 28: 255 – 266.
- Wang, L., J. Lyons, P. Rasmussen, P. Seelbach, T. Simon, M. Wiley, P. Kanehl, E. Baker, S. Niemela, and P.M. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest ecoregion, U.S.A.. *Canadian Journal of Fisheries and Aquatic Sciences*. 60: 491 – 505.
- Wang, L., J. Lyons, and P. Kanehl. 2002. Effects of watershed best management practices on habitat and fish in Wisconsin streams. *Journal of the American Water Resources Association*. 38:663 – 680.
- Wang, L., D.M. Robertson, and P.J. Garrison. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. 2007. *Environmental Management*. 39:194 – 212.



**Tables**

**Table 1:** Water quality and habitat based selection criteria were used to determine the reference pool. Primary criteria were used in the initial reference site selection conducted by the West Virginia DEP. Secondary criteria were used in order to selection additional reference sites from the pool of previously sampled sites following the methods of Anderson and Petty (2015, Chapter 2).

<b>Reference Criteria</b>	<b>Primary</b>	<b>Secondary</b>
<b>NPDES Point Source</b>	None	--
<b>Dissolved Oxygen (mg/L)</b>	$\geq 5.0$	--
<b>pH (Std. Units)</b>	6.0 – 9.0	6.0 – 9.0
<b>Specific Conductance (<math>\mu\text{mhos/cm}</math>)</b>	< 500	< 500
<b>Fecal Coliform Bacteria (colonies/100mL)</b>	<800	--
<b>State WQ Violations</b>	none	--
<b>U. S. EPA-RBP VBHA metric scores:</b>		
<b>Epifaunal substrate, channel alteration, sediment deposition</b>	$\geq 11$	--
<b>Bank vegetative protection, riparian vegetative zone width</b>	$\geq 6$	--
<b>Total RBP habitat score</b>	$\geq 130$	--
<b>Probability of Reference</b>	--	$\geq 0.90$

**Table 2:** Summary statistics and samples sizes (# Sites) for all reference and non-reference sites within each biomonitoring region.

<b>Reference</b>	<b># Sites</b>	<b>Drainage Area (km<sup>2</sup>)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>
<b>Mon CARV</b>				
Reference	83	8.07 – 357.6 (56.21)	554 – 1428 (831.09)	0 – 208.06 (42.25)
Non-Reference	155	7.47 – 309.4 (43.69)	397 – 1419 (701.28)	0 – 120.08 (29.57)
<b>Ohio CA</b>				
Reference	40	7.34 – 307.98 (55.22)	301 – 897 (525.68)	0 – 88.75 (32.47)
Non-Reference	206	7.29 – 363.50 (64.89)	216 – 880 (521.62)	0 – 112.59 (31.19)
<b>Ohio Mon WAP</b>				
Reference	40	7.38 – 301.6 (74.35)	268 – 602 (385.74)	0 – 78.36 (31.11)
Non-Reference	217	7.43 – 384.85 (91.24)	201 – 603 (361.08)	0 – 137.51 (31.05)
<b>Upper Kanawha</b>				
Reference	44	8.93 – 383.0 (87.72)	447 – 1330 (911.32)	0 – 135.93 (37.35)
Non-Reference	152	7.01 – 392.9 (84.72)	267 – 1314 (785.79)	0 – 134.68 (37.62)

**Table 3:** Boosted Regression Tree model results for selected metrics within each biomonitoring region.

	<b>Drainage Area</b>	<b>Elevation</b>	<b>Swim Distance</b>		<b>CV</b>
<b>Metrics</b>	<b>(km<sup>2</sup>)</b>	<b>(m)</b>	<b>(km)</b>	<b>Dev. Exp.</b>	<b>Dev.</b>
<b>Mon CA-RV</b>					
% Benthic-Tol	30.42	34.44	35.14	0.225	0.000
% Fish-Tol	26.38	--	73.62	0.157	0.059
% IN	53.78	22.30	23.92	0.296	0.127
Benthic	80.48	19.52	--	0.442	0.305
Cyprinid	82.00	18.00	--	0.346	0.191
Richness-Tol	82.56	17.44	--	0.637	0.515
CGS	71.45	--	28.55	0.280	0.129
<b>Ohio CA</b>					
% IN	64.13	25.55	10.32	0.813	0.594
% Tol	65.77	23.16	11.06	0.787	0.532
DMS	73.71	26.29	--	0.664	0.449
Richness-Tol	72.27	27.73	--	0.789	0.654
NGL	73.07	26.93	--	0.677	0.413
RGS	46.47	53.53	--	0.555	0.239
<b>Ohio-Mon WAP</b>					
% OH-CAAN	30.56	69.44	--	0.462	0.192
% Tol	60.56	39.44	--	0.483	0.241
Cyprinid-BNDSEAT	83.92	16.08	--	0.678	0.506
DMS	94.15	--	5.85	0.736	0.589
Richness-Tol	88.04	11.96	--	0.713	0.532
Int	95.73	2.37	1.90	0.454	0.270

NGL-Tol	89.90	3.58	6.52	0.505	0.301
RGS	88.19	6.54	5.27	0.305	0.146
<b>Upper Kanawha</b>					
% IN	71.21	13.57	15.22	0.563	0.282
% Tol	53.76	15.84	53.76	0.444	0.048
Benthic	89.22	5.00	5.78	0.476	0.355
CGS+RGS	84.93	15.07	--	0.631	0.451
Cyprinid	69.35	30.65	--	0.661	0.354
Richness	76.46	23.54	--	0.753	0.381
Int	71.54	28.46	--	0.518	0.245

**Table 4:** Boosted Regression Tree model results (Predictive Models) and Monte Carlo (MC) Simulation results for the top trait-based metrics within each biomonitoring region. The 95<sup>th</sup> percentile of the MC simulations were used a thresholds to determine if predictive models were used for each metric type.

Metrics	Predictive Models		Monte Carlo Simulation			
	Total Deviance Explained	C.V. Deviance Explained	Average Total Dev. Explained	Average C.V. Dev. Explained	p-value	95th Percentile
<b>Mon CARV</b>						
% LSR-Tol	0.6585	0.3659	0.0430	0.0009	0.0001	0.2135
IP-SEAT	0.6571	0.5201	0.04186	0.0010	0.0000	0.2095
<b>Ohio CA</b>						
% IN	0.8125	0.5938	0.0629	0.0004	0.0000	0.3246
Richness-Tol	0.7888	0.6535	0.04828	-0.0004	0.0000	0.2587
<b>Ohio Mon WAP</b>						
% Cyprinid-BNDSEAT	0.8636	0.6818	0.0623	0.0008	0.000	0.3218
IP-Tol	0.8077	0.556	0.0606	0.0014	0.0000	0.3087
<b>UK</b>						
% NGL-Tol	0.7297	0.5676	0.0576	-0.0001	0.0001	0.3032
IP	0.76	0.5209	0.0522	0.0009	0.0000	0.2769

**Table 5:** Selected metrics within each biomonitoring regions were correlated against specific conductance (SPC), pH, structure density (Str. Density) and % surface mining. Metrics were adjusted (observed/expected) using their expected condition given natural variation based on BRT models.

<b>WV IBI Metrics</b>	<b>SPC</b>	<b>pH</b>	<b>Str. Density</b>	<b>% Surface Mining</b>
<b>Mon CA-RV</b>				
%Benthic-Tol	-0.3372	0.0038	-0.1889	-0.2998
%Fish-Tol*	-0.1885	-0.1611	-0.2581	-0.1969
% IN	-0.0248	-0.0019	-0.0844	-0.0823
Benthic	-0.4105	0.4176	-0.0201	-0.2526
Cyprinid	-0.3409	0.4245	0.0535	-0.2183
Richness-Tol	-0.4244	0.4186	-0.0734	-0.2823
CGS	-0.4599	0.3925	-0.1759	-0.2896
<b>Ohio CA</b>				
% IN	-0.0344	0.1135	0.2385	-0.1293
% Tol	-0.1034	-0.2026	-0.1618	0.0751
DMS	-0.4101	0.0824	0.3841	-0.4665
Richness-Tol	-0.2914	0.1395	0.4861	-0.4221
NGL	-0.1998	0.2118	0.5174	-0.3369
RGS	-0.3037	0.1822	0.4361	-0.3862
SW-Trophic	0.0478	0.1111	0.3389	-0.2742
<b>Ohio-Mon WAP</b>				
% OH-CAAN	0.0085	0.1291	0.0636	-0.0014
% Tol	0.0506	0.1695	0.0407	-0.0778
Cyprinid-BNDSEAT	-0.1432	-0.0389	-0.0947	-0.1571
DMS	-0.2628	-0.1036	-0.2826	-0.0868
Richness-Tol	-0.2715	-0.1390	-0.1916	-0.1696

Int	-0.2368	-0.0572	-0.2490	-0.1090
NGL-Tol	-0.2027	-0.0823	-0.1949	-0.1250
RGS*	-0.1607	-0.0386	-0.2291	-0.1307
<b>Upper Kanawha</b>				
% IN	-0.0538	0.0427	-0.1605	0.0330
% Tol	0.1412	0.0257	0.2410	-0.0070
Benthic	-0.1560	0.1113	-0.1576	0.0138
CGS+RGS	-0.1769	0.1158	-0.0968	0.0729
Cyprinid	-0.0464	0.1301	0.0370	-0.0213
Richness	-0.0185	0.1366	-0.0139	-0.0129
Int	-0.1785	0.1928	0.0001	-0.2415

\*Trait-based metric was not adjusted using BRT model predictions.

Figures

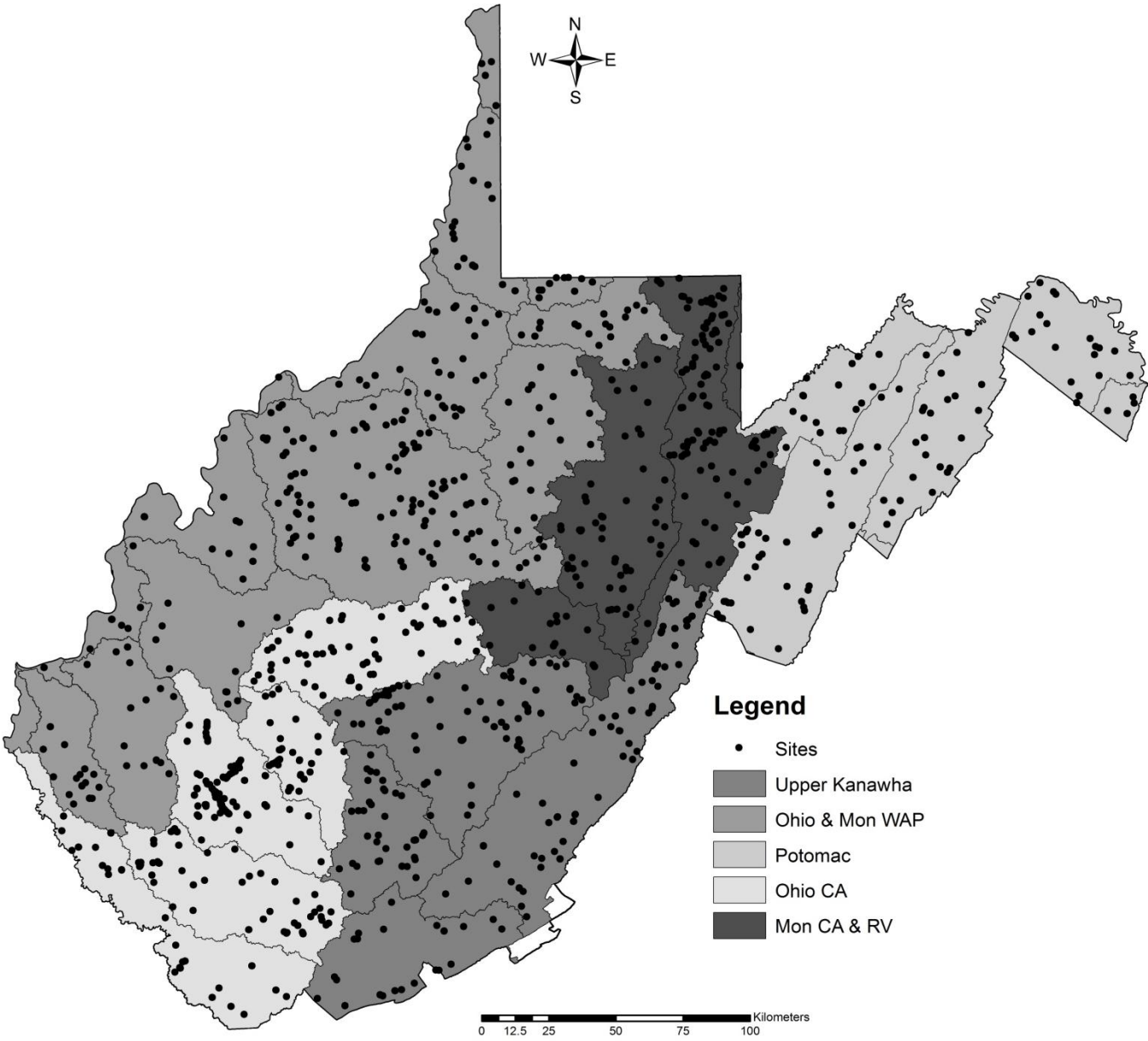
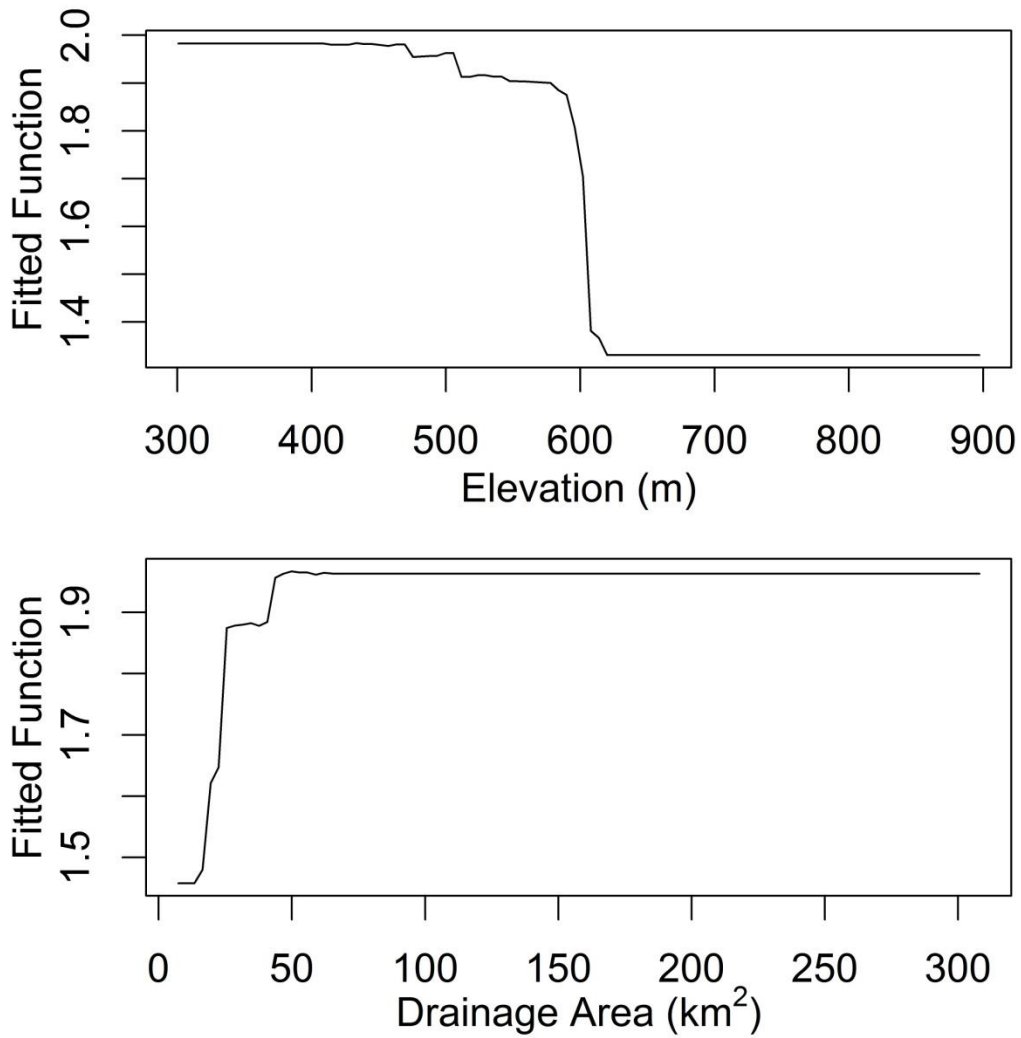


Figure 1: Fish biomonitoring regions of West Virginia with sampling locations (sites) indicated.





**Figure 2:** Fitted function curves for elevation (m) and drainage area (km<sup>2</sup>) from fitted boosted regression tree model for predicting rock-gravel spawner richness in the Ohio CA biomonitoring region.

## Appendix A

**Table 1:** Full Boosted Regression Tree model results for each metrics evaluated in the Mon CA-RV region.

Metrics	Drainage Area (km <sup>2</sup> )	Elevation (m)	Swim Distance (km)	Dev. Exp.	CV Dev.	se
P_Benthic	48.02	--	51.98	0.200	0.014	0.014
P_Benthic_CACO	49.02	--	50.97	0.290	0.044	0.010
P_Benthic2.DEP	30.42	34.44	35.14	0.225	0.000	0.006
P_BND_CACO_SEAT	38.71	--	61.29	0.360	0.200	0.005
P_Catfish	--	--	--	--	--	--
P_Catostomidae	--	--	--	--	--	--
P_CavitySpawn	31.79	36.03	32.17	0.209	0.023	0.005
P_CavitySpawn2.DEP	30.84	35.27	33.89	0.205	0.023	0.005
P_CGS_RGS	--	47.09	52.91	0.213	0.021	0.007
P_CGS_RGS2.DEP	58.05	23.13	18.82	0.037	0.000	0.013
P_Cold	51.63	23.37	24.99	0.372	0.167	0.016
P_Cold_SATR_ONMY	48.44	24.63	26.93	0.423	0.225	0.012
P_Cold2.DEP	51.28	23.11	25.62	0.244	0.073	0.017
P_Cottid	34.29	32.88	32.82	0.250	0.023	0.005
P_Cyprinid	39.32	29.22	31.46	0.214	0.036	0.007
P_Cyprinid_BNDSEAT	67.89	16.67	15.44	0.537	0.352	0.004
P_Cyprinid_NBNDSEAT	70.36	15.89	13.75	0.509	0.359	0.004
P_Cyprinid2.DEP	71.36	16.55	12.09	0.406	0.250	0.003
P_CyprinidN	58.57	19.92	21.51	0.231	0.062	0.007
P_CyprinidN2.DEP	77.03	22.97	--	0.406	0.219	0.003
P_DMS	--	--	--	--	--	--
P_DMS2.DEP	30.02	38.37	31.61	0.150	0.000	0.005
P_Fish2.DEP	26.38	--	73.62	0.157	0.059	0.006
P_Game	--	--	--	--	--	--
P_Game2.DEP	--	--	--	--	--	--
P_GameC	--	--	--	--	--	--
P_GSS	--	--	--	--	--	--
P_GSS2.DEP	--	--	--	--	--	--
P_IN	53.78	22.30	23.92	0.296	0.127	0.009
P_Int.DEP	24.38	39.36	36.27	0.167	0.000	0.011
P_Int_Benthic.DEP	--	--	--	--	--	--
P_Int_Cyprinid.DEP	63.33	17.85	18.82	0.400	0.100	0.004
P_Int_LSR.DEP	69.77	30.23	--	0.400	0.200	0.003
P_Int_NGL.DEP	--	--	--	--	--	--
P_Int_RGS.DEP	72.15		27.85	0.500	0.250	0.002
P_IP	19.22	15.64	65.14	0.120	0.120	0.007

Metrics	Drainage Area (km2)	Elevation (m)	Swim Distance (km)	Dev. Exp.	CV Dev.	se
P_IP_BenthicNG	--	--	--	--	--	--
P_IP_NonGameNB	81.71	9.46	8.82	0.317	0.171	0.113
P_IP_SEAT	13.60	14.24	72.16	0.093	0.093	0.006
P_IP2.DEP	15.35	16.52	68.13	0.118	0.020	0.004
P_LSR	63.98	18.27	17.75	0.342	0.190	0.006
P_LSR2.DEP	69.16	--	30.84	0.659	0.366	0.006
P_McC_CGS	38.33	30.01	31.66	0.212	0.039	0.006
P_McC_CGS2.DEP	--	--	--	--	--	--
P_MO	35.85	--	64.15	0.231	0.058	0.007
P_Mod.DEP	--	44.58	55.42	0.291	0.091	0.008
P_Native	31.11	21.05	47.85	0.420	0.140	0.005
P_Native2.DEP	30.12	--	69.88	0.120	0.000	0.006
P_NGL	66.96	--	33.04	0.333	0.333	0.001
P_NGL2.DEP	82.16	17.84	--	0.333	0.333	0.001
P_OH	16.88	15.54	67.59	0.080	0.080	0.005
P_OH_CAAN	26.81	13.46	59.73	0.205	0.045	0.005
P_OH_CAAN_CACO	32.13	13.81	54.07	0.209	0.047	0.007
P_OH_NG	16.78	14.29	68.94	0.080	0.000	0.004
P_OH2.DEP	--	--	--	--	--	--
P_Percidae	--	--	--	--	--	--
P_RGS	74.09	12.37	13.54	0.578	0.391	0.005
P_RGS2.DEP	69.83		30.17	0.651	0.395	0.005
P_Sunfish	--	--	--	--	--	--
P_Tol.DEP	--	27.84	72.16	0.177	0.059	0.003
P_Tol_Benthic.DEP	38.89	--	61.11	0.231	0.077	0.005
P_Tol_Cyprinid.DEP	32.64	--	67.36	0.160	0.040	0.004
R_Benthic	80.48	19.52	--	0.442	0.305	0.212
R_Benthic_CACO	82.77	17.23	--	0.450	0.289	0.209
R_Benthic2.DEP	82.47	10.56	6.97	0.555	0.426	0.113
R_CavitySpawn	63.35	19.03	17.62	0.231	0.462	0.123
R_CGS_RGS	83.47	16.53	--	0.457	0.347	0.172
R_CGS_RGS2.DEP	84.69	8.45	6.86	0.487	0.361	0.187
R_Cold	31.79	27.27	40.94	0.309	0.060	0.095
R_Cold2.DEP	51.03	--	48.97	0.223	0.041	0.087
R_Cyprinid	82.00	18.00	--	0.346	0.191	0.079
R_Cyprinid_BNDSEAT	84.84	8.08	7.09	0.419	0.296	0.274
R_Cyprinid_NBNDSEAT	88.97	11.03	--	0.418	0.300	0.175
R_Cyprinid2.DEP	84.66	--	15.34	0.364	0.251	0.105
R_CyprinidN	82.47	17.53	--	0.387	0.261	0.216
R_CyprinidN2.DEP	83.98	8.94	7.08	0.320	0.202	0.129
R_DMS	80.97	10.32	8.71	0.498	0.317	0.176

Metrics	Drainage Area (km <sup>2</sup> )	Elevation (m)	Swim Distance (km)	Dev. Exp.	CV Dev.	se
R_DMS2.DEP	79.87	10.73	9.40	0.508	0.368	0.119
R_FISH	76.11	23.89	--	0.580	0.437	0.231
R_Fish2.DEP	82.56	17.44	--	0.637	0.515	0.176
R_Game	52.60	47.40	--	0.505	0.341	0.109
R_Game2.DEP	52.73	47.27	--	0.384	0.202	0.112
R_GameC	53.05	46.95	--	0.321	0.118	0.108
R_GSS	58.30	41.70	--	0.254	0.096	0.108
R_GSS2.DEP	--	--	--	--	--	--
R_IN	85.86	14.14	--	0.536	0.414	0.351
R_Int.DEP	76.13	12.84	11.02	0.361	0.208	0.072
R_Int_Benthic.DEP	81.91	10.97	7.11	0.464	0.267	0.130
R_Int_LSR.DEP	88.94	--	11.06	0.279	0.175	0.053
R_Int_RGS.DEP	86.56	13.44	--	0.469	0.355	0.062
R_IP	77.18	22.82	--	0.630	0.443	0.202
R_IP_BenthicNG	86.83	7.87	5.30	0.468	0.350	0.158
R_IP_NonGameNB	81.47	18.53	--	0.309	0.193	0.178
R_IP_SEAT	79.26	20.74	--	0.658	0.534	0.095
R_IP2.DEP	79.84	20.16	--	0.648	0.514	0.120
R_LSR	79.29	20.71	--	0.363	0.227	0.179
R_LSR2.DEP	85.80	14.20	--	0.415	0.309	0.130
R_McC_CGS	71.45	--	28.55	0.280	0.130	0.077
R_McC_CGS2.DEP	74.01	--	25.99	0.342	0.210	0.088
R_MO	52.54	22.99	24.47	0.175	0.036	0.089
R_Mod.DEP	79.83	20.17	--	0.623	0.505	0.083
R_Native	79.07	20.93	--	0.545	0.397	0.252
R_Native2.DEP	82.02	17.98	--	0.585	0.465	0.180
R_NGL	74.71	25.29	--	0.396	0.279	0.165
R_NGL2.DEP	75.58	24.42	--	0.444	0.288	0.127
R_OH	54.98	26.64	18.39	0.273	0.107	0.112
R_OH_CAAN	40.79	30.54	28.66	0.223	0.026	0.096
R_OH_CAAN_CACO	47.36	33.60	19.04	0.173	0.042	0.065
R_OH_NG	57.31	25.75	16.94	0.251	0.072	0.130
R_Percidae	87.72	12.28	--	0.491	0.376	0.187
R_RGS	83.93	16.07	--	0.474	0.355	0.191
R_RGS2.DEP	86.01	13.99	--	0.491	0.385	0.170
R_Tol.DEP	63.59	36.41	--	0.331	0.183	0.099
R_Tol_Cyprinid.DEP	68.72	31.28	--	0.251	0.112	0.148

**Table 2:** Full Boosted Regression Tree model results for each metrics evaluated in the Ohio CA region.

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
P_Benthic	31.94	47.03	21.03	0.500	0.121	0.009
P_Benthic_CACO	30.64	48.46	20.89	0.397	0.103	0.011
P_Benthic2.DEP	--	--	--	--	--	--
P_BND_CACO_SEAT	58.4	31.56	10.04	0.784	0.602	0.008
P_Catfish	--	--	--	--	--	--
P_Catostomidae	--	--	--	--	--	--
P_CavitySpawn	--	--	--	--	--	--
P_CavitySpawn2.DEP	49.35	50.65	--	0.188	0.000	0.007
P_CGS_RGS	41.61	13.78	44.61	0.180	0.020	0.013
P_CGS_RGS2.DEP	57.38	42.62	--	0.625	0.300	0.006
P_Cold	16.09	60.58	23.32	0.549	0.294	0.014
P_Cold_SATR_ONMY	14.89	66.79	18.32	0.500	0.208	0.015
P_Cold2.DEP	--	--	--	--	--	--
P_Cottid	--	--	--	--	--	--
P_Cyprinid	--	--	--	--	--	--
P_Cyprinid_BNDSEAT	63.36	36.64	--	0.812	0.635	0.006
P_Cyprinid_NBNDSEAT	63.06	36.94	--	0.800	0.647	0.004
P_Cyprinid2.DEP	74.88	25.12	--	0.737	0.526	0.004
P_CyprinidN	--	--	--	--	--	--
P_CyprinidN2.DEP	73.58	26.42	--	0.737	0.526	0.005
P_DMS	--	--	--	--	--	--
P_DMS2.DEP	--	--	--	--	--	--
P_Fish2.DEP	66.73	22.62	10.65	0.787	0.489	0.005
P_Game	--	--	--	--	--	--
P_Game2.DEP	--	--	--	--	--	--
P_GameC	--	--	--	--	--	--
P_GSS	88.46	--	11.54	0.429	0.262	0.008
P_GSS2.DEP	15.92	84.08	--	0.500	0.000	0.000
P_IN	64.13	25.55	10.32	0.813	0.594	0.005
P_Int.DEP	93.51	6.49	--	0.467	0.333	0.003
P_Int_Benthic.DEP	--	--	--	--	--	--
P_Int_Cyprinid.DEP	85.85	14.15	--	0.417	0.167	0.002
P_Int_GSS.DEP	--	--	--	--	--	--
P_Int_LSR.DEP	86.32	13.68	--	0.500	0.250	0.002
P_Int_NGL.DEP	--	--	--	0.000	0.000	0.000
P_Int_RGS.DEP	89.45	--	10.55	0.286	0.071	0.004
P_IP	24.47	45.09	30.44	0.463	0.146	0.007
P_IP_BenthicNG	--	--	--	--	--	--
P_IP_NonGameNB	28.77	52.42	18.8	0.395	0.070	0.008

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
P_IP_SEAT	66.43	23.24	10.33	0.790	0.597	0.004
P_IP2.DEP	65.28	23.73	10.99	0.500	0.783	0.004
P_LSR	8.91	91.09	--	0.431	0.255	0.011
P_LSR2.DEP	60.37	39.63	--	0.778	0.578	0.004
P_McC_CGS	--	--	--	--	--	--
P_McC_CGS2.DEP	41.89	58.1	--	0.412	0.235	0.002
P_MO	24.95	43.11	31.94	0.488	0.122	0.008
P_Mod.DEP	54.58	31.53	13.89	0.680	0.320	0.005
P_Native	76.11	--	23.89	0.565	0.326	0.007
P_Native2.DEP	64.54	24.05	11.41	0.792	0.542	0.005
P_NGL	55.29	34.39	10.32	0.800	0.400	0.001
P_NGL2.DEP	76.25	--	23.75	0.500	0.250	0.001
P_OH	--	--	--	--	--	--
P_OH_CAAN	--	--	--	--	--	--
P_OH_CAAN_CACO	--	--	--	--	--	--
P_OH_NG	24.36	46.56	29.08	0.439	0.146	0.008
P_OH2.DEP	--	--	--	0.000	0.000	0.000
P_Percidae	--	--	--	--	--	--
P_RGS	57.01	42.99	--	0.613	0.307	0.012
P_RGS2.DEP	57.53	42.47	--	0.600	0.300	0.005
P_Sunfish	--	--	--	--	--	--
P_Tol.DEP	65.77	23.16	11.06	0.787	0.532	0.004
P_Tol_Benthic.DEP	19.94	49.65	30.41	0.478	0.174	0.010
P_Tol_Cyprinid.DEP	64.71	22.38	12.91	0.781	0.512	0.005
R_Benthic	60.86	39.14	--	0.673	0.467	0.181
R_Benthic_CACO	61.61	38.39	--	0.725	0.482	0.242
R_Benthic2.DEP	69.37	30.63	--	0.721	0.478	0.238
R_Catostomidae	74.85	25.15	--	0.182	0.037	0.089
R_CavitySpawn	45.66	39.44	14.89	0.494	0.224	0.185
R_CGS_RGS	42.51	57.49	--	0.606	0.331	0.223
R_CGS_RGS2.DEP	49.77	50.23	--	0.574	0.285	0.292
R_Cyprinid	41.7	45.11	13.19	0.721	0.422	0.118
R_Cyprinid_BNDSEAT	48.12	36.04	15.84	0.683	0.482	0.280
R_Cyprinid_NBNDSEAT	48.03	36	15.97	0.677	0.415	0.430
R_Cyprinid2.DEP	40.28	19.85	39.87	0.715	0.353	0.192
R_CyprinidN	41.24	45.99	12.77	0.721	0.470	0.094
R_CyprinidN2.DEP	40.12	20.23	39.65	0.692	0.321	0.212
R_DMS	73.71	26.29	--	0.664	0.449	0.167
R_DMS2.DEP	75.75	24.25	--	0.642	0.408	0.245
R_FISH	61.7	38.3	--	0.745	0.544	0.174
R_Fish2.DEP	72.27	27.73	--	0.789	0.654	0.167

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
R_Game	72.8	27.2	--	0.446	0.236	0.128
R_Game2.DEP	84.15	15.85	--	0.587	0.424	0.111
R_GameC	66.97	33.03	--	0.418	0.181	0.121
R_GSS	25.86	74.14	--	0.517	0.297	0.154
R_GSS2.DEP	40.29	59.71	--	0.662	0.458	0.200
R_IN	66.41	33.59	--	0.746	0.533	0.316
R_Int.DEP	85.45	9.4	5.15	0.497	0.273	0.116
R_Int_Benthic.DEP	53.54	34.08	12.38	0.661	0.326	0.108
R_Int_RGS.DEP	54.76	24.59	20.64	0.473	0.204	0.132
R_IP	72.2	27.8	--	0.774	0.619	0.134
R_IP_BenthicNG	72.38	27.62	--	0.676	0.479	0.220
R_IP_NonGameNB	51.96	34.12	13.92	0.736	0.500	0.155
R_IP_SEAT	74.14	25.86	--	0.773	0.541	0.234
R_IP2.DEP	76.91	23.09	--	0.783	0.615	0.151
R_LSR	49.06	50.94	--	0.707	0.465	0.195
R_LSR2.DEP	62.55	37.45	--	0.729	0.557	0.138
R_McC_CGS	38.25	61.75	--	0.564	0.301	0.100
R_McC_CGS2.DEP	44.67	55.33	--	0.523	0.287	0.236
R_MO	15.64	84.36	--	0.313	0.045	0.074
R_Mod.DEP	64.46	35.54	--	0.763	0.595	0.198
R_Native	58.36	41.64	--	0.742	0.545	0.202
R_Native2.DEP	67.89	32.1	--	0.768	0.550	0.566
R_NGL	73.07	26.93	--	0.677	0.413	0.199
R_NGL2.DEP	68.85	14.7	16.45	0.683	0.447	0.298
R_OH	23.61	76.39	--	0.503	0.208	0.123
R_OH_CAAAN	22.08	77.92	--	0.426	0.171	0.203
R_OH_CAAAN_CACO	20.37	79.63	--	0.556	0.305	0.191
R_OH_NG	19.86	80.14	--	0.280	0.040	0.053
R_Percidae	53.48	31.83	14.69	0.645	0.337	0.179
R_RGS	46.47	53.53	--	0.555	0.239	0.321
R_RGS2.DEP	45.08	43.79	11.13	0.597	0.308	0.264
R_Tol.DEP	32.29	67.71	--	0.414	0.153	0.103
R_Tol_Benthic.DEP	--	--	--	--	--	--
R_Tol_Cyprinid.DEP	38.14	61.86	--	0.590	0.314	0.092

**Table 3:** Full Boosted Regression Tree model results for each metrics evaluated in the Ohio-Mon WAP region.

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
P_Benthic	90.26	4.29	5.45	0.286	0.143	0.003
P_Benthic_CACO	86.77	5.47	7.76	0.250	0.100	0.003
P_Benthic2.DEP	97.00	1.76	1.24	0.167	0.083	0.006
P_BND_CACO_SEAT	94.31	3.25	2.44	0.767	0.651	0.006
P_Catfish	--	--	--	--	--	--
P_Catostomidae	--	--	--	--	--	--
P_CavitySpawn	--	--	--	--	--	--
P_CavitySpawn2.DEP	--	--	--	--	--	--
P_CGS_RGS	31.79	54.11	14.09	0.095	0.048	0.006
P_CGS_RGS2.DEP	32.17	67.83		0.556	0.278	0.003
P_Cold	69.00		31.00	0.375	0.125	0.005
P_Cold_SATR_ONMY	65.80	10.27	23.93	0.375	0.125	0.005
P_Cold2.DEP	--	--	--	--	--	--
P_Cottid	--	--	--	--	--	--
P_Cyprinid	90.71	4.56	4.72	0.182	0.000	0.006
P_Cyprinid_BNDSEAT	93.58	--	6.42	0.837	0.698	0.003
P_Cyprinid_NBNDSEAT	89.08	5.01	5.91	0.864	0.682	0.004
P_Cyprinid2.DEP	82.66	17.34	--	0.556	0.333	0.003
P_CyprinidN	90.51	4.79	4.70	0.182	0.000	0.006
P_CyprinidN2.DEP	81.97	18.03	--	0.556	0.333	0.003
P_DMS	9.41	90.59	--	0.231	0.000	0.008
P_DMS2.DEP	7.33	92.67	--	0.231	0.077	0.007
P_Fish2.DEP	59.28	40.72	--	0.483	0.276	0.008
P_Game	--	--	--	--	--	--
P_Game2.DEP	--	--	--	--	--	--
P_GameC	--	--	--	--	--	--
P_GSS	98.28	--	1.72	0.517	0.276	0.009
P_GSS2.DEP	--	--	--	--	--	--
P_IN	75.63	24.37	--	0.711	0.489	0.007
P_Int.DEP	57.88	42.12	--	0.615	0.385	0.002
P_Int_Benthic.DEP	--	--	--	--	--	--
P_Int_Cyprinid.DEP	52.06	47.94	--	0.571	0.143	0.002
P_Int_GSS.DEP	--	--	--	--	--	--
P_Int_LSR.DEP	59.48	40.52	--	0.429	0.143	0.002
P_Int_NGL.DEP	68.25	31.75	--	0.500	0.500	0.000
P_Int_RGS.DEP	33.50	66.50	--	0.556	0.333	0.002
P_IP	37.30	62.70	--	0.381	0.191	0.003
P_IP_BenthicNG	1.74	97.17	1.09	0.167	0.083	0.007
P_IP_NonGameNB	75.50	24.50	--	0.118	0.000	0.002



Metrics	Drainage Area (km2)	Elevation (m)	Swim Distance (km)	Dev. Exp.	CV Dev.	se
P_IP_SEAT	--	--	--	--	--	--
P_IP2.DEP	63.11	36.89	--	0.516	0.258	0.006
P_LSR	--	--	--	--	--	--
P_LSR2.DEP	77.55	22.45	--	0.667	0.444	0.002
P_McC_CGS	--	--	--	--	--	--
P_McC_CGS2.DEP	27.73	72.27	--	0.500	0.250	0.001
P_MO	36.72	63.28	--	0.381	0.095	0.003
P_Mod.DEP	--	--	--	--	--	--
P_Native2.DEP	60.66	39.34	--	0.483	0.207	0.005
P_NGL	41.00	38.71	20.29	0.375	0.125	0.005
P_NGL2.DEP	76.81	23.19	--	0.667	0.333	0.001
P_OH	19.88	80.12	--	0.438	0.125	0.005
P_OH_CAAN	30.56	69.44	--	0.462	0.192	0.003
P_OH_CAAN_CACO	35.98	64.02	--	0.500	0.250	0.004
P_OH_NG	36.95	63.05	--	0.381	0.143	0.004
P_OH2.DEP	31.78	48.83	19.39	0.500	0.000	0.001
P_Percidae	--	96.42	3.58	0.333	0.167	0.001
P_RGS	57.47	42.53	--	0.333	0.100	0.005
P_RGS2.DEP	26.19	73.81	--	0.556	0.333	0.002
P_Sunfish	--	--	--	--	--	--
P_Tol.DEP	60.56	39.44	--	0.483	0.241	0.006
P_Tol_Benthic.DEP	68.76	20.27	10.97	0.412	0.118	0.003
P_Tol_Cyprinid.DEP	54.16	45.84	--	0.444	0.222	0.006
R_Benthic	85.99	14.01	--	0.598	0.429	0.077
R_Benthic_CACO	89.30	10.70	--	0.654	0.451	0.127
R_Benthic2.DEP	94.19	5.81	--	0.627	0.460	0.137
R_Catostomidae	45.16	40.25	14.60	0.444	0.146	0.094
R_CavitySpawn	83.32	16.68	--	0.680	0.466	0.072
R_CavitySpawn2.DEP	82.28	9.68	8.03	0.743	0.528	0.030
R_CGS_RGS	78.96	21.04	--	0.568	0.320	0.076
R_CGS_RGS2.DEP	84.36	15.64	--	0.538	0.329	0.096
R_Cyprinid	75.03	24.97	--	0.709	0.492	0.057
R_Cyprinid_BNDSEAT	83.92	16.08	--	0.678	0.506	0.140
R_Cyprinid_NBNDSEAT	83.49	16.51	--	0.690	0.512	0.119
R_Cyprinid2.DEP	69.83	13.70	16.47	0.571	0.336	0.200
R_CyprinidN	72.16	27.74	--	0.748	0.528	0.080
R_CyprinidN2.DEP	67.59	16.93	15.48	0.619	0.361	0.156
R_DMS	94.15	--	5.85	0.736	0.589	0.036
R_DMS2.DEP	96.66	--	3.34	0.715	0.582	0.050
R_FISH	80.91	19.09	--	0.714	0.467	0.206
R_Fish2.DEP	88.04	11.96	--	0.713	0.532	0.253

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
R_Game	47.72	52.28	--	0.300	0.052	0.348
R_Game2.DEP	51.13	38.47	10.39	0.446	0.125	0.145
R_GameC	38.31	61.69	--	0.379	0.100	0.225
R_GSS	78.16	21.84	--	0.427	0.266	0.078
R_GSS2.DEP	89.93	--	10.07	0.553	0.380	0.101
R_IN	86.21	8.51	5.28	0.788	0.574	0.205
R_Int.DEP	95.73	2.37	1.90	0.454	0.270	0.139
R_Int_Benthic.DEP	94.72	3.19	2.07	0.370	0.225	0.094
R_Int_LSR.DEP	67.99	27.74	4.27	0.302	0.085	0.223
R_Int_NGL.DEP	90.05	6.32	3.63	0.510	0.230	0.185
R_Int_RGS.DEP	87.67	12.33	--	0.261	0.051	0.103
R_IP	86.34	13.66	--	0.750	0.581	0.187
R_IP_BenthicNG	90.28	5.67	4.05	0.739	0.538	0.108
R_IP_NonGameNB	88.79	6.18	5.03	0.716	0.574	0.116
R_IP_SEAT	86.10	13.90	--	0.753	0.564	0.285
R_IP2.DEP	81.60	11.10	7.30	0.808	0.556	0.197
R_LSR	72.07	19.36	8.57	0.634	0.342	0.126
R_LSR2.DEP	86.13	9.49	4.39	0.642	0.433	0.135
R_McC_CGS	72.68	27.32	--	0.402	0.173	0.073
R_McC_CGS2.DEP	87.09	12.91	--	0.409	0.247	0.095
R_MO	53.80	46.20	--	0.393	0.129	0.076
R_Mod.DEP	80.33	19.67	--	0.754	0.567	0.121
R_Native	76.54	23.46	--	0.747	0.538	0.257
R_Native2.DEP	86.58	13.42	--	0.721	0.562	0.212
R_NGL	74.69	10.24	15.06	0.523	0.273	0.110
R_NGL2.DEP	89.90	3.58	6.52	0.505	0.301	0.268
R_OH	8.15	84.21	7.64	0.232	0.071	0.081
R_OH_CAAAN	86.94	6.90	6.16	0.191	0.034	0.101
R_OH_CAAAN_CACO	26.11	73.89	--	0.281	0.069	0.117
R_OH_NG	50.74	38.46	10.81	0.374	0.129	0.080
R_OH2.DEP	21.45	67.06	11.49	0.320	0.263	0.146
R_Percidae	94.54	--	5.46	0.650	0.481	0.041
R_RGS	88.19	6.54	5.27	0.305	0.146	0.058
R_RGS2.DEP	82.79	8.72	8.48	0.301	0.148	0.064
R_Sunfish	37.65	48.52	13.83	0.230	0.006	0.265
R_Tol.DEP	52.07	47.93	--	0.614	0.185	0.072
R_Tol_Benthic.DEP	65.71	34.29	--	0.316	0.092	0.065
R_Tol_Cyprinid.DEP	67.32	32.68	--	0.552	0.185	0.095

**Table 4:** Full Boosted Regression Tree model results for each metrics evaluated in the Upper Kanawha region.

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
P_Benthic	--	21.49	78.51	0.578	0.310	0.012
P_Benthic_CACO	--	22.15	77.85	0.565	0.391	0.013
P_Benthic2.DEP	--	27.52	72.48	0.424	0.212	0.005
P_BND_CACO_SEAT	33.11	13.27	53.62	0.620	0.139	0.017
P_Catfish	--	--	--	--	--	--
P_Catostomidae	--	--	--	--	--	--
P_CavitySpawn	--	16.41	83.59	0.576	0.455	0.006
P_CavitySpawn2.DEP	--	14.99	85.01	0.588	0.471	0.005
P_CGS_RGS	10.10	14.19	75.70	0.389	0.167	0.009
P_CGS_RGS2.DEP	--	--	--	--	--	--
P_Cold	29.99		70.01	0.359	0.141	0.019
P_Cold_SATR_ONMY	28.99		71.01	0.422	0.172	0.015
P_Cold2.DEP	--	--	--	--	--	--
P_Cottid	--	19.05	80.95	0.272	0.091	0.010
P_Cyprinid	50.03	28.19	21.78	0.340	0.085	0.011
P_Cyprinid_BNDSEAT	71.09	11.70	16.21	0.713	0.483	0.007
P_Cyprinid_NBNDSEAT	90.99	9.01	--	0.229	0.086	0.005
P_Cyprinid2.DEP	78.67	10.39	10.94	0.656	0.406	0.006
P_CyprinidN	13.11	32.60	54.28	0.296	0.000	0.010
P_CyprinidN2.DEP	84.55	15.45	--	0.208	0.083	0.005
P_DMS	--	15.57	84.43	0.438	0.250	0.004
P_DMS2.DEP	12.98	17.90	69.12	0.500	0.281	0.006
P_Fish2.DEP	29.45	17.55	53.00	0.508	0.079	0.011
P_Game	38.56	12.04	49.40	0.457	0.143	0.013
P_Game2.DEP	38.58	16.56	44.86	0.516	0.032	0.019
P_GameC	40.70	19.54	39.76	0.452	0.161	0.014
P_GSS	10.91	19.84	69.24	0.259	0.093	0.014
P_GSS2.DEP	57.40	--	42.60	0.730	0.487	0.005
P_IN	71.21	13.57	15.22	0.563	0.282	0.007
P_Int.DEP	--	--	--	--	--	--
P_Int_Benthic.DEP		13.04	86.96	0.400	0.200	0.003
P_Int_Cyprinid.DEP	--	--	--	--	--	--
P_Int_GSS.DEP	--	--	--	--	--	--
P_Int_LSR.DEP	--	--	--	--	--	--
P_Int_NGL.DEP	--	--	--	--	--	--
P_Int_RGS.DEP	17.06	--	82.94	0.400	0.200	0.004
P_IP	18.12	19.69	62.19	0.462	0.077	0.010
P_IP_BenthicNG	--	22.96	77.04	0.471	0.294	0.006
P_IP_NonGameNB	50.99	25.91	23.10	0.646	0.354	0.011

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
P_IP_SEAT	27.26	17.95	54.78	0.583	0.222	0.008
P_IP2.DEP	28.03	18.18	53.79	0.549	0.143	0.010
P_LSR	27.16	25.96	46.88	0.506	0.230	0.011
P_LSR2.DEP	56.61	22.27	21.12	0.750	0.333	0.009
P_McC_CGS	--	--	--	--	--	--
P_McC_CGS2.DEP	--	--	--	--	--	--
P_MO	17.86	15.28	66.87	0.404	0.058	0.012
P_Mod.DEP	61.81	12.32	25.88	0.522	0.209	0.005
P_Native	--	--	--	--	--	--
P_Native2.DEP	--	27.71	72.29	0.222	0.056	0.009
P_NGL	60.72	--	39.28	0.722	0.556	0.004
P_NGL2.DEP	58.56	--	41.44	0.730	0.568	0.004
P_OH	23.09	--	76.91	0.346	0.039	0.011
P_OH_CAAN	19.25	--	80.75	0.455	0.146	0.015
P_OH_CAAN_CACO	16.87	13.02	70.11	0.529	0.196	0.012
P_OH_NG	17.28	18.78	63.95	0.442	0.096	0.010
P_OH2.DEP	--	--	--	--	--	--
P_Percidae	18.51	18.36	63.13	0.214	0.000	0.004
P_RGS	--	--	--	--	--	--
P_RGS2.DEP	--	--	--	--	--	--
P_Sunfish	19.36	--	80.64	0.500	0.000	0.001
P_Tol.DEP	53.76	15.84	53.76	0.444	0.048	0.013
P_Tol_Benthic.DEP	26.79	--	73.21	0.435	0.109	0.017
P_Tol_Cyprinid.DEP	26.76	16.45	56.79	0.475	0.051	0.015
R_Benthic	89.22	5.00	5.78	0.476	0.355	0.237
R_Benthic_CACO	89.00	6.80	4.22	0.528	0.425	0.212
R_Benthic2.DEP	90.46	--	9.54	0.592	0.443	0.135
R_CavitySpawn	46.47	37.18	16.35	0.536	0.223	0.131
R_CavitySpawn2.DEP	46.35	43.67	9.98	0.385	0.195	0.074
R_CGS_RGS	84.93	15.07	--	0.631	0.451	0.202
R_CGS_RGS2.DEP	90.64	9.36	--	0.606	0.466	0.203
R_Cold	30.22	47.11	22.67	0.537	0.255	0.195
R_Cold2.DEP	34.70	65.30	--	0.570	0.247	0.149
R_Cyprinid	69.35	30.65	--	0.661	0.354	0.225
R_Cyprinid_BNDSEAT	81.27	18.73	--	0.624	0.431	0.305
R_Cyprinid_NBNDSEAT	81.52	18.48	--	0.535	0.348	0.332
R_Cyprinid2.DEP	86.97	13.03	--	0.495	0.342	0.181
R_CyprinidN	79.65	20.35	--	0.435	0.248	0.144
R_CyprinidN2.DEP	80.07	19.93	--	0.448	0.264	0.174
R_DMS	76.33	13.48	10.19	0.464	0.224	0.214
R_DMS2.DEP	77.99	22.01	--	0.479	0.330	0.195

<b>Metrics</b>	<b>Drainage Area (km2)</b>	<b>Elevation (m)</b>	<b>Swim Distance (km)</b>	<b>Dev. Exp.</b>	<b>CV Dev.</b>	<b>se</b>
R_FISH	76.46	23.54	--	0.753	0.381	0.321
R_Fish2.DEP	83.35	9.61	7.04	0.682	0.455	0.313
R_Game	57.52	--	42.48	0.597	0.355	0.147
R_Game2.DEP	69.32	--	30.68	0.543	0.347	0.193
R_GameC	61.87	21.03	17.11	0.551	0.225	0.102
R_GSS	89.01	10.98	--	0.258	0.130	0.173
R_GSS2.DEP	86.15	--	13.85	0.509	0.317	0.258
R_IN	90.76	9.24	--	0.628	0.479	0.307
R_Int.DEP	71.54	28.46	--	0.518	0.245	0.244
R_Int_Benthic.DEP	85.27	--	14.73	0.539	0.281	0.122
R_Int_LSR.DEP	71.01	28.99	--	0.282	0.031	0.217
R_Int_RGS.DEP	51.14	24.55	24.31	0.393	0.021	0.185
R_IP	82.00	17.00	--	0.760	0.521	0.282
R_IP_BenthicNG	91.78	--	8.22	0.559	0.412	0.194
R_IP_NonGameNB	71.02	28.98	--	0.643	0.360	0.205
R_IP_SEAT	88.61	5.36	6.03	0.710	0.529	0.306
R_IP2.DEP	88.35	5.57	6.09	0.683	0.506	0.236
R_LSR	67.27	32.73	--	0.602	0.241	0.316
R_LSR2.DEP	91.16	5.70	3.14	0.438	0.280	0.444
R_McC_CGS	78.48	21.52	--	0.589	0.393	0.102
R_McC_CGS2.DEP	92.21	7.79	--	0.586	0.409	0.147
R_MO	--	--	--	--	--	--
R_Mod.DEP	68.32	13.25	18.42	0.718	0.498	0.155
R_Native	79.55	20.45	--	0.563	0.355	0.226
R_Native2.DEP	85.70	14.30	--	0.564	0.405	0.181
R_NGL	92.78	3.71	3.51	0.389	0.213	0.149
R_NGL2.DEP	91.38	--	8.62	0.513	0.324	0.210
R_OH	--	--	--	--	--	--
R_OH_CAAAN	--	--	--	--	--	--
R_OH_CAAAN_CACO	--	--	--	--	--	--
R_OH_NG	--	--	--	--	--	--
R_Percidae	88.14	--	11.86	0.459	0.318	0.089
R_RGS	67.37	32.63	--	0.609	0.342	0.322
R_RGS2.DEP	85.82	14.18	--	0.435	0.258	0.171
R_Tol.DEP	34.12	34.67	31.21	0.569	0.053	0.163
R_Tol_Cyprinid.DEP	29.33	39.79	30.88	0.623	0.097	0.093

**CHAPTER 4: LOCAL AND NEIGHBORHOOD SCALE CONTROLS ON FISH  
COMMUNITY STRUCTURE IN CENTRAL APPALACHIAN WATERSHEDS**

## Abstract

The hierarchical structure of watersheds and stream networks leads to a variety of landscape filters that can ultimately structure fish communities. Varying responses of measures of community structure demonstrate the need to evaluate multiple responses while taking into consideration the hierarchical importance of landscape scale processes. In West Virginia, current surface mining and residential development land-use practices have the potential to drastically alter the hierarchical filters governing species distribution and community assembly. In addition, natural variation due to large scale landscape patterns (i.e., ecoregional differences), spatial autocorrelation, and a long history of land-use alterations make understanding factors controlling fish community composition troublesome. Using a mixed-effects modeling approach, we evaluated the influence of local (cumulative segment-level watershed scale), neighborhood (HUC12 watershed scale), and natural landscape (i.e., drainage area, elevation, and swim distance) factors on species richness, diversity, community condition (West Virginia Index of Biotic Integrity), and composition (proportion of tolerant individuals). Both natural (i.e., drainage area and elevation) and local anthropogenic landscape factors were important in controlling species richness, Shannon-Weaver diversity, and WV IBI scores. Overall, increased measures of cumulative mining activity resulted in decreases in species richness, diversity, and IBI scores, while increases were observed with increases in cumulative local residential development. No local or neighborhood scale anthropogenic landscape variables were observed to influence the proportion of tolerant individuals instead, only a drainage area effect was detected. In a neighborhood context, the anthropogenic landscape structure did not significantly influence fish community variables. However, using the neighborhood as the random effects structure in the models incorporated the influences of potential dispersal process, the spatial configuration of sampling locations, and allowed for different relationship between local landscape structure and fish community variables on a HUC 12 basis. So, the quality of the neighborhood may not influence community structure *per se*, but sites within the same neighborhood respond to local landscape structure differently than sites within other neighborhoods. These results indicate the importance of

evaluating the hierarchical structure of watersheds for environmental management since not all stream communities respond similarly to anthropogenic stressors. In addition, the lack of measured anthropogenic controls on tolerant individuals may indicate community homogenization in these highly degraded watersheds.



## 1.0 Introduction

Numerous streams and rivers remain impaired despite the recognition of the threats and stressors that impact aquatic resources. Local threats, such as flow alterations, channelization, connectivity, and point-source pollution, and their influences on stream fish community health are well documented (Wang et al., 2003; Karr and Yoder, 2004; Pichon et al., 2006). Findings from research have directed numerous management efforts to manage and restore local instream habitats and riparian zone improvement to mediate the impacts of local stressors. However, streams and their associated communities are products of the larger landscape they occupy in addition to their local environments (Allan, 2004). In the riverscape context of stream health, the importance of changes at the catchment scale is increasingly considered as major threats and stressors to stream ecosystems (Allan, 2004). Furthermore, in order to interpret patterns in fish distribution and community assembly, the effects of stressors to aquatic communities needs to integrate both local and larger scale process.

Stream fish community composition is influenced by ecological “filters” at multiple spatial and temporal scales (Tonn et al., 1990; Poff, 1997). Local in-stream factors, such as biotic competition and predation, can affect species abundances while habitat characteristics (i.e., substrate, available cover, and water chemistry) and regional environmental factors (i.e., climate, landscape composition) can determine species presence. The overall distribution of a species can be determined by large scale historical processes dealing with speciation, extinction, and dispersal. In order to accurately evaluate the combined effects of anthropogenic changes on stream communities, multiple levels of community assembly need to be considered. For example, differences in community diversity, species presence, or species richness can be linked to historical processes as seen by differences between ecoregion (i.e., terrestrial, geographical, and vegetation patterns) and major drainage basin (i.e., zoogeography). Evaluations of local and regional influences are typically reserved for individual species distribution models or to evaluate species diversity. However, other important community attributes such as functional community

composition, respond to the hierarchical structure of stream networks differently than species richness or presence (Hoeinghaus et al., 2007).

Differing responses of species in the same community to anthropogenic landscape changes can be linked to species specific life history traits. In order to maintain a functionally stable community in heterogeneous landscapes, species specific responses to stressors must be highly variable (Gonzalez and Loreau, 2009). This makes evaluating life history traits at the community level important for providing insight into the processes that shape community structure. For example, an increased sediment load from landscape disturbances can have negative impacts on the relative abundances of clean-gravel spawning fish (Sutherland et al., 2002). The overall reduction of these specialist species due to increases in environmental disturbance leads to a community of organisms dominated by generalists species (Clavel et al., 2011).

Ongoing anthropogenic changes and increases in environmental stress can cause local species extinction or shifts in the distributions and abundances of species (Suttle et al., 2007). The shift of a forested watershed to an unforested watershed can lead to increased sedimentation and turbidity, which negatively influences primary production, as well as invertebrate and fish diversity (Dudgeon et al., 2000; Sutherland et al., 2002). Increased urban development can lead to an increase in the flashiness of stream flows, concentrations of nutrients and contaminants, and altered channel morphology ultimately leading to a reduced biotic richness dominated by tolerant taxa (Walsh et al., 2005). Increases in nutrients (i.e., nitrogen and phosphorous) due to agriculture have similar impacts on community structure by decreasing the percentage of intolerant taxa and reducing Index of Biotic Integrity scores (Wang et al., 2007). Similar to urbanization and agriculture, coal mining can cause an increase in hydrologic flashiness and sedimentation (Bernhardt and Palmer, 2011). In addition, regions with valley-fill operations can be shifted to a primary production dominated ecosystem as headwater streams are lost (Hill et al., 1995). Unlike other anthropogenic impacts, coal mining can also become a source of water contamination

indicated by elevated metals and sulfates (Hartman et al., 2005). Overall, coal mining contamination can reduce both benthic macroinvertebrate and fish-based IBI scores (Freund and Petty, 2007).

The increasing intensity of mountain top removal/valley fill (MTM/VF) coal mining and residential development within West Virginia has great potential to impact the 45,000 km of streams that drain the state. The Ohio Central Appalachian biomonitoring region in particular overlaps the majority of the MTM/VF region in West Virginia and has increasing residential development. However, these watersheds also harbor some of the highest numbers of native species in West Virginia (Stauffer et al., 1995) making them important in the evolution and speciation of North American freshwater fishes. Even though the effects of land-use practices such as MTM/VF mining and residential development have been studied for their impacts on stream communities (Merriam et al., 2013; Hitt and Chambers, 2014), the combined overall effects of land-use patterns on stream fish assemblages in this region are poorly understood (but see Daniel et al., 2014).

The objective of this research was to evaluate the influences of local and neighborhood level landscape conditions on local fish community structure. We measured local species richness, diversity (Shannon-Weaver diversity), fish community condition, and dominance of tolerant individuals. The West Virginia Index of Biotic Integrity (WVIBI; Chapter 1) was included in the analysis in order to evaluate if larger landscape processes were influencing local fish community condition and impairment status. The proportion of tolerant individuals was evaluated to see if large scale disturbances have forced communities from intolerant to tolerant species dominated assemblages. In order to accomplish this objective, a mixed modeling approach was utilized. A mixed modeling approach allowed for effects to be partitioned among spatial scales. Recognizing the presence of spatial autocorrelation in ecological data due to proximity of sampling locations is important since it may lead to statistical bias (i.e., lack independence). A mixed-modeling approach also allowed us to account for measures of spatial autocorrelation, if it was present.

## 2.0 Methods

### 2.1 Study Area

The Ohio Central Appalachian biomonitoring region (Figure 1; Chapter 1) consists of four 8-digit watersheds (Coal, Tug, Upper Guyandotte, and Upper Kanawha) and the lower portion of one additional 8-digit watershed delineated by a 12-digit watershed boundary (Elk). Land cover for this region is predominately forested. However, coal mining and residential development are the most common anthropogenic land-use practices. Residential development does not constitute a high percentage of the land-use but the topology of this region constricts development to narrow floodplains adjacent to streams (Figure 1). This region was selected as our study area due to its high anthropogenic impacts from residential development and overlap with the mountain top/valley fill surface mining region (Figure 1).

### 2.2 Fish community variables

Statewide fish community data were combined from various sampling sources including state and federal organizations, universities, and consulting companies. Sampling sites were selected for years 1997, 1998, and 2000 – 2013. Only electrofishing (backpack, parallel wires, and barge) sampling types were used (N=250). Fish community data consisted of identification of each fish captured to species and their abundances. Hybrid species and individuals not identified to species were removed from the sample. Sampling locations were then input to ArcGIS ArcMap 10.0 (Environmental Systems Research Institute, Redlands, California) and joined with segment level watersheds (1:24,000). Locations of sampling points were evaluated against the National Hydrography Dataset (NHD-24 K) to ensure site locations were attributed to the correct segment-level watershed. In order to reduce pseudoreplication, sampling locations were further reduced by selecting the most recent sampling event within each segment level watersheds and by using only wadeable streams (7 – 400 km<sup>2</sup>). Total species richness and a Shannon-Weaver Diversity index was calculated using package *vegan* (Oksanen et al., 2013) in R statistical program (R Core Team 2014). West Virginia Index of Biotic Integrity (WVIBI; Chapter 1) scores for the Ohio

Central Appalachian biomonitoring and the proportion of tolerant individuals were calculated based on species classification for the WVIBI.

### *2.3 Landscape attributes*

Landscape characteristics for all 1:24,000 segment-level watersheds (SLWs) within the state of West Virginia were quantified using spatial analysis functions in ArcGIS ArcMap 10.0 (Environmental Systems Research Institute, Redlands, California). In conjunction with flow tables, cumulative (i.e., all SLWs upstream of a given sampling location) measures of several landscape attributes for each segment-level watershed were quantified (Strager et al., 2009).

Land cover classifications were derived from the 2009 and 2010 National Agriculture Imagery Program (NAIP) orthophotography with a 1 meter pixel resolution at a scale of 1:10,000. Land cover types summarized included grass and agricultural (i.e., crops and pasture) lands and barren development. The mining-permit boundaries layer developed by the Technical Applications in GIS (TAGIS) office within WVDEP enabled the differentiation between all mining related land-cover forms (i.e., slurry impoundments and active and reclaimed mine lands) from non-mining land cover. All mining-related cover classes were summed into a measure of total surface mining. The density ( $\#/km^2$ ) of surface mining, underground mining, sewage and septic serviced structures, and National Pollution Discharge Elimination System (NPDES) permits were calculated from data obtained from WVDEP. The 2003 West Virginia Statewide Addressing and Mapping Board (WV SAMB) structures layer was used to calculate the density of residential and commercial structures ( $\#/km^2$ ). Natural landscape variables for each SLW were summarized including basin area ( $km^2$ ), mean elevation (m), and swim distance (km). Swim distance was defined as the minimum downstream distance (km) from the outflow of a SLW to the inflow of a SLW with a basin area  $\geq 200 km^2$  (Hitt and Angermeier, 2011). Local landscape measures consisted of the cumulative landscape at the segment level watershed scale, while neighborhood landscape consisted of the cumulative landscape at the outflow of the 12-digit scale watershed (HUC 12). Local and neighborhood-level landscape data were used in further statistical analyses.

## *2.4 Statistical analyses*

Principal Components Analysis (PCA) was used to summarize the patterns of co-variation in the cumulative separately for wadeable local and neighborhood-level watersheds. Specifically, PCA was utilized in order to reduce dimensionality and collinearity among variables. Cumulative land-use variables were transformed to approximate normality. The land-use variables used were structure density, NPDES permit density, septic and sewage density, underground and surface mine permit densities, % development, % grassland, % agriculture, and % surface mining. Density measures were  $\log_{10}(x+1)$  transformed while proportional measures were arc-sine square root transformed prior to analysis in order to approximate normality. Principal components (PCs) were retained for future analysis if the eigenvalues  $>1$  (McCune and Grace, 2002). Factor loadings were calculated as the correlation of PC scores to original landscape data and were regarded as statistically important contributors to the PC if the factor loading  $> |0.40|$  (McCune and Grace, 2002). The first two important PC axes (eigenvalues  $>1$ ) for the neighborhood and local analysis were used as predictor variables in all models along with mean elevation (m), drainage area ( $\text{km}^2$ ), and swim distance (km). Drainage area was log-10 transformed prior to analysis. Prior to model evaluation, each predictor variable was evaluated for redundancy with other variables by calculating a variable inflation factor (VIF; Zuur et al. 2009) using the vif function in package usdm (Naimi, 2013). A VIF value is used to detect collinearity between variables in which highly collinear variables were removed. A predictor variable with a VIF  $>2$  was removed from analysis.

Mixed-effects models were selected as the modeling framework due to the hierarchical, or nested, structure of the local and neighborhood landscape attributes and the fish community response variables (Venables and Dichmont, 2004). The community response variables were 1) taxonomic richness (i.e., number of species); 2) taxonomic diversity (Shannon-Weaver Diversity Index); 3) Index of Biotic Integrity scores; and 4) proportion of tolerant individuals.

The responses of fish to local and neighborhood land-use composition were analyzed using linear mixed-effects (LME) models, or generalized linear mixed models, for each response variable following

guidelines from Zuur et al. (2009). The first step in the model selection process was to evaluate the need to use a more complex model structure. Fully parameterized models (i.e., all predictor variables with meaningful interaction terms) were generated using both generalized least squares (GLS) regression and LME with a random intercept term only and compared using an Analysis of Variance (ANOVA). If models were significantly different ( $p < 0.05$ ), the model with the lowest Akaike Information Criterion (AIC) value was selected as the “best” model. If the linear mixed effect model was selected as the top model, the selection of the best random effects structure was evaluated. For this selection process, the same fully parameterized fixed effect structure was used with varying random effects structure, fitted using restricted maximum likelihood estimation (REML). For the mixed models, the HUC 12 watershed was assigned as the random intercept effect since individual sampling locations are nested within HUC 12 watersheds and some HUC12 watersheds have multiple sampling locations located within. All models compared had the same random intercept effect, but variables for the random slope effect differed between models. The best random effects structure was determined using AIC scores in which the model with the lowest value was retained.

Once a random effect was established, the model was re-fit using a maximum likelihood estimation (MLE) and the fixed effect structure was evaluated by sequentially dropping one variable, starting with interactions. Variables were dropped until only significant variables remained in the model and the resulting models were compared using ANOVA. The influence of spatial autocorrelation was also evaluated for each response variable by visually examining correlograms of the residuals for the GLS and final linear mixed effects models. Species richness response to local and neighborhood level landscape was evaluated with a generalized linear mixed model (GLMM), assuming a poisson distribution, following the same step-wise model selection process as the linear mixed models. Proportion of tolerant individuals (% Tol) was arc-sine square-root transformed prior to analysis in order to better approximate normality. Transformation was selected over using a binomial distribution because generalized linear mixed models currently do not have the capacity to accurately evaluate proportional data under this

distribution. Linear mixed effects and generalized linear mixed effects models were generated using the nlme (Pinheiro et al., 2014) and glmmADMB (Skaug et al., 2014) packages in R statistical program, respectively. Significance of results was determined using an alpha=0.05, but marginal differences ( $p < 0.10$ ) were noted.

### **3.0 Results**

#### *3.1 Local and Neighborhood Landscape Principal Components Analyses*

Local and regional landscape characteristics varied across the study sites. The local PCA identified three significant PC axes (eigenvalue  $>1$ ), which accounted for 63.4% of the variation in the Wadeable Segment Level Watersheds in this region (Table 1). For the local condition, PC1 accounted for 33.6% of the variation in the landscape structure among all Wadeable Segment Level Watersheds (Table 1). The first axis represents a residential and agricultural gradient that included structure density, NPDES permit density, septic and sewer serviced structure density, % development, % grassland, and % agriculture (Figure 2). PC2 accounted for an additional 17.1% of the variation while representing a mining and development gradient (Table 1). Variables that showed a contribution to this axis included surface and underground mine permit densities, % development, and % surface mining (Figure 2). The third axis represented an additional residential development axis while accounting for an additional 13% of the variation in the landscape dataset. Variables that showed significant contributions to this axis included NPDES permit density, septic and sewage serviced structure density, and % grassland.

The neighborhood PCA also identified three significant PC axes (eigenvalue  $>1$ ) accounting for a total of 77.1% of the variation exhibited in the landscape at the outflows of HUC 12 watersheds (neighborhoods) in this region (Table 1). PC1 characterizes a mining and residential development axis accounting for 40.3% of the variation in the landscape data set. All of the landscape variables analyzed, with the exception of % development, demonstrated a significant contribution to PC1 (Figure 2). Similarly, PC2 also demonstrated a gradient of residential development and mining while accounting for an additional



23.6% of the variation in the data set. Variables with significant contribution to PC2 included structure density, surface and underground mining permit densities, % surface mining, and % development (Figure 2). The final axis (PC3) accounted for an additional 13.2% of the variation in the landscape data set representing a residential development gradient. Only septic and sewage serviced structure density, % development, and % grassland contributed to the loadings of PC3. In the neighborhood and local PCAs the first two principal components accounted for the majority of the variation in their landscape data sets, therefore, only these axes were used in further analysis of fish community structure. All PC axes had loadings of anthropogenic landscape variables in the negative direction generating local and neighborhood human disturbance gradients.

### *3.2 Spatial autocorrelation of fish community variables*

The correlogram of the residuals of GLM model for species richness indicated that there was spatial autocorrelation (Moran's I correlation) after accounting for natural and anthropogenic landscapes variables (Figure 3). However, when a linear mixed model was used, the same spatial autocorrelation was no longer present in the residuals. The analysis of spatial autocorrelation of the residuals also indicated little spatial structure for the Shannon-Weaver diversity index (Figure 4). In addition, an evaluation of the model residuals for spatial autocorrelation for WV IBI scores indicated that a spatial structure was still present after a GLS was applied, but no longer existed in the linear mixed model residuals (Figure 5). The GLS correlogram for the proportion of tolerant individuals indicates small spatial autocorrelation after the model was applied (Figure 6). However, the same spatial autocorrelation is not present in the LME correlogram.

### *3.3 Responses of fish community to local and neighborhood land-use*

Model structures were evaluated using all natural landscape variables and anthropogenic land-use PC axes since none of the variables had a variable inflation factor  $\geq 2.0$ . General linear model (GLM; AIC=1533.0) and general linear mixed model (GLMM; AIC=1492.7) were compared using AIC, indicating that a mixed model structure was needed to evaluate the influence of neighborhood and local

landscape conditions on total species richness (Table 2). When differing random effects structure were evaluated, the model with a random slope (SLW PC2) and intercept (HUC 12) was selected as the top model (AIC=1490.35; Table 2). The stepwise selection of the fixed effects structure produced a top model (AIC=1478.98), which included elevation, drainage area, and SLW PC1 and PC2 after accounting for differing relationships of SLW PC2 within each HUC 12 watershed (Table 3). The results of the final model indicate that total species richness increases with drainage area and SLW PC2 (mining-development gradient). However, there were negative responses of species richness with mean elevation and SLW PC1 (development gradient). These results indicate that increases in mining intensity and elevation, leads to a decrease in overall species richness while increases in residential development and drainage area leads to increases in species richness.

The comparison of the generalized least squares (AIC=451.05) and linear mixed effects model (AIC=452.65) for Shannon-Weaver Diversity showed no significant difference ( $p=0.5233$ ; Table 2) between model structures. Even though there was no statistical difference between the GLS and LME, we chose to continue the analysis of Shannon-Weaver Diversity using a linear mixed effects model due to the hierarchical structure of the sampling design. The optimal random effects structure for this analysis was a random intercept model with HUC 12 as the grouping variable (Table 2). The top final model (AIC=408.48) consisted of mean elevation, drainage area, SLW PC1 and PC2, and neighborhood PC2 as fixed effects (Table 3). Shannon-Weaver diversity was found to decrease with increases in mean elevation and SLW PC1 (residential development gradient). However, increases in diversity were seen with increases in drainage area, SLW PC2 (mining disturbance gradient), neighborhood PC2 (mining and residential development gradient). These results indicate that increases in local mining intensity and elevation, leads to a decrease in overall species diversity while increases in local residential development and drainage area leads to increases in species diversity. In addition, increases in neighborhood-level mining and residential development leads to decreases in species diversity.

West Virginia Index of Biotic Integrity scores were initially evaluated with a generalized least squares regression and compared to a random intercept linear mixed model using Analysis of Variance (ANOVA). The ANOVA results indicate that the linear mixed model was significantly different ( $p=0.0114$ ; Table 2) from the generalized least squares model. These results signify that the assumption of independence is violated for the GLS and a more complex model (i.e., linear mixed models) was needed to help account for the presence of spatial structure. The results of model comparisons indicate that a model with a random intercept of the HUC 12 watershed while allowing for different SLW PC2 slopes within each HUC 12 performed the best of the models evaluated (Table 2). The final model for WVIBI scores indicate that SLW PC1 and PC2 and mean elevation are the main driving factors when differing relationships of SLWPC2 within each HUC 12 are accounted for (Table 3). Specifically, IBI scores were found to decrease with increases in both elevation and SLWPC1 while increases were observed with increases in SLWPC2. These results indicate that increases in local mining intensity and elevation, leads to a decrease in IBI scores while increases in local residential development leads to increases in IBI scores.

A marginal difference was observed between the generalized least squares and linear mixed effects models when the relationship of local and neighborhood landscape conditions were evaluated for the proportion of tolerant individuals ( $p=0.053$ ; Table 2). The optimal random effects structure was determined to be a random intercept model with HUC 12 as the grouping factor (AIC=308.83; Table 2). The only variable selected as a fixed effect was drainage area. An increase in drainage area resulted in a decrease in the proportion of tolerant individuals (Table 3). These results indicate that there was no local or neighborhood landscape influences on the distribution and proportion of tolerant individuals in this region. In addition, the marginal difference between GLM and LME models may indicate that there is no hierarchical structure in terms of the proportion of tolerant individuals within this region.

#### **4.0 Discussion**

Stream fish communities within Ohio Central Appalachian watersheds are structured by a combination of natural and anthropogenic landscape factors. Patterns in fish assemblage structure were partially dependent on the type of biological data used. Taxonomic and diversity measures were primarily driven by local landscape conditions in addition to natural variables. The multimetric index (WV IBI) was controlled by similar processes, except the effects of drainage area were removed. Finally, no local or regional anthropogenic landscape variables were determined to impact the proportion of tolerant individuals. For this analysis we recognized and accounted for 2 spatial scales apparent in riverine networks while evaluating the influence of both natural and anthropogenic impacts on the structure of fish communities (Swan and Brown, 2011).

The mixed effects modeling approach was better suited for our data due to the hierarchical structure of the landscape variables. The spatial autocorrelation present in some of the response variables could have resulted in statistical bias using a generalized linear model approach. Using a mixed effects structure, we accepted that the dynamics of local fish community structure is a function of larger landscape processes, including dispersal, resulting in a nested sampling design. For all models generated, the same random intercept grouping factor (HUC 12 watershed) was used. In this analysis, it is likely that sites within the same HUC 12 watershed are more similar, either due to spatial autocorrelation or due to the overall condition of the watershed. The use of a grouping variable allowed us to account for multi-level structure in the fish community data and avoids an assumption violation that sampling locations are independent of one another (Wagner et al., 2006).

The primary source of land-use change in the central Appalachians is a result of mountain-top mining (Bernhardt and Palmer, 2011). Other studies conducted in this region have also demonstrated a decrease in taxonomic and functional species richness (Hitt and Chambers, 2014), species occurrence (Hopkins and Roush, 2013), community trait assemblage diversity (Daniel et al., 2014), and biotic conditions of fish (USEPA, 2003) and invertebrate communities (Merriam et al., 2013) in streams receiving run-off from surface mining operations. Based on the models generated, we can expect nearly a four point

decrease in WV IBI scores with increases in the PC axis associated with surface mining activities. Similarly, we can also expect slight decreases in both species richness and diversity. However, in this region we also see an inflation of WV IBI scores, species richness, and diversity as a result of increases in residential development. Residential development within the Ohio Central Appalachian region is confined to narrow floodplains, resulting in a close association with streams, leaving little room for the proper installation of septic systems (Cook et al., 2013). The increased nutrients associated with residential development maybe augmenting aquatic systems, ultimately restructuring the food quality and quantity available to stream fishes (Wang et al., 2007). In a region dominated by current and legacy surface mining activities it is possible that the shift to a primary production dominated system with influxes of excess nutrients can lead to an inflation of fish production and potentially provide refuge from other surface mining related stressors.

Our results add to the growing literature evaluating the importance of natural landscape controls on fish community structure. The majority of the response variables analyzed exhibited responses to at least one natural landscape variable (drainage area, mean elevation, or swim distance). The influence of drainage area on fish community structure has been well documented (Angermeier and Schlosser, 1989; Osborne and Wiley, 1992). The WVIBI model was the only model that did not include drainage area in the fixed effects structure. This was to be expected since the WVIBI was developed with the effects of drainage area removed (Chapter 1). It is likely that elevation in this study acted as a surrogate for stream temperature in which higher elevation generally indicates cooler stream temperatures. Elevation was an important predictor variable in three (Richness, Shannon-Weaver diversity, and WVIBI) of the models evaluated. Decreases in species richness and Shannon-Weaver diversity correspond to current literature regarding longitudinal community changes with elevation and stream temperatures (Rahel and Hubert, 1991). The shift from a cold water species dominated community to a minnow-sucker dominated community over a longitudinal elevation gradient can happen over relatively short distances and are mainly due to the addition of new species downstream (Rahel and Hubert, 1991). The WV IBI was

developed exclusively for warm water streams and even though strictly cold water streams were not evaluated in this study, an elevation effect was still present.

Dispersal of fishes from adjacent streams, due to mass effects, can have major implications on biomonitoring programs by altering the local assemblage structure. Specifically, immigrating fishes can bias biomonitoring assessments towards false identification of local degradation or failure to detect degradation due to inflation of local species (Hitt and Angermeier, 2011). Swim distance (i.e., distance to a drainage area of  $\geq 200\text{km}^2$ ) was evaluated as a predictor variable during model development in an attempt to account for the influence of stream position and potential mass effects. However, this variable was not retained in any of the models evaluated. The use of random effects models accounted for some of the spatial autocorrelation present in our response variables. More importantly, using HUC 12 watershed as a grouping factor we built the importance of spatial structure into our models by indicating that sites within the same HUC 12 watershed may be more similar.

Pease et al. (2015) examined the functional-trait structure of stream fish assemblages and found that local reach and catchment scale environmental variables were significantly associated with functional trait composition. They also found strong larger scale ecoregional controls on both taxonomic and functional assemblage structure. However, overall environmental disturbance and land-use patterns were ecoregion dependent making it difficult to separate historical controls on fish distributions from current environmental controls. In the study presented here, broad scale physiographic controls on fish assemblage structure were already taken into consideration by using one biomonitoring region (Ohio Central Appalachians) which was determined by fish assemblage similarity (Chapter 1). After accounting for larger regional controls and recognizing the importance of the hierarchical structure of stream networks (Frissell et al., 1986), the numerical landscape characteristics of the neighborhood scale was found to be unimportant in structuring the majority of the response variables.

The lack of local and neighborhood level controls on the proportion of tolerant individuals could point to the increased homogenization of tolerant fish species within the Ohio Central Appalachian region.

Homogenization of ecosystems occurs when increased anthropogenic impacts decrease habitat suitability for a large number of specialized species while simultaneously increasing suitability for a small number of generalist species (Smart et al., 2006). Typically, biotic homogenization is reserved for evaluating the colonization and spread of exotic species over native species. The Ohio Central Appalachian region, and West Virginia as a whole, has few exotic species. Instead the expansion in the distribution of tolerant native species is becoming more prevalent. The evaluation of long-term trends of fish distribution will need to be conducted in order to determine the extent of homogenization of fish communities in highly impacted regions of West Virginia.

## Literature Cited

- Allan, J.D. 2004. Landscape and Riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*. 35: 257 – 284.
- Angermeier, P.L. and I.J. Schlosser. 1989. Species-area relationships for stream fishes. *Ecology*. 70:1450–1462.
- Bernhardt, E. S. and M. A. Palmer. 2011. The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians. *Annals of the New York Academy of Sciences*. 1223:39 – 57. DOI: 10.1111/j.1749-6632.2011.05986.x.
- Clavel, J., R. Julliard, and V. Devictor. 2011. Worldwide decline of specialist species: toward a global functional homogenization? *Frontiers in Ecology and the Environment* 9:222 – 228. DOI:10.1890/080216
- Cook, N.A., Krometis, L.A., and Sarver, E.A. 2013. Inventory of bacterial impairments in central Appalachia. *Environmental Considerations in Energy Productions*. 5:214 – 228.
- Daniel, W. M., D. M. Infante, R. M. Hughes, Y.P. Tsang, P. C. Esselman, D. Wiefelich, K. Herreman, A. R. Cooper, L. Wang, and W. W. Taylor. 2014. Characterizing coal and mineral mines as a regional source of stress to stream fish assemblages. *Ecological Indicators*. 50:50 – 61.
- Dudgeon, D. 2000. The ecology of tropical Asian rivers and streams in relation to biodiversity conservations. *Annual Review of Ecology and Systematics*. 31:239 – 263.
- Freund, J.G. and J.T. Petty. 2007. Response of fish and macroinvertebrate bioassessment indices to water chemistry in a mined Appalachian watershed. *Environmental Management*. 39:707 – 720.
- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management*. 10:199 – 214.
- Gonzalez, A. and M. Loreau. 2009. The causes and consequences of compensatory dynamics in ecological communities. *Annual Review of Ecology, Evolution, and Systematics*. 40:393 – 414.
- Hartman, K. J., J. D. Kaller, J. W. Howell, and J. A. Sweka. 2005. How much do valley fills influence headwater streams? *Hydrobiologia*. 532:1 – 19.
- Hill, W. R., M. G. Ryon, and E. M. Schilling. 1995. Light limitation in a stream ecosystem: responses by primary producers and consumers. *Ecology*. 76:1297 – 1309.
- Hitt, N. P. and P. L. Angermeier. 2011. Fish community and bioassessment responses to stream network position. *Journal of the North American Benthological Society*. 30:296 – 309.
- Hitt, N. P. and D. B. Chambers. 2014. Temporal changes in taxonomic and functional diversity of fish assemblages downstream from mountaintop mining. *Freshwater Science*. DOI: 10.1086/676997.
- Hoeinghaus, D.J., K.O. Winemiller, and J.S. Birnbaum. 2007. Local and regional determinants of stream fish assemblage structure: inferences based on taxonomic vs. functional groups. *Journal of Biogeography*. 34:324 – 338.
- Hopkins, R.L. and J.C. Roush. 2013. Effects of mountaintop mining on fish distributions in central Appalachia. *Ecology of Freshwater Fish*. doi: 10.1111/eff.12061.



- Karr, J.R. and C.O. Yoder. 2004. Biological assessment and criteria improve Total Maximum Daily Load Decision Making. *Journal of Environmental Engineering*. 130:594 – 604.
- McCune B and Grace JB. 2002. *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, Oregon.
- Merriam, E.R., J.T. Petty, M.P. Strager, A.E. Maxwell, and P.F. Ziemkiewicz. 2013. Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined central Appalachian watershed. *Freshwater Science*. 32:1246 – 1259.
- Naimi, Babak. 2013. usdm: Uncertainty analysis for species distribution models. R package version 1.1-12. <http://CRAN.R-project.org/package=usdm>.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. Stevens and H. Wagner. 2013. vegan: Community Ecology Package. R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>
- Osborne, L.L. and M.J. Wiley. 1992. Influence of tributary spatial position on the structure of warmwater fish communities. *Canadian Journal of Fisheries and Aquatic Sciences*. 49:671–681.
- Pease, A. A., J. M. Taylor, K. O. Winemiller, and R. S. King. 2015. Ecoregional, catchment, and reach-scale environmental factors shape functional-trait structure of stream fish assemblages. *Hydrobiologia*. 753:265 – 283. DOI 10.1007/s10750-015-2235-z.
- Pichon, C.L., G. Gorges, P. Boet, J. Baudry, F. Goreaud, and T. Faure. 2006. A spatially explicit resource-based approach for managing stream fishes in riverscapes. *Environmental Management*. 37: 322 – 335.
- Pinheiro J., D. Bates, S. DebRoy, D. Sarkar and R Core Team. 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-118, <URL:<http://CRAN.R-project.org/package=nlme>>.
- Poff, N.L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*. 16:391 – 409.
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rahel, F. J. and W. A. Hubert. 1991. Fish assemblages and habitat gradients in a Rocky Mountain-Great Plains stream: biotic zonation and additive patterns of community change. *Transactions of the American Fisheries Society*. 120:319 – 332.
- Skaug, J., D. Fournier, B. Bolker, A. Magnusson, and A. Nielsen. 2014. Generalized Linear Mixed Models using AD Model Builder. R package version 0.8.0.
- Smart, S. M., K. Thompson, R. H. Marrs, M. G. Le Duc, L. C. Maskell, and L. G. Firbank. 2006. Biotic homogenization and changes in species diversity across human-modified ecosystems. *Proceedings of the Royal Society B*. 273: 2659 – 2665. doi:10.1098/rspb.2006.3630.
- Stauffer, J.R., J.M. Boltz & L.R. White. 1995. *The fishes of West Virginia*. Academy of Natural Sciences of Philadelphia, Philadelphia.
- Strager, M. P., J. T. Petty, J. M. Strager, J. Barker-Fulton. 2009. A spatially explicit framework for quantifying downstream hydrologic conditions. *Journal of Environmental Management*. 90:1854 – 1861.

- Sutherland, A. B., J. L. Meyer, and E. P. Gardiner. 2002. Effects of land cover on sediment regime and fish assemblage structure in four southern Appalachian streams. *Freshwater Biology*. 47:1791 – 1805.
- Suttle, K. B., M. A. Thomsen, and M. E. Power. 2007. Species interactions reverse grassland responses to changing climate. *Science* 315: 640 – 642.
- Swan, C.M. and B.L. Brown. 2011. Advancing theory of community assembly in spatially structured environments: local vs regional processes in river networks. *Journal of the North American Benthological Society*. 30:232 – 234.
- Tonn, W.M., J.J. Magnuson, M. Rask, and J. Toivonen. 1990. Intercontinental comparison of small-lake fish assemblages: the balance between local and regional processes. *The American Naturalist*. 136:345 – 375.
- US EPA. 2003. Ecological assessment of streams in the coal mining region of West Virginia using data collected by the U.S. EPA and environmental consulting firms. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Washington, D.C.
- Venables, W.N. and C.M. Dichmont. 2004. GLMs, GAMs, and GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research*. 70:319 – 337.
- Wagner, T., D.B. Hayes, and M.T. Bremigan. 2006. Accounting for multilevel data structures in fisheries data using mixed models. *Fisheries*. 31:180 – 187.
- Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society*. 24:706 – 723.
- Wang, L., J. Lyons, P. Rasmussen, and P. Seelbach, T. Simon, M. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest ecoregion, U.S.A. *Canadian Journal of Fisheries and Aquatic Sciences*. 60:491 – 505.
- Wang, L., D. M. Robertson, and P. J. Garrison. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient development criteria. *Environmental Management*. 39:194 – 212. DOI 10.1007/s00267-006-0135-8.
- Zuur, A., E.N. Leno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer-Verlag, New York. DOI:10.1007/978-0-387-87458-6.

## Tables

**Table 1:** Principal Components Analysis results for segment level watershed (SLW) and HUC12 Neighborhood (Neigh.) scales conducted with cumulative landscape variables. All landscape variables were transformed prior to analysis in order to approximate normality. Density values were  $\log_{10}(x+1)$  transformed and proportional (%) variables were arc-sine square root transformed. Bold numbers indicate a significant ( $\geq|0.40|$ ) contribution to Principal Component axis.

Components	SLW PC1	SLW PC2	Neigh. PC1	Neigh. PC2
<b>Std. Deviation</b>	1.738	1.244	1.904	1.456
<b>Prop. Var.</b>	0.336	0.172	0.403	0.236
<b>Cum. Prop.</b>	0.336	0.508	0.403	0.638
<b>Structure Density</b>	<b>-0.751</b>	0.010	<b>-0.580</b>	<b>-0.501</b>
<b>NPDES Permit Density</b>	<b>-0.772</b>	0.144	<b>-0.872</b>	-0.242
<b>SM Permit Density</b>	0.087	<b>-0.672</b>	<b>0.480</b>	<b>-0.671</b>
<b>UM Permit Density</b>	0.039	<b>-0.509</b>	<b>0.416</b>	<b>-0.638</b>
<b>Septic/Sewage Density</b>	<b>-0.755</b>	0.198	<b>-0.849</b>	-0.122
<b>% Development</b>	<b>-0.594</b>	<b>-0.415</b>	-0.289	<b>-0.636</b>
<b>% Grassland</b>	<b>-0.716</b>	-0.148	<b>-0.613</b>	-0.335
<b>% Agriculture</b>	<b>-0.639</b>	-0.085	<b>-0.815</b>	0.079
<b>% SM</b>	0.081	<b>-0.759</b>	<b>0.531</b>	<b>-0.645</b>

**Table 2:** Model construction results from the step-wise selection process outlined by Zuur et al., 2009. Response variables were initially compared using generalized least squares (GLS) or generalized linear models (GLM; Richness) and linear mixed effects models (LME) or generalized linear mixed models (GLMM; Richness) using Analysis of Variance (ANOVA) and evaluating AIC values. The random structure of each model (random str.) was evaluated using a full model and varying random slope components, all with the same random intercept. The best random structure was determined using AIC. Fixed effect structure (fixed str.) was determined using a deletion test until only significant variables remained. These models were then compared using ANOVA and AIC values.

Response Variable	Model	Structure	df	AIC
<b>Richness</b>	GLM	Full model <sup>A</sup>	249	1533
	GLMM	Full model + HUC12	14	1492.7
	GLMM (random str.)	Full model + HUC12	14	1494.04
		+Neigh. PC1 HUC12	--	--
		+Neigh. PC2 HUC12	15	1496.04
		+SLW PC1 HUC12	--	--
		<b>+SLW PC2 HUC12</b>	15	1490.35
	GLMM (fixed str.)	-Interactions	10	1483.81
		-Neigh. PC2	9	1481.98
		-Swim Distance	8	1480.17
-Neigh. PC1		7	1478.98	
<b>S. W. Diversity</b>	GLS	Full Model	250	451.05
	LME	Full Model + HUC12	14	452.65
	LME (random str.)	<b>Full Model + HUC12</b>	15	477.84
		+Neigh. PC1 HUC12	17	481.61
		+Neigh. PC2 HUC12	17	481.84
		+SLW PC1 HUC12	--	--
		+SLW PC2 HUC12	--	--
	LME (fixed str.)	-Interactions	10	369.20
		-Swim Distance	9	367.47
		-Neigh. PC1	8	367.44
<b>WV IBI Scores</b>	GLS* (L ratio=6.4; p=0.0114)	Full Model	250	2208.59
	LME	Full Model + HUC12	14	2204.19
	LME (random str.)	Full Model + HUC12	15	2221.61
		+Neigh. PC1 HUC12	17	2225.14
		+Neigh. PC2 HUC12	17	2224.88
		+SLW PC1 HUC12	17	2219.68

		<b>+SLW PC2 HUC12</b>	17	2218.18
	LME (fixed str.)	-Interactions	12	2210.23
		-Neigh. PC2	16	2213.04
		-Drainage Area	15	2212.38
		-Neigh. PC1	14	2210.38
		-Swim Distance	13	2208.38
<b>% Tolerant Ind.</b>	GLS* (L ratio=3.74; p=0.053)	Full Model	250	284.44
	LME	Full Model + HUC12	14	282.70
	LME (random str.)	<b>Full Model + HUC12</b>	15	308.83
		+Neigh. PC1 HUC12	17	312.83
		+Neigh. PC2 HUC12	17	312.83
		+SLW PC1 HUC12	17	312.64
		+SLW PC2 HUC12	17	310.87
	LME (fixed str.)	-Interactions	10	191.02
		-Neigh. PC2	9	189.02
		-Elevation	8	187.02
		-Neigh. PC1	7	185.08
		-SLW PC1	6	184.22
		-SLW PC2	5	183.05
		-Swim Distance	4	182.73

<sup>A</sup>Full model structure: SwimDist\_KM\*MEAN+log10(CUMU\_DA\_KM2)+SLW PC1\*SLW PC2+Neigh. PC1\*Neigh. PC2+ SLW PC2\* Neigh. PC1+ SLW PC2\* Neigh. PC2+ SLW PC1\* Neigh. PC1+ SLW PC1\* Neigh. PC2

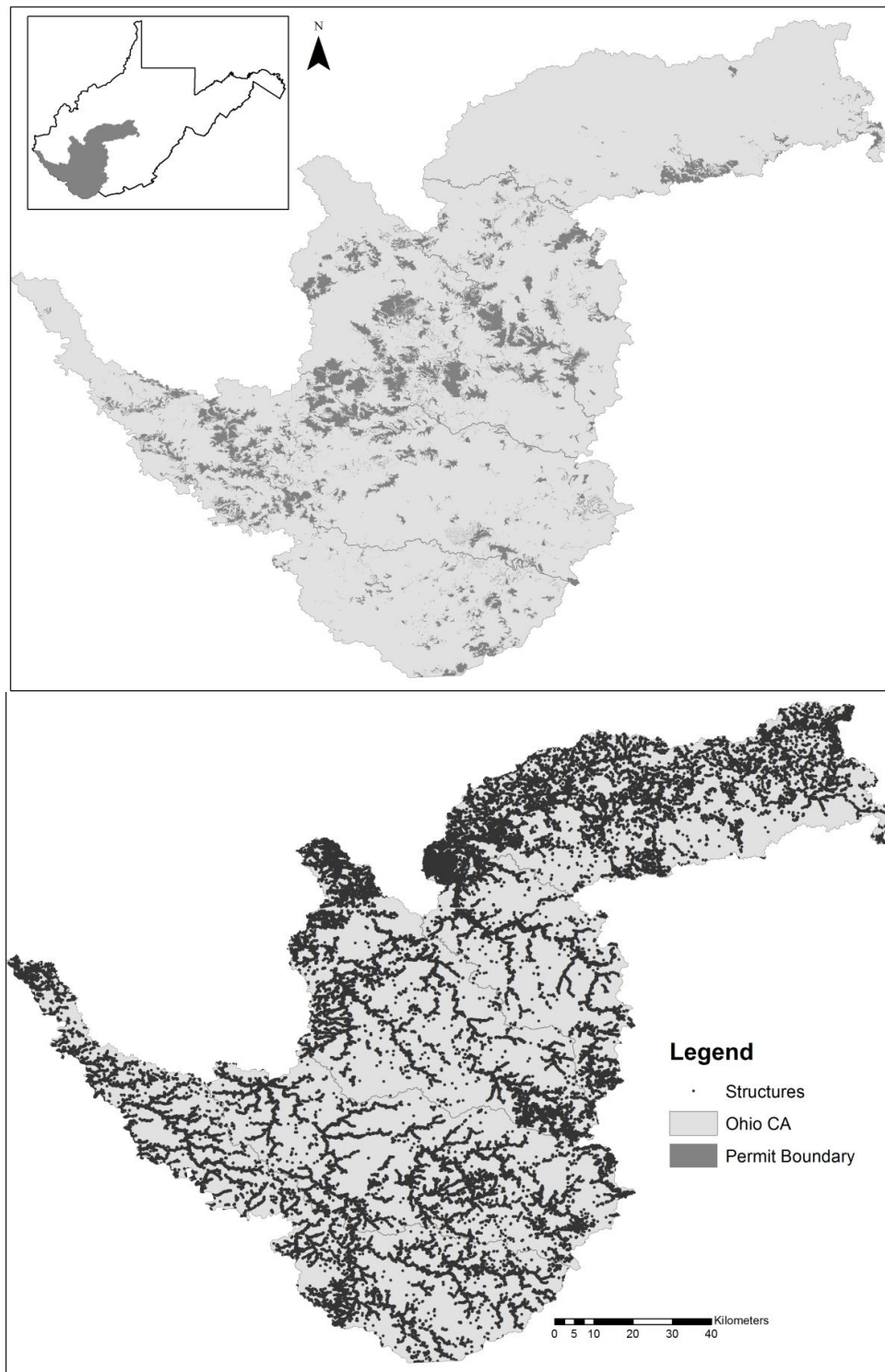
\*Indicates significant (p<0.05) or marginal (p<0.10) differences between model structures

**Table 3:** Final models selected using the step-wise approach outlined by Zuur et al., 2009. Species richness was evaluated with a generalized linear mixed model (GLMM) while other variables were evaluated with a linear mixed model. The test statistic (Statistic) for species richness was a z-value while the others used a t-value. Significance of each variables was determined using  $p < 0.05$ .

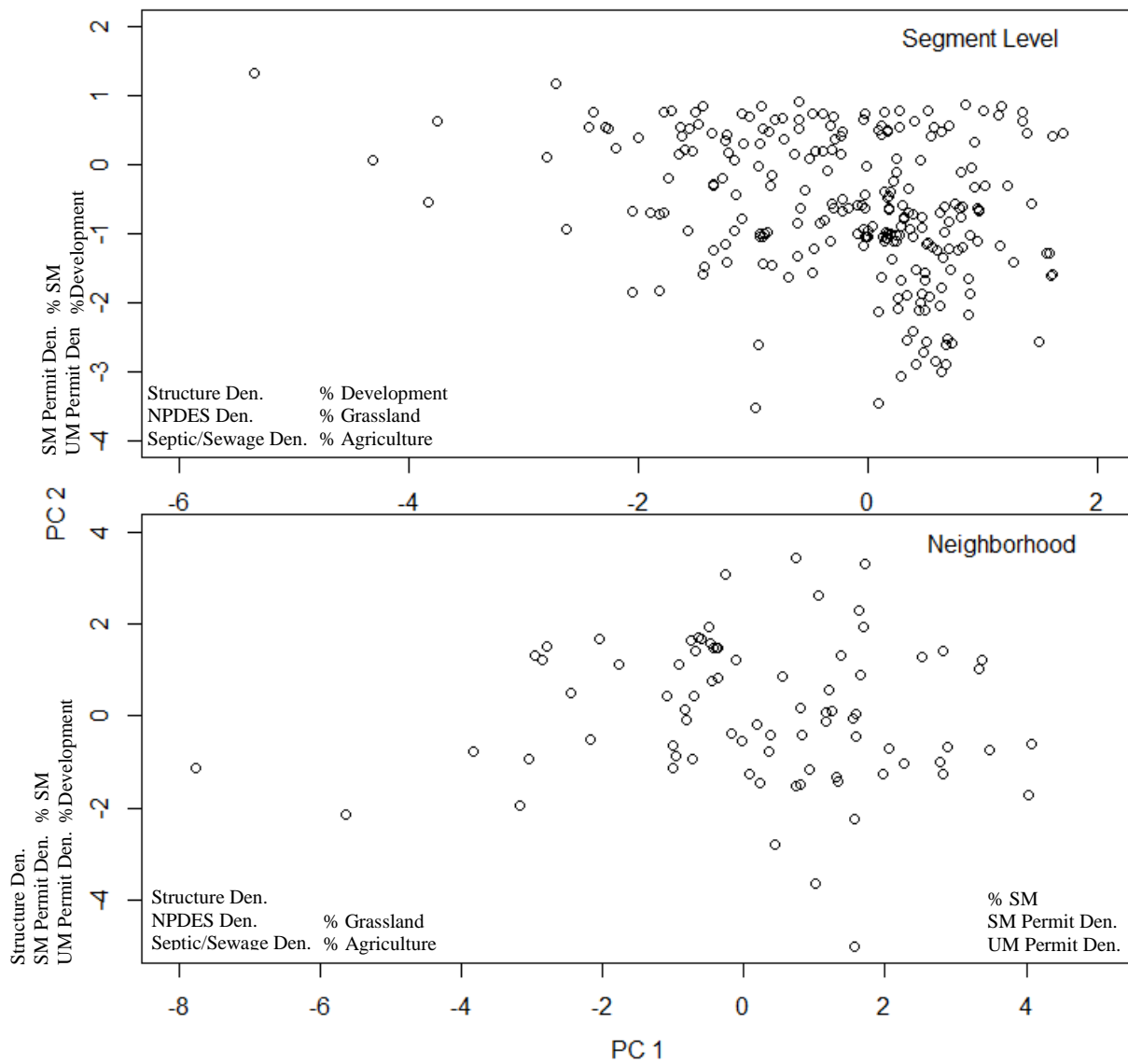
<b>Response</b>	<b>Variable</b>	<b>Coefficient</b>	<b>SE</b>	<b>DF</b>	<b>Statistic</b>	<b>P</b>
<b>Species Richness</b>	Intercept	1.550	0.178	--	8.70	0.0000
	Elevation	0.001	0.000	--	-5.52	0.0000
	Drainage Area	0.938	0.064	--	14.69	0.0000
	SLW PC1	-0.092	0.028	--	-3.27	0.0000
	SLW PC2	0.242	0.053	--	4.59	0.0000
<b>S-W Diversity</b>	Intercept	0.981	0.177	166	5.56	0.0000
	Elevation	-0.001	0.000	166	-4.98	0.0000
	Drainage Area	0.767	0.070	166	10.95	0.0000
	SLW PC1	-0.112	0.031	166	-3.59	0.0004
	SLW PC2	0.130	0.040	166	3.27	0.0013
	Neigh. PC 2	0.057	0.029	78	1.98	0.0513
<b>WV IBI</b>	Intercept	71.457	5.544	167	12.89	0.0000
	Elevation	-0.020	0.004	167	-4.96	0.0000
	SLW PC1	-3.977	1.348	167	-2.95	0.0036
	SLW PC2	6.510	2.072	167	3.14	0.0020
<b>% Tolerant Ind.<sup>A</sup></b>	Intercept	1.509	0.078	169	19.44	0.0000
	Drainage Area	-0.306	0.048	169	-6.38	0.0000

<sup>A</sup> Variable was arc-sine square root transformed prior to analysis.

## Figures

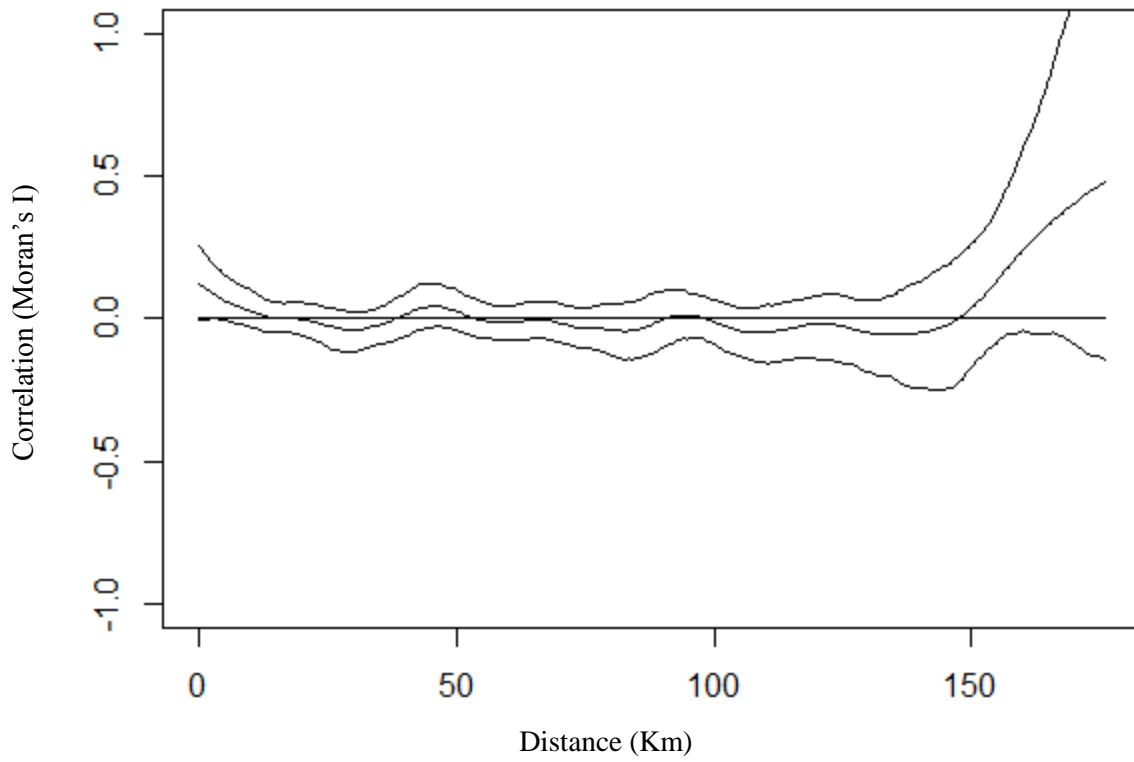


**Figure 1:** Extent of surface mining permit boundaries (Permit Boundary) and residential structures (Structures) in the Ohio Central Appalachian (Ohio CA) biomonitoring region in West Virginia.

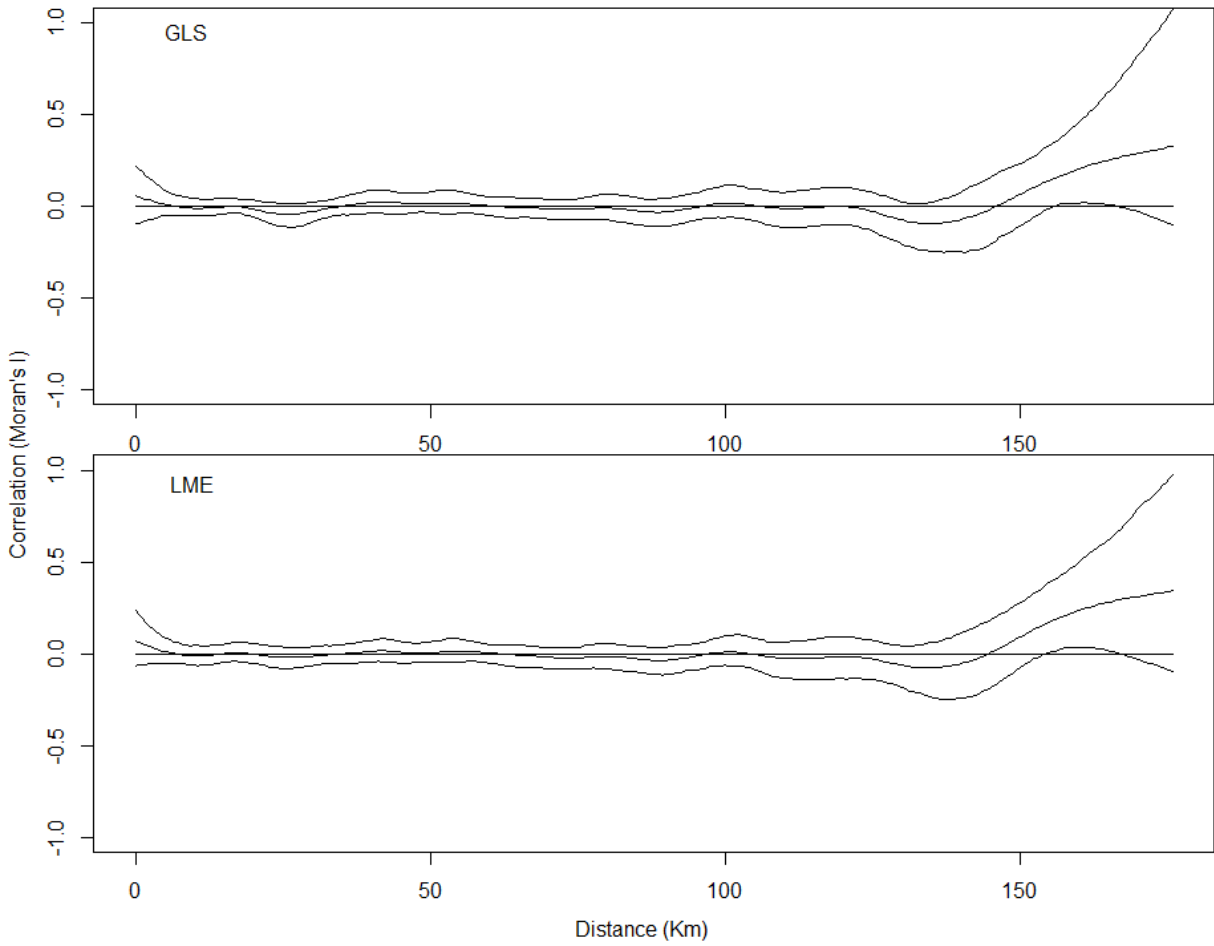


**Figure 2:** Fish sampling locations with loadings from segment level watershed (Segment Level) and HUC12 (Neighborhood) level Principal Components Analyses. Landscape variables were considered as significant contributors to the PC axes if loading  $\geq |0.40|$ .

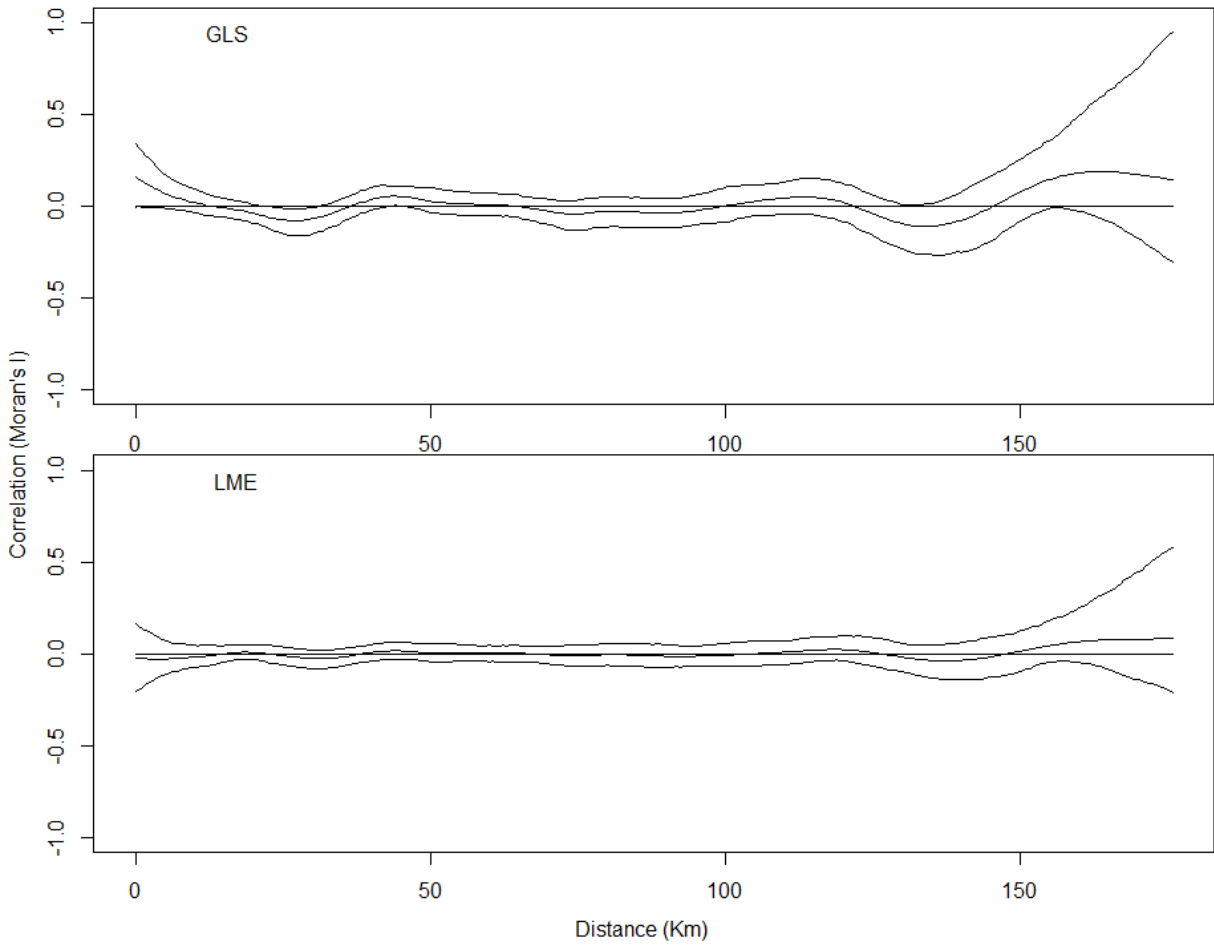




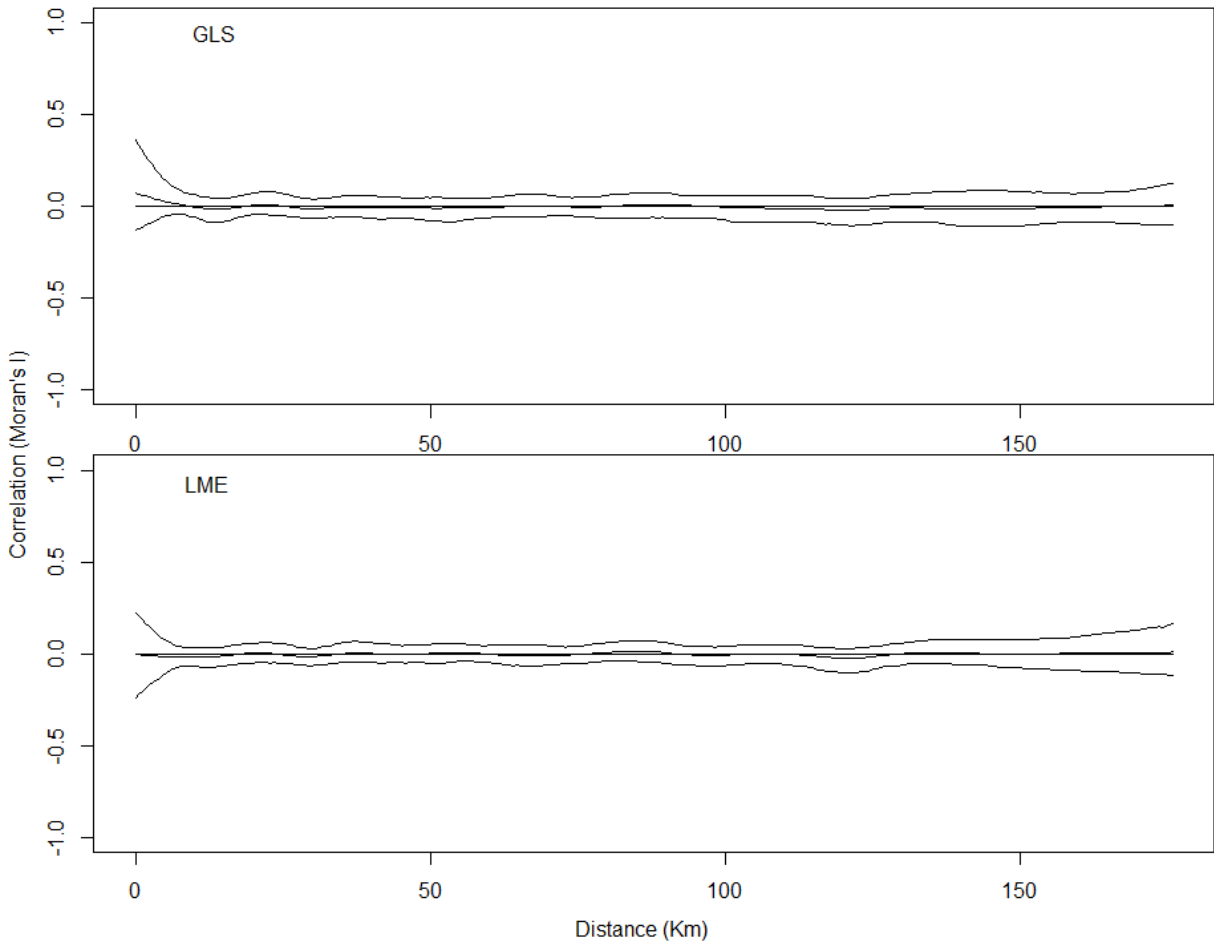
**Figure 3:** Correlogram of the residuals from the generalized linear model for species richness.



**Figure 4:** Correlograms comparing the residuals of the generalized least squares (GLS) and linear mixed effects models (LME) for Shannon-Weaver diversity. Both correlograms indicate no spatial autocorrelation in diversity after the models were applied.



**Figure 5:** Correlograms comparing the residuals of the generalized least squares (GLS) and linear mixed effects models (LME) for WV IBI scores. The GLS correlogram indicates small spatial autocorrelation after the model was applied. The same spatial autocorrelation is not present in the LME correlogram.



**Figure 6:** Correlograms comparing the residuals of the generalized least squares (GLS) and linear mixed effects models (LME) for the proportion of tolerant individuals. The GLS correlogram indicates small spatial autocorrelation after the model was applied. The same spatial autocorrelation is not present in the LME correlogram.