

Article

# Understanding Landscape Influences on Aquatic Fauna across the Central and Southern Appalachians

R. Daniel Hanks <sup>1,\*</sup> , Paul B. Leonard <sup>2</sup>  and Robert F. Baldwin <sup>1</sup> 

<sup>1</sup> Department of Forestry and Environmental Conservation, Clemson University, Clemson, SC 29634, USA; baldwi6@clemson.edu

<sup>2</sup> Office of Science Applications, U.S. Fish and Wildlife Service, Fairbanks, AK 99701, USA; paul\_leonard@fws.gov

\* Correspondence: rhanks@clemson.edu; Tel.: +1-(864)-656-4853

Received: 6 December 2019; Accepted: 7 January 2020; Published: 10 January 2020



**Abstract:** For the success of aquatic conservation efforts, it is imperative for there to be an understanding of the influences multiple stressors across the landscape have on aquatic biota, as it provides an understanding of spatial patterns and informs regional stakeholders. The central and southern Appalachians contain biodiversity hotspots for aquatic fauna. Therefore, we sought to create a comprehensive multimetric model that is based on the influence of abiotic factors on fish and aquatic macroinvertebrates that could predict watershed quality. Good spatial coverage exists for land use/land cover (LULC) and other physicochemical components throughout the region, yet biological data is unevenly distributed, which creates difficulties in making informed management and conservation decisions across large landscapes. We used boosted regression trees (BRT) to model a variety of biological responses (fish and aquatic macroinvertebrate variables) to abiotic predictors and by combining model outputs created a single score for both abiotic and biotic values throughout the region. The mean variance that was explained by BRT models for fish was 73% (range = 48–85%) and for aquatic macroinvertebrates was 81% (range = 76–89%). We categorized both predictor and response variables into themes and targets, respectively, to better understand large scale patterns on the landscape that influence biological condition of streams. The most important themes in our models were geomorphic condition for fish and water quality for aquatic macroinvertebrates. Regional models were developed for fish, but not for aquatic macroinvertebrates due to the low number of sample sites. There was strong correlation between regional and global watershed scores for fish models but not between fish and aquatic macroinvertebrate models. We propose that the use of such multimetric scores can inform managers, NGOs, and private land owners regarding land use practices, thereby contributing to large landscape scale conservation efforts.

**Keywords:** Aquatics; modeling; boosted regression trees; Appalachians

## 1. Introduction

Aquatic biodiversity is declining faster than terrestrial biodiversity globally and one of the contributing factors is the condition of the landscapes in which aquatic systems are embedded, coupled with the aquatic environments themselves [1,2]. There is a need to understand the drivers of this decline across a geography of concern in order to reverse or stabilize this trend and conserve aquatic species and ecosystems (e.g., the central and southern Appalachians). Biological sampling efforts can be geographically extensive (e.g., National Fish Habitat Partnership, state agency sampling, etc.), yet it is intractable to consider biological sampling of every stream segment or small watershed across large geographies. Therefore, there is a need to develop an understanding of the underlying

physical, chemical, and biological influences that affect the population and community condition, model those effects, and extend predictions to unsampled stream segments or watersheds.

Assessing the biological condition of rivers and streams has been performed at various spatial scales and grains. As data availability and computing power both increase, the ability to perform spatial analysis across large geographic extents and at small grain size is increasing. Examples of large scale spatial analyses include Esselman et al. [3], who used fish as indicators, the US Environmental Protection Agency's Wadeable Stream Assessment [4], which included benthic macroinvertebrates and fish as indicators, and Hill et al. [5], who modeled biological stream condition across the conterminous U.S. based on anthropogenic and natural watershed features. Multimetric indices (MMIs) are commonly used in studies that attempt to understand anthropogenic influences on the condition of biological aquatic communities over vast geographies [6–9]. Multimetric indices use taxonomic and functional metrics that are known to be sensitive to gradients of anthropogenic disturbance as a means of assessing the health of aquatic systems.

Landscape-scale analyses are important, as they allow for a more complete understanding of anthropogenic stressors on biological and ecological needs of species, populations, and communities across broad spatial extents. We define landscape scale as the spatio-temporal extent that allows for multigenerational population processes to occur [10]. Such landscape scale analyses require extensive data that are representative across taxonomic groups through both space and time, and physicochemical conditions must be represented, which range from natural to extensive anthropogenic disturbance [11]. It is often difficult to ensure that data sets are consistent across regions, representative taxa (response variables), and physicochemical factors (predictor variables). These difficulties arise due to biological and environmental datasets being expensive to obtain, varying goals and objectives from entities collecting data, and varying monetary support for such collections. This often results in either poor spatial coverage and spatially clumped data or both. Data sets are often combined from various collection programs to overcome these difficulties. However, this can result in data existing at varying taxonomic levels, problems with standardization of sampling protocols, and biases that are associated with site selection. For example, variability in site densities may result in biases towards trends that are prevalent in heavily sampled areas. Angermeier and Smogor [12] found that sampling effort influenced accuracy and precision of estimates of biological community attributes.

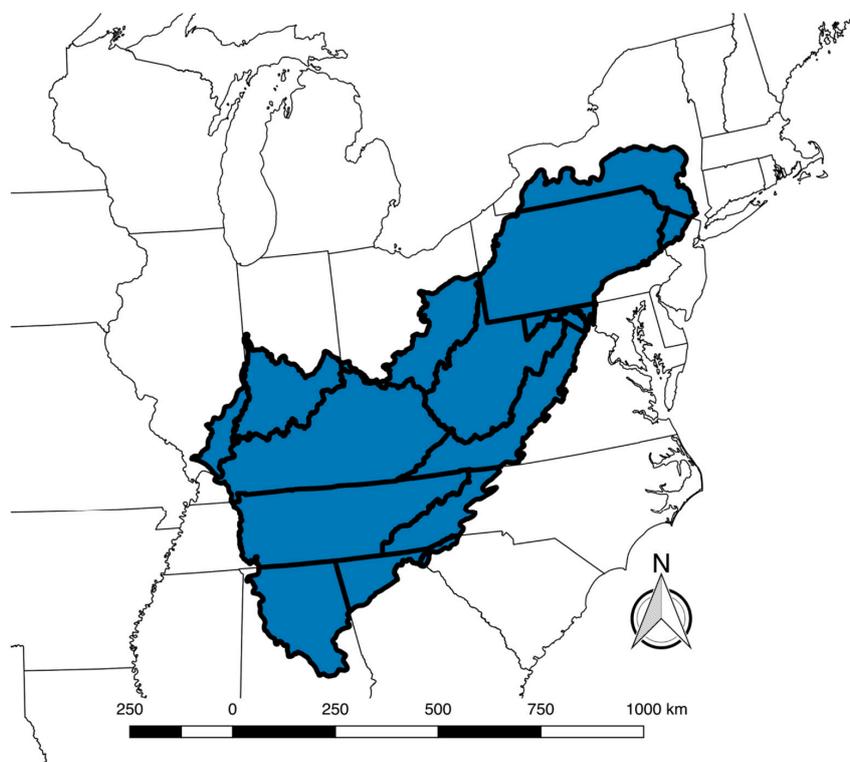
Anthropogenic induced stressors often correlate strongly with biological conditions of aquatic systems and, thus, landscape-scale analyses typically aim to use land cover and land use practices over large geographies as surrogates for local environmental conditions to provide insight into biological conditions of streams and rivers. The use of such analyses allow for broad geographic scale understanding of the existing gradients of poor-to-good biological condition and the environmental factors that affect them. This aids in identifying the landscape-scale stressors that further the ability of conservation practitioners, managers, and regulators to take actions that positively impact aquatic fauna through a greater understanding of large-scale stressors. However, anthropogenic stressors and their biological responses are heterogenous across large areas, but they are the result of local biogeographical interactions with both natural and anthropogenic stressors over expansive time periods [3].

Indices for evaluation of biological condition across large geographies have become standard procedures, enabling transferability across regions through the use of statistically robust methods [3,11,13,14]. Biological condition indices should sufficiently cover the range of environmental gradients, exhibit low temporal variability to repeated visits to the same site(s), and demonstrate high responsiveness to environmental degradation. Indicators that do not meet these requirements should be excluded from analyses [11,15]. The use of the screening methods provided by Whittier et al. [15] and Stoddard et al. [11] allow for a large set of contending predictor variables to be winnowed down to a smaller, manageable group of indicators used to describe the biological condition gradient in a consistent manner across regions.

Multimetric indices (MMI) combine a variety of biological indicators and indices. While using boosted regression trees (BRT; [16]), we developed watershed-based MMIs that combined fish and aquatic macroinvertebrates across the Appalachian Landscape Conservation Cooperative (App LCC) region (see methods for full description) using the methods that were developed by Stoddard et al. [11] and Whittier et al. [15]. Additionally, we aimed to identify large scale themes that structure biological communities and thus biological condition across this large geography. The App LCC encompasses biologically important aquatic ecosystems (e.g., Tennessee and Mobile Rivers) and it is predicted to undergo continued land conversion, resulting particularly in increased urbanization [17]. Using broad scale land use information to develop models to predict biological condition of streams and watersheds provides the opportunity to predictively map spatially explicit biological condition across large geographic regions [5,18]. While studies have been performed to understand the response of fish and aquatic macroinvertebrates for the conterminous United States [3–5], a study that is spatially explicit to the central and southern Appalachians that could inform aquatic conservation within the App LCC region is lacking. An understanding of the anthropogenic threats that exist in the Appalachians is an important endeavor, as more than half of the U.S. population lives within a day’s drive of some part of the Appalachians [19]. Thus, our objective was to model and map the current biological condition and develop an understanding of the threats to conservation across the region as an important step towards identifying conservation opportunities and needs, so that conservation practitioners can act accordingly.

## 2. Materials and Methods

Our study focused on the rivers, streams, and landscape encompassed by the App LCC. The App LCC comprised 592,129 km<sup>2</sup>, 266,307 HUC 12 watersheds, and intersects 15 states that range from central New York to the north, and central Georgia, Alabama, and Mississippi to the south [20] (Figure 1).



**Figure 1.** Overview map of the Appalachian Landscape Conservation Cooperative [21]. Map created by R. Daniel Hanks using ArcGIS v10.3 [22].

The eastern most extent reaches western New Jersey in the north and the western most portions of North and South Carolina, while reaching as far west as southern Illinois. The topography and habitat of the region are complex over such a large land area, which is comprised of six Level II ecoregions (U.S. EPA), 11 Freshwater Ecoregions, and seven major river basins and has been greatly influenced by the lack of glaciation and ancient geology, which gave rise to high levels of aquatic biodiversity. The App LCC region encompasses some of the most biologically diverse freshwater fish resources in the United States [23]. This is particularly true in the southwestern portion of the region [24]. As the 19th and 20th centuries have progressed, aquatic diversity has been impacted by anthropogenic activities that range from agricultural practices, land conversion, and extractive industries.

Through the aid of a steering committee comprised of 60 regional experts, which include academic scientists, state wildlife agency managers, non-governmental organization scientists, and conservation practitioners, we identified 52 predictor variables comprising six themes that are known to influence fish and aquatic macroinvertebrate communities (Table A1). The themes included streamflow, geomorphic condition, connectivity, water quality, non-point source pollution, and point source pollution, and they were established and evaluated to understand large scale thematic processes that influence aquatic biota across the App LCC. The predictor data were compiled from regional and national datasets at the National Hydrography Dataset version 2 (NHD + V2) HUC12 watershed level (Final predictors used in BRT models are shown in Table 1; a comprehensive list of predictor data with source is provided in Appendix A).

**Table 1.** List of final predictors used in boosted regression tree models and how each was grouped thematically.

Aquatic Habitat Metric	Themes	
Flow Alteration from Storage (total storage/mean annual flow)	Flow regime	
Density and type of dam		
Altered streamflow		
Agricultural water withdrawal		
Industrial water withdrawal		
Erosive Forces	Geomorphic condition	
Resistive forces		
Density of dams: Catchment		
Density of dams: Watershed	Connectivity	
Density of crossings: Catchment		
Density of crossings: Watershed		
Nitrogen	Water Quality	Overall Predictors Score
Phosphorus		
Dissolved Organic Carbon		
% Impervious Surface in Watershed, Active River Area, & Catchment	Non-point sources of pollution	
% Natural Cover in Watershed & Active River Area		
% Agriculture in Watershed, Active River Area, & Catchment		
Comprehensive Environmental Response, Compensation, and Liability Information System site density	Point sources of pollution	
Permit Compliance System site density		
Toxic release inventory site density in Watershed and Catchment		
Coal mine density		
Wind turbine density		
All mine density in Watershed and Catchment		
Natural gas well density		

Riparian corridors are known to be influential for developing ecological diversity [25] and in structuring aquatic communities; hence, we used The Nature Conservancy's Active River Area [26] as a buffer into which land cover variables were clipped for modeling. The active river area framework

provides an ecologically sound modelling mechanism to capture important variables that structure aquatic ecosystems [26]. We centered and scaled all of the predictor variables to overcome instances where predictor variables had wide ranges resulting single predictors dominating models.

Our fish and aquatic macroinvertebrate community data were assembled from federal databases of national datasets (United States Environmental Protection Agency’s National Rivers and Streams Assessment and the National Fish Habitat Partnership). We used these community datasets as the predictors to assess the status of biological communities within HUC12 watersheds (Table 2).

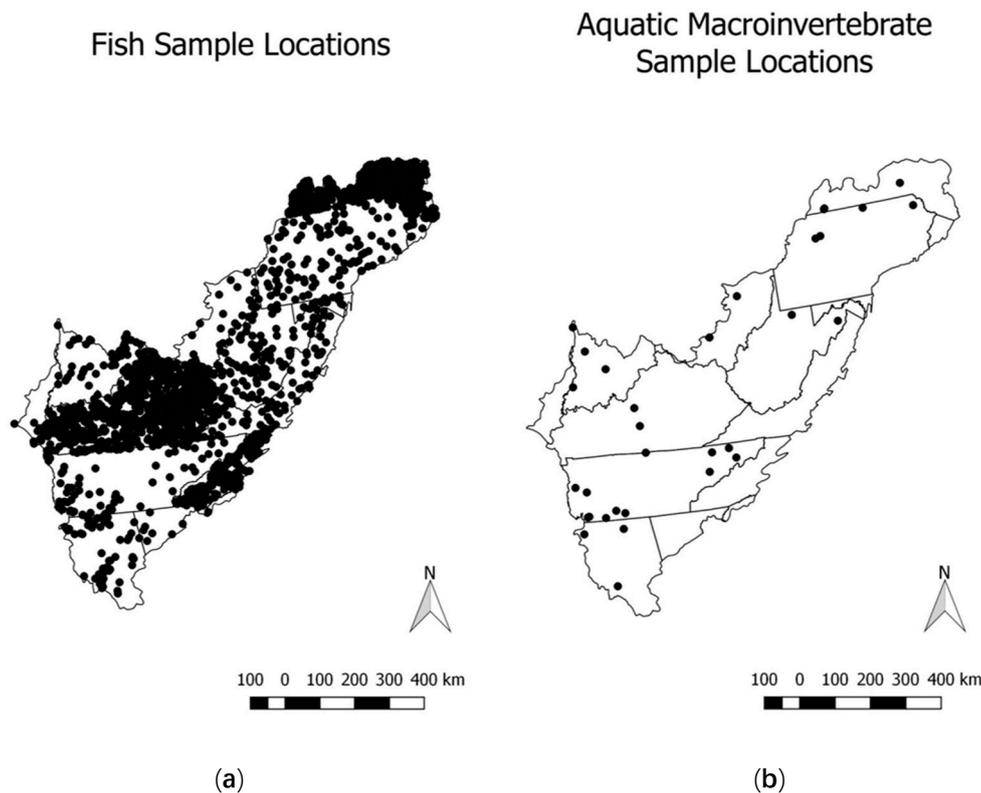
**Table 2.** List of final responses modeled with boosted regression trees and how each was condensed (i.e., averaged and grouped) into an average score to represent larger taxonomic resolution.

Biological Metric	Targets	Condensed Attributes	
Diversity	Shannon Diversity		
Invertivore Taxa			
Piscivore Taxa	Functional Group		
Herbivore Taxa			
Lithophilic Spawners		Fish Score	
Taxa Preferring Coarse Sediment	Fish Taxa Quality		Overall Response Score
Intolerant Taxa			
Tolerant Taxa			
EPT Taxa			
5 Dominant Taxa	Macroinvert Taxa Quality	Macroinvert Score	
Intolerant Taxa			
Tolerant Taxa			

The sample point locations were spatially joined to HUC12 watersheds by projecting into a conic equal area projection in Quantum GIS [27]. Our fish community dataset was comprised of 2991 sample sites across the region and was somewhat sparse (2991 locations within 266,309 HUC12 watersheds = 1.1%), but it exhibited good spatial distribution (Figure 2).

Our aquatic macroinvertebrate dataset was comprised of only 194 sampling locations and it had poor spatial coverage (194 locations of 266,309 HUC12 watersheds = 0.07%) and distribution across the region (Figure 2).

For fish, we calculated taxonomic richness and diversity and several functional group metrics while using the fish traits database that was provided by Frimpong and Angermeier [28] and the tolerance database of Stoddard et al. [29]. In total, we calculated 45 metrics for further analysis of fish responses to our predictor dataset; however, our final analysis used only nine of these metrics and these were further aggregated into three groupings that we called targets (e.g., Shannon diversity, functional group, fish taxa quality) (Table 2). For aquatic macroinvertebrates the metrics modeled as response variables were % EPT (i.e., percent Ephemeroptera/Plecoptera/Trichoptera), % of total abundance comprised by the five most dominant taxa, % of total abundance that were tolerant taxa, and % of the total abundance that were intolerant taxa (Table 2). These four metrics were aggregated into a target called macroinvertebrate taxa quality (Table 2).



**Figure 2.** Side-by-side comparison of response data for (a) fish and (b) aquatic macroinvertebrates showing spatial distribution of sample locations for both. Data source: United States Environmental Protection Agency’s National Rivers and Streams Assessment and the National Fish Habitat Partnership. Map created by R. Daniel Hanks using ArcGIS v10.3 [22].

Prior to model development, we successively screened all of the predictor variables (e.g., [3,11]). Screening included evaluating predictor variables for: 1) range, where predictor variables with small range (e.g., <4) were removed, 2) percent zero or NAs, where variables with >33% zero or NA were removed, and 3) percent values equal to the mode, where the variables with >75% of the values were the same were removed. After screening, the 52 predictor variables were reduced to 24, which comprised six themes (Table 1). Subsequently, we employed boosted regression trees (BRTs) to predict fish and aquatic macroinvertebrate response variables at the watershed scale (Table 2). Boosted regression trees are robust to missing data, have both explanatory and predictive power, and they are capable of handling complex relationships that are both non-linear and interactive without the need for prior data transformation. By iteratively fitting and combining simple regression models, BRT models improve model structure and predictive performance. Tree complexity and learning rate were systematically altered to identify the optimal model structure. We reduced model complexity by removing the redundant variables ( $r > 0.80$ ) and variables with minimum variation among sites. Global models were developed with the remaining variables and then further simplified while using scree plots of the predictor relative influence. Simplified models were retained if their cross-validation was greater than or equal to that of the global model. BRT models were developed with methods and code that were provided by Elith et al. [16] while using Program R: A language and environment for statistical computing [30]. Initial analyses were performed across the entire geography without accounting for regional differences (i.e., ecoregions), but we used Freshwater Ecoregions of the World [31] to understand these regional differences due to a clear need for the inclusion of regional differences within the models. Freshwater ecoregions attempt to place aquatic ecosystems in a biogeographical framework that is based on ecological and evolutionary patterns of fish distributions [31]. We aggregated the 11 ecoregions that comprised the App LCC into four due

to low sample site numbers in some Freshwater Ecoregions. The four aggregated ecoregions and their composite Freshwater Ecoregions were Atlantic Slope (Appalachian Piedmont, Chesapeake Bay, Northeast US and Southeast Canada Atlantic Drainages), Cumberland (Cumberland River Basin), Ohio (Teays-Ohio and Laurential Great Lakes River Basins), and Tennessee (Tennessee and Mobile Bay River Basins). The regional models were successfully run for fish, but we were unable to model aquatic macroinvertebrates regionally due to the lack of spatial coverage, and we therefore retained models for the entire geography for aquatic macroinvertebrates.

Predictions of response variables from BRT models were centered and scaled. As a multimetric approach, the predicted responses (biological metrics in Table 1) were then averaged within their target group (see Table 1) and further condensed via averaging into overall fish and aquatic macroinvertebrate scores. Boosted regression trees provide the relative influence of predictor variables in the modelling of responses, and we averaged these relative influences within our thematic framework (see above and Table 1) for each fish and aquatic macroinvertebrate response variable.

### 3. Results

The screening process reduced the total number of predictor variables to 37 from 52. When the predictor variables were correlated, we attempted to retain the variable that was most responsive to our models; however, this was not formally tested. The themes with the greatest number of predictor variables were non-point source pollution and point source pollution (each with eight variables). While the theme with the least number of predictor variables was geomorphic condition (two variables) (Table 1).

Below, we present the results from the boosted regression tree models where the models were run with all ecoregions together for both fish and aquatic macroinvertebrates. We refer to this as a “Global” model (see below). After which, we present the results where separate models were run for each Freshwater Ecoregion (defined in the materials and methods section). We refer to these models as regional models and, because there were not enough data points within the regional data sets for aquatic macroinvertebrates, they were developed for fish only.

#### 3.1. Global Models

Boosted regression tree models were developed for 45% (28/62) of the response variables; however, we only retained 12 (eight and four response variables for fish and aquatic macroinvertebrates, respectively) for our final multimetric index of biotic condition (Table 2). Response variables were collapsed into targets, where, for the fish category, we retained three targets (Shannon diversity, functional group, and taxa quality) and for aquatic macroinvertebrates, we retained macroinvertebrate taxa quality as the only target.

The range of variance that was explained by the BRT models was 48% for percentage of invertivore taxa to 85% for fish taxa diversity ( $\bar{x} = 73\%$ ) for fish and 76% for percentage of the five dominant taxa to 89% for percentage of tolerant taxa ( $\bar{x} = 81\%$ ) for aquatic macroinvertebrates. For the fish BRT models, the cross-validated deviance ranged from 32% for percentage of invertivore taxa to 69% for fish taxa diversity ( $\bar{x} = 52\%$ ), while it ranged from 24% for percentage of the five dominant taxa to 76% for percentage of tolerant taxa for aquatic macroinvertebrates ( $\bar{x} = 41\%$ ) (Table 3).

**Table 3.** Boosted regression tree result for fish (richness, diversity, % invertivore taxa, % piscivore taxa, % herbivore taxa, % lithophilic taxa, % taxa preferring coarse substrate, % tolerant taxa, and % intolerant taxa) and aquatic macroinvertebrates (% EPT taxa, % 5 dominant taxa, % tolerant taxa, and % intolerant taxa).

<i>Fish</i>	Diversity	% Invertivore Taxa	% Piscivore Taxa	% Herbivore Taxa	% Lithophilic Taxa	% Coarse Sediment Taxa	% Tolerant Taxa	% Intolerant Taxa
# of Trees	5950	5600	4700	3850	4600	8150	4550	5650
Learning Rate	0.01	0.001	0.01	0.01	0.01	0.001	0.01	0.01
Total Deviance	902.95	0.133	18.78	12.57	19.22	6.48	12.85	12.63
Residual Deviance	298.56	0.12	8.19	7.8	9.52	4.97	6.12	6.49
Variance Explained	85%	48%	79%	70%	78%	56%	79%	80%
CV Deviance (se)	433.96(6.15)	0.12(0.03)	11.31(0.22)	10.23(0.22)	12.79(0.17)	5.51(0.23)	8.85(0.31)	10.08(0.26)
CV Deviance Explained	69%	32%	64%	48%	61%	45%	51%	41%
<i>Aquatic Macroinvertebrates</i>	% EPT	% 5 Dominant Taxa	% Tolerant Taxa	% Intolerant Taxa				
# of Trees	2150	8300	2700	2600				
Learning Rate	0.001	0.001	0.001	0.0005				
Total Deviance	2.83	2.27	6.45	1.05				
Residual Deviance	1.61	1.68	1.81	0.56				
Variance Explained	81%	76%	89%	77%				
CV Deviance (se)	2.65(0.32)	2.20(0.25)	2.88(0.35)	0.85(0.12)				
CV Deviance Explained	27%	24%	76%	36%				

The lower cross validated deviances for aquatic macroinvertebrates relative to those of fish are likely due to the lack of spatial data coverage in the aquatic macroinvertebrate data (Figure 2b).

The predictor variables that had the greatest relative influence were freshwater ecoregion and network percent impervious surface for fish and aquatic macroinvertebrates based on the results of the BRT models, respectively. Those with the least relative influence were how confined a stream reach was (CONF\_CL) and watershed mine density for fish and aquatic macroinvertebrates, respectively (Table 4).

The theme with the greatest cumulative relative influence for fish was geomorphic condition (46.6%) and for aquatic macroinvertebrates was water quality (41.7%), while point source pollution was the least important theme for both fish (1.4%) and aquatic macroinvertebrates (0.8%). The relative thematic influence was variable between response variables; where, for example, the flow theme was the most influential for % EPT (27.4%), but the most influential theme was water quality for % intolerant taxa (87.9%) (Table 5).

**Table 4.** Summary of the relative influence of each predictor variable on the response variables for both fish and aquatic macroinvertebrates.

Variable	Fish									Aquatic Macroinvertebrates			
	Richness	Diversity	% Invertivore Taxa	% Piscivore Taxa	% Herbivore Taxa	% Lithophilic Taxa	% Coarse Sediment Taxa	% Tolerant Taxa	% Intolerant Taxa	% EPT	% 5 Dominant Taxa	% Intolerant Taxa	% Tolerant Taxa
ALK_CL	0.3	0.3	0.1	0.8	0.4	1.7	0.6	0.5	0.7	0.4	13.2	0.2	0
bfi_avg	2.9	5.4	6	3.8	1.9	2.5	3.1	3.7	4.2	1.4	1.7	3.5	1.8
CONF_CL	0.4	0.2	0	0.1	0.3	0.2	0.1	0.6	0.2	5.7	0.1	1.4	0
DamDensWs	0.6	0.5	0.3	0.6	1.1	0.7	0.8	1	0.6	0.1	0.6	0.2	0.1
elevcm_avg	11.2	2	3.8	4.8	4.6	8.8	10	4	4.5	3.5	1.8	2.6	2.2
FWECOREG	4.8	29.3	4.5	6.7	12.1	4.2	5.5	7.3	4.9	11.2	9.4	2.2	1.9
imp11_per	1.9	1.3	1.3	2	1.8	1.9	2.1	2.2	3.6	2.8	4.3	12.2	2.6
imp11_perN	3	2.1	1.7	4.4	4.2	5.3	5.2	6.5	4.9	6.7	15.7	5.1	9
IncrTempMM	9.3	5.3	3.5	16.8	9.2	4.3	6.1	5.7	3.5	2.3	1.2	8.4	2
kfact_avg	2.7	1.2	4.3	3.9	2.4	3.7	3.9	3.1	2.6	0.9	1.1	3.4	2.1
kfactup_av	2.7	3.2	2.5	3.6	2.8	2.3	2.6	2.6	3.2	0.9	1.5	0.8	2.3
MineDensWs	0.4	0.4	0.3	0.4	0.4	0.4	0.8	0.4	0.3	0	0	0	0
NCat	2.8	1.8	18.2	3.7	3.4	2.9	4	3.9	4.3	4.3	7.9	4.6	2.2
nid_storN	5.9	2.5	5.1	2.9	5	5.7	5.8	1.9	2.3	2.4	2.7	0.6	14.7
nlcd11_ag	1	0.8	0.2	1.4	1.3	1	2.3	1.7	2.1	2.8	3.4	2.1	1.5
nlcd11_agN	1.7	1.3	0.5	2.1	2.5	2.2	3.3	2.8	2.6	3.9	1.5	4.3	1.5
nlcd11_t	1.3	1	0.5	1.5	1.7	1.4	1.5	1.8	2.1	3.5	1.3	3.2	0.8
NPDESdensW	0.4	0.3	0	0.3	0.7	0.3	0.4	0.4	2.1	0	0	0	0.1
P2O5Cat	2	1.7	2.3	1.5	1.8	1.5	1.6	2	2.5	2.9	0.8	3.3	0.7
P2O5Ws	1.8	2.1	0.8	1.3	2.2	1.4	2	2.6	2.8	4.3	0.9	1.3	1.5
Pct_Ag	3.3	2.1	2	1.3	2	2	1.5	1.7	2.8	4.6	2.5	1.7	4.6
Pct_Imp	2.5	1.5	1.6	2.2	2.3	3.4	1.5	2	2.4	2	2.3	1.7	5.4
Pct_Nat	3	1.7	1.9	2.3	3	3.4	4.1	4	3.4	5	0.7	5.4	2
rchr_avg	3.5	4.7	8.3	3.4	4.5	2.8	4.5	6.8	4.9	2.3	2.3	2.5	5.5
RdCrsCat	0.9	0.9	0.2	1.1	1.1	1	0.9	2.2	1.1	1.6	2.2	1.9	0.3
RdCrsWs	1.3	1.6	0.8	1.7	1.9	1.4	1.9	2.9	2.3	2.4	1.8	2.7	2.9
rf30_avg	7.1	7.2	10.8	6.9	6.3	4.3	4.9	4.1	5.3	1.9	1.2	6.5	2.2
sand_avg	2.4	4.7	2.8	3.3	3.1	2.5	3.9	3.6	4.2	4.8	3.6	7.1	9.5
silt_avg	2.4	3.3	1.6	3	3.5	2.8	2.1	2.7	4.4	3	2.6	1.1	4.9
SIMP_NUM	3.1	1.2	1.3	2.2	3.3	0.9	1.5	1.9	1.3	1.1	1.8	2.5	1.5
slope_avg	2.7	2	2.3	1.7	2.4	3	2.4	2.2	3.2	6.6	3	1.9	0.8
slope_avgN	3	2.9	2.7	3.6	3.9	3.1	3.8	4.1	5	2.2	3.6	3.8	1
SLP_CL	2.7	0.8	0.2	1.9	0.7	2.9	2	0.6	1.2	0.1	1.2	0.2	3.5
SZ_CL	4	0.8	6.6	1.8	1.1	2.5	2	1.8	2.3	1.1	2	0.6	8.2
TEMP_CL	0.4	0.4	0	0.1	0.3	10.9	0.4	0.1	0.5	0	0.2	0.1	0
TNC6HYDP	0.3	0.4	0.8	0.1	0.2	0.2	0.1	3.8	0.3	0	0.1	0.3	0.5
TRIDensWs	0.3	1	0.1	0.7	0.6	0.5	0.9	0.6	0.5	0.8	0	1.1	0

**Table 5.** Summary of the average relative influence for each theme on response variables modeled with the boosted regression trees for both fish and aquatic macroinvertebrates. The average thematic % relative influence is the percent relative contribution of each theme to the BRT model. Themes are comprised of flow condition (NID storage, dam density, altered streamflow, agricultural and industrial water withdrawals), geomorphic condition (erosive and resistive forces), connectivity (dam and road crossing density at the watershed and catchment levels), water quality (total nitrogen load, total phosphorous load, and dissolved organic carbon), non-point source pollution (% agricultural land cover, % impervious surface, and % natural land cover in the watershed, active river area, and catchment), and point source pollution (CERCLA site density, permit compliance system site density, toxic release site density, coal mine, wind turbine and natural gas well density, and mine density).

Variable	Fish								Aquatic Macroinvertebrates					
	Diversity	% Invertivore Taxa	% Piscivore Taxa	% Herbivore Taxa	% Lithophilic Taxa	% Coarse Sediment Taxa	% Tolerant Taxa	% Intolerant Taxa	% EPT	% 5 Dominant Taxa	% Intolerant Taxa	% Tolerant Taxa		
Flow	15	26.3	17.2	21.8	25.2	24.8	17.9	16.4	20.6	27.4	14.6	11	7	15.0
Geomorphic	62.5	48.4	51.5	47.6	38.7	46.3	38.7	39.4	46.6	23.4	36.7	6.6	17.3	21.0
Connectivity	4.4	2.3	6	6.7	7	3.7	10.9	9.3	6.3	15.7	13.2	2.3	8	9.8
Water Quality	12	18.5	9.9	6.7	6.1	7.4	14.6	11.5	10.8	12.8	9.2	87.9	56.7	41.7
Non-point Pollution	7.5	7.3	15.7	16.8	22.2	17.5	18.7	22.4	16.0	20.9	27.8	6.8	19.2	18.7
Point Pollution	0.6	0.3	1.4	1.4	1.8	1.5	1.8	3	1.5	1.8	0	0.1	1.2	0.8

### 3.2. Regional Models

Regional differences were evident in the results of the relative influence of predictor variables on responses. For example, the variable with the greatest relative influence for richness in the Atlantic Slope data was July temperature (22.8%), for the Cumberland River Basin, was ground flow recharge (17.6%), for the Ohio River Basin, was average catchment elevation (21.4%), and for the Tennessee River Basin, it was dam storage capacity (8.2) (Table 6).

**Table 6.** Summary of the relative influence within each region of each predictor variable on the response variables for fish. A = Atlantic Slope, C = Cumberland River Basin, O = Ohio River Basin, and T=Tennessee River Basin.

Variable	Richness				Diversity				% Invertevore Taxa			
	A	C	O	T	A	C	O	T	A	C	O	T
ALK_CL	0.1	2.1	0.4	0.1	0.1	2.3	0.6	0.3	0.6	0.5	0.3	0.3
bfi_avg	1.7	6.8	3.4	3.2	3.2	1.2	2.1	25.8	2.7	2.6	0.7	25.8
CONF_CL	0.2	0.1	0.4	0.1	0.2	0.4	0.3	0.1	0.1	0.1	0.1	0.1
DamDensWs	0.4	0.5	0.5	0.6	1	0.5	0.8	0.1	0.8	0.3	0.8	0.1
elevcm_avg	4.1	4.1	21.4	6	4.6	5.3	3.7	4.3	4.2	2.7	10.8	4.3
FWECOREG	3.1	0	0.9	0	3.1	0	2.2	0	0.1	0	0.2	0
imp11_per	1.1	4.2	0.9	3.3	1.9	1.3	1.8	3	2.8	5.4	0.5	3
imp11_perN	2.2	4.4	2.9	3.6	3.7	2.5	3.3	2.7	8.6	4.9	0.8	2.7
IncrTempMM	22.8	4.6	11.3	6.7	18.4	1.7	4.7	2.9	5.3	1.2	4.3	2.9
kfact_avg	3.1	2.1	1.2	4.1	4	2.3	1.4	1.1	4.5	1.9	1.2	1.1
kfactup_av	1.8	1.3	4.8	2.3	1.8	2.5	4	1.5	3.5	1.2	1.7	1.5
MineDensWs	0.2	0.3	0.6	0	0.3	0.1	0.7	0	0.5	0.1	0	0
NCat	2.7	3	1.1	4	1.9	2.5	2.6	1.1	3.2	9.3	21.6	1.1
nid_storN	5.3	4.5	6.9	8.2	8	0.9	3.3	3.1	12	0.9	8.1	3.1
nlcd11_ag	1.9	2.1	0.6	0.8	1.6	1.1	1.2	1.3	1.8	1	0.8	1.3
nlcd11_agN	2.5	3.9	1.2	1.1	1.5	1.3	1.9	2	2.2	2.8	0.3	2
nlcd11_t	1.3	5.2	0.4	1.4	0.9	2.9	1.6	1	1.7	8.9	0.1	1
NPDESDensW	0.5	0.4	0.3	0.3	0.4	0.4	0.4	0.1	0.5	0	0	0.1
P2O5Cat	2.8	2.1	1.2	2.5	1.5	1.5	1.8	2.9	2.2	3.6	0.4	2.9
P2O5Ws	2.9	1.4	1.9	1.6	3	2.2	2.6	1.4	2	2.1	1	1.4
Pct_Ag	2.6	1.8	2.5	4.8	2.3	2.8	2.5	4.7	3.1	1.4	0.8	4.7
Pct_Imp	2.1	2.7	1.1	1.6	2.6	2.7	1.7	2.3	1.8	5	1.9	2.3
Pct_t	2.1	2.6	2.7	2.9	2.6	1.9	2.8	1.7	2.2	6	0.9	1.7
rchrng_avg	3.9	17.6	2.1	3.4	2.7	5.3	6	5.2	3.5	7.1	1	5.2
RdCrsCat	0.8	0.8	0.4	1	0.8	1	1.4	0.6	1.6	2	0.1	0.6
RdCrsWs	1	2.1	0.4	1.3	2.2	2.8	3.2	2.2	2	3.2	0.2	2.2
rf30_avg	3.9	5.5	4.9	5.4	2.9	26.4	18.6	3.2	5	2.2	7.4	3.2
sand_avg	2.3	2.6	1.6	5.2	2.3	6.1	7	2.7	4.5	8.6	2.2	2.7
silt_avg	5.6	1.4	1.2	2.8	6.4	8.4	3.2	2.9	3.3	2.5	1.5	2.9
SIMP_NUM	3.9	2.6	1.9	5.1	1.8	1.5	1.2	3.4	2.3	3.3	2.6	3.4
slope_avg	1.8	1.5	2.2	4.8	1.9	0.9	2.9	2.9	2.4	3.4	3.5	2.9
slope_avgN	2	2.7	3.6	2.3	3	4.2	3.5	0.9	3.2	2.2	0.9	0.9
SLP_CL	1.8	0.7	3.1	1.9	2.5	1.9	1.3	0.4	0.9	2	0.1	0.4
SZ_CL	3.1	1.8	8.1	6.8	3.2	0.6	1.3	0.8	4.2	1.6	21.1	0.8
TEMP_CL	1.2	0.1	1.1	0	1.1	0.5	0.4	0	0.1	0.1	0.1	0
TNC6HYDP	0.2	0.2	0.3	0.2	0	0	0.5	9.5	0.2	0	1.7	9.5
TRIDensWs	0.6	0.2	0.4	0.6	0.7	0.2	1.4	1.9	0.7	0.1	0.1	1.9

Table 6. Cont.

Variable	% Piscivore Taxa				% Herbivore Taxa				% Lithophilic Taxa			
	A	C	O	T	A	C	O	T	A	C	O	T
ALK_CL	0.5	0	1.1	0.4	0.6	6.6	0.6	0.4	0.4	8	0.8	1
bfi_avg	2.4	5	2.7	6.1	2.7	2.1	2.1	2.7	1.5	2.2	2	2.6
CONF_CL	0.1	0.2	0.5	0.1	0.1	0.2	0.1	0.1	0.8	0.1	0.1	0.3
DamDensWs	0.8	1.1	0.5	0.5	0.8	1.1	0.5	0.8	0.7	1.3	0.5	0.4
elevcm_avg	1.9	1.7	6.4	9.6	4.2	4.3	4.9	4.7	17	3.8	16.9	3.9
FWECOREG	5.4	0	5.3	0.2	0.1	0	0.2	0.2	1.3	0	1.9	0
imp11_per	1.8	4.2	3	1.3	2.8	1.7	1.9	2.6	1.5	7.4	2	3.2
imp11_perN	2.1	4.4	5.8	6.2	8.6	2.8	11	8.7	11.7	3.6	5.6	6.1
IncrTempMM	23.5	5	12.5	7.1	5.3	6.5	5.5	5.8	2	6.8	4	3.5
kfact_avg	3	1.3	4	1.8	4.5	2.7	4.5	4.1	1.2	2.2	4.2	2.3
kfactup_av	5.8	3.7	5.8	3	3.5	2.7	3.4	3.5	3.5	1.3	4.3	3.8
MineDensWs	0.1	0.1	0.2	0.2	0.5	0.2	0.3	0.6	1.3	0.9	0.3	0.5
NCat	4	2.1	2.8	4.7	3.2	2.7	2.1	2.9	1.9	1.4	2.9	3.4
nid_storN	1.2	5.8	1.7	9.3	12	1.9	18.2	13.6	6.2	1.9	3	12.6
nlcd11_ag	0.9	3	1.3	1.9	1.8	1.6	0.8	1.1	1.6	3.3	1.4	1.2
nlcd11_agN	1.5	5.4	2.1	1.9	2.2	2.2	1.2	1.9	2.3	3.6	2.9	2.1
nlcd11_t	0.8	4.6	1.7	1.3	1.7	2.8	1	1.6	1.3	3.7	1.6	2.3
NPDESDensW	0.1	0.6	0.4	0.1	0.5	0.8	0.3	0.4	0	0.5	0.2	0.3
P2O5Cat	2.1	2	1.3	1.9	2.2	2.3	1.6	2	3.3	2.1	1.6	1.8
P2O5Ws	1.3	1.9	1.7	1.6	2	4.4	1.2	1.6	1.8	2.6	1.5	2.2
Pct_Ag	0.8	1.5	1.6	1.5	3.1	1.1	3.3	2.9	1.5	1.1	2.3	1.3
Pct_Imp	1.9	5.7	2.9	2.4	1.8	1.5	1.9	2.1	2	2.5	4.3	3
Pct_t	3.4	1.6	2.9	2.1	2.2	3	2	2.4	1.9	1.8	3.9	3
rcharg_avg	3.4	9.8	4	2.2	3.5	3.3	2.9	3.7	1.9	4.7	2.9	4.6
RdCrsCat	1.3	1.5	1.4	1.2	1.6	1	0.7	1.2	1.5	1.1	1	1.1
RdCrsWs	1.7	2	2.1	1.8	2	1.5	1.1	1.5	2.2	2	1.6	1.2
rf30_avg	6.8	6	5.9	3.4	5	8.7	5.5	5.3	4.6	4.7	4	3.5
sand_avg	3.6	3	2.9	5.2	4.5	4.7	4.2	4.4	6.2	8	2.1	4.3
silt_avg	6.4	5.3	2.6	5.1	3.3	10.7	2.2	2.9	4.8	2.7	2.2	2.3
SIMP_NUM	1.6	4.9	2	1.8	2.3	1.9	2.9	1.9	1	2.6	1.1	1.2
slope_avg	2.4	1.1	2.3	3.3	2.4	4.1	2.4	2.6	2	1.6	2.9	2.1
slope_avgN	4.4	2.2	4.4	3.5	3.2	4.4	3.1	3.4	2.9	3.5	3.3	4.8
SLP_CL	0.9	1.4	1.5	1.7	0.9	1.8	0.4	0.9	2.8	3.6	3.2	0.7
SZ_CL	1.8	1.1	1.4	5.2	4.2	0.8	5.2	4.5	2.4	2.2	1.5	10.7
TEMP_CL	0.4	0	0.3	0.1	0.1	0	0	0	0.5	0.2	5	1
TNC6HYDP	0	0.3	0.1	0.1	0.2	0.4	0.1	0.2	0	0.1	0.1	0.7
TRIDensWs	0.1	0.6	0.9	0.3	0.7	1.4	0.5	0.7	0.2	1	0.6	1

Table 6. Cont.

Variable	% Coarse Sediment Taxa				% Tolerant Taxa				% Intolerant Taxa			
	A	C	O	T	A	C	O	T	A	C	O	T
ALK_CL	0.5	0.3	0.6	0.7	1.1	0.1	1	0.2	2.1	0.2	1.2	0.3
bfi_avg	1.5	3.5	2.4	10.8	1.7	2.3	3.3	7.2	1.4	4	3.7	3.9
CONF_CL	0.2	0.1	0.1	0.2	0.2	0.2	1.8	0.3	1.1	0.4	0.1	0.2
DamDensWs	2.6	2.1	0.3	0.6	0.1	0.8	0.5	0.3	0.3	0.5	0.2	0.7
elevcm_avg	8.6	4.7	17.5	4.6	2.8	3.4	3.7	6.5	6.8	1.6	8.2	3.7
FWECOREG	0.1	0	2.7	0	4.4	0	0.7	9.9	3.6	0	0.5	0
imp11_per	4.1	6.5	1.7	1.7	2.1	7	3.2	1.9	3.2	1.4	3.2	5
imp11_perN	4.4	2.5	6.5	8.3	5.1	2.6	7.7	9.1	7.9	2.9	4.1	4.1
IncrTempMM	4.5	6.3	6	7.8	16.6	1.7	8.1	6.1	4.7	2.1	4.7	5.4
kfact_avg	4.9	2.6	4.6	1.4	2.7	4.5	2.7	2.3	3.9	1.9	2.3	1.4
kfactup_av	4.3	5	3.4	1.6	7.8	0.9	2.6	2	2.7	3.2	2.9	4.3
MineDensWs	0.3	0.6	0.6	0.2	0	0.7	0.4	0.3	0	0	0.7	0
NCat	1.9	1.9	1.5	2.3	4.8	3.1	6.3	1.9	4.2	7.3	3.8	4.6
nid_storN	18.7	1.3	5.5	7.6	1	0.2	2	2.8	2.8	0.7	1.6	4.2
nlcd11_ag	1	2.7	1.3	1.1	0.5	1.8	1.3	1.9	1.2	0.9	1.2	2.1
nlcd11_agN	1.2	3.6	4.2	0.9	0.9	5.8	1.6	2.2	1.5	7.2	1.1	4.6
nlcd11_t	0.6	2.2	0.8	2.5	0.5	4.2	1.3	2	0.9	3.8	0.9	1.5
NPDES DensW	1.2	1.4	0.1	0	0	2.8	0.2	0.1	0.2	0.9	1.8	0
P2O5Cat	2.3	0.8	0.9	1	1.3	3.2	2.1	2.7	1.9	1.1	1.1	1.6
P2O5Ws	2.3	2	1.3	1.4	2.2	5	2.8	2.7	1.5	4.1	1.7	1.9
Pct_Ag	1.3	1.9	1.1	1.9	1.9	2.8	0.8	1	4.2	3.2	1.3	2.7
Pct_Imp	1.4	2.1	0.9	1.7	1.1	2.8	1.6	1.9	3.5	1.9	1.9	1.7
Pct_t	3	1.4	4.3	2.9	5.3	4.8	4	1.4	8.1	4.8	2.7	3.3
rchrng_avg	2.2	9.1	7.4	5.7	3.9	6.7	4.8	4.2	4.6	2.5	11.1	7.6
RdCrsCat	2.8	1.2	0.4	0.5	1.5	3.1	1.5	1	1.2	0.5	0.7	1.4
RdCrsWs	1.5	1.9	1.4	1.1	0.4	3.3	5.5	2.1	3.2	1.9	1.1	1.6
rf30_avg	1.8	3.8	3.5	6.6	6.8	8.9	3.5	4.4	2.7	8.3	7.7	10.2
sand_avg	3.7	8.1	4.6	7.1	3.6	3.9	4.7	2.8	2.4	18.6	4.4	1
silt_avg	5.8	9.2	1.3	1.3	3	2	2.4	1.6	5.4	2.8	4	2.7
SIMP_NUM	0.8	1.5	0.8	3.1	0.7	3	3.5	1	1.3	1.9	0.9	0.9
slope_avg	4.3	3.2	2.1	0.9	6.9	0.8	2	3	2.5	2.4	1.8	5.2
slope_avgN	2.6	3.3	3.7	2.7	5.9	1.1	3.4	9.4	3.8	5	7.9	6.9
SLP_CL	1.4	0.9	2.6	0.8	0.3	0.3	0.7	1.3	1.3	0.5	1.1	1.3
SZ_CL	1.7	0.7	2.1	8.8	2.6	2.4	0.8	2.4	3.7	0.2	6.3	2.7
TEMP_CL	0	0.1	0.9	0	0.1	0.4	0.3	0.1	0.2	0	1.1	0.2
TNC6HYDP	0	0.3	0.1	0.1	0.4	1.9	6.2	0.3	0	0.1	0.5	0.3
TRIDensWs	0.4	0.9	1.1	0	0	1.4	1.1	0.1	0.1	1.5	0.3	0.7

Additionally, the relative influence varied widely within a given region for the same predictor variable, but for different response variables. For example, k-factor (a measure of soil erodibility) had the 24th most influence on diversity in the Atlantic Slope, but it was the fourth most influential variable in the % herbivore taxa model. The standard deviation that was associated with the relative influence of predictor variables for a specific region was also variable; where, for example, for invertivore diversity the standard deviation was 2.4, 2.6, 5.1, and 4.3 for the Atlantic Slope, the Cumberland River Basin, the Ohio River Basin, and the Tennessee River Basin, respectively.

The mean relative influence for themes were also somewhat variable across regions; where, for example, flow was the most influential theme for each region for the BRT model for the % of species that prefer coarse sediment. However, for the % of tolerant fish species, the flow theme represented the fourth most influential theme for the Atlantic Slope ( $\bar{x} = 1.7\%$ ), third most for the Cumberland River Basin ( $\bar{x} = 2.5\%$ ), second most for the Ohio River Basin ( $\bar{x} = 2.7\%$ ), and first most for the Tennessee River Basin ( $\bar{x} = 3.6\%$ ) (Table 7).

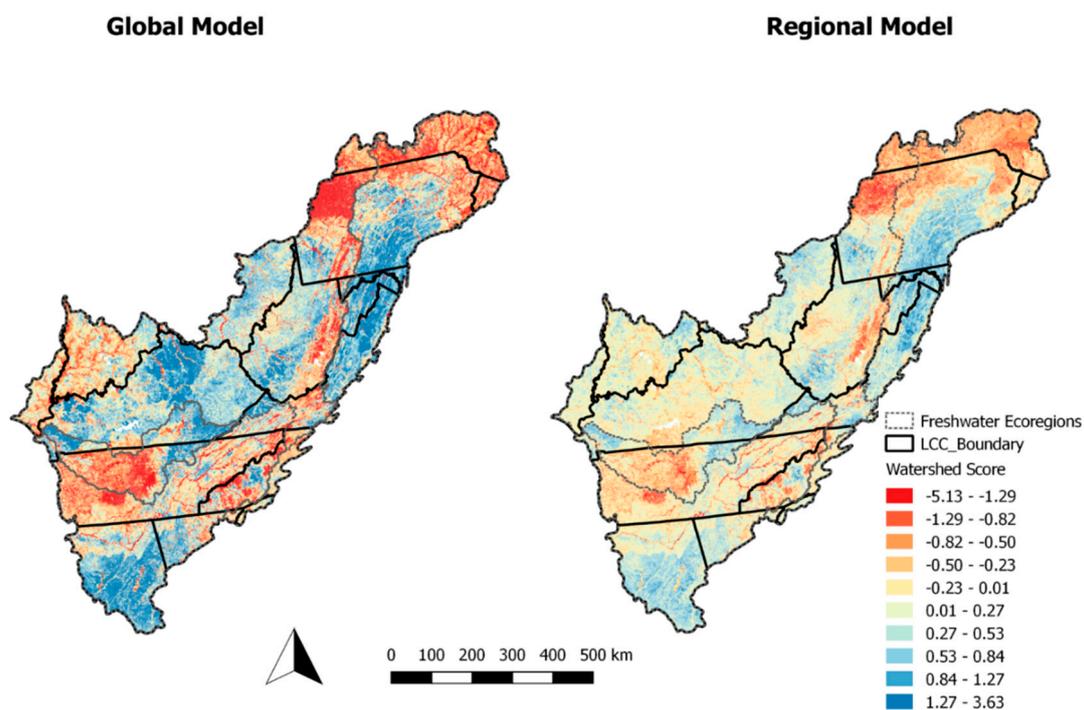
**Table 7.** Summary of the regional average relative influence for each theme on response variables modeled with the boosted regression trees for fish. The average thematic % relative influence is the percent relative contribution of each theme to the BRT model. Themes are comprised of flow condition (NID storage, dam density, altered streamflow, agricultural and industrial water withdrawals), geomorphic condition (erosive and resistive forces), connectivity (dam and road crossing density at the watershed and catchment levels), water quality (total nitrogen load, total phosphorous load, and dissolved organic carbon), non-point source pollution (% agricultural land cover, % impervious surface, and % natural land cover in the watershed, active river area, and catchment), and point source pollution (CERCLA site density, permit compliance system site density, toxic release site density, coal mine, wind turbine and natural gas well density, and mine density).

Theme	Richness				Diversity				% Invertivore Taxa				% Piscivore Taxa				% Herbivore Taxa				
	A	C	O	T	A	C	O	T	A	C	O	T	A	C	O	T	A	C	O	T	
Flow	2.8	7.4	3.2	3.8	3.7	2.0	3.0	8.6	4.7	2.7	2.6	8.6	1.9	5.4	2.2	4.5	4.7	2.1	5.9	5.2	
Geomorphic	2.9	3.0	3.6	3.9	2.9	10.4	8.0	1.9	4.4	1.8	3.4	1.9	5.2	3.6	5.2	2.7	4.4	4.7	4.5	4.3	
Connectivity	0.4	0.5	0.5	0.6	1.0	0.5	0.8	0.1	0.8	0.3	0.8	0.1	0.8	1.1	0.5	0.5	0.8	1.1	0.5	0.8	
Water Quality	1.7	1.7	0.9	1.6	1.3	1.7	1.5	1.1	1.6	3.1	4.7	1.1	1.6	1.2	1.4	1.7	1.6	3.2	1.1	1.4	
Non-point Pollution	1.8	3.0	1.4	2.2	1.9	1.8	1.9	2.1	2.7	3.9	0.7	2.1	1.5	3.4	2.4	2.1	2.7	1.9	2.6	2.6	
Point Pollution	0.4	0.3	0.4	0.3	0.4	0.2	0.8	0.7	0.6	0.1	0.1	0.7	0.1	0.4	0.5	0.2	0.6	0.8	0.4	0.6	
Theme	% Lithophilic Taxa				% Coarse Sediment Taxa				% Tolerant Taxa				% Intolerant Taxa				Averages				
	A	C	O	T	A	C	O	T	A	C	O	T	A	C	O	T	All	A	C	O	T
Flow	2.6	2.5	2.1	5.0	6.2	4.0	3.9	6.2	1.7	2.5	2.7	3.6	2.3	1.9	4.2	4.1	3.9	3.4	3.4	3.3	5.5
Geomorphic	3.1	2.7	4.2	3.2	3.7	3.8	3.8	3.2	5.8	4.8	2.9	2.9	3.1	4.4	4.3	5.3	4.0	3.9	4.4	4.4	3.3
Connectivity	0.7	1.3	0.5	0.4	2.6	2.1	0.3	0.6	0.1	0.8	0.5	0.3	0.3	0.5	0.2	0.7	0.7	0.8	0.9	0.5	0.4
Water Quality	1.5	2.8	1.3	1.7	1.4	1.0	0.9	1.1	1.9	2.3	2.4	1.5	2.0	2.6	1.6	1.7	1.7	1.6	2.2	1.8	1.4
Non-point Pollution	2.6	3.0	2.7	2.5	1.9	2.6	2.3	2.3	1.9	3.5	2.4	2.4	3.4	2.9	1.8	2.8	2.4	2.3	2.9	2.0	2.3
Point Pollution	0.5	0.8	0.4	0.6	0.6	1.0	0.6	0.1	0.0	1.6	0.5	0.2	0.1	0.8	0.9	0.2	0.5	0.4	0.7	0.5	0.4

Generally, geomorphic condition was the most influential theme followed by flow, non-point source pollution, water quality, connectivity, and point source pollution when averaged across all of the response variables. With little variation, the themes were consistent in terms of their relative influence across regions; for example, non-point source pollution was the fourth most influential theme for the Atlantic Slope, the Ohio River Basin, and the Tennessee River Basin, but the fifth most influential for the Cumberland River Basin.

### 3.3. Fish Models

The model results suggest there are distinct areas of the App LCC region that exhibit characteristics of poor biological condition (i.e., watershed score) (Figure 3).



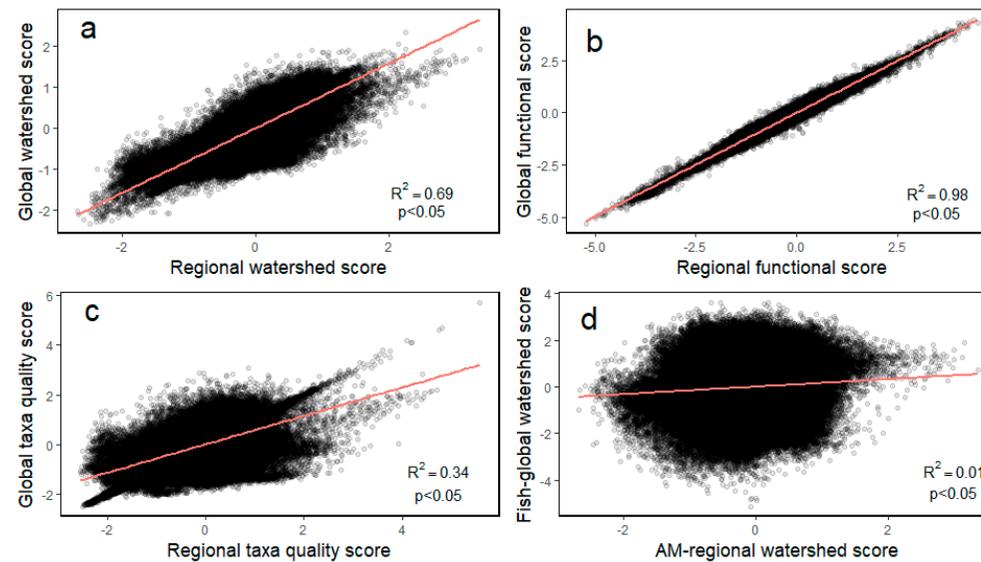
**Figure 3.** Comparison of the watershed score for the global and regional models. The watershed scores were centered and scaled. The regional watershed scores were centered and scaled within each respective region within the Appalachian Landscape Conservation Cooperative (LCC) region. The global watershed scores were centered and scaled across the entire Appalachian LCC region. Map created by R. Daniel Hanks using ArcGIS v10.3 [22].

For example, western Tennessee (i.e., the Southeastern Plains) and the northern-most portion of the region (i.e., the Atlantic Highlands). Additionally, large rivers tended to exhibit poor biological condition. This was especially true in areas with high agricultural land use (i.e., east Tennessee). Regions where the models displayed relatively good biological condition were largely in areas free from agricultural land use. However, large rivers exhibited poor biological condition, even in areas with good biological condition.

Pearson correlations between the global and regional models were variable and ranged from 0.33 (% intolerant taxa) to 0.99 (% invertivore taxa) for pairwise comparisons. The overall indices for watershed scores (i.e., biological condition) were moderately correlated (0.69). The targets were highly correlated for the functional group (0.98) but, were relatively poorly correlated for taxa quality (0.34) (Table 8) (Figure 4).

**Table 8.** Pearson correlation coefficients for the global and regional models for all response variables. Correlations for the same response variables for each model are highlighted in gray.

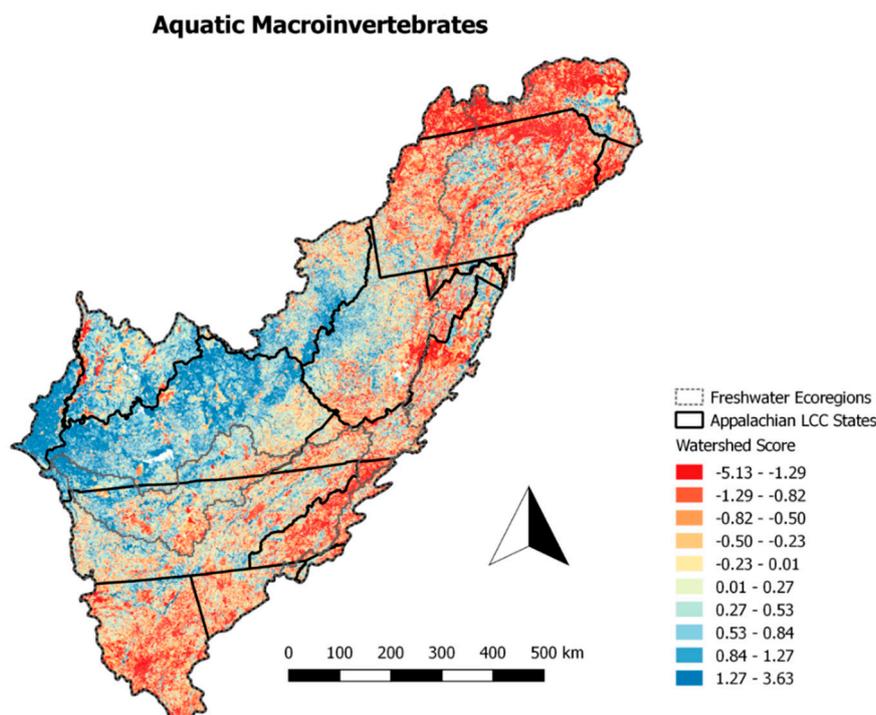
	Global Model										
	Diversity	% Invertivore Taxa	% Piscivore Taxa	% Herbivore Taxa	% Lithophilic Taxa	% Coarse Sediment Taxa	% Tolerant Taxa	% Intolerant Taxa	Functional Group	Taxa Quality	Overall Watershed Score
<b>Regional Model</b>											
Diversity	0.99	0.08	0.05	0.05	0.10	0.12	0.24	0.08	0.04	0.14	0.37
% Invertivore Taxa	0.08	1.00	0.01	0.13	0.10	0.06	0.07	0.25	0.59	0.06	0.18
% Piscivore Taxa	0.06	0.00	0.99	0.21	0.06	0.02	0.48	0.10	0.62	0.21	0.52
% Herbivore Taxa	0.04	0.14	0.22	0.98	0.35	0.26	0.25	0.11	0.69	0.41	0.56
% Lithophilic Taxa	0.18	0.18	0.15	0.43	0.68	0.30	0.05	0.00	0.24	0.40	0.27
% Coarse Sediment Taxa	0.19	0.07	0.05	0.53	0.32	0.44	0.16	0.06	0.34	0.42	0.45
% Tolerant Taxa	0.25	0.07	0.47	0.25	0.05	0.02	1.00	0.20	0.41	0.48	0.62
% Intolerant Taxa	0.19	0.18	0.07	0.03	0.06	0.09	0.08	0.33	0.04	0.11	0.11
Functional Group	0.04	0.59	0.63	0.68	0.21	0.09	0.42	0.02	0.99	0.30	0.65
Taxa Quality	0.19	0.06	0.18	0.51	0.37	0.26	0.50	0.24	0.39	0.58	0.61
Overall Watershed Score	0.39	0.25	0.49	0.60	0.26	0.23	0.62	0.23	0.69	0.57	0.86



**Figure 4.** Plots show Pearson correlation between (a) global and regional watershed scores, (b) global and regional functional scores, (c) global and regional taxa quality scores, and (d) global watershed scores for fish and regional watershed score for aquatic macroinvertebrates (AM). Plots generated using Program R [30].

### 3.4. Aquatic Macroinvertebrate Model

We were unable to develop regional models for aquatic macroinvertebrates due to the data set being sparse in some regions. The global model for aquatic macroinvertebrates showed poor ability to identify areas with a good and bad biological condition. Only very large geographic regions were distinguishable from one another. For example, the western portion of the App LCC region was identified as having relatively higher biological condition when compared to the eastern and northern portions of the region. Where the fish models identified large rivers as having poor biological condition, the aquatic macroinvertebrate models were unable to identify the biological condition at this scale (Figure 5).



**Figure 5.** Global watershed scores from the boosted regression tree models for aquatic macroinvertebrates. Watershed scores were centered and scaled. Map created by R. Daniel Hanks using ArcGIS v10.3 [22].

Additionally, the Pearson correlation coefficient between the aquatic macroinvertebrate watershed score and those of the global (Figure 4) and regional fish watershed scores was 0.01 for both.

## 4. Discussion

Freshwater ecosystems are highly threatened worldwide. In North America, the Appalachian LCC region is a particularly valuable aquatic ecosystem due, in large part, to its high aquatic biodiversity. It is also a region of great conservation concern due to current and projected rapid land cover change (i.e., urbanization), biodiversity loss, habitat fragmentation, and climate change [17,20]. Models, such as the ones presented here, and others (e.g., [3,5]), which provide insight into how land use affects aquatic resources (e.g., fish and aquatic macroinvertebrates), are valuable tools for identifying areas of greatest concern for restoration activities and those areas in greatest need of increased and/or continued protection [32,33].

Multimetric indices may be beneficial in identifying a biological condition across large geographies. Our results suggest that multimetric scores (i.e., watershed scores) corroborate one another at the global and regional scales (see Table 8), which indicate that the predictor variables we used were

responsive in a consistent manner across scales. This is an important finding, as the ability to rely on a consistent set of landscape variables enables cross-regional and cross-scale comparisons, which can be particularly important when data sets are assembled from multiple sources [3].

The individual predictors that most commonly had high relative influence for fish metrics (i.e., was ranked in the top ten percent of predictors within an individual metric) were primarily related to geomorphic condition (e.g., erosive and resistive forces) and flow regime (e.g., storage and recharge capacity). However, those with the highest relative influence for aquatic macroinvertebrates were primarily related to sources of pollution (e.g., road crossings, nitrogen, and silt). Our results reinforce those of Hill et al. [5], who found that agriculture and urbanization were the primary drivers of landscape impacts on aquatic macroinvertebrates. The differences in driving factors for fish and aquatic macroinvertebrates suggest that, at a large spatial scale, such as that presented here, the important land use factors that influence aquatic communities may differ somewhat and mitigation efforts to overcome these large-scale stressors likely need to have some taxonomic specificity.

The relative influence of individual predictors to our boosted regression tree models was variable between regions in our regional models. This indicates the high variability between regions regarding what environmental factors influence community structure. These findings are not surprising, as they are congruent with other large-scale biological condition assessments [3,5] and have been found to be a shortcoming of MMIs [34,35]. While individual predictors were relatively variable between regions, our models suggested themes (i.e., combined predictors) were more consistent; yet, some discrepancies between regions still existed. We chose to condense predictor variables into themes as a higher order mode of understanding of how these groups of stressors impacted fish and aquatic macroinvertebrates. The themes that exhibited the greatest influence on fish watershed scores across the entire region were geomorphic condition and flow regime. The geomorphic condition in our analysis was a measure of soil stability and erosivity. This finding is unsurprising, as aquatic taxa are known to be sensitive to sedimentation. Surprisingly, point source pollution and connectivity were the least influential. This is likely due to these two factors being measured somewhat sparsely (i.e., few watersheds have NPDES sites) on the landscape and not necessarily in close proximity to biological samples. For example, NPDES density, which is a measure of point source pollution, is likely to exhibit strong influence on aquatic biota in the immediate area, but these sites are scattered across the landscape and, thus, across a large region, such as the App LCC region, they were less influential in our models than predictors that have greater spatial representation (e.g., soil erosivity).

A high correlation between the global and regional fish multimetric indices suggests that the predictor variables used were responsive across both scales. While there was high correlation between global and regional fish multimetric indices (i.e., watershed score), there was very low correlation between multimetric indices from either of the fish models and the global aquatic macroinvertebrate model. The low correlation between the fish watershed scores and aquatic macroinvertebrate watershed scores is most likely due to the low number of sites ( $N = 334$ ), which results in poor spatial coverage for our aquatic macroinvertebrate data set, as compared to the good spatial distribution and relatively high number of sites ( $N = 8355$ ) in our fish data set. Correlations between the fish and aquatic macroinvertebrate indices are likely to be higher when the same, or nearby, sites are sampled for both. For example, at 18 sites in the River Raisin watershed in Michigan, correlations between fish and aquatic macroinvertebrate indices ranged from 0.21 to 0.51 [36]. Additionally, stressors have been found to affect fish and aquatic macroinvertebrates similarly. Free flowing stream sections of the Fox River, IL exhibited higher index scores for both fish and aquatic macroinvertebrates when compared to sections of the river above low-head dams [37].

Esselman et al. [3] suggested regional stratification for analyses that span large geographies, while also including multiple biological response variables, and using statistical modeling techniques that are non-linear would lead to improved multimetric indices. This is precisely the approach that we took modeling biological condition across the 15 state App LCC region. However, one major difference between our study and previous studies is that previous work to develop multimetric indices across

broad geographies included regional influences by using Omernick ecoregions [3,5,15] (Whittier et al., 2007; Esselman et al. 2013; and, Hill et al. 2017). Omernick ecoregions are a spatial framework that includes aquatic and terrestrial ecosystems [38]; however, we used Freshwater Ecoregions of the World for our analysis, because they were developed on a framework that focused exclusively on aquatic ecosystems with fish as the focal taxonomic group [31].

Another difference between our work and previous work is that we aggregated both predictor and response variables into themes and targets respectively. By aggregating predictor variables into themes, we were able to identify categories of predictors that were influential in structuring fish and aquatic macroinvertebrate communities in aggregate (e.g., watershed score for fish), as well as for individual response variables (e.g., % EPT). Response variable aggregation allows for a greater understanding of how individual and aggregate predictors (i.e., themes) structure fish and/or aquatic macroinvertebrate communities in a more functional manner. For example, two targets (e.g., fish functional group and fish taxa quality) may respond differently to the same stressor. Understanding which higher order stressors (theme) affect our targets (fish functional group) might lend insight into more effective regional management and conservation strategies. Our results suggest that regional variability of the relative influence of predictor variables should garner significant consideration as the stressors that affect fish and aquatic macroinvertebrates were not the same from ecoregion to ecoregion. Furthermore, care should be taken when using large scale modeling efforts such as ours to make management recommendations across large geographies that span more than one ecoregion, as thematic importance was variable between regions.

While we were able to develop regional models for fish indices, unfortunately, we were unable to develop regional aquatic macroinvertebrate models due to low numbers of sample sites. As aquatic macroinvertebrates are known to be good indicators of biological condition [39–41], we believe that there needs to be increased systematic sampling efforts that result in an increased number of sites that are spatially dispersed across multiple freshwater ecoregions and that include known areas of both high and low biological integrity. In doing so, models such as the one provided here (see also Hill et al. 2017 and Esselman et al. 2013 [3,5] can be improved in their ability to provide managers with information to help guide local plans within a regional context. For example, mitigation efforts that aim to provide ecological lift to a watershed might look for areas that have been identified as having poor condition but surrounded by watersheds with good condition or, alternatively, these models could be used to identify good quality watersheds that are in need of protection from multiple encroaching stressors.

## 5. Conclusions

We combined and used multiple biotic and abiotic datasets to model stream condition across the Appalachian LCCs 15 state boundary (i.e., central and southern Appalachians). We developed separate fish and aquatic macroinvertebrate stream condition models highlighting the existence of both similarities and differences in how environmental factors structure these two aquatic taxonomic communities while using BRTs. Overall, correlation existed between fish and aquatic macroinvertebrate models and global and regional fish models. Furthermore, there was high correlation between global and regional fish models for most of the fish metrics evaluated. We were unable to create aquatic macroinvertebrates regional models due to low sample size regionally; however, for fish, separate models for each of the Freshwater Ecoregions varied in what environmental factors were most important in structuring aquatic communities, which exemplified the need to carefully consider ecoregional factors when developing cross-ecoregion spatial plans and management suggestions. Biological samples are difficult and expensive to collect, and effort varies between taxa, and this results in uneven distributions in both the number of samples and spatial distribution between taxa making it difficult to create MMIs that can better represent the condition of aquatic resources across broad geographies. Careful selection of biological sample locations that are representative of the spectrum spanning from high quality to low quality environmental conditions that are spatially well dispersed

can greatly enhance the ability of models such as ours to inform spatial and management plans. Additionally, common sample locations for various taxa would be useful to improve the comparison of environmental responses among and between taxa.

We recommend a centralized database where these records are made available for public download, where efforts such as ours can be more easily replicated, improved upon, and then performed at various spatial scales. With an extensive and systematic sampling approach that includes both fish and aquatic macroinvertebrates and it is integrated across jurisdictional boundaries the ability to develop accurate models of large geographies will improve. Such improved spatial models will lend greater confidence in such indices, which will only increase the ability of managers, conservation practitioners, and scientists alike to address and meet regional conservation goals.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “conceptualization, R.D.H., P.B.L. and R.F.B.; methodology, R.D.H. and P.B.L.; formal analysis, R.D.H.; writing—original draft preparation, R.D.H.; writing—review and editing, R.D.H., P.B.L., and R.F.B.; visualization, R.D.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Appalachia Landscape Conservation Cooperative through the Wildlife Management Institute.

**Acknowledgments:** We would like to thank the U.S. Fish and Wildlife Service for funding this research along with the Appalachian Landscape Conservation Cooperative staffs, partners, and collaborators that provided great insight and guidance throughout this project. We also acknowledge the Department of Forestry and Environmental Conservation at Clemson University. We would like to acknowledge the US Environmental Protection Agency’s work and for providing access to the StreamCat data set. Technical Contribution No. 6820 of the Clemson University Experiment Station.

**Conflicts of Interest:** The authors declare that they have no competing interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A

**Table A1.** Full list of predictors, showing each variable’s thematic grouping and the source for each data set.

Attribute	Metric	Variable Name	Source
General	Catchment area	areasqkm	TNC
	Upstream Drainage Area (Net Catch Area)	TotDASqKm	TNC
Flow Regime	Base Flow	bfi_avg	TNC
	Base Flow	bfi_avg_N	TNC
	Recharge	rchrng_avg	TNC
	Recharge	rchrng_avgN	TNC
	Total Storage/mean annual flow	nid_storN	TNC
	Density of Large Dams	DamDensCat	StreamC
	Density of Large Dams	DamDensWs	StreamC
	Municipal Water Withdrawal	dom_wdc	NFHAP
	Ag Water Withdrawal	ag_wdc	NFHAP
	Industrial Water Withdrawal	ind_wdc	NFHAP
	Ag Water Withdrawal Normalized w/Summer flow	ag_wdc_Snorm	
	Ag Water Withdrawal Normalized w/Annual Flow	ag_wdc_norm	
	Municipal Water Withdrawal Normalized w/ Annual Flow	dom_wdc_norm	
	Industrial Water Withdrawal Normalized w/Annual Flow	ind_wdc_norm	
Connectivity	Density of Dams Upstream	upnetdensr	NFHAP
	Density of Dams Downstream	downmaindens	NFHAP
	Density of Road Crossings (Tiger 2010)	RdCrsCat	StreamC
	Density of Road Crossings (Tiger 2010)	RdCrsWs	StreamC
Water Q (Pop)	Athro Nitrogen	n_yield	NFHAP
	Anthro Phosphorus	p_yield	NFHAP
	Anthro Sediment Yield	s_yield	NFHAP

Table A1. Cont.

Attribute	Metric	Variable Name	Source
Water Q (LU)	Impervious (NLCD 2011)	imp11_per	TNC
	Impervious (NLCD 2011)	imp11_perN	TNC
	Riparian Buffer (NLCD 2011) 100m	PctDecid2011CatRp100	StreamC
	Riparian Buffer (NLCD 2011) 100m	PctDecid2011WsRp100	StreamC
	Riparian Buffer (NLCD 2011) 100m	PctConif2011CatRp100	StreamC
	Riparian Buffer (NLCD 2011) 100m	PctConif2011WsRp100	StreamC
	Riparian Buffer (NLCD 2011) 100m	PctMxFst2011CatRp100	StreamC
	Riparian Buffer (NLCD 2011) 100m	PctMxFst2011WsRp100	StreamC
	% Crop (NLCD 2011)	nlcd11_81N nlcd11_82N	TNC
% Crop (NLCD 2011)	nlcd11_81 nlcd11_83	TNC	
Water Q (Pollution)	CERC	cerc_denC	NFHAP
	Permit Compliance	pcs_denC	NFHAP
	Toxic Release Sites	tri_denC	NFHAP
	Coal Mine Density	all_col_denC	NFHAP
Other	Temperature	IncrTempMM07	TNC
	Temperature	IncrTempMM08	TNC
	Elevation	elevcm_avg	TNC
	Elevation	elevcm_avgN	TNC
	Slope	slope_avg	TNC
	Slope	slope_avgN	TNC
	R-factor	rf30_avg	TNC
	R-factor	rf30_avgN	TNC
	K-factor	kfact_avg	TNC
	K-factor	kfact_avgN	TNC
	Sand	sand_avg	TNC
	Sand	sand_avgN	TNC
Silt	silt_avg	TNC	
Silt	silt_avgN	TNC	
Landscape Condition	Percent Agricultural Land within ARA (HUC 8)	Perc_Ag_HUC8	
	Percent Natural Land within ARA (HUC 8)	Perc_Nat_HUC8	
	Percent Impervious Surface within ARA (HUC 8)	Perc_Imp_HUC8	
	Percent Agricultural Land within ARA	Perc_Ag	
	Percent Natural Land within ARA	Perc_Nat	
	Percent Impervious Surface within ARA	Perc_Imp	

## References

- Vörösmarty, C.J.; McIntyre, P.B.; Gessner, M.O.; Dudgeon, D.; Prusevich, A.; Green, P.; Glidden, S.; Bunn, S.E.; Sullivan, C.A.; Liermann, C.R.; et al. Global threats to human water security and river biodiversity. *Nature* **2010**, *467*, 555–561. [[CrossRef](#)]
- Dudgeon, D.; Arthington, A.H.; Gessner, M.O.; Kawabata, Z.I.; Knowler, D.J.; Lévêque, C.; Naiman, R.J.; Prieur-Richard, A.H.; Soto, D.; Stiassny, M.L.J.; et al. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biol. Rev. Camb. Philos. Soc.* **2006**, *81*, 163–182. [[CrossRef](#)]
- Esselman, P.C.; Infante, D.M.; Wang, L.; Cooper, A.R.; Wieferich, D.; Tsang, Y.-P.; Thornbrugh, D.J.; Taylor, W.W. Regional fish community indicators of landscape disturbance to catchments of the conterminous United States. *Ecol. Indic.* **2013**, *26*, 163–173. [[CrossRef](#)]
- USEPA. *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*; USEPA: Washington, DC, USA, 2006.
- Hill, R.A.; Fox, E.W.; Leibowitz, S.G.; Olsen, A.R.; Thornbrugh, D.J.; Weber, M.H. Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. *Ecol. Appl.* **2017**, *27*, 2397–2415. [[CrossRef](#)]
- Harris, J.H.; Silveira, R. Large-scale assessments of river health using an Index of Biotic Integrity with low-diversity fish communities. *Freshw. Biol.* **1999**, *41*, 235–252. [[CrossRef](#)]

7. Hering, D.; Moog, O.; Sandin, L.; Verdonschot, P.F.M. Overview and application of the AQEM assessment system. *Hydrobiologia* **2004**, *516*, 1–20. [[CrossRef](#)]
8. Pont, D.; Hugueny, B.; Beier, U.; Goffaux, D.; Melcher, A.; Noble, R.; Rogers, C.; Roset, N.; Schmutz, S. Assessing river biotic condition at a continental scale: A European approach using functional metrics and fish assemblages. *J. Appl. Ecol.* **2006**, *43*, 70–80. [[CrossRef](#)]
9. Pont, D.; Hughes, R.M.; Whittier, T.R.; Schmutz, S. A Predictive Index of Biotic Integrity Model for Aquatic-Vertebrate Assemblages of Western U.S. Streams. *Trans. Am. Fish. Soc.* **2009**, *138*, 292–305. [[CrossRef](#)]
10. Hanski, I. Metapopulation dynamics. *Nature* **1998**, *396*, 41–49. [[CrossRef](#)]
11. Stoddard, J.L.; Herlihy, A.T.; Peck, D.V.; Hughes, R.M.; Whittier, T.R.; Tarquinio, E. A process for creating multimetric indices for large-scale aquatic surveys. *J. N. Am. Benthol. Soc.* **2008**, *27*, 878–891. [[CrossRef](#)]
12. Angermeier, P.L.; Smorgor, R.A. Estimating number of species and relative abundances in stream-fish communities: Effects of sampling effort and discontinuous spatial distributions. *Can. J. Fish. Aquat. Sci.* **1995**, *52*, 936–949. [[CrossRef](#)]
13. Bramblett, R.G.; Johnson, T.R.; Zale, A.V.; Heggem, D.G. Development and Evaluation of a Fish Assemblage Index of Biotic Integrity for Northwestern Great Plains Streams. *Trans. Am. Fish. Soc.* **2005**, *134*, 624–640. [[CrossRef](#)]
14. McCormick, F.H.; Hughes, R.M.; Kaufmann, P.R.; Peck, D.V.; Stoddard, J.L.; Herlihy, A.T. Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands Region. *Trans. Am. Fish. Soc.* **2001**, *130*, 857–877. [[CrossRef](#)]
15. Whittier, T.R.; Hughes, R.M.; Stoddard, J.L.; Lomnický, G.A.; Peck, D.V.; Herlihy, A.T. A Structured Approach for Developing Indices of Biotic Integrity: Three Examples from Streams and Rivers in the Western USA. *Trans. Am. Fish. Soc.* **2007**, *136*, 718–735. [[CrossRef](#)]
16. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
17. Terando, A.J.; Costanza, J.; Belyea, C.; Dunn, R.R.; McKerrow, A.; Collazo, J.A. The Southern Megalopolis: Using the Past to Predict the Future of Urban Sprawl in the Southeast U.S. *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
18. Carlisle, D.M.; Wolock, D.M.; Meador, M.R. Alteration of streamflow magnitudes and potential ecological consequences: A multiregional assessment. *Fornt. Ecol. Environ.* **2011**, *9*. [[CrossRef](#)]
19. Ford, T.R. *The Southern Appalachian Region: A Survey*; University of Kentucky Press: Lexington, KY, USA, 2015.
20. Leonard, P.B.; Baldwin, R.F.; Hanks, R.D. Landscape-scale conservation design across biotic realms: Sequential integration of aquatic and terrestrial landscapes. *Sci. Rep.* **2017**, *7*, 1–12. [[CrossRef](#)]
21. Baldwin, R.F.; Trombulak, S.C.; Leonard, P.B.; Noss, R.F.; Hilty, J.A.; Possingham, H.P.; Scarlett, L.; Anderson, M.G. The Future of Landscape Conservation. *Bioscience* **2018**, *68*, 60–63. [[CrossRef](#)]
22. ESRI. *ArcGIS Desktop: Release 10*; Environmental Systems Research Institute: Redlands, CA, USA, 2011.
23. Sutherland, A.B.; Meyer, J.L.; Gardiner, E.P. Effects of land cover on sediment regime and fish assemblage structure in four southern Appalachian streams. *Freshw. Biol.* **2002**, *47*. [[CrossRef](#)]
24. Neves, R.J.; Angermeier, P.L. Habitat alteration and its effects on native fishes in the upper Tennessee River system, east-central USA. *J. Fish. Biol.* **1990**, *37*, 45–52. [[CrossRef](#)]
25. Naiman, R.J.; Decamps, H.; Pollock, M. The role of riparian corridors in maintaining regional biodiversity. *Ecol. Appl.* **1993**, *3*, 209–212. [[CrossRef](#)] [[PubMed](#)]
26. Smith, M.P.; Schiff, R.; Olivero, A.; MacBroom, J.G. *The Active River Area: A Conservation Framework for Protecting Rivers and Streams*; The Nature Conservancy: Boston, MA, USA, 2008.
27. Team, Q.D. *QGIS Geographic Information System*; QSGEO: Chicago, IL, USA, 2016.
28. Frimpong, E.A.; Angermeier, P.L. Fish traits: a database of ecological and life-history traits of freshwater fishes of the United States. *Fisheries* **2009**, *34*, 487–495. [[CrossRef](#)]
29. Stoddard, J.L.; Peck, D.V.; Larsen, D.P.; Van Sickle, J.; Hawkins, C.P.; Hughes, R.M.; Whittier, T.R.; Lomnický, G.; Herlihy, A.T.; Kaufmann, P.R. *Environmental Monitoring and Assessment (EMAP): Western Streams and Rivers Statistical Summary*; US Environmental Protection Agency, Office of Research and Development: Washington, DC, USA, 2005.
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: St. Louis, MI, USA, 2018.

31. Abell, R.; Thieme, M.L.; Revenga, C.; Bryer, M.; Kottelat, M.; Bogutskaya, N.; Coad, B.; Mandrak, N.; Balderas, S.C.; Bussing, W.; et al. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *Bioscience* **2008**, *58*, 403–414. [[CrossRef](#)]
32. Carlisle, D.M.; Falcone, J.; Meador, M.R. Predicting the biological condition of streams: Use of geospatial indicators of natural and anthropogenic characteristics of watersheds. *Environ. Monit. Assess.* **2009**, *151*, 143–160. [[CrossRef](#)] [[PubMed](#)]
33. Villeneuve, B.; Souchon, Y.; Usseglio-Polatera, P.; Ferréol, M.; Valette, L. Can we predict biological condition of stream ecosystems? A multi-stressors approach linking three biological indices to physico-chemistry, hydromorphology and land use. *Ecol. Indic.* **2015**, *48*, 88–98. [[CrossRef](#)]
34. Mazor, R.D.; Rehn, A.C.; Ode, P.R.; Engeln, M.; Schiff, K.C.; Stein, E.D.; Gillett, D.J.; Herbst, D.B.; Hawkins, C.P. Bioassessment in complex environments: Designing an index for consistent meaning in different settings. *Freshw. Sci.* **2016**, *35*, 249–271. [[CrossRef](#)]
35. Cao, Y.; Hawkins, C.P.; Olson, J.; Kosterman, M.A. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *J. N. Am. Benthol. Soc.* **2007**, *26*, 566–585. [[CrossRef](#)]
36. Lammert, M.; Allan, J.D. Assessing biotic integrity of streams: Effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates. *Environ. Manag.* **1999**, *23*, 257–270. [[CrossRef](#)]
37. Santucci, V.J.; Gephard, S.R.; Pescitelli, S.M. Effects of Multiple Low-Head Dams on Fish, Macroinvertebrates, Habitat, and Water Quality in the Fox River, Illinois. *N. Am. J. Fish. Manag.* **2005**, *25*, 975–992. [[CrossRef](#)]
38. Omernik, J.M.; Griffith, G.E. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environ. Manag.* **2014**, *54*, 1249–1266. [[CrossRef](#)] [[PubMed](#)]
39. Pond, G.J.; Passmore, M.E.; Borsuk, F.A.; Reynolds, L.; Rose, C.J. Downstream effects of mountaintop coal mining: Comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *Am. Benthol. Soc.* **2008**, *27*, 717–737. [[CrossRef](#)]
40. Pond, G.J.; McMurray, S.E. *A Macroinvertebrate Bioassessment Index for Headwater Streams in the Eastern Coalfield Region, Kentucky*; Kentucky Department for Environmental Protection, Division of Water: Frankfort, KY, USA, 2002.
41. Gerritsen, J.; Burton, J.; Barbour, M.T. *A Stream Condition Index for West Virginia Wadeable Streams*; US EPA: Washington, DC, USA, 2000; Volume 3.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).