MODELING STREAM FISH BIOGEOGRAPHY TO ASSESS WATER QUALITY IN A SPECIES-RICH REGION

by

STEPHEN PAUL MAURANO

(Under the Direction of Mary Freeman)

ABSTRACT

Natural resource managers commonly monitor fish to assess the water quality condition of streams. I have evaluated a biological assessment metric, based on observed and expected fish species richness, for application in Georgia. A multivariate species distribution model was built using a Random Forest machine-learning algorithm to predict expected fish taxa based on a stream's environmental characteristics, such as elevation, slope, and flow. The ratio of observed to expected taxa richness was then used to estimate whether a stream was close to, or far from, a least-disturbed condition. The model was useful in the species-rich northern half of the state above the Fall Line, but inadequate in the less speciose southern half. This research analyzed the biogeography of Georgia's fishes, demonstrated a tradeoff when including rare species in bioassessment, exhibited the sensitivity of fluvial taxa to human disturbance, and provided an additional method for assessing stream conditions in the state.

INDEX WORDS: Stream, Fish, Water Quality, Bioassessment, Biogeography, Biocriteria, Georgia, Machine-Learning, Random Forest, Species Distribution Model

MODELING STREAM FISH BIOGEOGRAPHY TO ASSESS WATER QUALITY IN A SPECIES-RICH REGION

by

STEPHEN PAUL MAURANO

B.S. Biology and B.S. Environmental Science, Santa Clara University, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTERS OF SCIENCE

ATHENS, GEORGIA

2015

MODELING STREAM FISH BIOGEOGRAPHY TO ASSESS WATER QUALITY IN A SPECIES-RICH REGION

by

STEPHEN PAUL MAURANO

Major Professor:	Mary Freeman
Committee:	C. Rhett Jackson
	Seth Wenger

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

ACKNOWLEDGEMENTS

I am immensely grateful for the mentoring of my advisor, Mary Freeman; for the guidance from my committee members, Rhett Jackson, and Seth Wenger; for the opportunity to collaborate with GAWRD colleagues Patti Lanford, Paula Marcinek, and Jane Argentina; for the input of colleagues in the Freeman Lab, the Odum School of Ecology, and the greater UGA community, particularly Megan Hagler, John Drake, and Michael Paul; the support of colleagues at the EPA, especially Annie Godfrey and Jim Giattina; and most importantly, for the encouragement from my wife, Carly Maurano.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Literature Review

Water quality evaluation is an inherently high dimensional problem. We might consider any of a number of physical and chemical habitat variables as well as biological conditions reflecting abiotic habitat characteristics (Jackson & Pringle, 2010). Furthermore, each aspect of water quality features temporal variability. Comprehensive assessments of water quality are therefore costly, and simplified metrics that assess the condition of waterbodies are fundamental to aquatic resource management.

Metrics for quantifying the aquatic community provide the means for managers to define biological criteria goals (biocriteria) and to assess the biological status of waterbodies (bioassessment) (Karr, 1981; Yoder & Rankin, 1998). These metrics can help characterize the severity, or identify the cause, of environmental damage, or evaluate the effectiveness of restoration actions (Barbour, Gerritsen, Snyder, & Stribling, 1999). Ideally, the metrics employed will be simple enough for routine monitoring, and their responses to anthropogenic impacts will be sensitive, integrative, and predictable (Dale & Beyeler, 2001).

In streams, fish communities can provide a measure of ecological status that is responsive to changing water quality, habitat, and biotic interactions (Karr, 1991). Fish can temporally integrate water quality conditions, reflecting cumulative exposures over their lives to chemical, physical, or radiological stressors. Fish can manifest the synergistic effects of multiple contaminants, at biologically relevant levels, and express sub-lethal effects (reduced growth, depressed reproduction, or increased external anomalies) (Karr, Fausch, Angermeier, Yant, & Schlosser, 1986). The fish community can reflect the effects of interspecific interactions and changes across the food web, since different fish species inhabit a variety of trophic levels (e.g., herbivores, insectivores, planktivores, piscivores, and omnivores) (Karr & Chu, 1997). Moreover, fish can commonly be identified and released at the field site, and therefore their monitoring can potentially be less time-consuming than other assemblages requiring more extensive laboratory sorting and identification (e.g., diatoms or macroinvertebrates). Finally, fish species are often the focus of aquatic conservation, and changes in the fish community can be interpreted based on the respective species' life histories, which aids in communicating impacts to the general public.

Conversely, the bioassessment approach is observational and lacks experimental control to isolate the effects of covariates along natural gradients (geology, climate, and biogeography), and human disturbance gradients (e.g., land use, and habitat fragmentation) (Hawkins, Norris, Gerritsen, et al., 2000). Biological monitoring can also be influenced by variability over time (seasonality, climate cycles, or stochastic weather events), variability within a site (depending on which microhabitats are accessible and chosen for sampling), variability in detection probability across sites and species, and sampling error (inconsistent application of methodology or identification errors). The physiological and behavioral complexity of fishes can also make it difficult to relate fish community composition to specific anthropogenic stressors. Genetic adaptations can confer increased toxicant resistance in a fish population, (Klerks & Weis, 1987) so that damage at the subcellular, tissue, or organismal level may not be reflected at the population, community and ecosystem level. Likewise, fish can selectively avoid pollutant plumes (Beitinger, 1990; Giattina & Garton, 1983) and move beyond the boundaries of local sampling units (Hitt & Angermeier, 2008, 2011), potentially confounding the presumed relationship between their presence and local environmental quality. Although rare species are of particular conservation concern, they are frequently excluded from bioassessment calculations due to an inherent scarcity of occurrence data (Cao, Larsen, & Thorne, 2001; Cao, Williams, & Williams, 1998; Marchant, 2002). Finally, a structural metric like biodiversity is calculated from a single fish community sample reflecting ecological conditions at a given point-in-time and therefore may respond differently than functional metrics, which collect repeated measures to derive rates such as productivity (Palmer & Febria, 2012).

Cognizant of these strengths and limitations, aquatic assessment programs generally use point-in-time samples of fishes, macroinvertebrates, and diatoms to infer biological conditions. Biological monitoring results are then assessed under environmental regulations in areas such as Australia (Davies, Wright, Sutcliffe, & Furse, 2000), Canada (Borisko, Kilgour, & Stanfield, 2007), the European Union (Hering et al., 2010), South Africa (Dickens & Graham, 2002), the United Kingdom (J. F. Wright, Sutcliffe, & Furse, 2000), and the United States (Barbour et al., 1999). In the U.S., the Federal Water Pollution Control Act (Clean Water Act) includes biotic integrity as a goal in its first line, stating that, “The objective of this Act is to restore and maintain the chemical, physical, and biological integrity of the Nation’s waters.” (U.S.C., 2002). More specifically, the European Union Water Directive Framework specifies elements for classifying the ecological status of different waterbody types, including guidelines for biological monitoring design and frequency, and plans for calibrating and presenting bioassessment results (European Community, 2000).

Pursuant to these regulatory goals, most bioassessment approaches have adopted one of two general methodologies for quantifying biotic integrity (Figure 1). The first is the “multimetric” approach, which combines multiple taxa metrics (e.g., biota abundances, ratios, and functional groups) into a single index to quantifying biotic integrity. The second is the “multivariate” approach, which uses multiple environmental variables to predict expected species richness for a site, and then uses the ratio of observed to expected species richness (O/E) to quantify biotic integrity. Given adequate species data for a site, either methodology can be used to calculate a measure of biotic condition (Roset, Grenouillet, Goffaux, Pont, & Kestemont, 2007). A literature search of Google Scholar (http://scholar.google.com/) was conducted to survey the most highly cited bioassesment publications returned by the search terms “Index of Biotic Integrity” and "RIVPACS.” These terms are most closely associated with the two most common bioassessment approaches: IBI, the Index of Biotic Integrity, and RIVPACS, the River InVertebrate Prediction And Classification System, which are explained in further detail below (R T Clarke, Furse, Wright, & Moss, 1996; Karr, 1981). Relevant publications were considered to be those describing the development and evaluation of bioassessment models, excluding review papers and evaluations of methods that were not specific to a given assemblage or location (avian publications were also excluded). Where there were multiple publications regarding the same bioassessment index, the most highly cited publication was retained. This analysis returned 51 publications based on multimetric approaches compared to 14 based on multivariate approaches (see Appendix A, Table A.1). 80% of the bioassessment indices described in these publications were for streams or rivers (with the remainder for lakes, wetlands, and estuaries), while approximately 60% were for fish, 30% for macroinvertebrates, and 10% for the remaining assemblages (periphyton, phytoplankton, zooplankton, macrophytes, and aquatic amphibians) (Figure 2).

The multimetric approach is widely used in the form of the Index of Biotic Integrity method (Karr, 1981). The IBI approach summarizes the observed species into a single index based on multiple traits such as trophic characteristics, habitat guilds, and phylogeny. These metrics are chosen based on their responsiveness to human disturbances and are intended to measure multiple aspects of an aquatic community's structural integrity (e.g., presence of intolerant species, proportion of omnivores or top carnivores, richness and composition of taxa) (Karr et al., 1986). IBIs have been developed for a variety of assemblages, and waterbody types, and the methodology has also been considered in terrestrial environments (Andreasen, O'Neill, Noss, & Slosser, 2001). The common elements of these IBIs include grouping based on natural assemblage variability, defining a reference condition for these groups, screening metrics based on discrimination efficiency, and selecting the most sensitive and complementary metrics (Roset et al., 2007).

The term IBI is used here in reference to any multimetric bioassessment approach, both those adapted from Karr 1981, as well as alternatives that use different site classification, metric selection, or model calibration approaches to derive a multimetric index. The application of the IBI outside of the ecoregion where it was initially developed, with minimal adaptation, has been critiqued, and it has been suggested that round-robin testing (independent biota sampling and identification for inter-comparison) should be used to determine the variance of sampling methods, and to better quantify the accuracy and precision of a biological index (Seegert, 2000). IBI scores can also be strongly influenced by a sample's location in a drainage network, so IBIs generally need to be developed independently for headwaters, tributaries, and main-channels (Osborne et al., 1992). To account for these natural gradients in species distributions (Gorman & Karr, 1978), an approach has been explored that combines some of the advantages of multimetric and multivariate bioassessment models. These “predictive-IBIs” directly model expected IBI metric scores based on natural variables such as temperature, channel gradient, and geology (D. Pont et al., 2006; Didier Pont, Hughes, Whittier, & Schmutz, 2009).

The alternative approach to IBIs is the Observed-Expected (O/E) method, which reports a richness ratio of species observed to those that were expected based on the environmental characteristics of the sample location (e.g., stream size, elevation, basin, and ecoregion). This approach assumes that the observed taxa richness (O) will depart from the expected taxa richness (E) as the aquatic community becomes increasingly dissimilar from the biota observed at least-disturbed locations (Figure 3). The multivariate O/E method was pioneered by the RIVPACS approach for application to stream benthic macroinvertebrates in the United Kingdom (R T Clarke et al., 1996). The O/E approach is particularly useful when dealing with large numbers of taxa whose life histories or tolerances may not be sufficiently known to select IBI metrics. It is also well-suited for assessments over heterogeneous regions based on its ability to continuously correct for variation in species richness across environmental gradients (Carlisle, Hawkins, Meador, Potapova, & Falcone, 2008). Multivariate predictive models can also benefit from continued advancements in modeling species distributions, which help refine the expected taxa richness used in the denominator of the O/E metric (Buisson, Blanc, & Grenouillet, 2008; Flebbe, Roghair, & Bruggink, 2006; J. R. Leathwick, Rowe, Richardson, Elith, & Hastie, 2005; Olden & Jackson, 2002; D Pont, Hugueny, & Oberdorff, 2005).

The O/E approach has a number of potential deficiencies as well. On a pragmatic level, the empirical modeling approach may require high biodiversity and large data sets to train effective models. From a more theoretical perspective, it’s ambiguous whether species richness is the most sensitive indicator to anthropogenic stress, or whether O/E values greater than one (i.e., those “enriched” with species) indicate that a site is degraded (Meador & Carlisle, 2009). While a comparison of the fishes observed to those expected can be communicated intuitively and enables comparisons between regions, the expression of the metric as a ratio, rather than a magnitude, may be problematic when the modelled biodiversity is low. For example, if managers select the O/E threshold for determining healthy waterbodies as 0.8 or greater, failure to detect just one out of four taxa would result in the conclusion that the waterbody is impaired.

An underlying assumption of both the multimetric and multivariate methods is that streams with depauperate biotic communities reflect extirpations due to anthropogenic stressors in the catchment. This follows from the idea that a least-disturbed reach will support an indigenous biological community with a particular array and abundance of species. If natural variation in community composition has been accounted for (correcting for factors such as stream size, elevation, and biogeographic constraints), then the residual variation is attributable to human disturbance. Not all of the “expected” species are likely to be present in a given stream site at all times, nor is probability of detection likely to be 100%. Each species belonging to the expected community, therefore, has some probability of being present and detected in any given undisturbed reach, at any given time. As capacity of the system to support species declines (for example, because of land use intensification), occurrence probabilities for those species also decline, resulting in lower observed richness. Therefore, human influences on streams are manifested as changes in species abundances, and by definition, these changes reflect a loss of biotic integrity. The goal of bioassessment, therefore is to estimate whether a site is close to, or far from a least-disturbed condition, based on observations of biotic assemblages.

Both the multimetric and multivariate assessment methods define the expected healthy biotic community based on reference conditions, while attempting to account for natural variation among streams. A reference condition is defined as a biotic community with structure and function that would occur in the absence of anthropogenic impact, based on observations of minimally disturbed sites and historical conditions, or in modified areas, based on least-disturbed sites and best attainable conditions (Stoddard & Larsen, 2006). Streams distant from human land use, infrastructure and other impacts are commonly identified as reference streams. These biotic communities, from presumably less anthropogenically influenced locations, are used as benchmarks for measuring the severity of disturbances elsewhere (Hawkins, Olson, & Hill, 2010). In IBI assessments, stream types, and the corresponding reference sites, are generally determined based on predefined geographic groupings such as ecoregions (Omernik, 1987). The O/E approach relies on the selection of reference sites as well, but its multivariate methodology uses statistical methods to group and model “virtual” (Borja et al., 2004) or “predicted” (Hughes, Gore, Brossett, & Olson, 2009) reference communities. These methods predict which species are likely to be found at a location in the absence of human impacts, given species habitat requirements (such as water temperature) and biogeographic constraints (such as basin boundaries). In addition to defining reference sites spatially, (e.g., using regional reference sites) alternative approaches are also possible, including defining a reference condition temporally (e.g., using historical or paleoecological data), or using other lines of evidence (e.g., using experimental laboratory data or best professional judgment) (Reynoldson, Norris, Resh, Day, & Rosenberg, 1997). Ecologists have proposed refinements and alternatives to these methods. For example, some bioassesment advancements has focused on species-specific modeling (Olden, 2003), probabilistic approaches for measuring species richness (Oberdorff & Pont, 2001), variance component modeling to summarize the magnitude of among-site, among-year, site-year, and residual variance (Zuellig, Carlisle, Meador, & Potapova, 2012), and standardizing reference site selection (Hawkins et al., 2010), as well as alternatives that don't require reference sites for regions with pervasive anthropogenic impacts (Chessman & Royal, 2004). Multimetric and multivariate models can also be used as complements, as in a recently developed bioassessment index in the state of California which calculates both IBI and O/E scores for stream macroinvertebrates, and then averages both into the final combined index (Rehn, Mazor, & Ode, 2015).

Fishes are commonly included in stream assessment programs, potentially because they are considered to be of interest to the general public to whom managers may report assessment results. Overwhelmingly, managers have employed IBI approaches for assessing stream fish communities – of the 32 widely cited bioassessment indices for stream fish reviewed in a literature search, all but 5 applied an IBI approach. The opposite is true for stream macroinvertebrate assessment programs, for which 9 out of the 14 widely cited publications employed an O/E approach.

Although applications of multivariate O/E approaches to stream fishes are limited, a number of papers have explored their use. The first example related to the multivariate approach used occurrence data for the 34 most common freshwater fish species in France at 650 reference sites to develop logistic regression equations predicting species occurrence, validated with 88 reference sites, and then compared the log likelihood for the observed and expected taxa in 88 disturbed sites to inform metric selection for an IBI (Oberdorff & Pont, 2001). This publication analyzed a large dataset and proposed a useful probabilistic method for calculating the deviation between observed and expected assemblages, but the research goal was the derivation of an IBI tool, so the development of O/E metric was not pursued. A second multivariate fish modelling application the following year used fish occurrences from 142 reference sites in a region of New

Zealand to develop a discriminant function model that successfully assigned 67% (20 of 30) of validation sites to the correct group – although the model only included 13 fish species, with usually less than five species expected (Joy & Death, 2002). A third publication mirrored this approach using 72 reference sites in Eastern Australia (withholding 10 additional reference sites for external validation and 48 disturbed test sites to measure model sensitivity) and built a discriminant function model that was also limited by the depauperate ichthyofauna of only 24 fish species, so that the average expected richness predicted by the final model was six species. A similar approach was developed in the more species-rich Appalachian streams in the U.S., using 73 sites reference sites and a discriminant function model that predicted 4-22 fish taxa at sites, although the model lacked an independent validation set and performance was judged based on internal validation (Carlisle et al., 2008). This approach was expanded the following year using 228 reference sites (and 38 validation sites) from fish surveys made from 1993-2004 in 28 basins in the Eastern U.S. to build two separate (north and south) discriminant function models (Meador & Carlisle, 2009).

This later paper also examined average taxa O/E response, and found that nearly two-thirds of the Southeastern U.S. fish species were observed less than expected at disturbed sites. This is particularly relevant in a region where 28% of freshwater fish taxa are considered extinct, endangered, threatened, or vulnerable (Warren et al., 2000; Warren, Angermeier, Burr, & Haag, 1997), including endemic species with narrow geographic ranges and restricted habitat requirements (Meador, Coles, & Zappia, 2005). One state in the U.S. of particular concern is Georgia, which has the third most diverse fish fauna in the U.S., of which 17% are considered at risk (Stein, 2002), with 58 taxa under state or federal protection (GAWRD, 2015).

Bioassessment refinements in such locations could improve our understanding of protected stream fish distributions and provide a metric to guide habitat conservation and restoration.

Although few in number, these stream fish O/E publications addressed a variety of locations (Australia, France, New Zealand, and the U.S.). Modeling approaches generally used discriminant function models, with one exception of a logistic regression framework. The datasets used to train, validate, and test the models ranged from one to twelve years of sampling data, 72 to 650 reference sites, 0 to 171 validation sites (using internal validation methods), and 30 to 283 test sites. The number of taxa modelled varied as well, from 13 to 60 fish taxa and the models used from 7 to 11 predictor variables (Table 1). All models included elevation as a variable, all but one included drainage area, and all but two included stream gradient or slope (Table 2). These modeling efforts set the stage for further applications and refinements, particularly opportunities for exploring alternative modeling approaches that leverage continually growing biological monitoring datasets, with a need to develop these tools for the species-rich, but threatened, Southeastern U.S. fish communities.

Research Objective

Stream fish communities have been monitored and assessed since 1998 in the state of Georgia using a multimetric IBI approach directed by the stream survey group (Stream Team) within the Georgia Wildlife Resource Division (GAWRD) (Marcinek & Lanford, 2013). These monitoring results inform water quality management, including assessment and restoration under the Clean Water Act, by the Georgia Environmental Protection Division (GAEPD), which is a sister agency within the Georgia Department of Natural Resources (GADNR). As of 2015, agency biologists are exploring potential refinements in Georgia’s IBI, and are interested in testing the utility of an O/E for Georgia's stream fishes. The objective of this research has been to collaborate with the Stream Team in their analyses and refinement of the state's stream fish assessment methods by developing a complementary O/E model. This O/E model could be incorporated as a measure into Georgia's bioassessment methodology and also could assist in the selection of metrics for a revised IBI.

The methods for building the O/E model are described in Chapter 2. The approach began with linking a dataset comprising samples of fishes from wadeable streams statewide to geographic databases of environmental and anthropogenic variables. Samples classified with the least amount of human land use and infrastructure in their drainages were then identified as least-disturbed sites, under the assumption that the samples would best reflect the natural diversity of stream fish communities. This subset of samples was grouped into community types via statistical clustering methods, and then used to train a multivariate model for predicting taxa occurrences in relation to naturally variable stream characteristics. Together, the clustering and prediction techniques created a species distribution model for estimating taxa-specific probabilities of occurring and being caught in a given site, taking into consideration the site's physical characteristics and location. Taxa occurrence probabilities expected by this model were compared with the taxa observed in the sample, and the end product was a ratio of taxa richness.

The results in Chapter 3 discuss how the performance of the model was validated with subsets of data which had been withheld during the model calibration phase. In the future, new samples can be assessed on the same basis, comparing the observed taxa against those expected by the model, to provide a metric for quantifying stream conditions.

Finally the discussion in Chapter 4 considers how this research has been useful for examining the relative strengths and weaknesses of the IBI and O/E approaches for fish bioassessment. Georgia has a particularly diverse fish fauna relative to other regions of the U.S. (Master, Flack, & Stein, 1998), and this biodiversity provided an opportunity to apply an O/E model, commonly used for invertebrates, to Georgia's species-rich fish taxa. Insights into the biogeography and sensitivity Georgia's fish taxa are discussed along with suggestions for continued stream bioassessment refinements.

Table 1. Summary of six stream fish multivariate models using observed and expected taxa richness. Years indicates the number of years of fish samples in the data set used to train the model. Model indicates the statistical framework (logistic regression, discriminant function or Random Forest) used to model species occurrence probabilities. Predictors indicates the number of environmental predictors used to train the model. Reference, validation, and test indicate the number of samples used in the training (least-disturbed sites), validation (set aside sites), and application (stressed sites) of the model. Models with zero validation sites used internal validation methods. Taxa indicates the approximate number of taxa in the local fish assemblage potentially available for modeling. Results are based on a Google Scholar (http://scholar.google.com/) literature search for “RIVPACS” modeling of stream fish communities.

Location	**Years**	**Model**	**Predictors**	**Reference**	**Validation**	**Test**	**Taxa**	
France	10	Logistic Regression	9	650	88	88	34	[1]
New Zealand	1	Discriminant Function	11	142	0	30	13	[2]
Eastern Australian	4	Discriminant Function	7	72	10	48	24	[3]
Appalachia, US	10	Discriminant Function	10	73	0	54	?	[4]
Eastern US	12	Discriminant Function	9	228	38	80-116	58 - 60	[5]
Georgia, US	16	Random Forest	21	272	171	283	189	[6]

[1] Oberdorff & Pont, 2001; [2] Joy & Death, 2002; [3] Kennard, Pusey, Arthington, Harch, & Mackay, 2006; [4] Carlisle et al., 2008;
[5] Meador & Carlisle, 2009; [6] Maurano, 2015.

Table 2. Predictor variables included in the six stream fish multivariate models reviewed. Results are based on a Google Scholar (http://scholar.google.com/) literature search for the term “RIVPACS”. Publications summarized below, in addition to this thesis, are Oberdorff & Pont, 2001; Joy & Death, 2002; Kennard, Pusey, Arthington, Harch, & Mackay, 2006; Carlisle et al., 2008; and Meador & Carlisle, 2009.

Variable	**Number of Models Included In (of six reviewed)**
Elevation	6
Drainage Area	5
Gradient / Slope	4
Distance from Headwaters	3
Stream Depth	3
Stream Width	3
Air Temperature	2
Basin	2
Distance Inland	2
Ecoregion	2
Flow / Velocity	2
Geology / Soil	2
Latitude	2
Longitude	2
Precipitation	2
Water Temperature	2
Base Flow Index	1
Drainage Density	1
Embeddedness	1
Flow Type (Riffle, Run, Pool…)	1
Reach Length	1
Sample Date / Year / Day of Year	1
Substrate Size	1

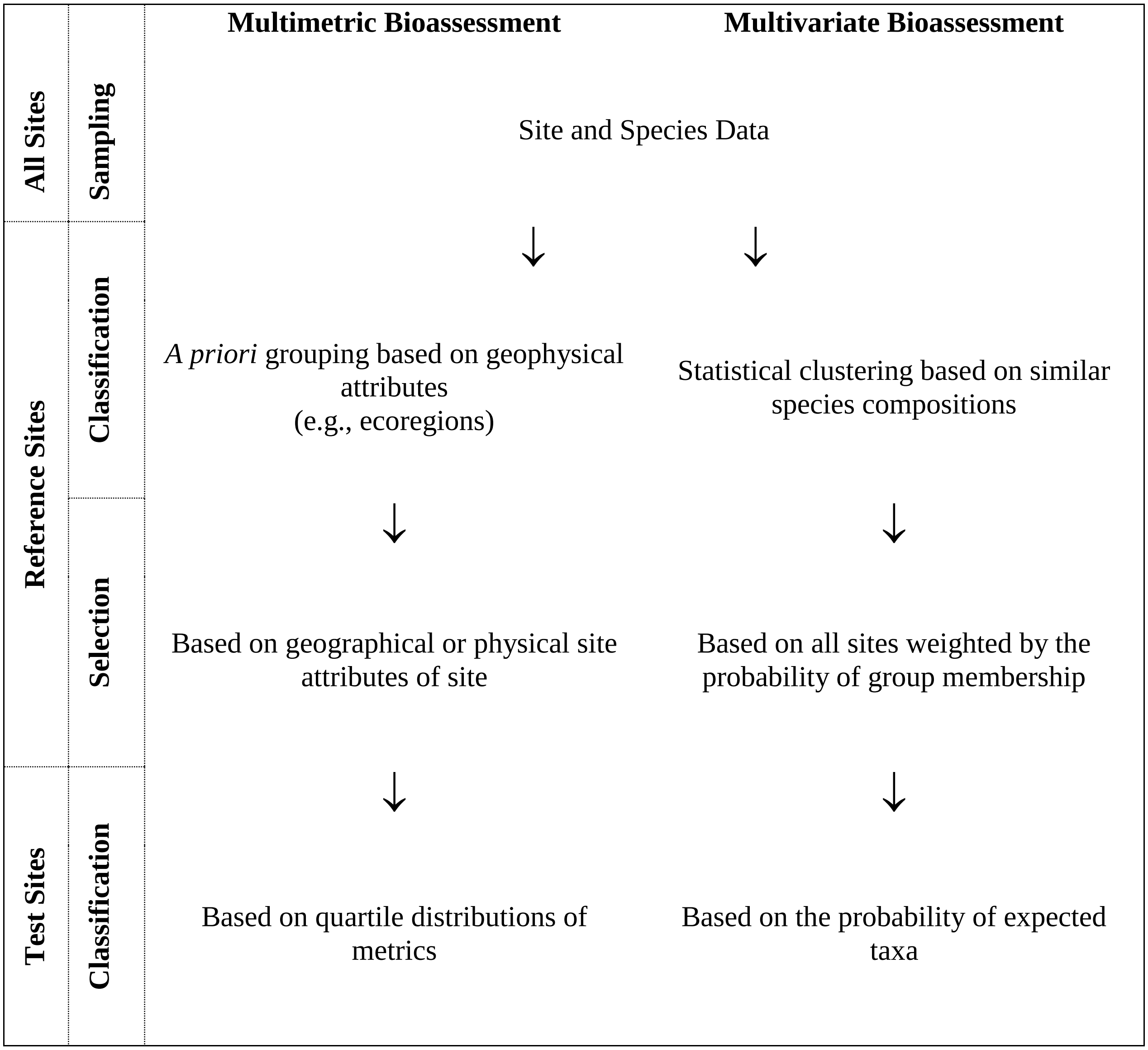


Figure 1. Comparison of the processes for developing and applying multimetric versus multivariate bioassessment metric methodologies. Adapted from Reynoldson et al., 1997.

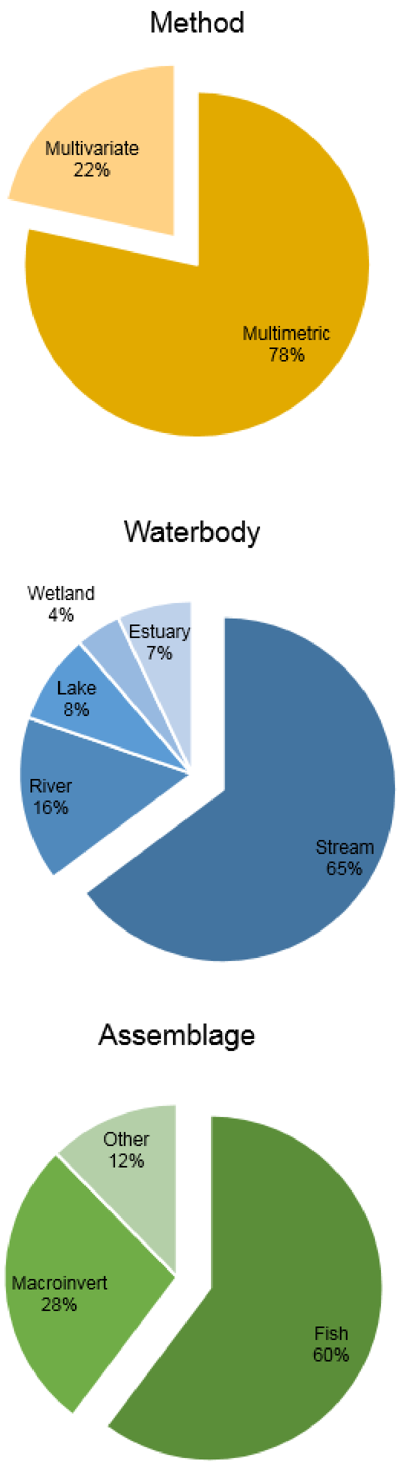
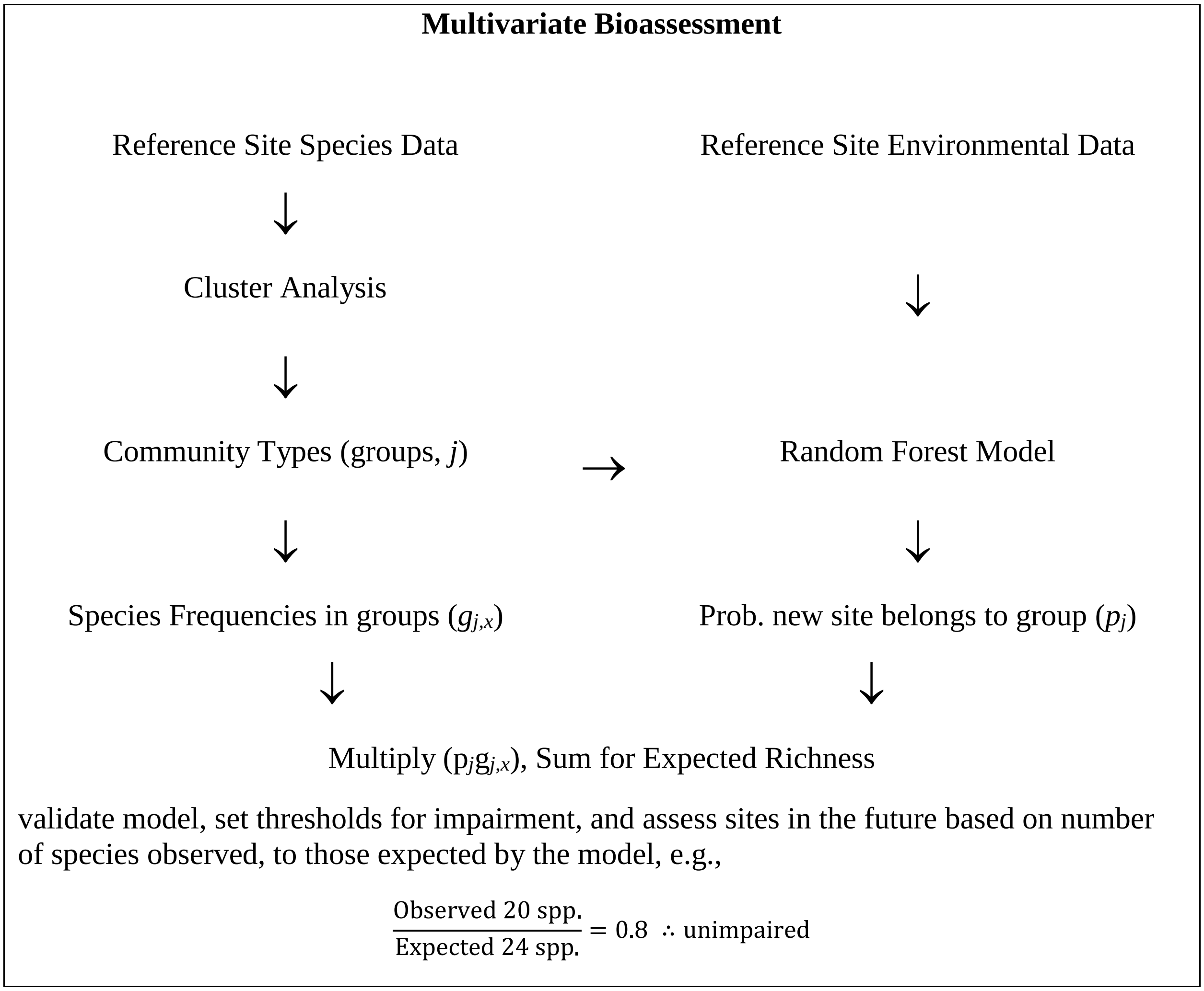


Figure 2. Graphical display of literature search results for bioassessment publications. Results are from a literature search of Google Scholar (http://scholar.google.com/) which returned 51 publications based on multimetric approaches compared to 14 based on multivariate approaches. Results offset from the pie graph indicate the application of this research: multivariate stream fish bioassessment.



validate model, set thresholds for impairment, and assess sites in the future based on number of species observed, to those expected by the model, e.g.,

$$\frac{\text{Observed 20 spp.}}{\text{Expected 24 spp.}} = 0.8 \;\therefore \text{unimpaired}$$

Figure 3. Multivariate model construction process. Sites are limited to those in least-impacted “reference” condition. “Community types” are derived via cluster analysis, and then predicted via Random Forest model using site environmental variables. The probability that a new site belongs to each group is then multiplied by the species frequencies in those groups, and summed for richness.

CHAPTER 2

METHODS

Field Data Collection

Fish were sampled by GAWRD at wadeable stream sites statewide (Figure 4, See Appendix B, Table B.1). Samples were collected from sixteen basins (Figure 5) and six ecoregions (Figure 6), and the product of these two factors created numerous biogeographic units (Figure 7). Surveys were conducted from 1998 to 2013, generally during the April through October period (Figure 8). Fish sampling combined electrofishing and seining techniques in a single upstream pass for 35 times the mean stream width (up to 500 meters sampled) to ensure inclusion of all major habitat types (pools, riffles, runs, woody debris, undercut banks, large rocks, thick root mats, etc.) (GAWRD, 2005).

The young-of-the-year (less than 25 mm total length) individuals were excluded during sample processing since capture efficiency tends to be low and variable (Freeman & Crawford, 1988; Holland-Bartels & Dewey, 1997; Moore & Gregory, 1988; Parsley, Palmer, & Burkhardt, 1989) and those cohorts had not been subject to site conditions for a prolonged period of time, and therefore might not have reflected the long-term conditions in the reach (Schleiger, 2000; Schlosser, 1985). Fish species and abundance in the catch were recorded along with the occurrence of external anomalies (deformations, erosions, lesions, and tumors). Site characteristics including location, elevation, rapid geomorphic and habitat assessment variables, and physicochemical water quality parameters also were recorded.

Geospatial Data Processing

I compiled these field data into a geospatial database and dropped repeated samples (retaining the earliest) to maintain independence of samples. Data quality checks were conducted with the guidance of the Stream Team to drop samples that hadn't meet sampling protocols (e.g., a collection that took place on a braided river site rather than stream), fix data entry errors (e.g., transposed data), and correct species misidentifications. Additional site variables were then derived from a Geographic Information Systems (GIS) analysis (Table 3, Table 4) and associated with the field data. The GIS variables were proxies for the environmental factors (independent of human impacts) that define fish niches and constrain species distributions and abundances: biogeography, geomorphology, temperature, temporal variation, stream size, and connectivity (Poff, 1997). The species distribution model then used these variables to predict species occurrences, in order to account for the influence of biogeography, geology, geomorphology, and temporal variation.

Drainage basin areas were derived using the ArcGIS 10.2 watershed delineation and flow accumulation toolset, by applying the flow direction, sink, and watershed tools (ESRI, 2014). The basin and physiographic province in which sites were located was determined via spatial joins with Level III Ecoregions of the Coterminous United States (Omernik, 1987) and Georgia Department of Natural Resource Basins (GAEPD, 2003) (See Appendix C, Table C.1). Sixteen basins were used in the analysis, as defined by GADNR, based on level six (basin) and eight (subbasin) digit hydrologic unit codes (HUCs) (Seaber, Kapinos, & Knapp, 1987). Temporal variables for the effects of seasonality (day of the year) and inter-annual variation (year) were included. In order to associate the samples with National Hydrograph Dataset (NHD) Plus

Version 2 attributes (Dewald et al., 2012), sites were snapped with a one kilometer tolerance (Figure 9) via the ArcGIS Editing toolbox snap tool to the South Atlantic and Tennessee Vector Processing Units downloaded from http://www.horizon-systems.com/NHDPlus/. NHD flowlines were joined to attribute tables by one of three geoprocessesing methods: via latitude and longitude coordinates, via NHD common identification numbers, or via linear referencing with the NHD network. Reach data were then extracted for mean annual stream velocity, mean annual stream flow, mean annual stream temperature, modified Strahler stream order, stream slope, and arbolate sum of upstream stream kilometers (Table 5).

In order to identify reference sites, variables measuring human disturbance upstream of the sample site were calculated and aggregated. National land cover database data (Fry et al., 2011; C. G. Homer et al., 2015; C. Homer et al., 2007) downloaded from http://www.mrlc.gov/ were associated by the year closest to the sample date (2001, 2006, 2011). Land cover was weighted by a landscape development intensity index that measures levels of human activity as nonrenewable energy input (e.g., electricity, fuel, fertilizer, pesticide, public water supply, and irrigation), in solar energy joules per hectare per year (empirical coefficients corrected for different qualities and normalized on a scale of 1-10) (Brown & Vivas, 2005). Coefficients published for use with the NLCD coverage were used so, for example, land uses of open water, wetland, and forest were assigned a value of 1.00, cultivated crops 4.54, and high density development 9.42 (Gara & Micacchion, 2010). Additionally, dam density, road density, and stream crossings were generated for each site catchment. Dam density was obtained from the SouthEast Aquatic Connectivity Assessment Project coverage at http://maps.tnc.org/seacap/. Road density was generated from 2014 U.S. Census Bureau TIGER/Line Shapefiles downloaded from https://www.census.gov/geo/maps-data/data/tiger.html, and stream crossing density was tabulated from intersections with the NHD+v2 network. Sites with the lowest combined percentile scores, stratified by basin and ecoregion, were identified as least-disturbed sites for model calibration. Samples were subset into least-disturbed (n = 272), stressed (n = 283), validation (n = 171), and no designation (n = 548). The "least-disturbed" sites were treated here as synonymous with the term "least impacted" sites (Davis & Simon, 1995). Of these least-disturbed sites, 86% were used in model calibration and 14% were set aside for model validation. These impact designations were reviewed graphically with respect to the most commonly occurring species (Figure 10) and with respect to environmental gradients (Figure 11) and spatial distribution that might influence taxa richness (Figure 12). Model construction and diagnostics were performed in R version 3.2.2 (R Core Team, 2015). Scripts for building and applying the model were based on a RIVPACS-type approach (Ralph T. Clarke, Wright, & Furse, 2003; J. F. Wright et al., 2000) and adapted from a macroinvertebrate implementation by the US Environmental Protection Agency Western Ecology Division (Van Sickle, Hawkins, Larsen, & Herlihy, 2005; Van Sickle, Huff, & Hawkins, 2006; Van Sickle, Larsen, & Hawkins, 2007) downloaded from http://www.epa.gov/wed/pages/models/rivpacs/rivpacs.htm. A subset of draft revised IBI scores for the Apalachicola-Chattahoochee-Flint Basin portion of the Piedmont Ecoregion (ACF-PDM) were derived by the Stream Team using multimetric bioassessment techniques and used to tentatively compare results between draft IBI and O/E indices.

Fish Community Types

Over two-hundred fish taxa were observed in Georgia streams, and several methods were applied to reduce the community data. First, the following species were combined to the genus level because congeneric species occurring allopatrically in different river systems or because of uncertainty in their field identification: *Campostoma*, *Cottus*, *Gambusia*, *Hybopsis*, *Nocomis*, *Pimephales*, *Rhinichthys*, *Semotilus*, *Pteronotropis*, and “Redeye Bass Species” (*Micropterus* spp.) (See Appendix C, Table C.2). Second, the matrix of catch-abundances was converted to detections and non-detections, in response to the potentially high variability in fish counts caused by incomplete and variable detection of individuals – potential biases which are difficult to mitigate in a dataset with relatively few repeat samples. Third, rare taxa (present at less than five percent of sites) were screened from the dataset because they often have higher variability in detection probability, which may lead to over-splitting when discriminating between site groups (Hawkins, Norris, Hogue, & Feminella, 2000). Fourth, taxa from samples at least impacted sites were clustered into six groups with the flexible-Beta method (β = -0.6), based on Sørensen distance (Bray & Curtis, 1957) using the “agnes” function in the “cluster” package (Cluster Analysis Extended Rousseeuw et al) in R. Flexible-beta is one of six methods for the function “agnes” (Agglomerative Nesting) in the "cluster" package, where at each iteration, the two nearest clusters are combined based on one of several potential methods: average of dissimilarities of points in one cluster versus points in another cluster, nearest neighbor, furthest neighbor, or flexible, which is specified by the Lance-Williams formula (Murtagh & Legendre 2011). The flexible beta approach which was employed is a generalization of a weighted average method and was chosen because it can be used to balance clustering and chaining in the final dendrogram. Beta (β) = -0.6 was selected after visually examining plots from β = -1 (overclustered) to 0.5 (badly chained) in increments of 0.25 (Figure 13).

Model Construction

A Random Forest (RF) model was built to predict the probability that a site belongs to one of the previously defined community types. RF is a data mining technique based on a machine learning algorithm that builds an ensemble of classification or regression trees through binary recursive partitioning, and outputs the mode (classification) or mean prediction (regression) of the “forest” of trees (Breiman, 2001). Classification trees are trained on independent predictors in a learning dataset to project a dependent categorical outcome. The algorithm uses the predictors to repeatedly split the data from a “parent” node into two parts, based on an “impurity function” that maximize the homogeneity of the “child” nodes (Lewis, 2000; Loh, 2011). The RF methodology selects about 2/3 of the samples (with replacement, i.e., bootstrapping) to build a “forest” of trees, but randomly selects only a small number of the predictor variables to use at each split. Increasing the number of predictor variables used at each split increases the strength of an individual tree, but also increases the correlation between trees, resulting in a tradeoff in the overall error rate. Each tree is then used to predict the observations from the original dataset that weren’t included in the bootstrap training sample (called “out-of-bag” (OOB) observations), which provides cross-validated accuracies and error rates that can be averaged over all trees (Breiman, 2001). Whereas a number of statistical methods measure variable importance indirectly using metrics of a model’s statistical significance or Akaike’s Information Criterion, the RF method derives novel variable importance measures (Cutler et al., 2007). The value for each predictor variable is then randomly permuted for the OOB observations, passed down the tree, and the difference in misclassification rates between the permuted and unpermuted OOB data provides a measure of variable importance (Breiman, 2001).

This modeling approach was chosen based on its performance in other ecological applications, particularly species distribution and bioassessment modeling, where RF has demonstrated high classification accuracy, ability to model complex interactions among predictor variables, and flexibility for handling disparate data types, with relatively few tuning parameters and multiple variable importance metrics (Cutler et al., 2007). In a species distribution modeling application in the Eastern U.S., the RF model was more predictive than regression tree analysis and multivariate adaptive regression splines (Prasad, Iverson, & Liaw, 2006). RF also had the highest predictive accuracy in a Northern European species distribution modeling application, based on the area under the curve performance metric, when compared to generalized linear and additive models, multivariate adaptive regression splines, artificial neural networks, a general boosting method, classification tree analysis, and mixture discriminant analysis (Marmion, Parviainen, Luoto, Heikkinen, & Thuiller, 2009). Specific to aquatic ecosystems, the approach better predicted the biological condition of stream benthic macroinvertebrate communities in the Chesapeake Bay watershed when compared to classification and regression trees, conditional inference trees, and ordinal logistic regression (Maloney, Weller, Russell, & Hothorn, 2009). The approach has also been extended to predict the macroinvertebrates in streams across a large portion of the Eastern U.S. (Carlisle, Falcone, & Meador, 2009). In a similar application in the Western United Sates, the advantages of ensemble regression tree approaches compared to multiple linear regression were highlighted, including the ability to include more variables in the model building phase, easier testing for interaction effects, and the availability of model diagnostics such as partial dependency plots (Waite et al., 2012). A closely related ensemble regression tree approach, boosted regression trees, has successfully been applied to predict algal and macroinvertebrate communities in National Water-

Quality Assessment sites across the U.S. (Waite, 2014). Conversely, applying RF to model species distribution has potential drawbacks, as the models can be difficult to visualize or interpret, the measures of performance and variable importance have been critiqued, and the model may be prone to over-fitting data, resulting in a less transferable model (Lobo, Jiménez-valverde, & Real, 2008; Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Wenger & Olden, 2012). These limitations were considered when selecting the modeling approach and when interpreting results, and the RF approach was ultimately chosen since it represented a novel application for stream fish O/E assessments, an exploratory machine-learning approach with few prior assumptions required, a wide window for tuning parameters, and it had strong performance in a range of related applications.

The RF model was built via the “randomForest” function in the "randomForest" package (Breiman & Cutler's Random Forests for classification & regression). The parameters “ntree” for number of trees and "mtry" for number of variables tried at each split were tuned to OOB error (Liaw & Wiener, 2002). The parameter “ntree” was iterated in increments of 100 (up to 10,000) searching for an asymptote in OOB error, but this resulted in multiple OOB minima at widely different “ntree” values, so the default value of 500 was retained. Higher values of "mtry" increase correlation between trees (increasing forest OOB error rate) but also increase the strength of individual trees (decreasing OOB error). Setting “mtry” as the square root of the number of predictors is considered the “Random Forest” method while the use all predictors is known as the “bagging” method. The “tuneRF” function within the randomForest package was used to search for an optimum “mtry” number of variables to be randomly selected at each split, based on OOB error (in steps of 1.5, searching until improvement was <0.01) (Figure 14).

Variables were removed from the model if they were sparsely populated in the dataset or had low mean decreases in either the "accuracy" or "Gini" variable importance metrics (Figure 15). The mean decrease in accuracy metric quantifies changes in OOB error from variable permutations, as previously described. Gini measures homogeneity in a child node and ranges from zero (homogeneous) to one (heterogeneous), so decreases in Gini reflect increases in purity. Mean decreases in Gini from each parent to child node are summed over all trees, and normalized for each variable (Breiman, 2001). Although the variable importance measures were sensitive to the tuning parameters chosen, their rankings were generally stable. The model domain was limited to predicting species found in the majority of samples (i.e., overall occurrence, called "probability of capture (*Pc*)" greater than 0.5) to decrease the variability in predictions (Hawkins, Norris, Hogue, et al., 2000). In order to decrease variability in O/E scores, *Pc* was iterated from near zero (including all taxa) to 0.5 (including only taxa found in >50% of samples) (Figure 16). The predicted relationships between the site environmental variables and group membership were examined in partial dependence plots where the Y-axis is the mean value of logit(*p*) and *p* is the predicted probability of being in a group, with the mean taken over all other combinations of the other predictors (Cutler et al., 2007).

The final model was used to assess fish assemblage integrity at each site. In order to predict a site's identity in one of the aforementioned six community groups, the site's environmental variables were run down the RF trees, and their votes summed and divided by the total number of trees. Then the probabilities of a site belonging to each community group were multiplied by the taxa frequencies in that group (previously defined in the clustering and model calibration stage from taxa occurrences at least impacted sites). The product of the group probabilities matrix and taxa frequencies matrix was summed for each taxa, to calculate a probability of occurrence for that taxa at the site. The predicted probabilities of those taxa above the *Pc* threshold (previously defined as 0.5, that is, predicted to occur at 50% or more of the sites in that community group), were summed for each site. This taxa richness expected by the model was used as the denominator for the final metric. The numerator was the summed richness observed for the same subset of taxa (i.e., those above the *Pc* threshold). The final metric, therefore, was a ratio of observed to expected taxa richness (over the domain of frequent taxa).

The O/E scores and model were analyzed with several methods. The distribution of calibration and validation sites were examined for normality, and the O/E scores were plotted to examine their deviation from one (the theoretical value for least-impacted sites). O/E scores were graphed against the environmental predictor variables (e.g., drainage basin area, elevation, etc.) to confirm that the model wasn't confounded by unexplained variation in these natural gradients. Outliers in O/E scores were mapped to examine potential spatial patterns or biases in the model.

Model performance was measured primarily by two metrics: the standard deviation of O/E, and the ability of O/E scores to discriminated between sites that had been designated least-impacted and those designated as stressed. The effects of varying probability of capture on O/E scores were examined empirically and hypothetically. Empirically, the *Pc* values were iterated in the model and the resulting O/E discrimination efficiency was examined. This interaction was also examined through simulations with theoretical data sets of varying taxon occurrences. Because setting *Pc* at 0.5 resulted in low (<10) expected taxon richness for most regions of the state, the effect of low richness on O/E score variability and on discrimination between least impacted and stressed conditions was examined using data simulations performed in R. Specifically, for taxa richness of 3, 6, 9 and 18 (representing the range of expected richness generated for clusters in this analysis), 10,000 random binomial samples were generated using a mean probability of occurrence and capture of 0.75 (i.e., midway between the *Pc* threshold for taxa inclusion of 0.5 and 1.0) at least impacted sites. To simulate samples from stressed sites, 10,000 random binomial samples were generated using a mean probability of occurrence and capture of 0.60 (i.e., a 20% reduction in mean occurrence compared to least impacted sites). Values for O/E (where expected richness was mean probability of occurrence and capture, 0.75, times number of number of taxa) were generated for the 10,000 samples at each level of taxa richness and plotted to examine score variability and discrimination between least impacted and stressed sites.

O/E scores were also plotted against the potential revised IBI scores in the Apalachicola-Chattahoochee-Flint Basin portion of the Piedmont Ecoregion (ACF-PDM; provided by P. Marcinek and J. Argentina, GADNR), to compare the potential assessment outcomes of the two candidate bioassessment metrics. The scores were plotted for all subsets of the data (calibration, validation, no designation, and stressed), and the discrimination of least-impacted versus stressed sites was examined by basin and ecoregion. In addition to site O/E scores, taxa O/E scores were calculated at stressed sites (based on a O/E model that included all taxa, with *Pc* > 0), to examine taxa that occurred more or less frequently than expected in response to anthropogenic disturbances (Meador & Carlisle, 2009). Taxa O/E summed and averaged expectations across sites, as compared to a site O/E, which summed expectations across taxa. The average taxa O/E scores were analyzed against traits assigned by the Stream Team on the basis of other fish traits databases, local knowledge, regional taxonomy books, and peer reviewed publications. Where adequate information was available, the Stream Team assigned taxa feeding designations based on the fishes' foraging methods: generalist feeder (may prefer a certain source but will readily consume plants, invertebrates, fish; i.e., omnivore), herbivore (consumes plant matter including, but not limited to, aquatic vegetation, algae, detritus and plankton), parasitic (feeds off of fish), and predator (adults consume fish and invertebrates; juveniles may feed primarily on invertebrates due to limited gape). Similarly, taxa were assigned to habitat preference categories: habitat generalist which will thrive in impoundments; fluvial (stream) specialist which relies on flow and will not thrive in impoundments; and swamp specialist, strongly associated with swamps, backwater, and ditches, often associated with vegetation, and life cycles tied to flood plain inundation. Finally, taxa were grouped into qualitative tolerance categories, based on their presumed tolerance (high, medium, or low) to anthropogenic disturbance. Response of taxa O/E score were analyzed with summary statistics, and graphed with density and strip plots.

Table 3. Variables retained in the Random Forest (RF) model to predict the probability that a site belongs to one of the defined community types. Variables were used to train the RF model, and retained based RF's internal variable importance metrics.

Variable	**Description**	
ArbolateSum	Arbolate sum in kilometers of stream upstream of the bottom of the NHDFlowline feature	4
Atlantic_Slope	Atlantic Slope Drainage	2
Bas_Coo	Coosa Basin	2
Bas_Ten	Tennessee Basin	2
Date	Sampling Date	1
DayOfYear	Sampling Day of Year	1
DBA_km2	Drainage Basin Area in square kilometers	5
Eco_BRM	Blue Ridge Ecoregion	3
Eco_PDM	Piedmont Ecoregion	3
Eco_SCP	Southern Coastal Plain	3
Eco_SEP	Southeastern Plains Ecoregion	3
Elevation_m	Elevation in meters	1
Flow_m3ps	Mean Annual Stream Flow at downstream end with gage adjustment in cubic meters per second	4
Gulf_Slope	Gulf Slope Drainage	2
Lat	Latitude in decimal degrees	1
Long	Longitude in decimal degrees	1
Mean_Ann_Temp_C	Mean Annual Stream Temperature in Celsius	4
RL_m	Reach Length in meters	1
Slope	Unitless slope of reach	4
Velocity_mps	Mean Annual Stream Velocity at downstream end with gage adjustment in meters per second	4
Year	Sampling Year	1

Sources: (1) Field data collection; (2) Georgia DNR Basins (GAEPD, 2003) (3) Ecoregions of the Coterminous United States (Omernik, 1987); (4) National Hydrography Plus Version 2 (Dewald et al, 2012). (5) ArcGIS 10.2 watershed delineation and flow accumulation toolset (ESRI, 2014).

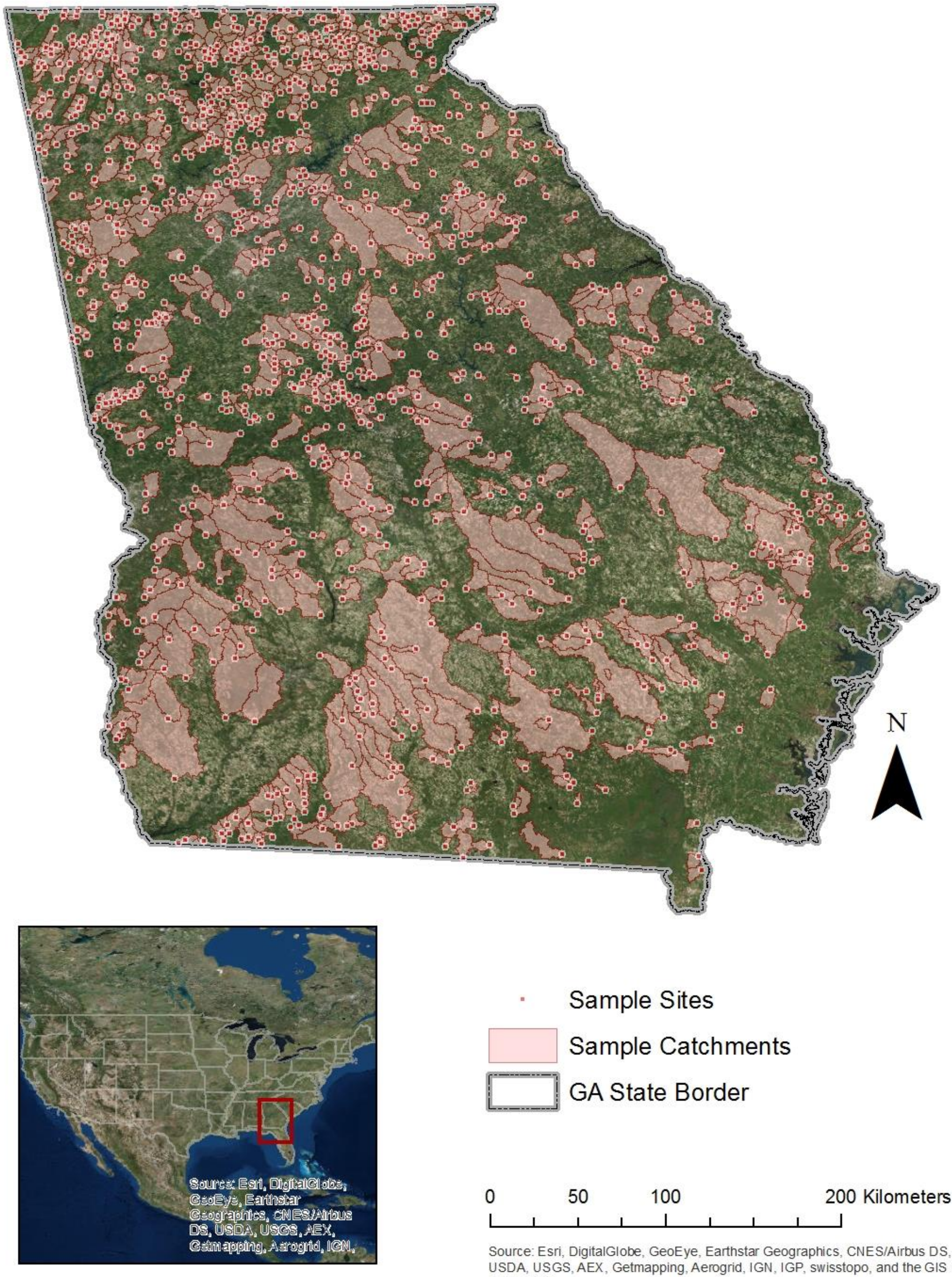
Table 4. Variables dropped from Random Forest (RF) model for predicting site membership probability. Candidate variables were dropped from the model based on their low variable importance metrics, relative to the variables retained and used to train the model.

Variable	**Description**	
Bas_Alt	Altamaha Basin	2
Bas_Apa	Apalachicola Basin	2
Bas_Auc	Aucilla Basin	2
Bas_Cht	Chattahoochee Basin	2
Bas_Fli	Flint Basin	2
Bas_Och	Ochlockonee Basin	2
Bas_Ocm	Ocmulgee Basin	2
Bas_Oco	Oconee Basin	2
Bas_Oge	Ogeechee Basin	2
Bas_Sat	Satilla Basin	2
Bas_Sav	Savannah Basin	2
Bas_Stm	St. Marys Basin	2
Bas_Suw	Suwanee Basin	2
Bas_Tal	Tallapoosa Basin	2
Eco_RGV	Ridge and Valley Ecoregion	3
StreamOrder	Modified Strahler Stream Order	4

Sources: (2) Georgia DNR Basins (GAEPD, 2003); (3) Ecoregions of the Coterminous United States (Omernik, 1987); (4) National Hydrography Plus Version 2 (Dewald et al, 2012).

Table 5. In order to associate the database of field fish monitoring with other site environmental data, the sample sites were joined to National Hydrograph Dataset (NHD) Plus Version 2 attributes by one of three geoprocesses (via latitude and longitude coordinates, via NHD common identification numbers, or via linear referencing with the NHD network). The parameters for the join are specified in the first three columns, and the data that were extracted for site environmental characteristics are summarized in the final two columns.

From Feature	**Join Field**	**To Attribute**	**Extracted Data**	**Data Summary**
NHDFlowline.shp	ComID	EROM_MA0001	Q0001E, V0001E	Stream Flow, Stream Velocity
NHDFlowline.shp	ComID	PlusFlowlineVAA	StreamOrde, StreamCalc	Stream Order
NHDFlowline.shp	ComID	ElevSlope	Slope	Stream Slope
Catchment.shp	FeatureID	IncrTempMA.txt	TempV	Catchment Temperature
NHDFlowline.shp	ReachCode via linear reference	IBI_Site_NHD_Snapped_Table	n/a	Stream Hydrography
Site	Spatial Join	Ecoregion_Level3	Ecoregion	Level III Ecoregions
Site	Spatial Join	WBD_Subwatershed	Hydrologic Unit Code	GADNR Basins



Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 4. Locations and drainages of GADNR wadeable stream samples from 1998-2013. Sampling effort was focused in subregions in certain years. Nearly twice as many samples were collected in the Gulf Slope (n = 941) as the Atlantic slope (n = 563).

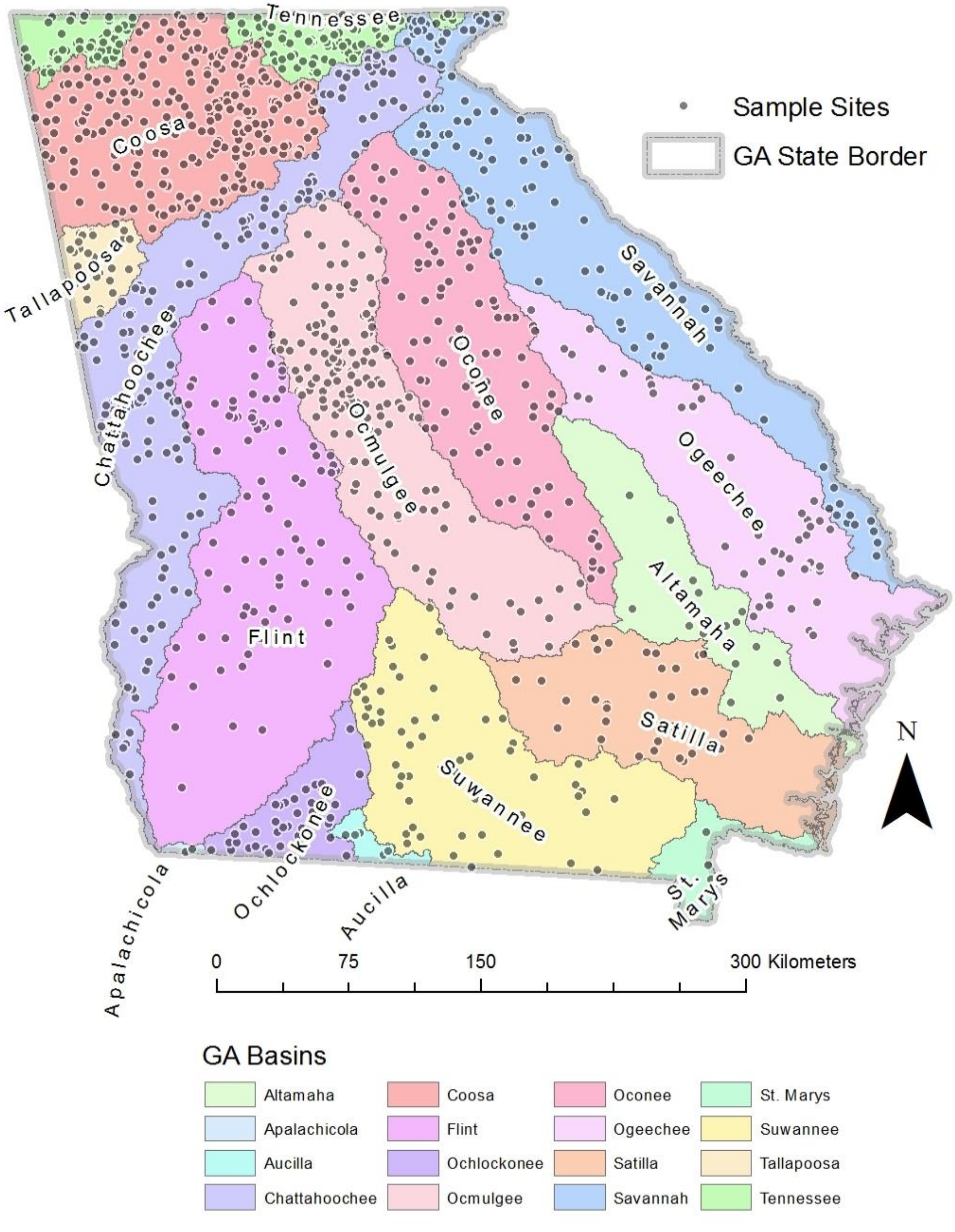


Figure 5. GADNR wadeable stream fish samples, 1998-2013, overlain on the sixteen basins in the state. Sampling effort was spread relatively evenly over the state, although there are fewer samples in some basins below the Fall Line. The Coosa, Chattahoochee, and Ocmulgee Basins had the greatest number of samples (n > 200) over the monitoring period.

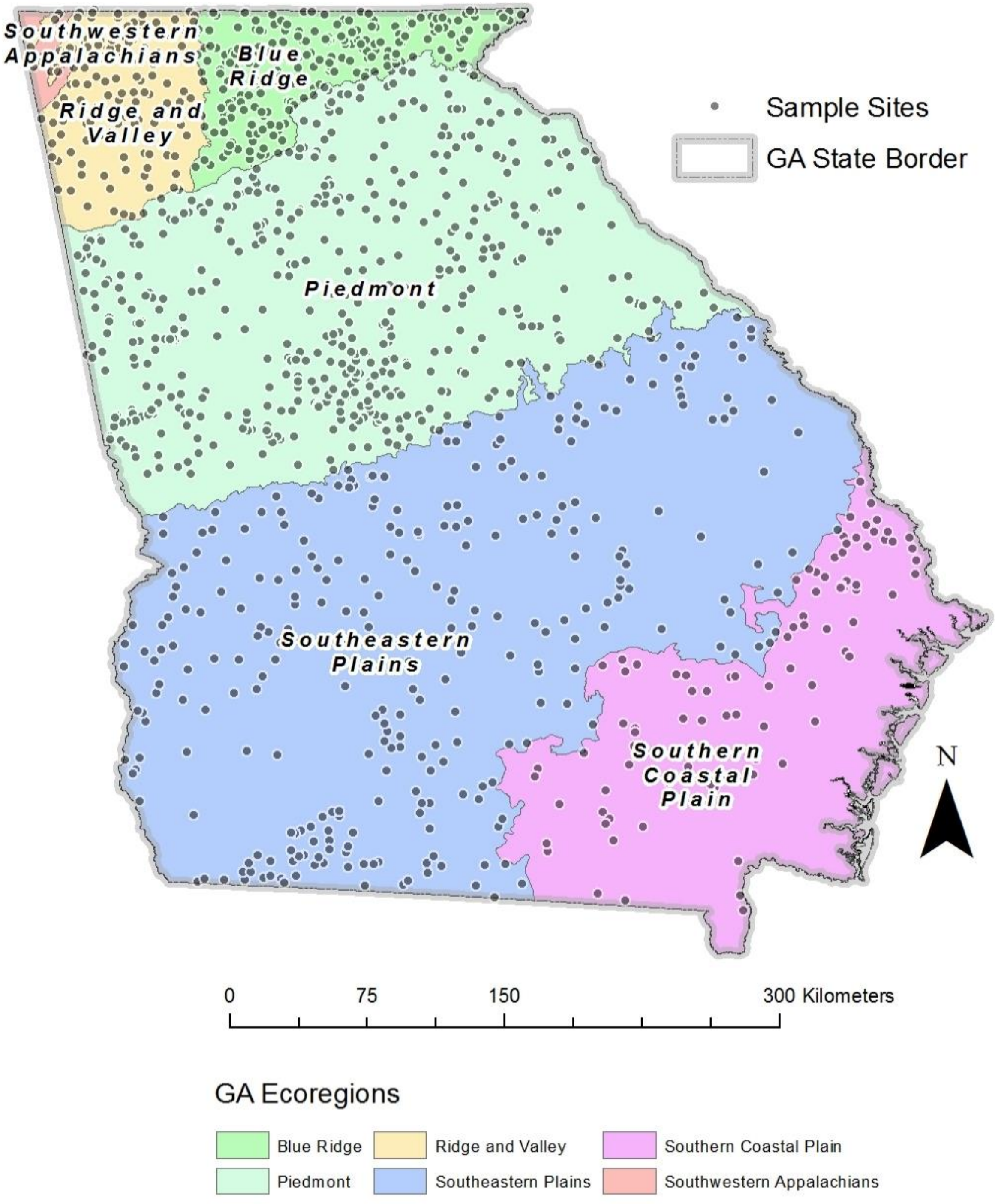


Figure 6. GADNR wadeable stream fish samples, 1998-2013, overlain on the six ecoregions in the state. Sampling density was greatest in the northern mountainous regions of the state and least in the Southern Coastal Plain.

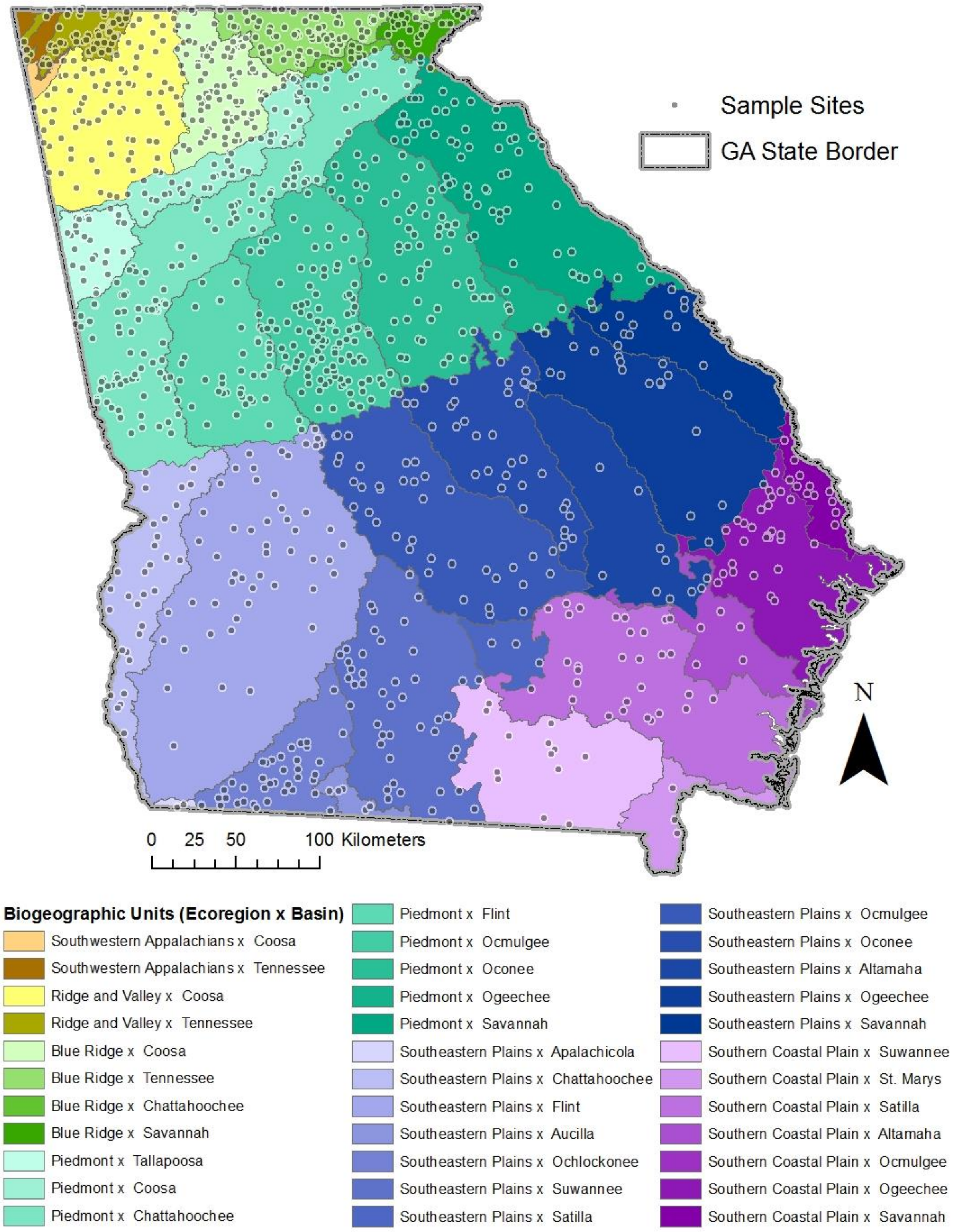
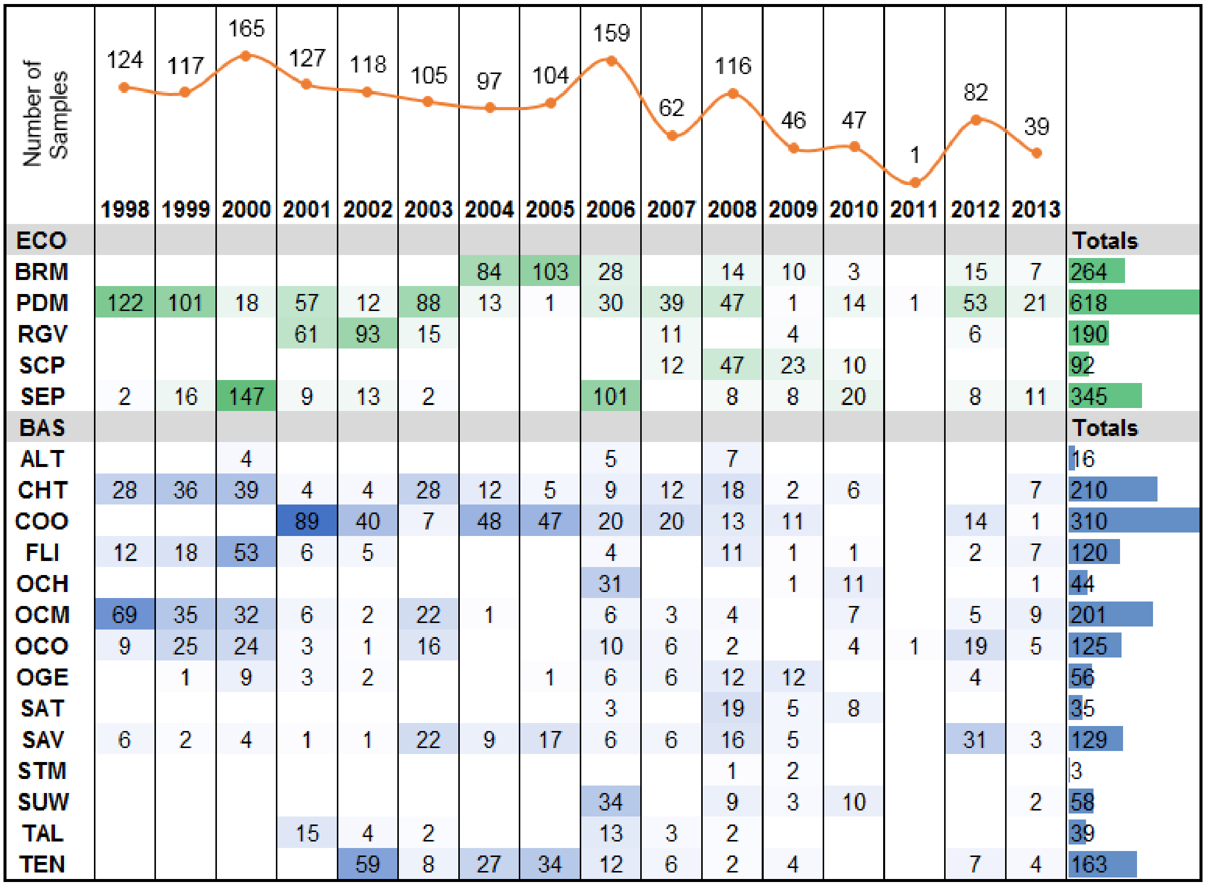


Figure 7. Biogeographic units created from the overlay of Ecoregions on Basins (clipped to Georgia state border and generated via geometric union in ESRI ArcGIS Analysis Toolbox Overlay Toolset). Points are locations of GA wadeable stream fish samples, 1998-2013.



Number of Samples	124	117	165	127	118	105	97	104	159	62	116	46	47	1	82	39	
	1998	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	
ECO																	**Totals**
BRM							84	103	28		14	10	3		15	7	264
PDM	122	101	18	57	12	88	13	1	30	39	47	1	14	1	53	21	618
RGV				61	93	15				11		4			6		190
SCP										12	47	23	10				92
SEP	2	16	147	9	13	2			101		8	8	20		8	11	345
BAS																	**Totals**
ALT			4						5		7						16
CHT	28	36	39	4	4	28	12	5	9	12	18	2	6			7	210
COO				89	40	7	48	47	20	20	13	11			14	1	310
FLI	12	18	53	6	5				4		11	1	1		2	7	120
OCH									31			1	11			1	44
OCM	69	35	32	6	2	22	1		6	3	4		7		5	9	201
OCO	9	25	24	3	1	16			10	6	2		4	1	19	5	125
OGE		1	9	3	2			1	6	6	12	12			4		56
SAT									3		19	5	8				35
SAV	6	2	4	1	1	22	9	17	6	6	16	5			31	3	129
STM											1	2					3
SUW									34		9	3	10			2	58
TAL				15	4	2			13	3	2						39
TEN					59	8	27	34	12	6	2	4			7	4	163

Figure 8. Temporal trends in Georgia Wildlife Resource Division wadeable stream fish sampling from 1998-2013. Ecoregions and Basins are color coded green and blue respectively with greater opacity indicating higher number of samples, and totals are presented in bar charts. Sampling effort rotated among basins and ecoregions, focusing on different parts of the state at different time periods.

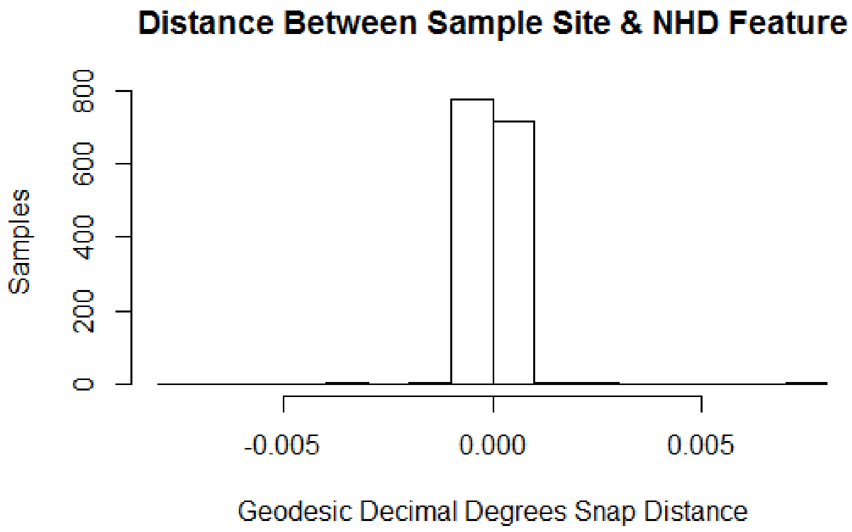


Figure 9. Most sample sites were close to, or spatially coincident with, the hydrography network and features used for deriving environmental variables. Samples sites were snapped (with a 1 kilometer tolerance) to National Hydrograph Dataset (NHD) Plus Version 2 http://www.horizon-systems.com/NHDPlus/.

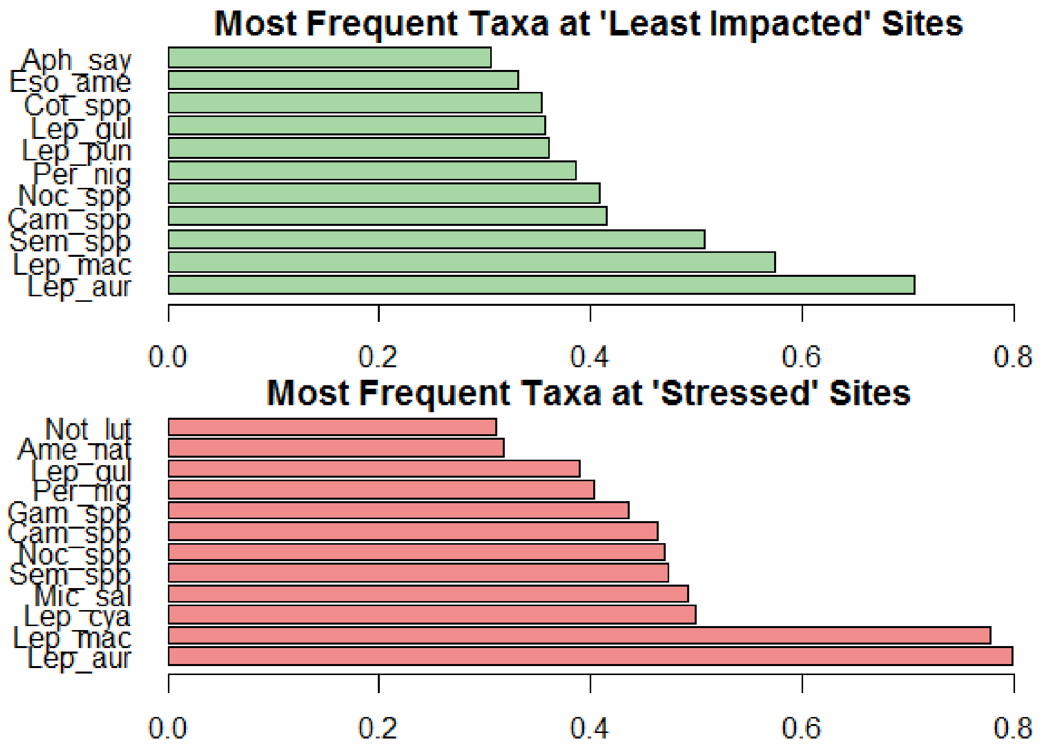


Figure 10. Summary of taxa that were found most frequently at sites which had been designated least impacted and stressed sites, based on anthropogenic influence in the catchment.

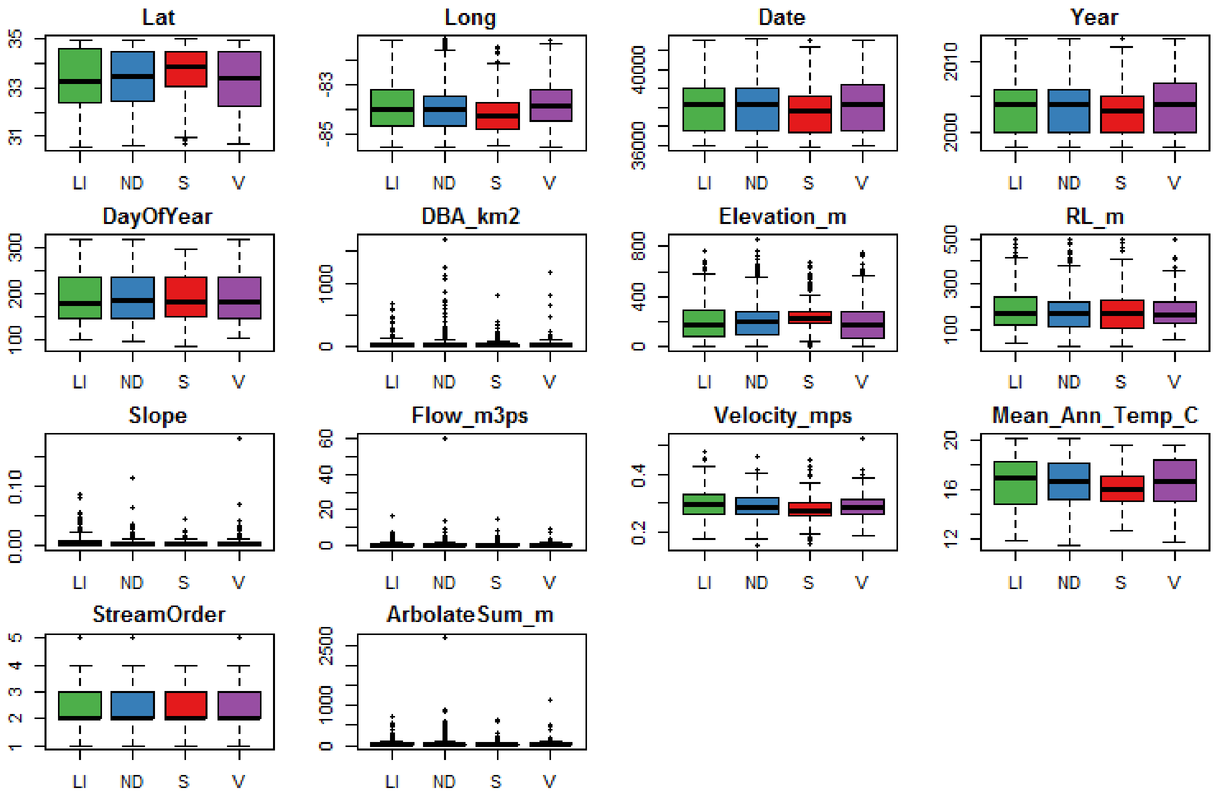
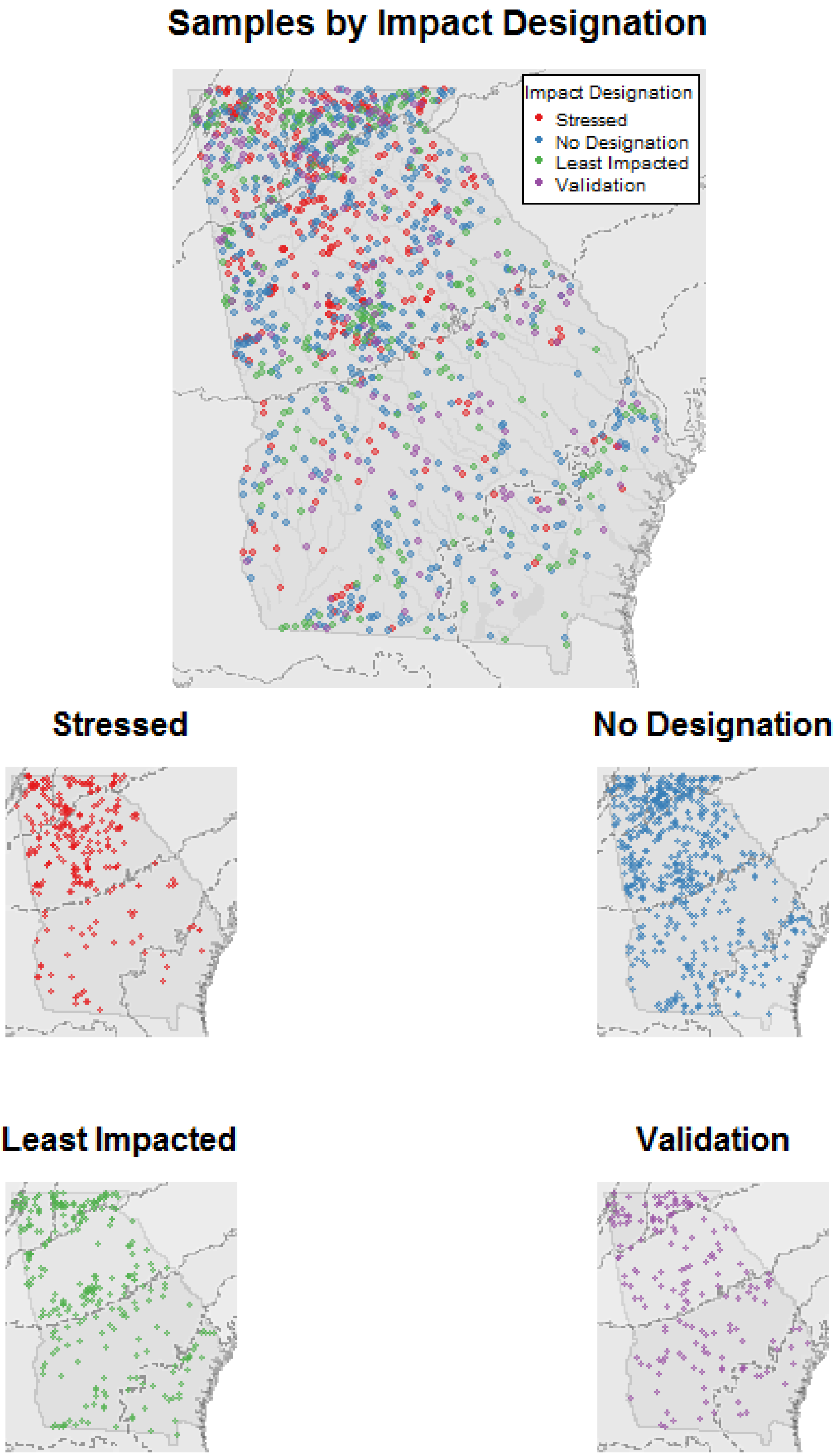


Figure 11. Box and whisker plot diagrams displaying the distribution of predictor variables by impact designation. LI denotes “least impacted”; ND denotes, “no designation”, “S” denotes “stressed”, and V denotes “validation” samples.


Samples by Impact Designation

Stressed

No Designation

Least Impacted

Validation

Figure 12. Spatial distribution of reference sites used for community clustering and taxa prediction (above) and mapped separately by impact designation (below).

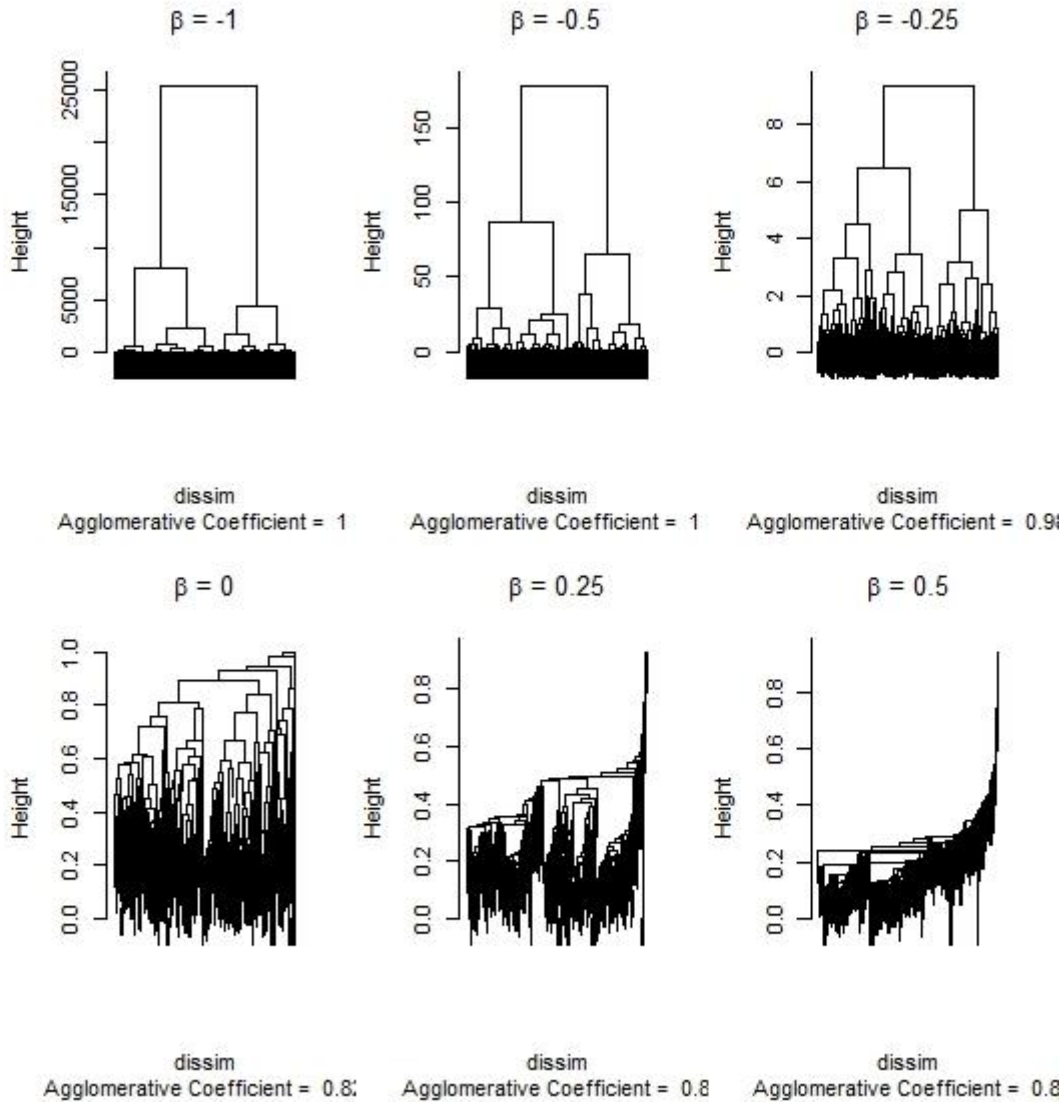


Figure 13. Results from balancing the chaining and clustering of the dendrogram via the flexible beta method for cluster analysis. Positive β's resulted in chaining, while β's approaching -1 resulted in overclustering of the dataset. Final β selected for cluster analysis was -0.6.

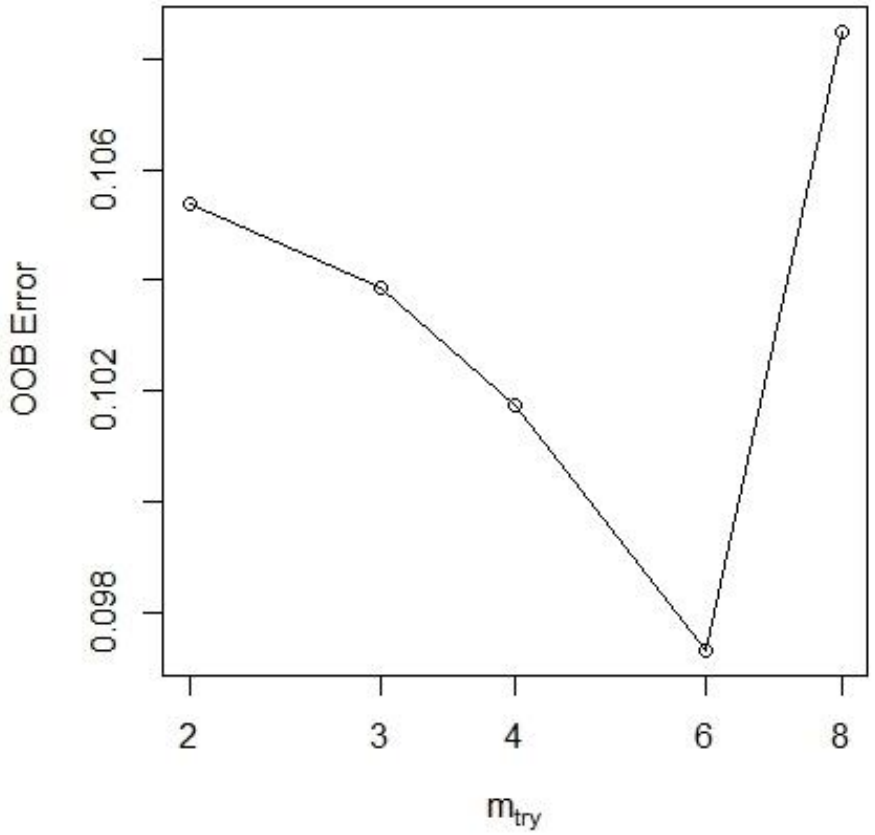


Figure 14. Example of tuning graph for selecting mtry, the number of variables to select at each split by the Random Forest (RF) model. Optimum mtry value is based on out-of-bag error rate, selected by left and right searches by the “tuneRF” function in steps of 1.5, searching until improvement was <0.01.

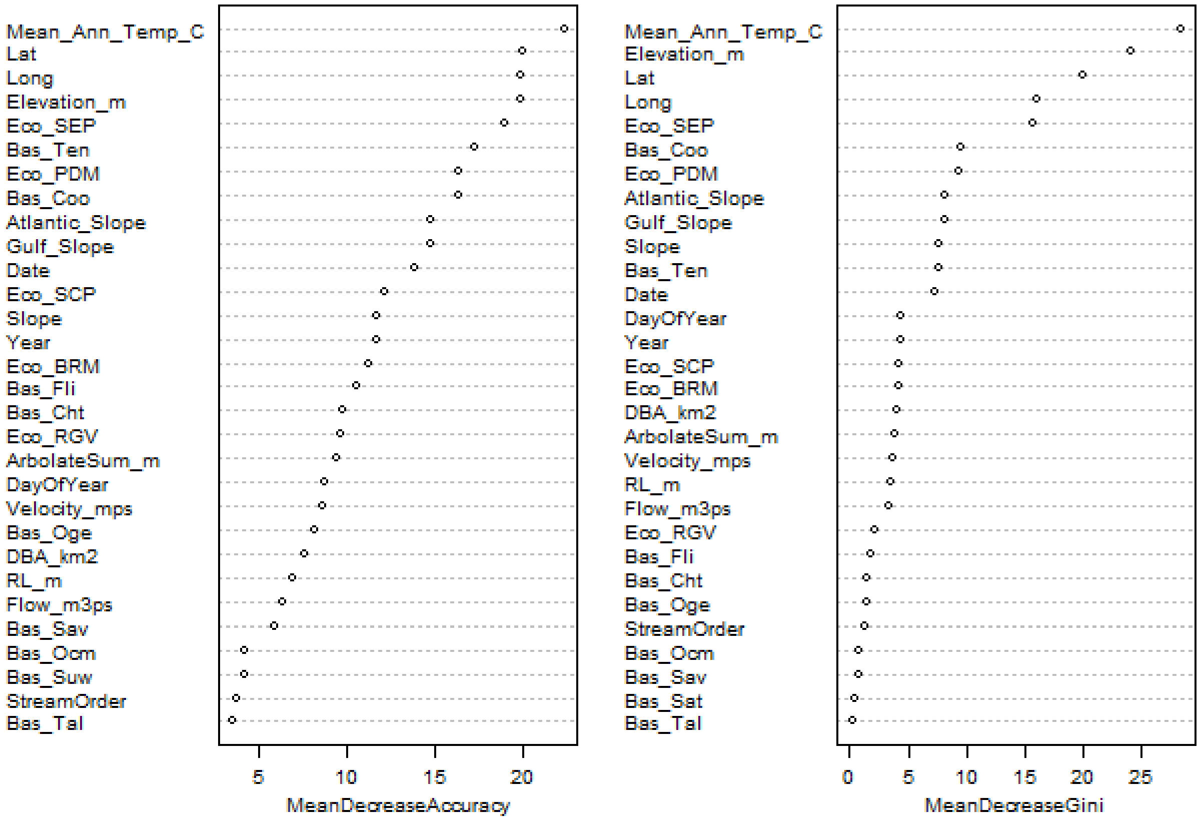


Figure 15. Mean decrease in accuracy and Gini as measures of variable importance for classification in the Random Forest model with all candidate variables. These metrics were used to select the most predictive variables for inclusion in the final model. To derive mean decrease in accuracy, the values of a predictor variable are randomly permuted for the “out-of-bag” OOB observations (those withheld while bootstrapping the dataset during construction of the forest), passed down the tree, and the difference in misclassification rates between the permuted and unpermuted OOB data provides a measure of variable importance. Gini measures homogeneity and ranges from zero (homogeneous) to one (heterogeneous), so decreases in Gini reflect increases in purity. Mean decreases in Gini from each parent to child node are summed over all trees, and normalized for each variable.

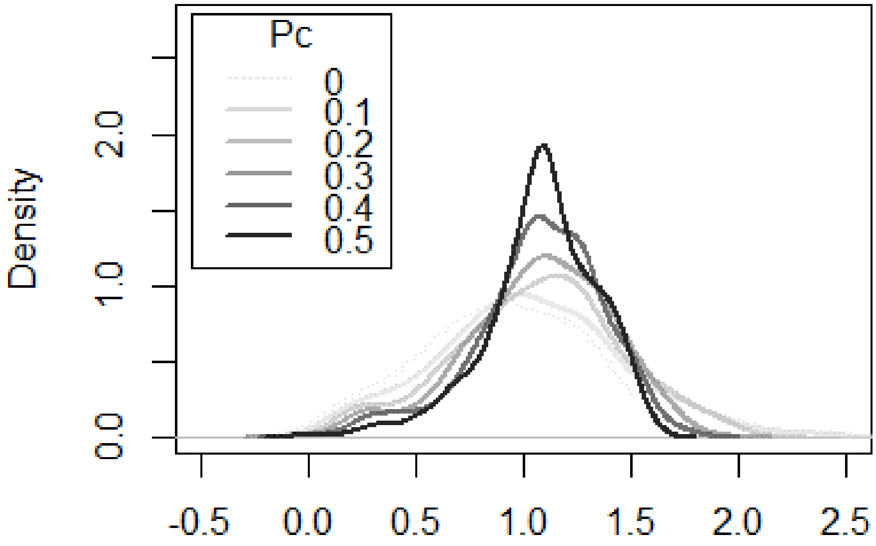


Figure 16. Density plot of O/E for calibration sites iterating probability of capture from near zero (include all taxa) to 0.5 (only include taxa found in >50% of samples for a given group). Greater *Pc* increased the model domain over a wider range of taxa, but increased the dispersion of O/E scores, even amongst least-disturbed calibration sites.

CHAPTER 3

RESULTS

Multivariate Model of Georgia Fish Assemblages

The clustering of fish communities was cut at six groups based on visual examination of the dendrogram, after observing the concordance of the groups with basin and ecoregional boundaries, and in order to ensure that adequate data was available to train the model for each group (Figure 17, Figure 18). One notable exception to this concordance is the boundary between the Southern Coastal Plain and the Southeastern Plains Ecoregions, where the grouping correlated poorly, regardless of the number of clusters selected (Figure 19). The six groups, derived from the fish communities observed at least impacted sites, represented theoretical community types of stream fishes in Georgia and reflected basin and ecoregional biogeographic influences: the Piedmont Ecoregion, Southeastern Plains Ecoregion, Coosa-Tallapoosa Basins, Tennessee Basin in the Ridge and Valley Ecoregion, the Tennessee Basin in the Blue Ridge Ecoregion, and the Southern Coastal Plain Ecoregion. The names assigned to the groups are generalizations reflecting the locations where the majority of the samples were located. The pruning of the tree at this level resulted in theoretical fish communities groups that most intuitively aligned with these biogeographic regions (Figure 20).

The final RF model included twenty-one predictor variables (Table 6). The most predictive variable was mean annual temperature, followed by elevation and geographic coordinates (Figure 21). Variables were retained to account for drainages (Coosa and Tennessee

Basins, and Atlantic and Gulf Slopes), ecoregions (South Eastern Plain, Piedmont, Blue Ridge Mountains, and Southern Coastal Plain), and temporal variation (Date, Year, and Day of Year). Physical site characteristics in the final model included slope, reach length, stream flow, and velocity. The measure of upstream drainage network length was retained, but stream order was not. The remaining basins and ecoregional candidate variables were dropped from the model.

The responses of the community groups to some of the environmental variables were intuitive and reflected in the partial dependency plots. For example, membership in the groups above the Fall Line was more likely to be predicted by the model for sites at higher elevations, whereas membership in groups below the Fall Line were more likely to be predicted for sites at lower elevations (Figure 22). Other modelled relationships exhibited in the partial dependency plots are difficult to interpret ecologically. For example, the response of the groups to drainage basin area may be indicative of overfitting of the data or the limited range of stream sizes used to calibrate the model.

The final RF model’s OOB error was 6.62%, making 0 to 7% classification errors for the Piedmont Ecoregion, Coosa-Tallapoosa and Tennessee Basins (including both the Ridge and Valley and Blue Ridge Ecoregions), and approximately a 12% and 21% error for the Southeastern Plains and Southern Coastal Plain Ecoregions (Table 7). The predictive model had an observed/expected (O/E) richness ratio of 1.09 for calibration (least impacted sites), indicating a slight bias in the model to under predict species richness (Figure 23), which was also evident in histogram, density, and Q-Q plots (Figure 24). The uncertainty in model predictions for Southeastern Plains, Southern Coastal Plains, and Tennessee Basin in the Blue Ridge Ecoregion translated into wider uncertainty in O/E scores for reference sites in those locations (Figure 25).

Model performance was also measured by the standard deviation (σ) of observed to expected richness, with a smaller σ indicating that the model accounted for a larger proportion of the variability associated with natural environmental gradients. Replicate sample σ is expected to represent a theoretical lower bound for variance in this type of model. In contrast, a null model which assumes fixed occurrence probabilities for individual taxa across reference sites (explaining none of the variability from natural gradients) is expected to represent the upper bound of model σ (Van Sickle et al., 2005). The model had a σ of 0.262 which is bracketed by a σ of 0.251 from replicate samples and a σ of 0.554 from the null model. This indicates that the model was able to explain a substantial amount of the variability in species richness that was due to natural environmental variation. The O/E scores were also visually examined, via scatterplot with LOWESS (locally weighted scatterplot smoothing) line, to determine the degree to which they accounted for natural gradients in the model (Figure 26). Trends were not evident for most variables included in the model, although there was an upward trend in O/E scores with increasing stream size (stream flow, order, and upstream arbolate stream kilometers). No spatial patterns were evident in the highest and lowest O/E scores for the least impacted sites used in model calibration (Figure 27). The model was also tested against the subset of validation data, which exhibited a slight bias in mean O/E scores (1.024), and a slightly larger σ (0.331) (Figure 28).

Discrimination of the O/E scores between least impacted and stressed sites was weak across the range of *Pc* (Figure 29). Greater *Pc* decreased σ in O/E scores up to about a 0.3 level, but a 0.5 threshold was ultimately selected for its slightly greater discrimination (the stressed site median O/E below the lower quartile O/E of least impacted sites) and for greater consistency with numerous O/E models and reviews (Hawkins, Norris, Hogue, et al., 2000; Moss, Wright, Furse, & Clarke, 1999; Van Sickle et al., 2007; J. F. Wright et al., 2000). Standard deviation of O/E scores decreased from 0.484 to 0.302 when iterating *Pc* from near 0 to 0.6. Above this *Pc* level, model results became unstable since no taxa were predicted to occur at or above this frequency in a number of groups.

The effect of varying the *Pc* value in the model was also examined through simulations with a theoretical data set. Mean O/E scores for simulated samples from groups of 3, 6, 9 or 18 taxa with a mean probability of occurrence of 0.75 (least impacted) or 0.6 (stressed) sites averaged 1.0 and 0.8 regardless of level of taxa richness. However, standard deviation of scores decreased with increasing taxa richness (e.g., from 0.34 to 0.13 at least impacted sites). Ranges of O/E scores overlapped extensively between least impacted and stressed sites at all levels of taxa richness, although with 18 taxa the range of scores at stressed sites at least did not encompass the full range of possible scores (Figure 30). Lower *Pc* and lower expected richness both increased variability in scores even at least impacted sites.

Further diagnostics were conducted for the ACF-PDM subregion in Georgia. The revised O/E scores in the ACF-PDM demonstrated greater discrimination than statewide. However, the model was not able to discriminate as well (Figure 31), nor did it exhibit a strong relationship with (Figure 32) the state’s draft multimetric IBI index. The performance of the O/E model statewide was driven by the range in performance in each of the various basins and ecoregions (Figure 33, Figure 34). For example, variability was lowest in the Ridge and Valley Ecoregion, but widest in the Southern Coastal Plain Ecoregion. Additionally, the distributions of O/E scores were examined to evaluate whether they were bimodal (Figure 35). If so, extremely high and low O/E scores (i.e., sites enriched and depleted in species) may not have been evident when viewing average scores, but the O/E scores did not appear to be distributed in this manner.

Due to the wide variability in O/E scores at least impacted sites, and their contribution of relatively few samples to the model, several basins and ecoregions were removed from the results. The Saint Marys, Apalachicola, and Aucilla basins had too little data to be evaluated and were removed. The Ochlockonee, Suwanee, Tallapoosa, Ocmulgee basins and Southern Coastal Plain, Southeastern Plains Ecoregion ecoregions did not have a positive relationship or multiple linear regression coefficient of determination greater than 0.3 and were also removed (Figure 36, Figure 37). Removing these poorly performing basins and ecoregions improved the relationship between observed and expected scores at all validation sites from an r^2 of 0.57 to 0.67 (Figure 38).

Functional Group Responses to Disturbance

The analysis of taxa O/E scores indicated increases or decreases in some taxa at stressed sites. Of the five families with 10 or more taxa, the taxa O/E of Centrarchidae, on average, just slightly exceeded one (1.05), indicating that these taxa tend to occur about as often as predicted by the model at stressed sites. Ictaluridae and Catostomidae had lower average O/E scores at stressed sites (0.87 and 0.85 respectively), as did Cyprinidae (0.78) and Percidae (0.63) (Figure 39). Species defined as feeding generalists increased at stressed sites (1.15 average taxa O/E) while invertivores decreased (0.72) (Figure 40). Most striking were the increases in habitat generalists (1.22), versus the decreases in fluvial and swamp specialists (0.67 and 0.54, respectively) (Figure 41). The *a priori* designations of tolerance responded as expected with mean O/E taxa scores increasing notably for high tolerance taxa (1.33), decreasing slightly for medium (0.76), and decreasing dramatically for low tolerance taxa (0.53). The responses of the most extreme “increasers” and “decreasers” typified trends at the family level, as the most extreme decreaser taxon, *Etheostoma jessiae* (0.08), is an invertivorous fluvial specialist, while the most extreme increaser taxon, *Micropterus punctulatus* (2.60) is a predatory habitat generalist (See Appendix D, Table D.1). The only parasitic taxon analyzed, *Ichthyomyzon castaneus,* was a slight decreaser (0.74). Of Georgia's protected fish species in the model, some had marginal changes such as *Cyprinella xaenura* (1.05) and *Etheostoma scotti* (0.91), and others were decreasers, such as *Etheostoma duryi* (0.70) and *Percina sciera* (0.70). However, many of the protected species demonstrated extreme responses as out of the 136 taxa analyzed, they ranked as the 21st greatest decreaser, *Percina smithvanizi* (0.32), 17th greatest decreaser, *Etheostoma tallapoosae* (0.28), 6th greatest decreaser, *Notropis hypsilepis* (0.19), and 2nd greatest decreaser, *Moxostoma sp. Apalachicola redhorse* (0.14).

Table 6. Importance values for the 21 variables used in the final Random Forest model. Ranks (1-21) are shown for each community cluster (1 through 6) with the highest ranks indicating the greatest importance for assigning samples to that cluster. Gini measures homogeneity and ranges from zero (homogeneous) to one (heterogeneous), so decreases in Gini reflect increases in purity. Mean decreases in Gini from each parent to child node are summed over all trees, and normalized for each variable. Mean decrease in accuracy is derived from permuting values from a single predictor variable for the out-of-bag observations, running the permuted data down the forest trees, and averaging mean decrease in accuracy between the permuted and unpermuted variable across all observations, normalized by the standard deviation of the differences.

Variable	🟥 **1**	🟦 **2**	🟩 **3**	🟪 **4**	🟧 **5**	🟫 **6**	**MeanDecrease Accuracy**	**MeanDecrease Gini**
Mean_Ann_Temp_C	14	13	19	15	21	14	26	31
Elevation_m	9	12	16	14	19	13	21	25
Lat	16	13	14	15	13	13	21	21
Long	13	5	18	15	7	11	22	18
Eco_SEP	15	20	12	9	10	10	20	16
Bas_Coo	11	8	17	11	0	6	17	10
Atlantic_Slope	14	5	15	10	0	-1	16	9
Eco_PDM	16	9	11	10	10	9	17	10
Gulf_Slope	13	3	13	8	1	-2	14	8
Date	10	6	7	8	7	5	14	7
Slope	4	4	6	6	10	1	10	8
Bas_Ten	6	5	13	20	6	3	18	8
Eco_BRM	8	6	4	6	12	4	11	5
Year	6	3	5	7	6	5	11	4
Eco_SCP	4	-1	3	2	2	14	12	4
DayOfYear	6	-4	6	5	1	9	10	5
ArbolateSum_m	5	3	6	0	5	4	10	4
Velocity_mps	7	2	7	3	4	1	10	4
DBA_km2	7	3	2	0	4	-1	8	4
Flow_m3ps	4	3	4	1	3	2	7	3
RL_m	5	5	5	2	2	-5	6	4

Table 7. Random Forest model classification confusion matrix. Colors correspond to the mapping of the six theoretical Georgia stream fish community types. Final column reflects classification error. Overall out-of-bag error (a cross validation measure of model performance) was 6.62%.

			1	2	3	4	5	6	Error
Piedmont Ecoregion	🟥	1	54	0	2	0	0	0	3.57%
Southeastern Plains Ecoregion	🟦	2	4	73	1	0	0	5	12.05%
Coosa-Tallapoosa Basins	🟩	3	0	0	71	0	0	0	0.00%
Tennessee Basin x Ridge and Valley Ecoregion	🟪	4	0	0	0	14	0	0	0.00%
Tennessee Basin x Blue Ridge Ecoregion	🟧	5	0	0	2	0	27	0	6.90%
Southern Coastal Plain Ecoregion	🟫	6	0	4	0	0	0	15	21.05%

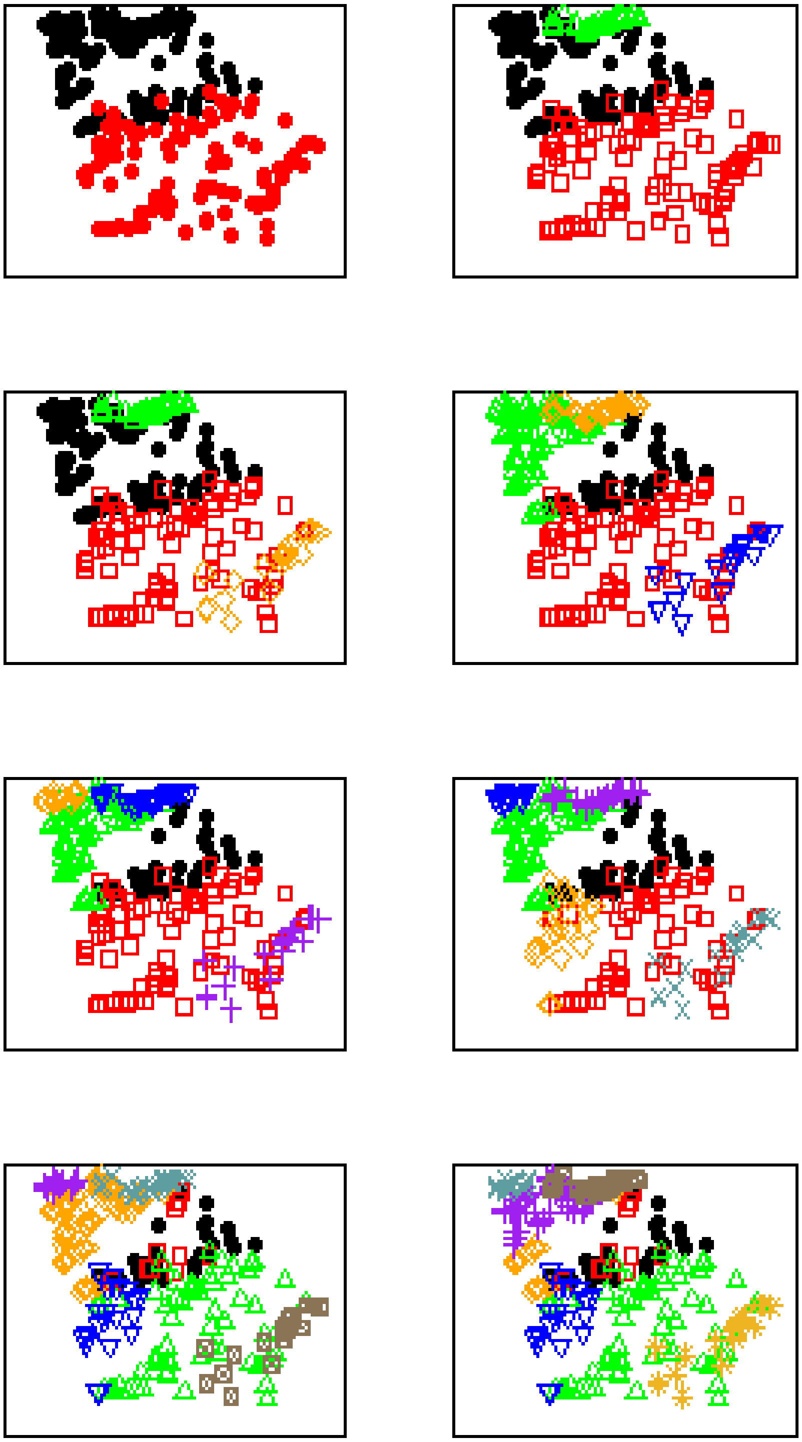


Figure 17. Alternate clustering of theoretical stream fish communities in Georgia, based on selection of 2 (upper left) to 9 (lower right) clusters of sample data, as indicated by differing colors.

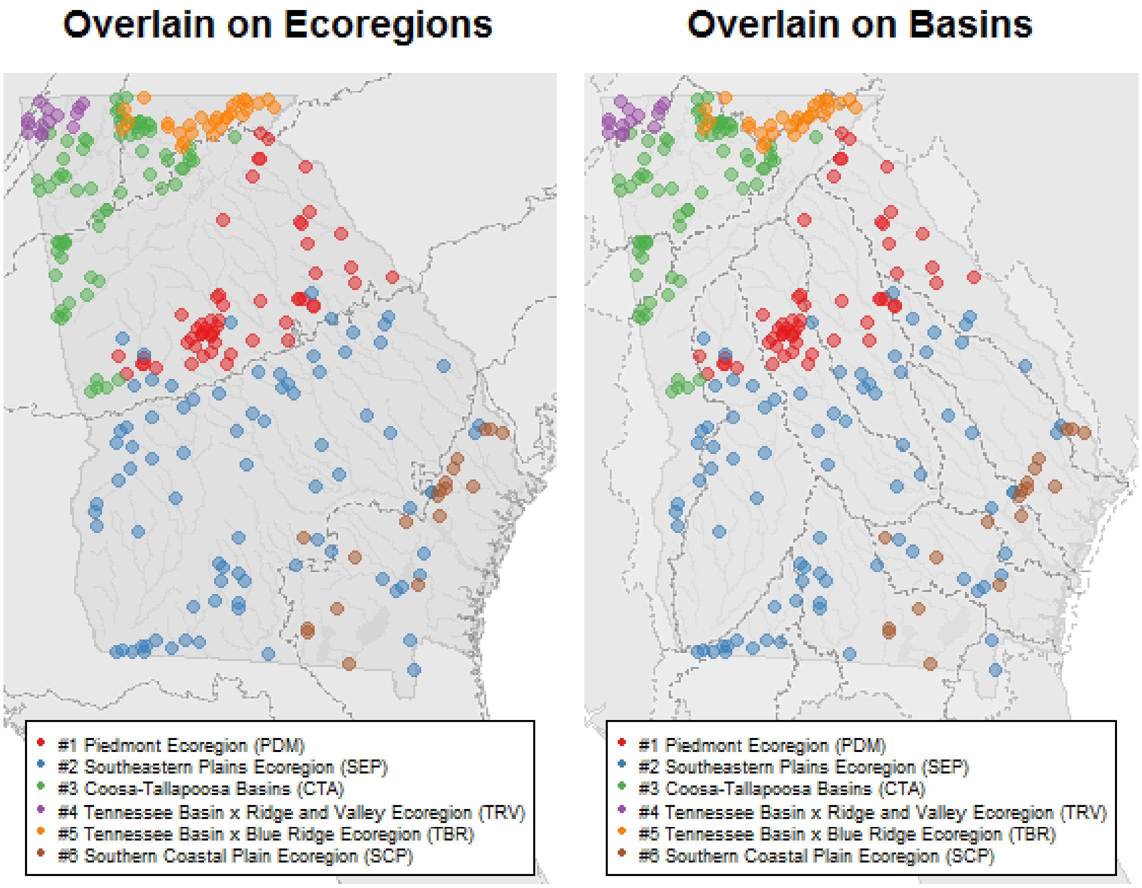


Figure 18. Georgia stream fish groups overlain on ecoregion (left) and basin (right), based on observations at least impacted sites. The Piedmont was not well represented among least impacted sites due to the extensive urbanization from the greater metro Atlanta area. Names assigned to the groups are generalizations reflecting the locations where the majority of the samples were located.

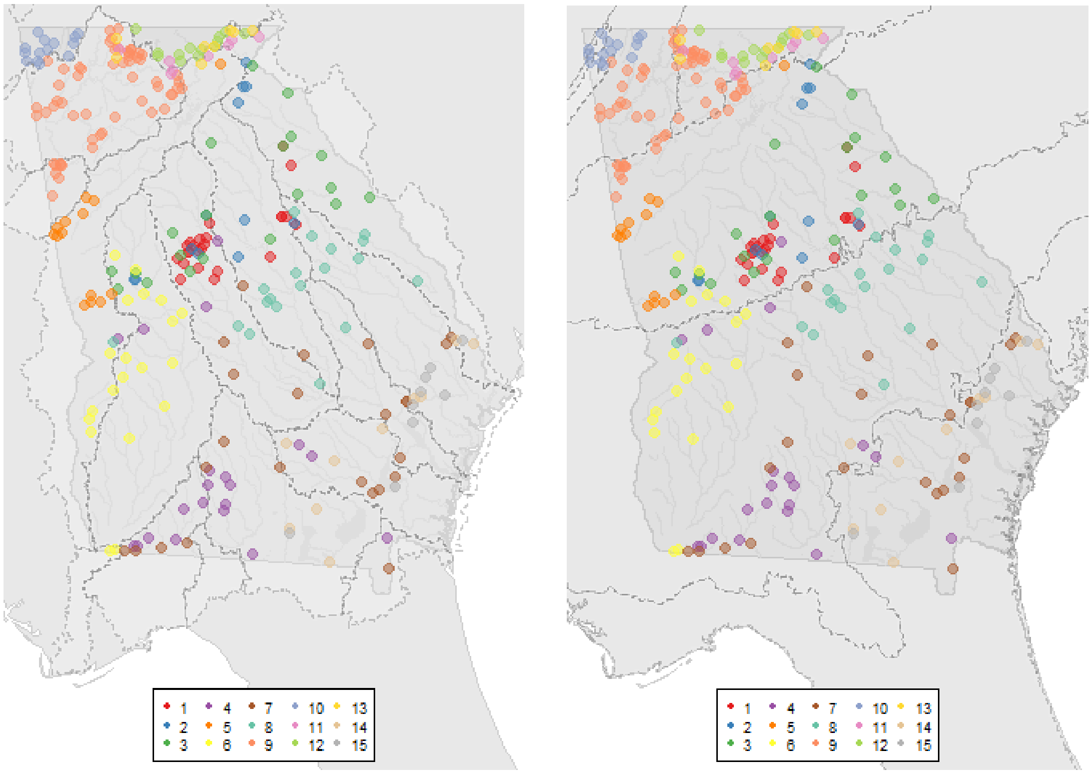


Figure 19. An alternate division of theoretical stream fish communities in Georgia with 15 groups, overlain on basins (left) and ecoregions (right), to demonstrate the effects of dividing the groups at a higher level.

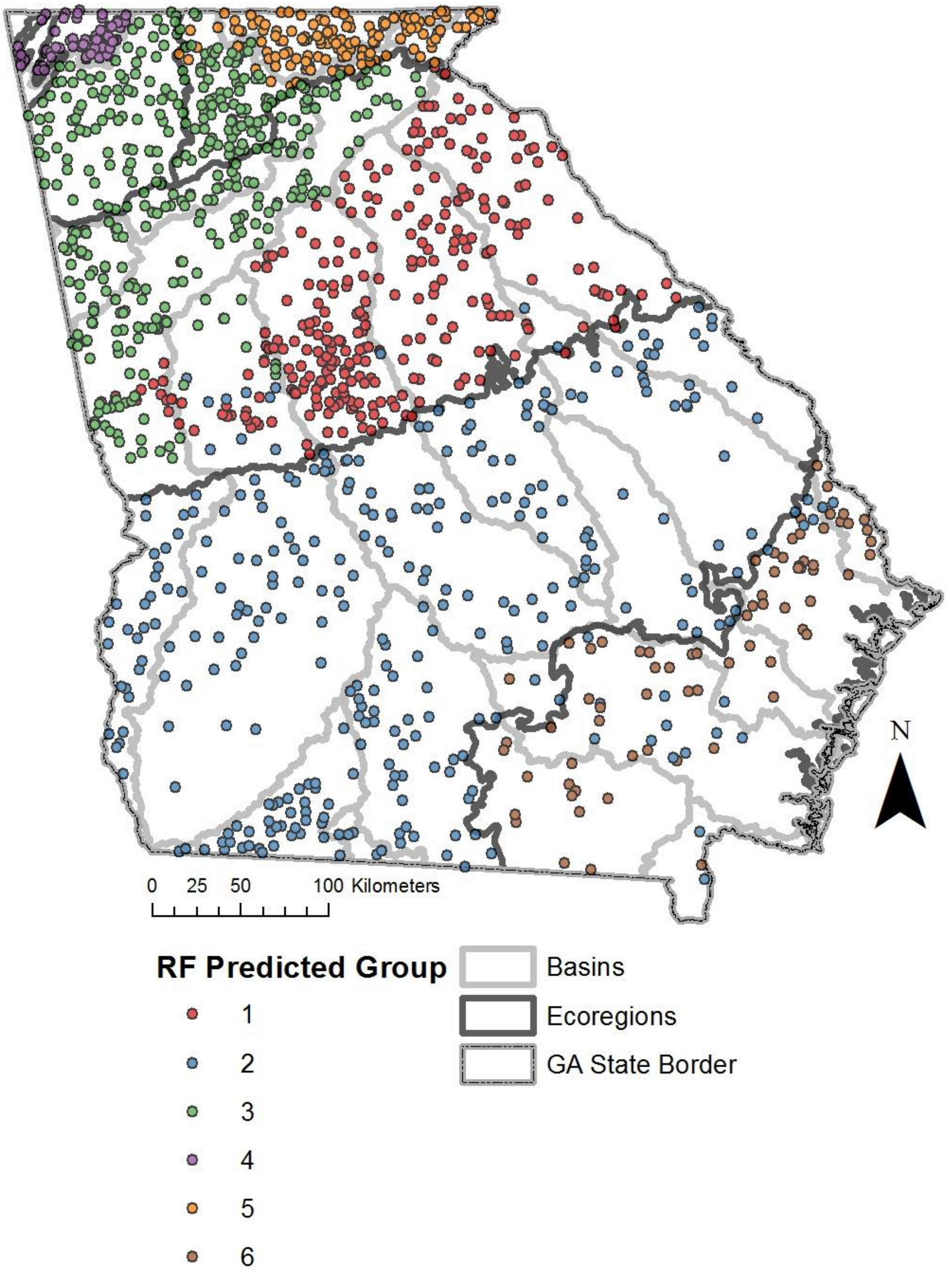


Figure 20. There was strong concordance between the Random Forest predicted group membership and Georgia’s biogeographic regions (basin and ecoregions), with the notable except of samples in the Southern Coastal Plain.

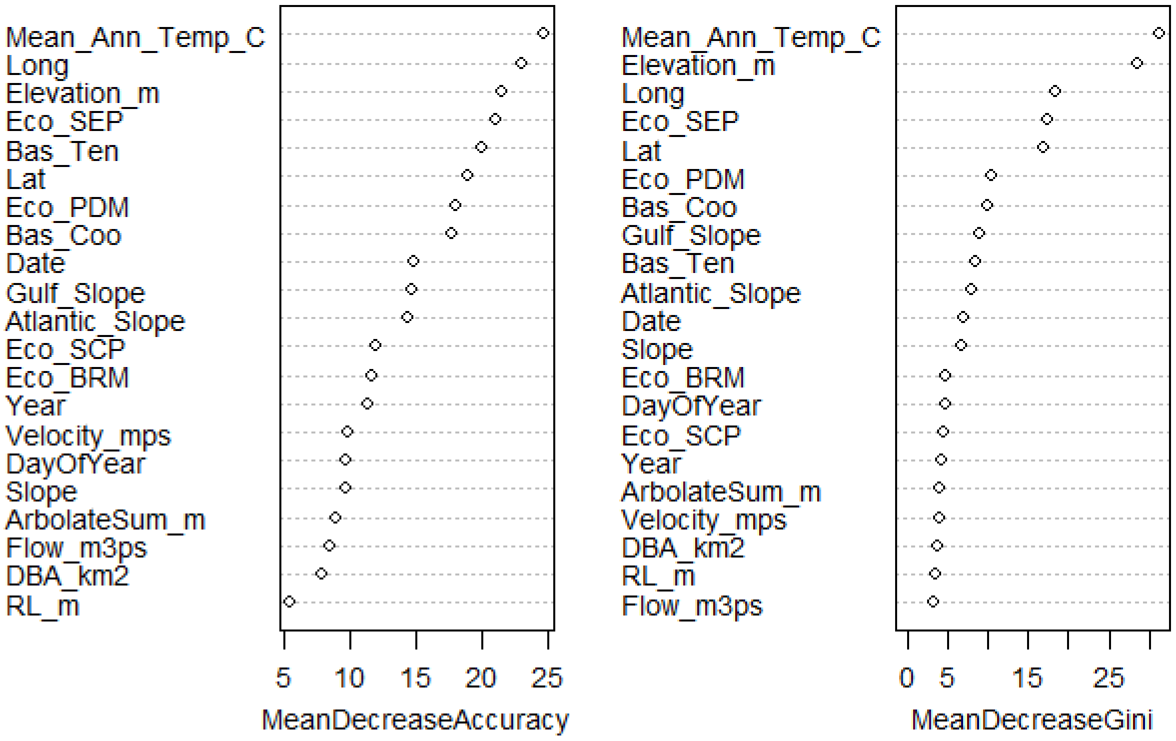


Figure 21. Mean decrease in accuracy and Gini as measures of variable importance for classification in the final Random Forest model. 500 trees were included in the final forest and 4 variables were randomly selected as classifiers at each split.

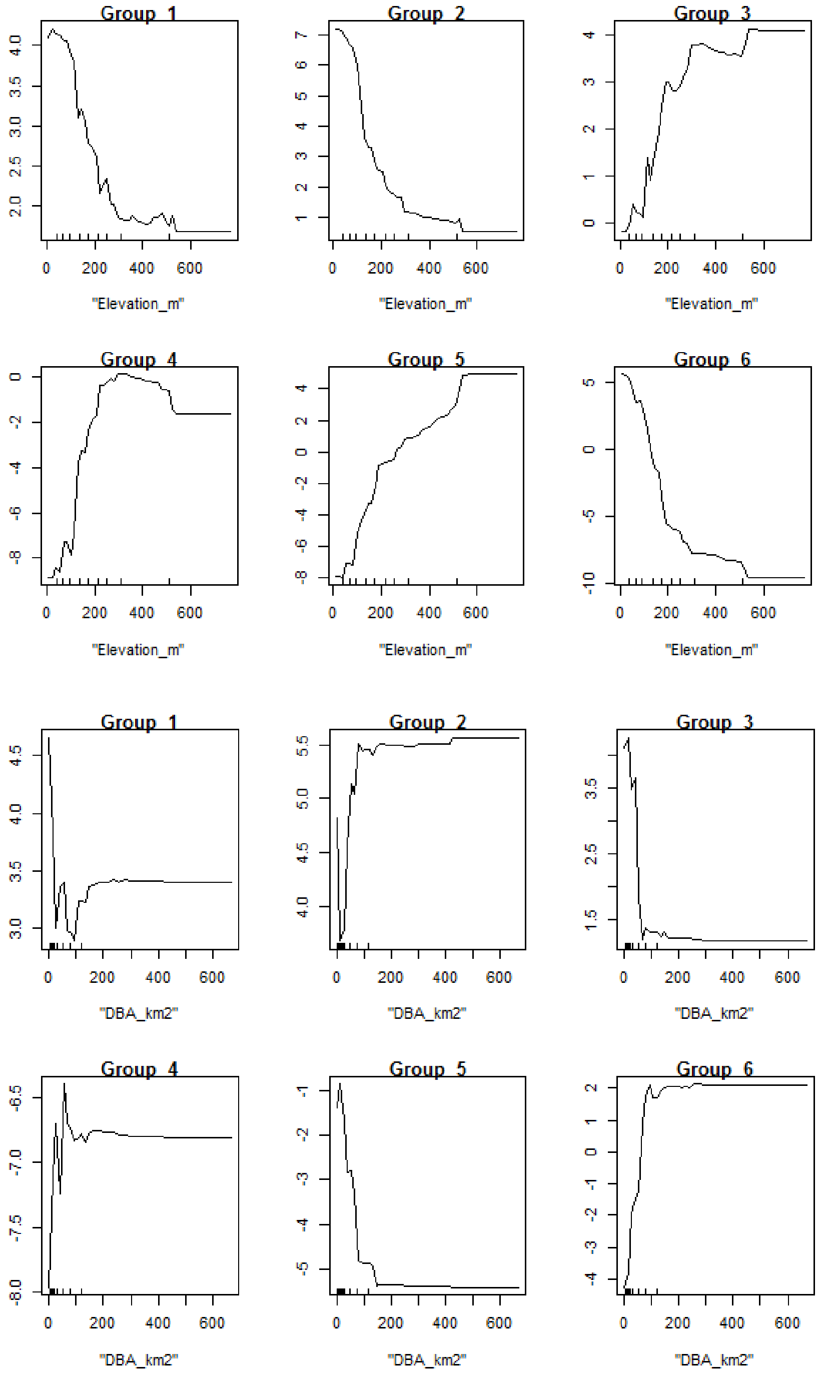


Figure 22. Partial dependence plot for drainage basin area (DBA) in square kilometers (top) and elevation in meters (bottom) where Y-axis is the mean value of logit(p) and p is the predicted probability of being in a group and the mean is taken over all other combinations of the other predictors. Community types displayed are Group 1 (Piedmont Ecoregion), Group 2 (Southeastern Plains Ecoregion), Group 3 (Coosa-Tallapoosa Basins), Group 4 (Tennessee Basin x Ridge and Valley Ecoregion), Group 5 (Tennessee Basin x Blue Ridge Ecoregion), and Group 6 (Southern Coastal Plain Ecoregion).

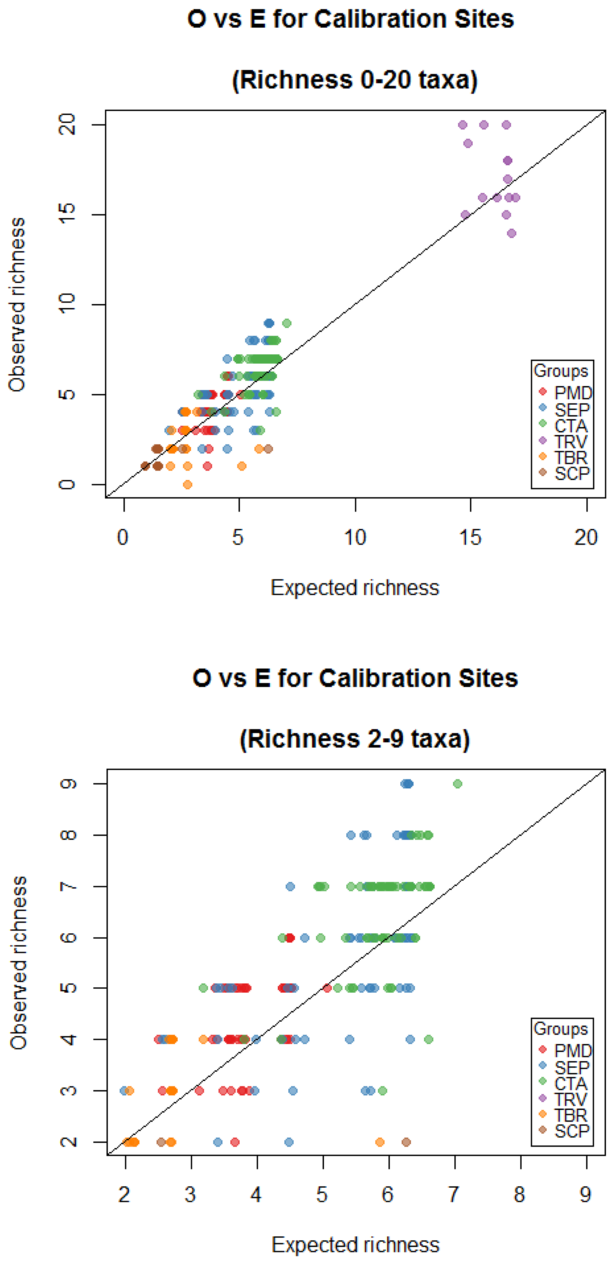
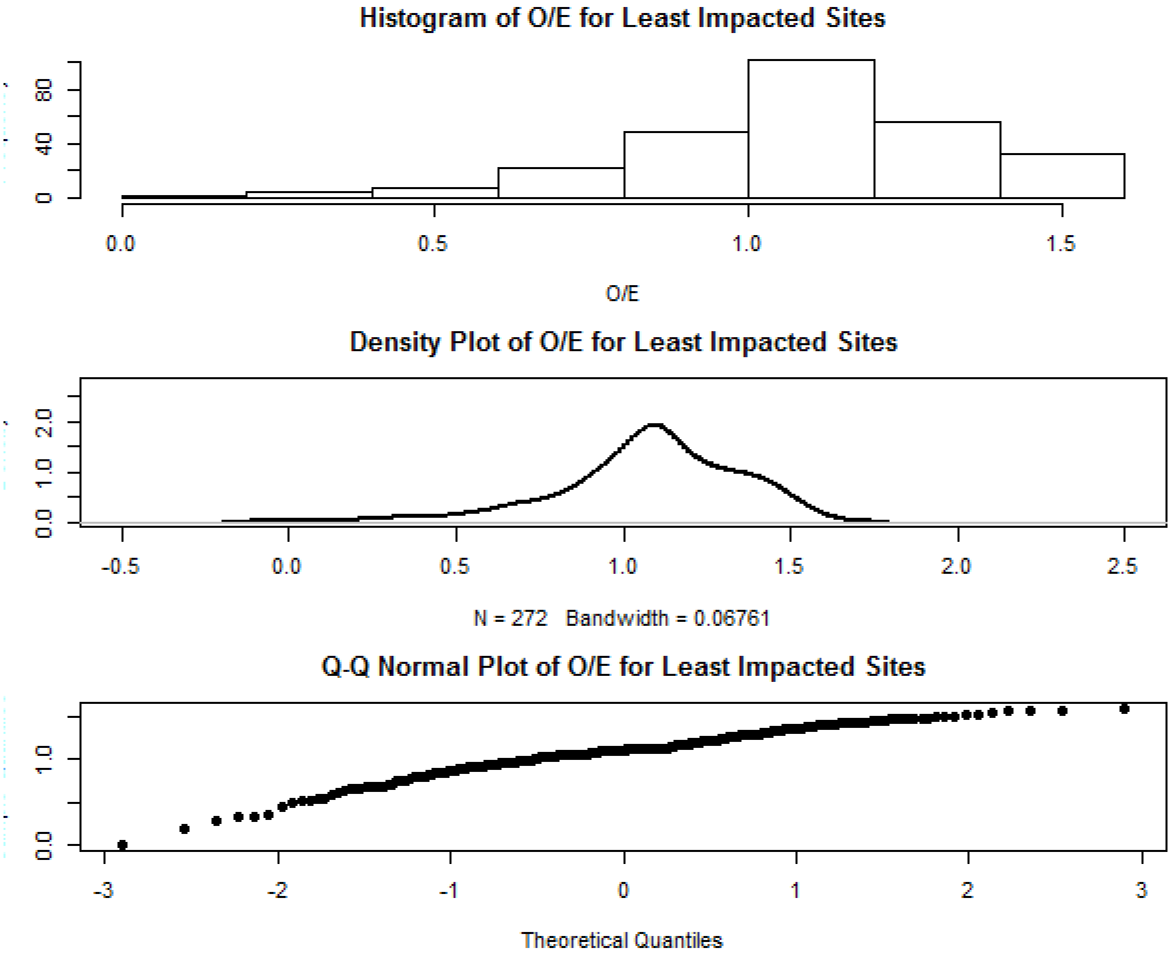


Figure 23. Observed versus expected species richness for calibration sites. Choice of axes based on Piñeiro, Perelman, Guerschman, & Paruelo, 2008. Top figure displays across the range of modelled richness (0-20 taxa), while the bottom figure displays a subset (2-9 taxa). Points are color coded by theoretical Georgia stream fish communities: PDM (Piedmont Ecoregion), SEP (Southeastern Plains Ecoregion), CTA (Coosa-Tallapoosa Basins), TRV (Tennessee Basin x Ridge and Valley Ecoregion), TBR (Tennessee Basin x Blue Ridge Ecoregion), and SCP (Southern Coastal Plain Ecoregion). The predictive model had an observed/expected (O/E) richness ratio of 1.085, indicating a slight bias in the model to under predict species richness.


Histogram of O/E for Least Impacted Sites

Density Plot of O/E for Least Impacted Sites

Q-Q Normal Plot of O/E for Least Impacted Sites

Figure 24. Histogram, density and Q-Q plots of O/E scores for least impacted sites indicating a slight bias in the model to under predict species richness.

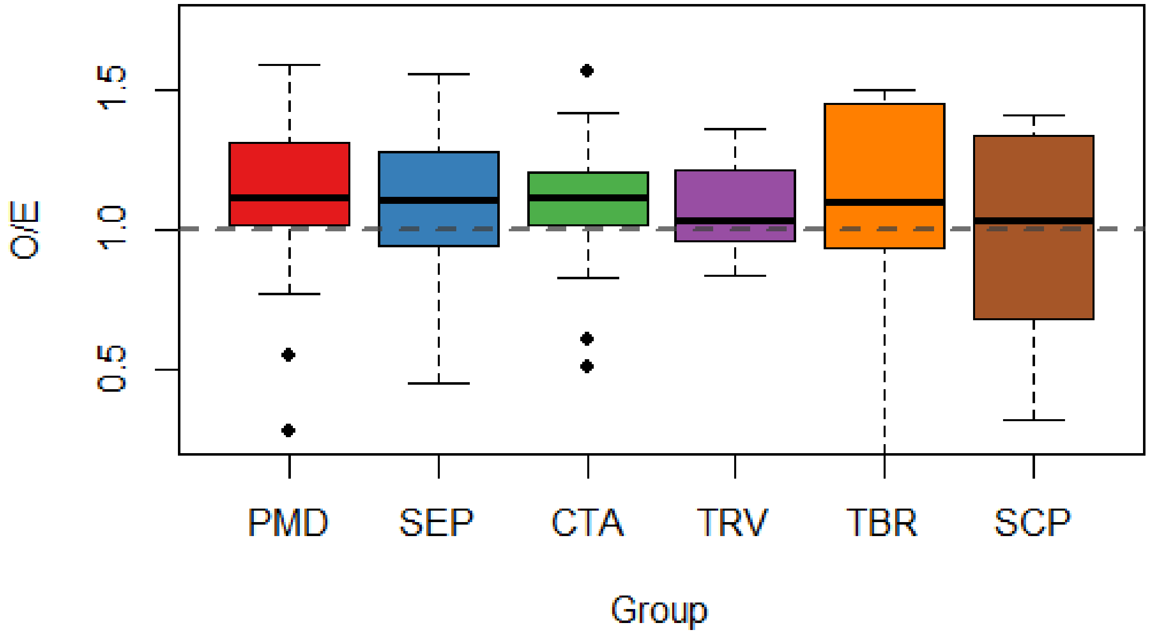


Figure 25. Median and range of O/E scores for Least Impacted Sites by community group: PDM (Piedmont Ecoregion), SEP (Southeastern Plains Ecoregion), CTA (Coosa-Tallapoosa Basins), TRV (Tennessee Basin x Ridge and Valley Ecoregion), TBR (Tennessee Basin x Blue Ridge Ecoregion), and SCP (Southern Coastal Plain Ecoregion). The widest range in O/E scores were observed among the Tennessee-Blue Ridge and Southern Coastal Plain groups.

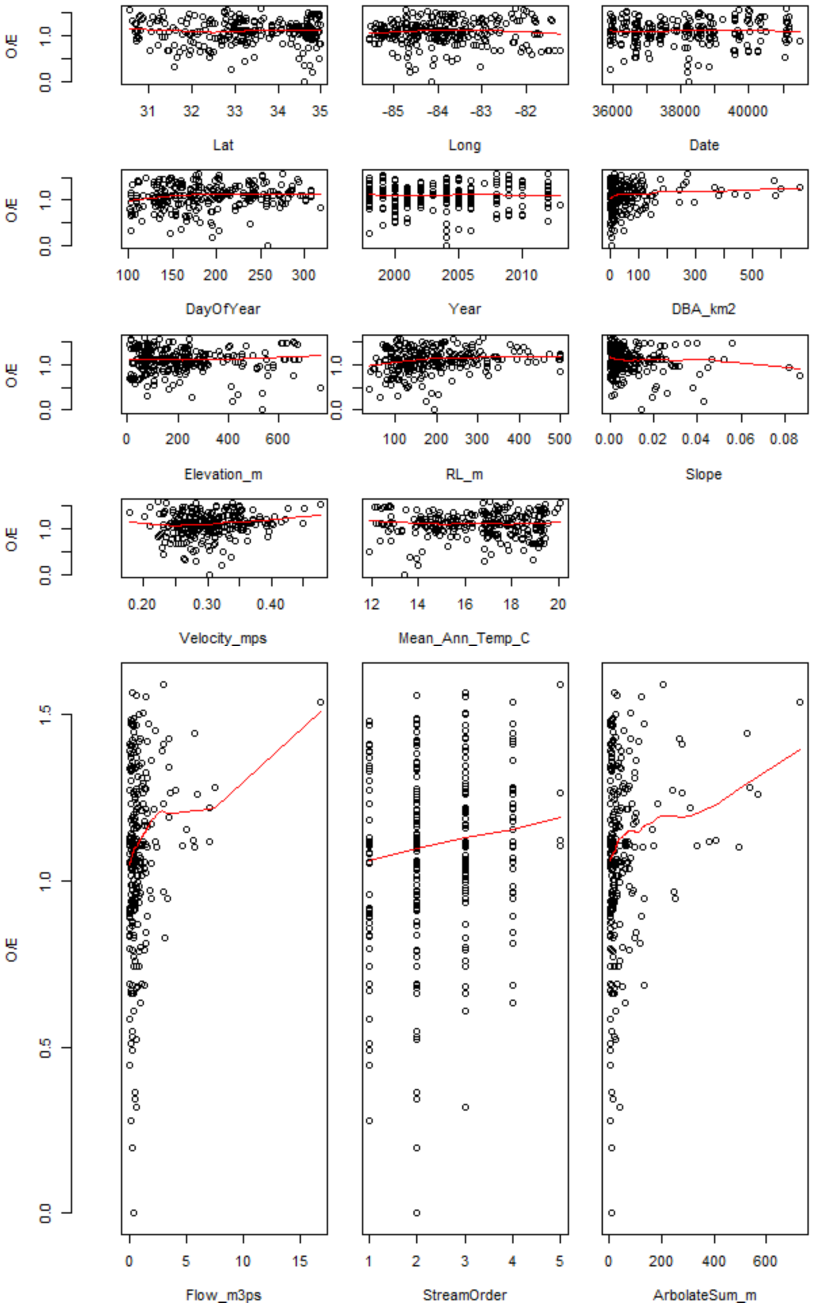


Figure 26. Variation in O/E scores for all sites across natural gradients displayed via scatterplot with LOWESS line. Top panel displays eleven variables without evident trends, bottom panel displays three variables with possible trends, although they appear to be influenced primarily by outlying points.

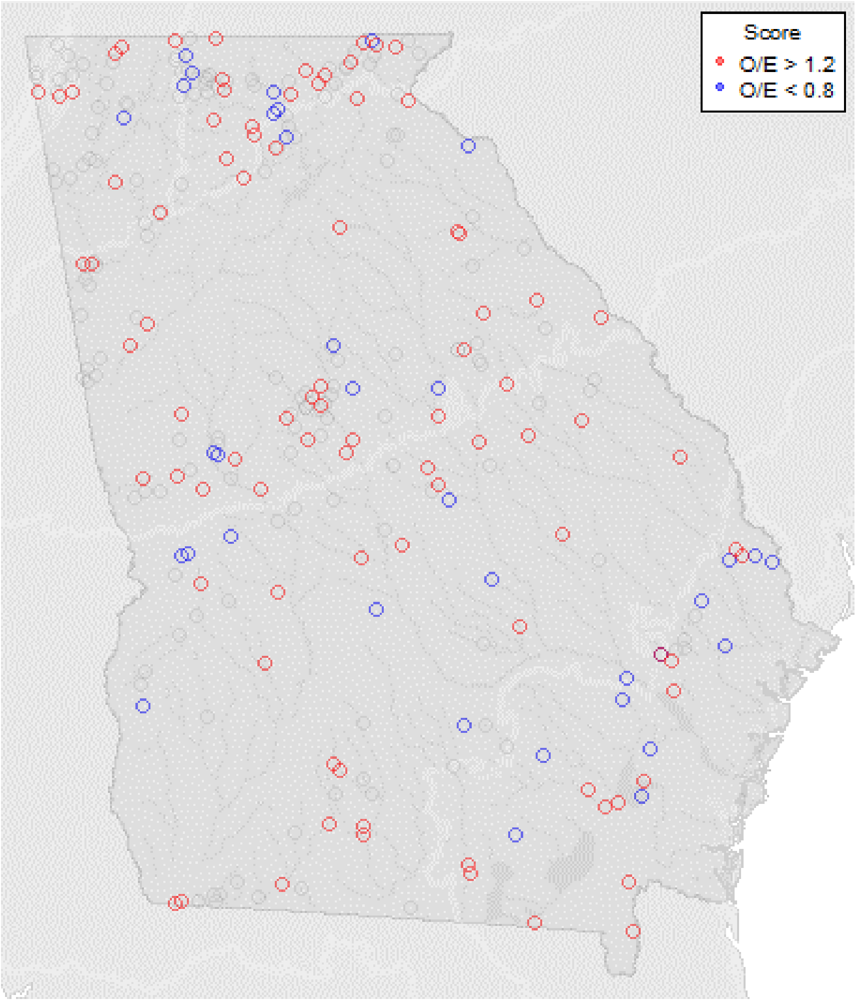


Figure 27. Most extreme O/E scores for least impacted reference sites are graphed spatially. No patterns were evident in the spatial distribution of the highest and lowest O/E scores.

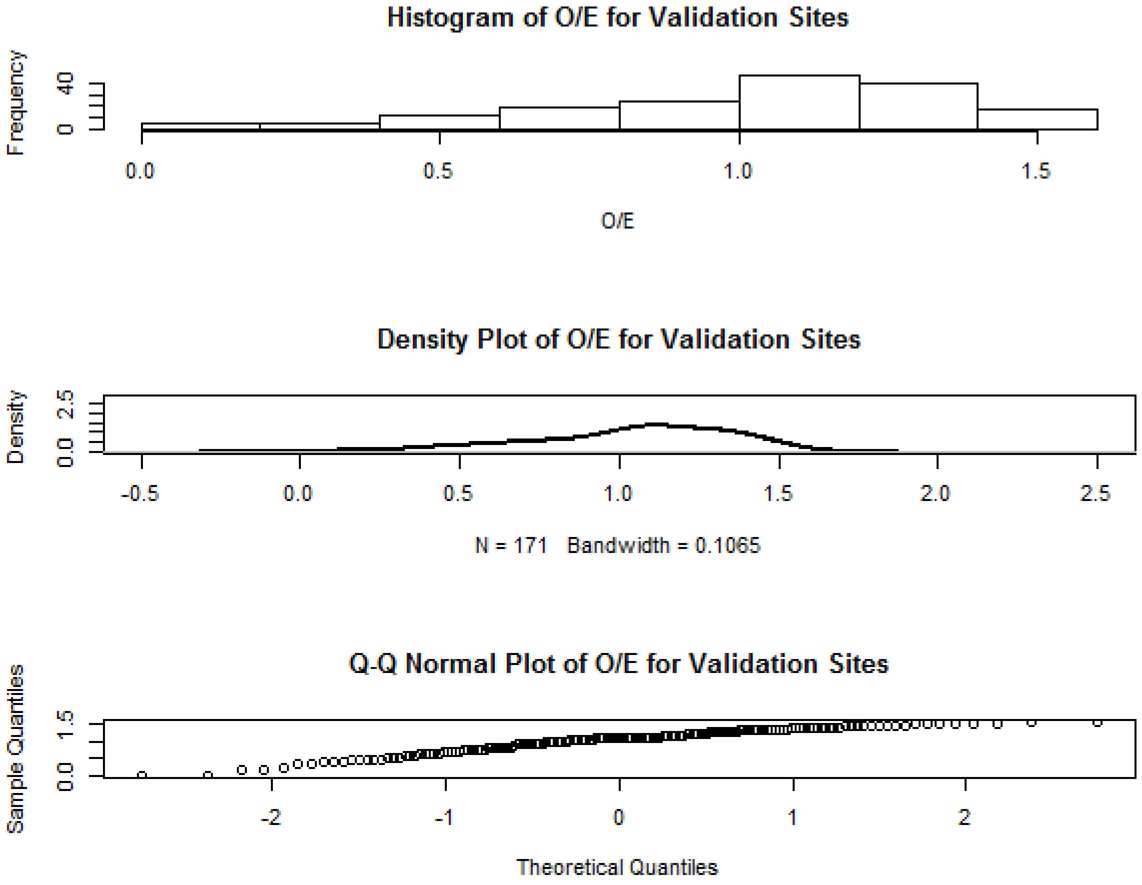


Figure 28. Histogram, density and Q-Q plots of O/E scores for validations sites indicating a bias in the model to under predict species richness.

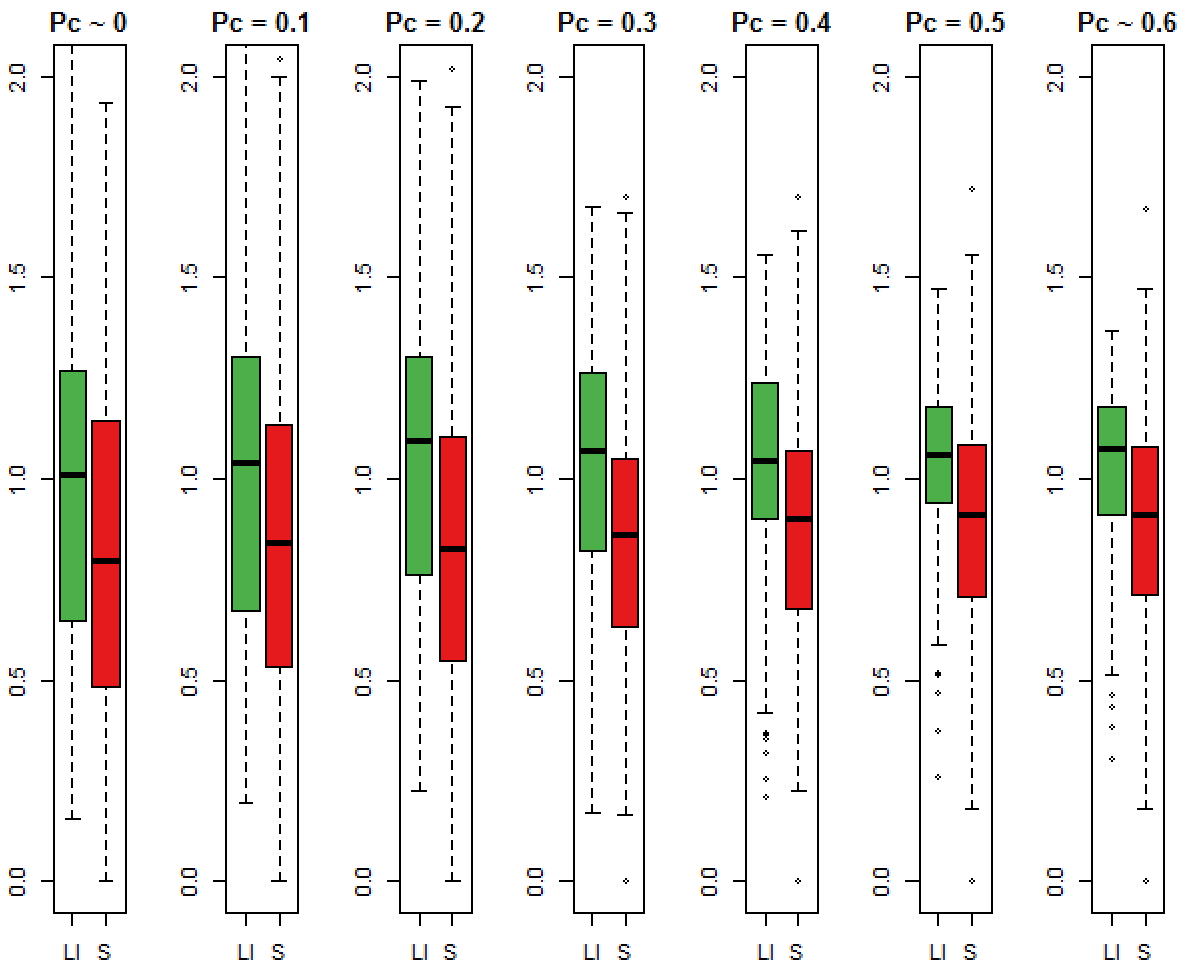


Figure 29. Discrimination of O/E scores for least impacted versus stressed sites, iterating the probability of capture (*Pc*) from near 0 to 0.6. Inclusion of rare taxa (as *Pc* approaches 0) increased the model domain over a wider range of fish diversity, but decreased the ability to differentiate between least impacted and stressed sites.

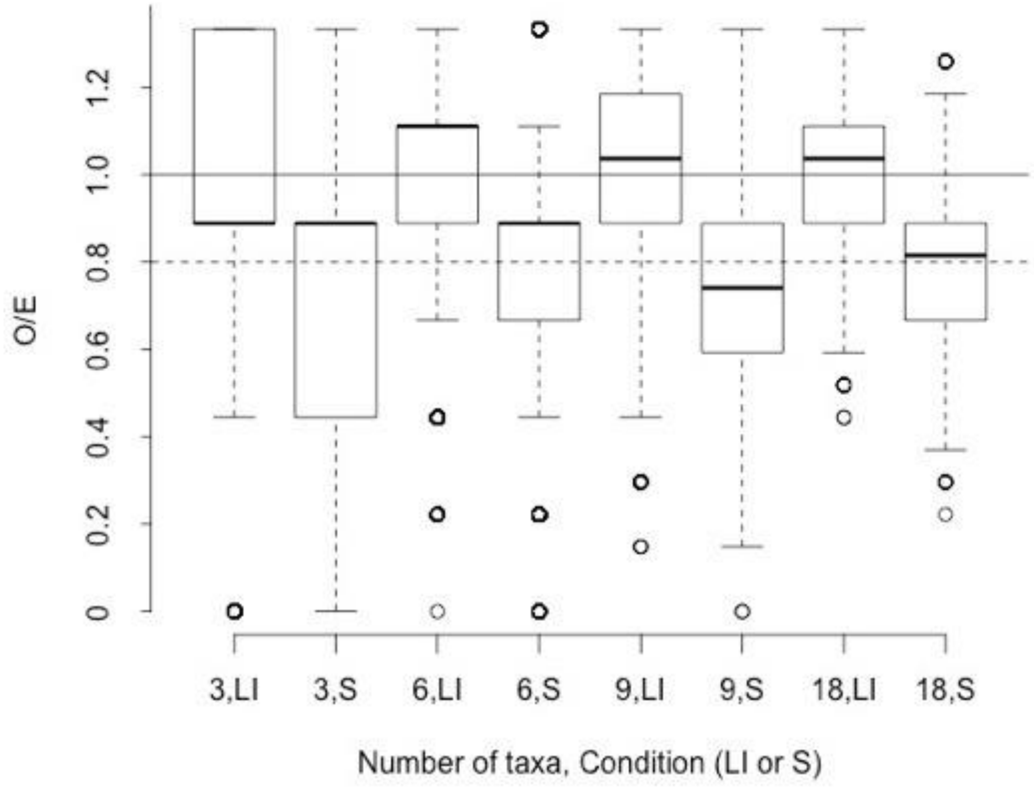


Figure 30. Simulated O/E scores for 10,000 samples taken from least impacted (mean species occurrence and capture, *Pc* = 0.75) and stressed (mean *Pc* =0.6) sites having 3, 6, 9, or 18 species included in analyses. Horizontal lines show mean scores for least impacted (solid) and stressed (dashed) sites.

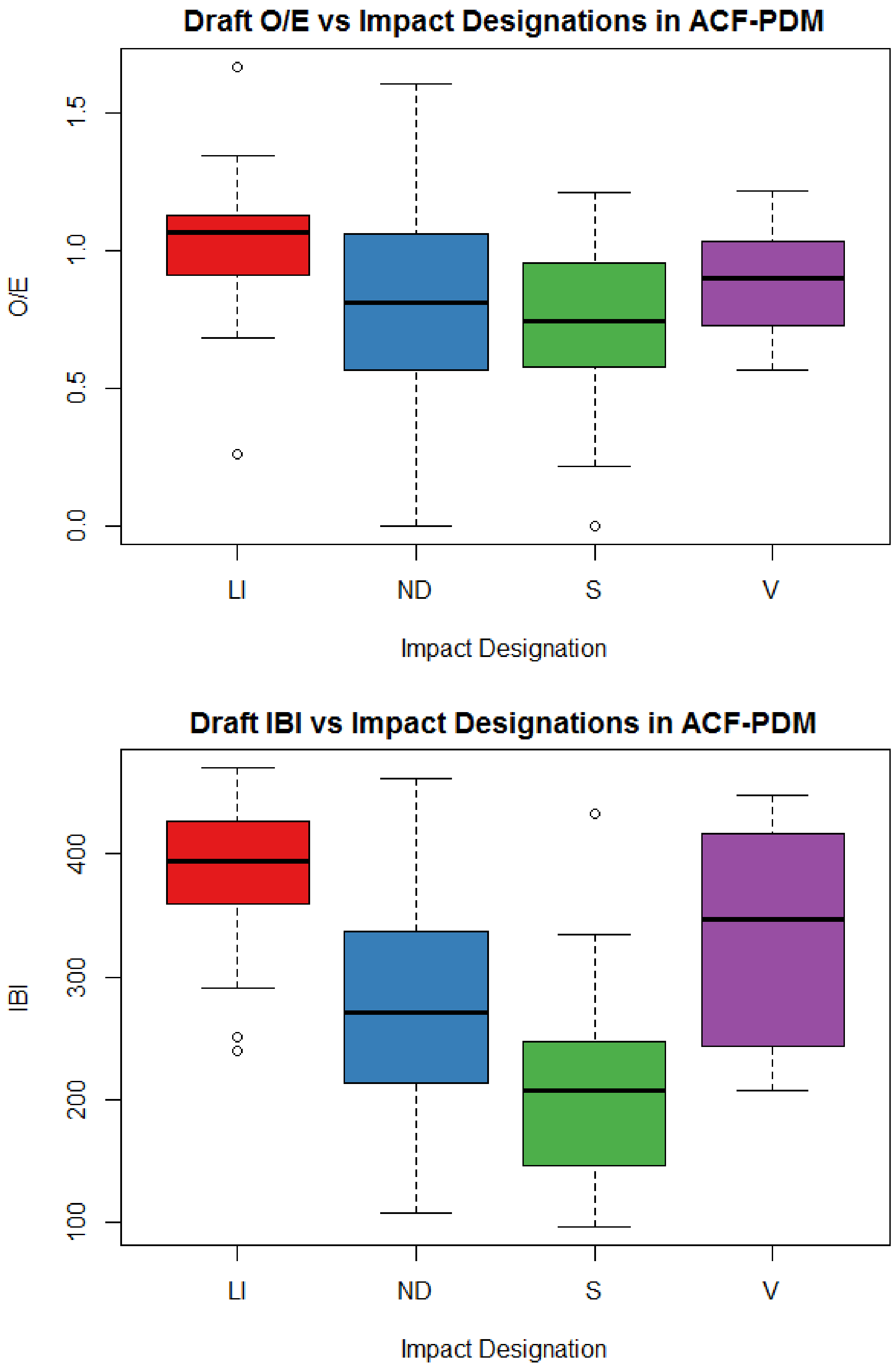


Figure 31. Discrimination of bioassessment scores by impact designation. The relative discrimination efficiency of the O/E approach (top) is compared with IBI discrimination for the Apalachicola-Chattahoochee-Flint Basin portion of the Piedmont Ecoregion (ACF-PDM) region.

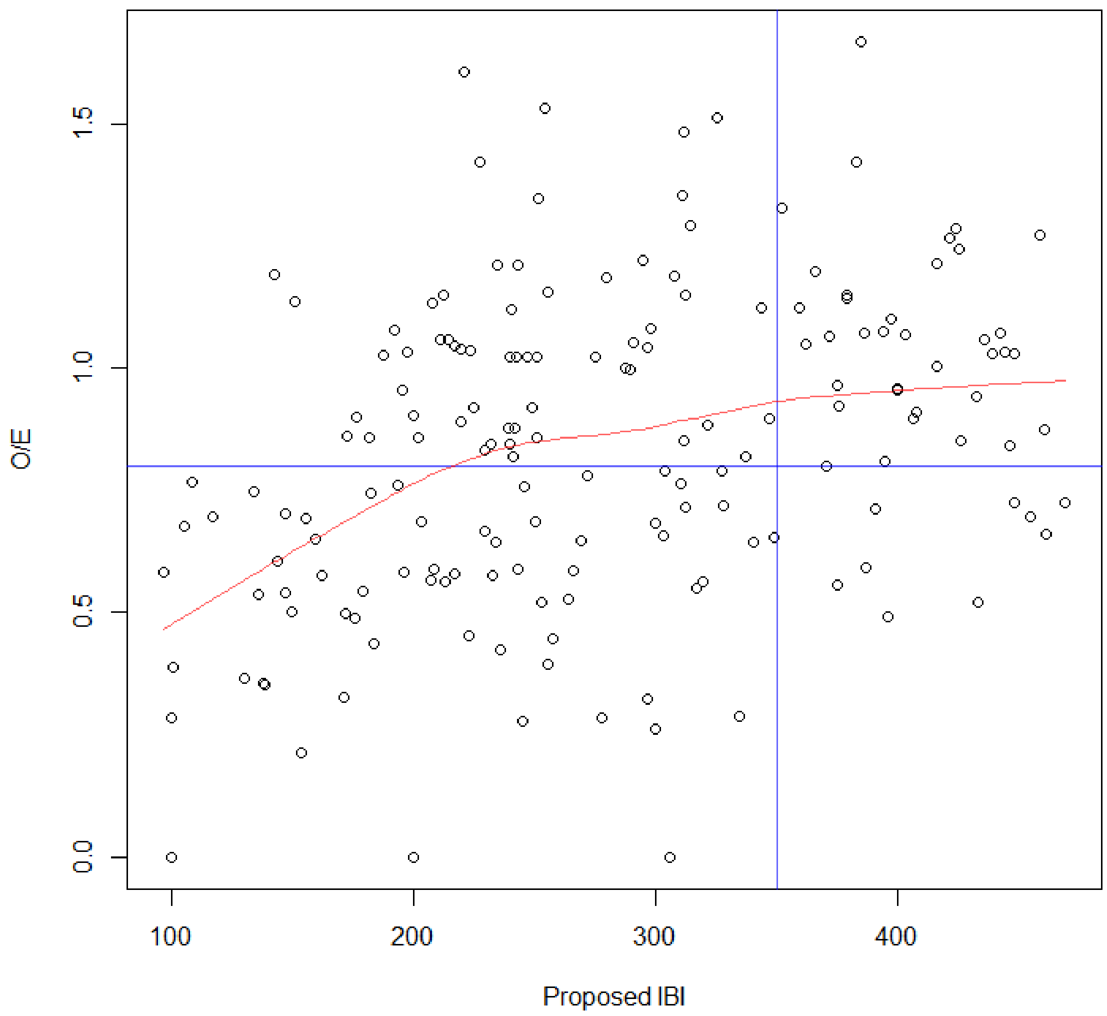


Figure 32. Draft revised O/E Scores and draft revised IBI scores for Apalachicola-Chattahoochee-Flint Basin portion of the Piedmont Ecoregion (ACF-PDM) region. Red line displays a locally weighted scatterplot smoothing (LOWESS) non-parametric regression (nearest neighbor) line. Blue lines represent examples of impairment thresholds: a horizontal line for O/E scores ≤0.8 and a vertical line for IBI scores ≤350.

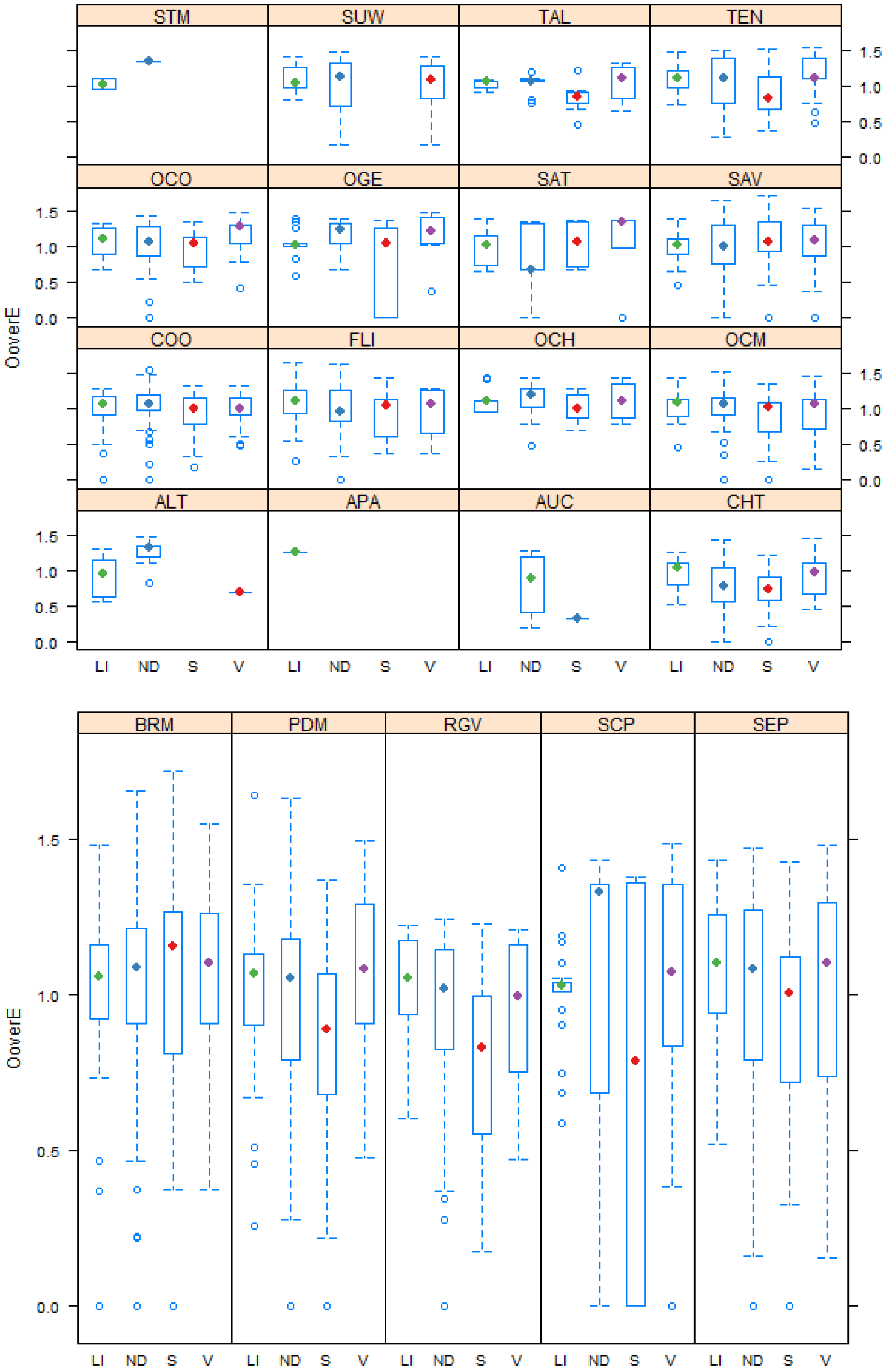


Figure 33. Variability in the range of O/E Scores for all sites by impact designation depended largely on the basin (top) and ecoregion (bottom). "LI" indicates least impacted sites, "ND" sites with no-designation, "S" stressed sites and "V" validation sites.

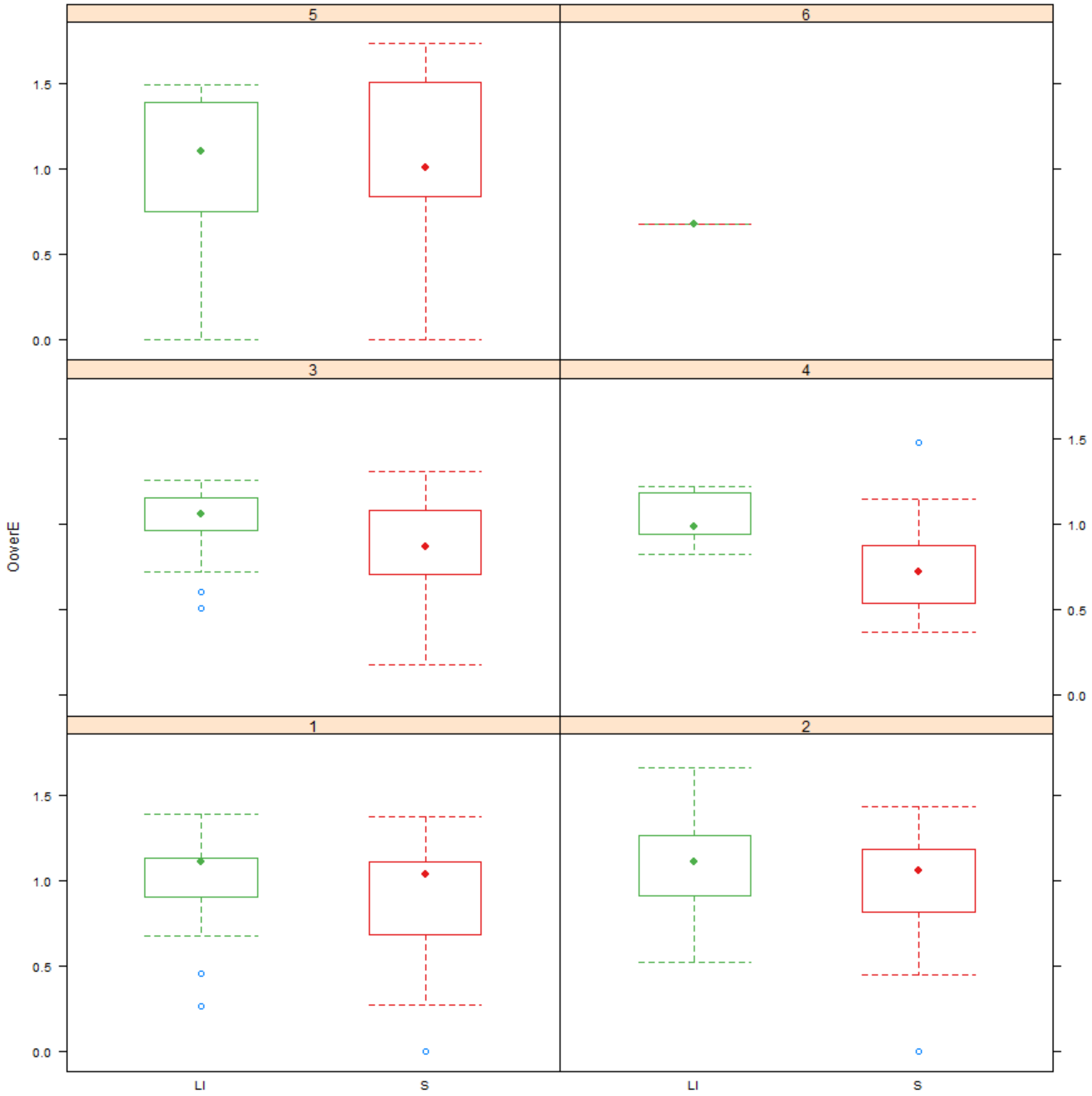


Figure 34. Variability in the range of O/E Scores for least impacted and stressed sites are displayed by likely group assignment (the group with the most votes from the RF model). A site may have been assigned, with nearly equal probability, to one or more other groups. “LI” indicates least impacted sites and “S” stressed sites. Group assignments are 1 (Piedmont Ecoregion), 2 (Southeastern Plains Ecoregion), 3 (Coosa-Tallapoosa Basins), 4 (Tennessee Basin x Ridge and Valley Ecoregion), 5 (Tennessee Basin x Blue Ridge Ecoregion), 6 (Southern Coastal Plain Ecoregion). Note that the model assigned very few sites to the sixth group, the Southern Coastal Plain Ecoregion.

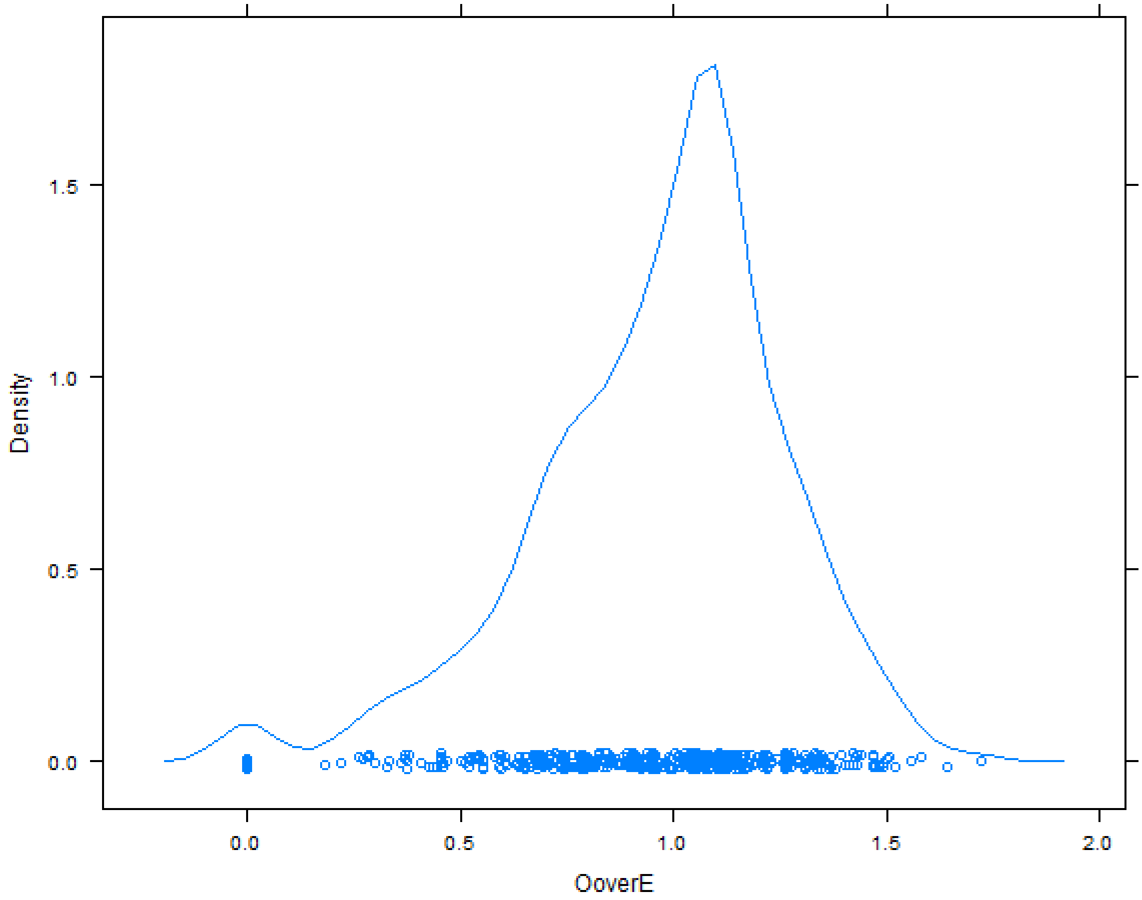


Figure 35. The distributions of O/E scores from all samples did not appear to be bimodal with a large percentage of high and low O/E scores (i.e., sites enriched and depleted in species).

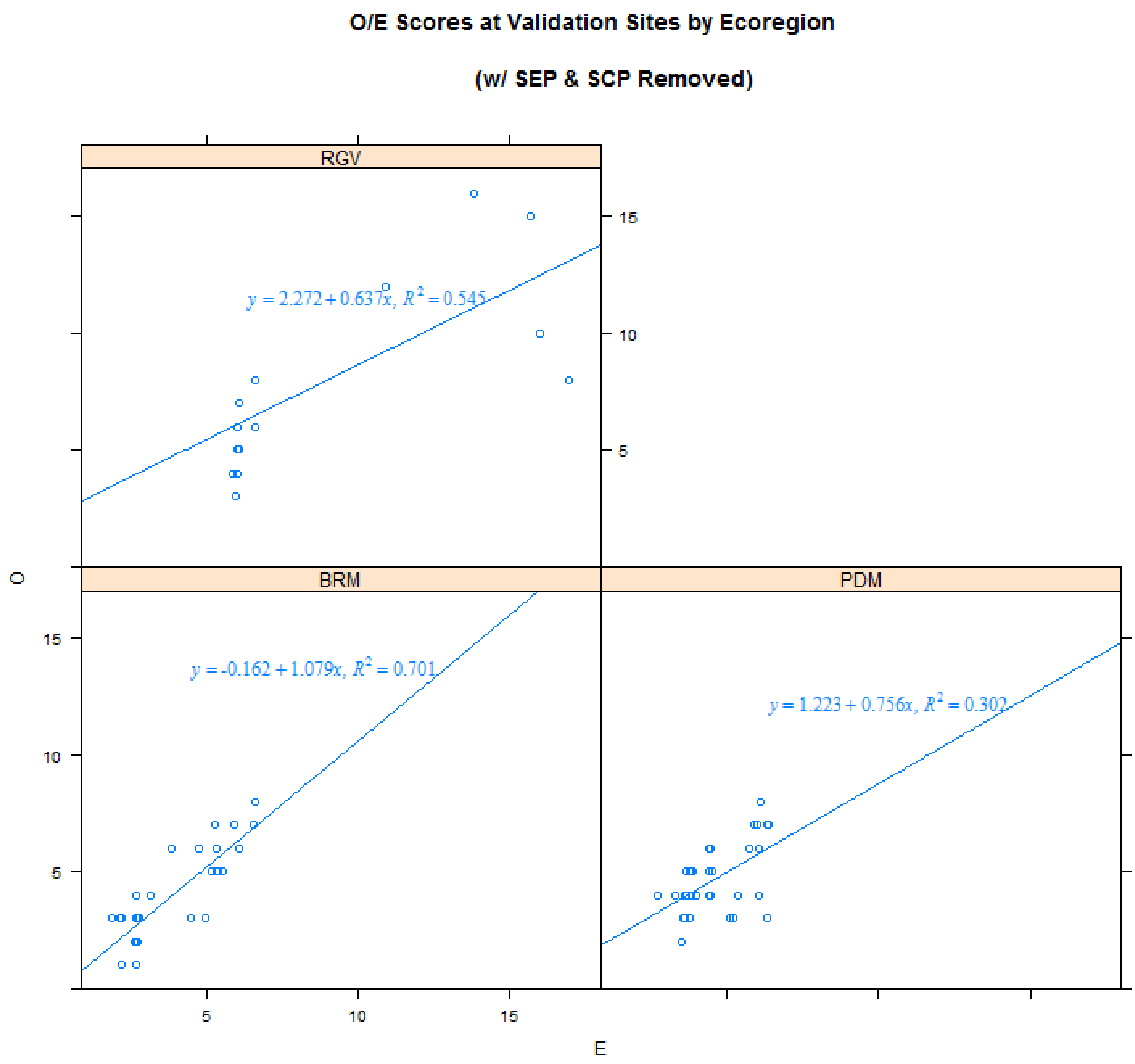


Figure 36. Ecoregion relationships between observed and expected fish taxa richness at validation sites. Those considered strongest have a positive multiple linear regression slope and r^2 greater than 0.30.

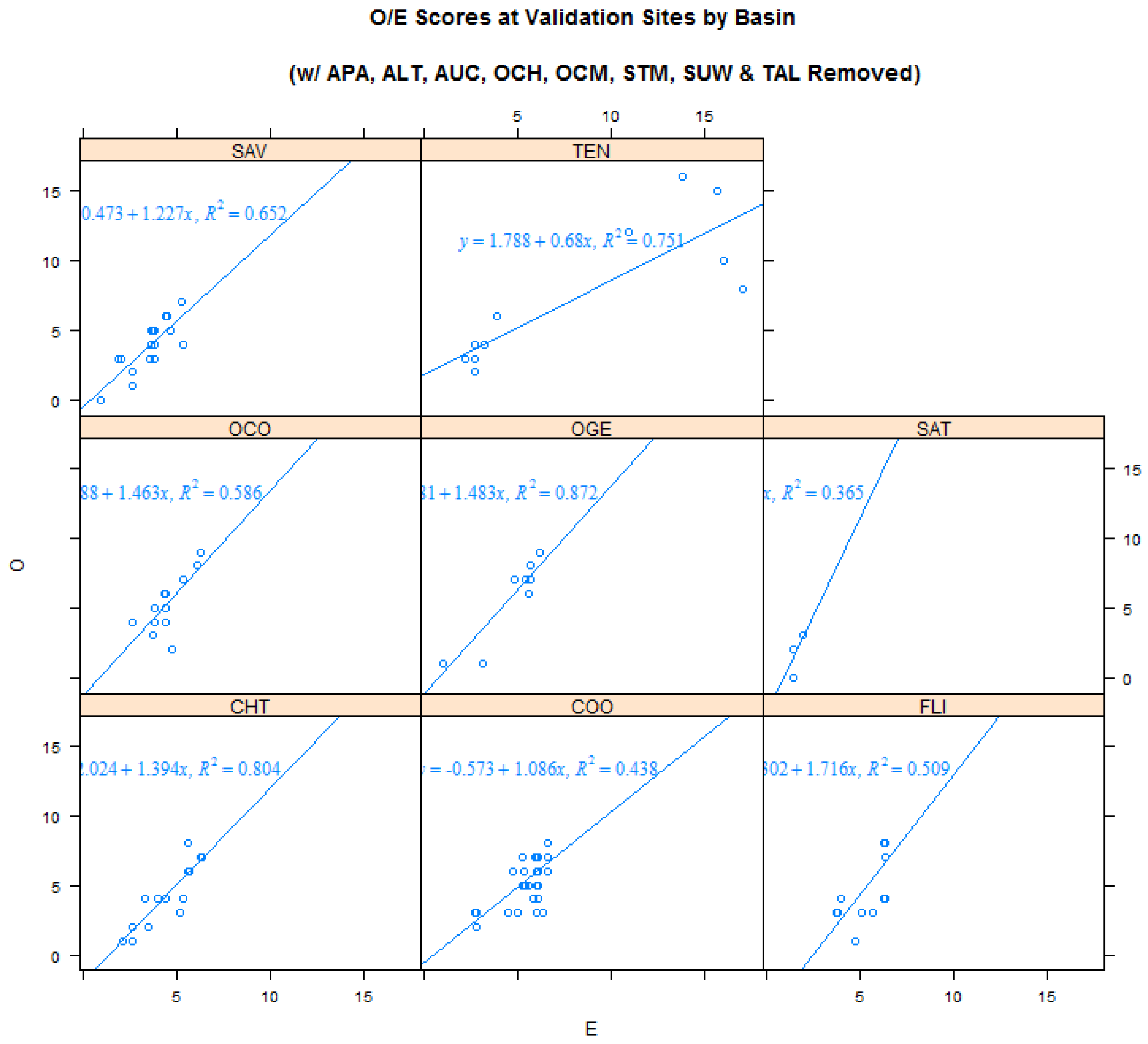


Figure 37. Basins with the strongest relationship between the observed and expected fish taxa richness at validation sites, defined as having a positive multiple linear regression slope and r^2 greater than 0.30.

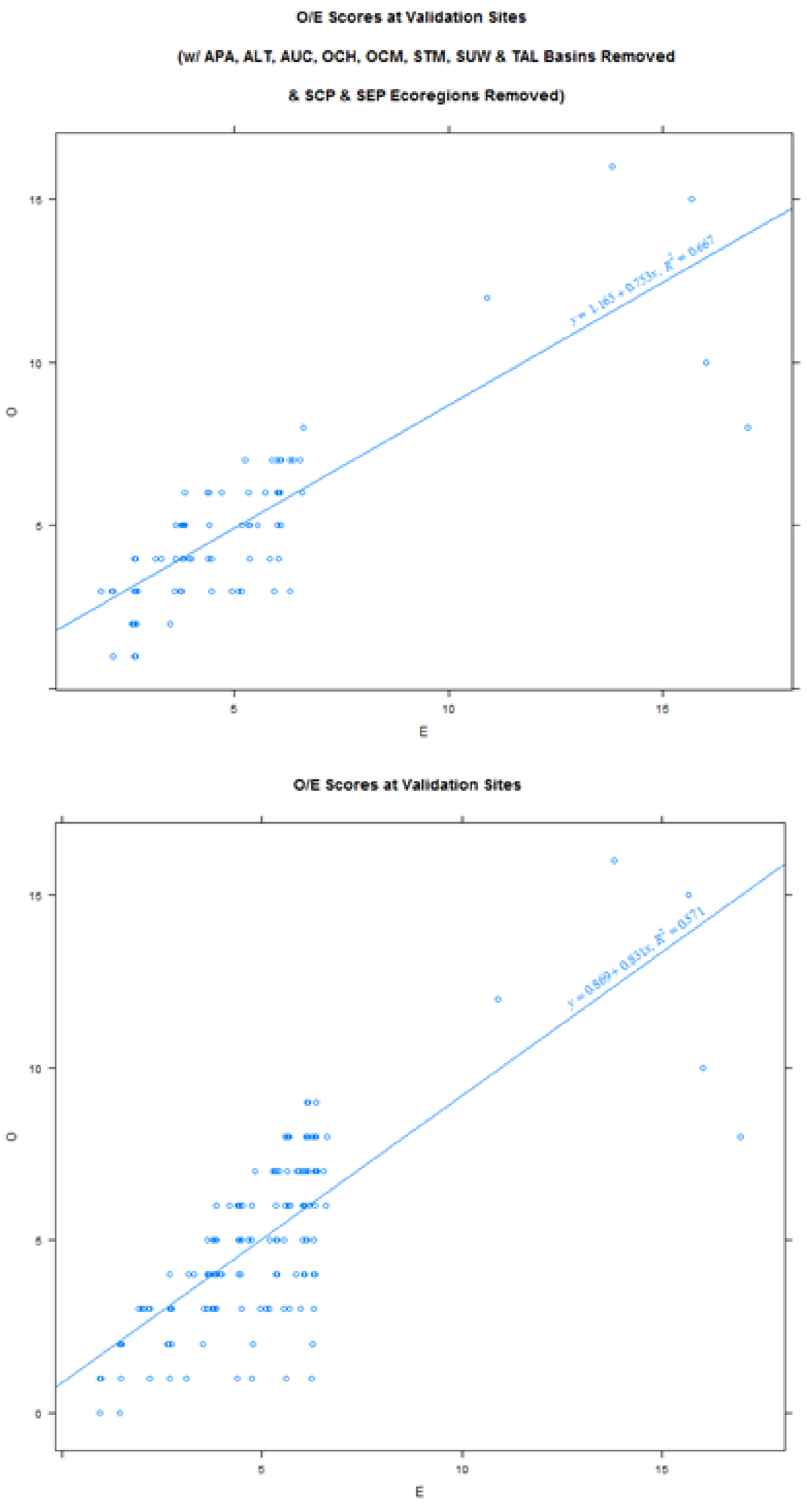


Figure 38. O/E Scores for validation sites showed a stronger relationship with poorly performing basins and ecoregions removed (above, r^2 = 0.67), rather than included (below, r^2 = 0.57).

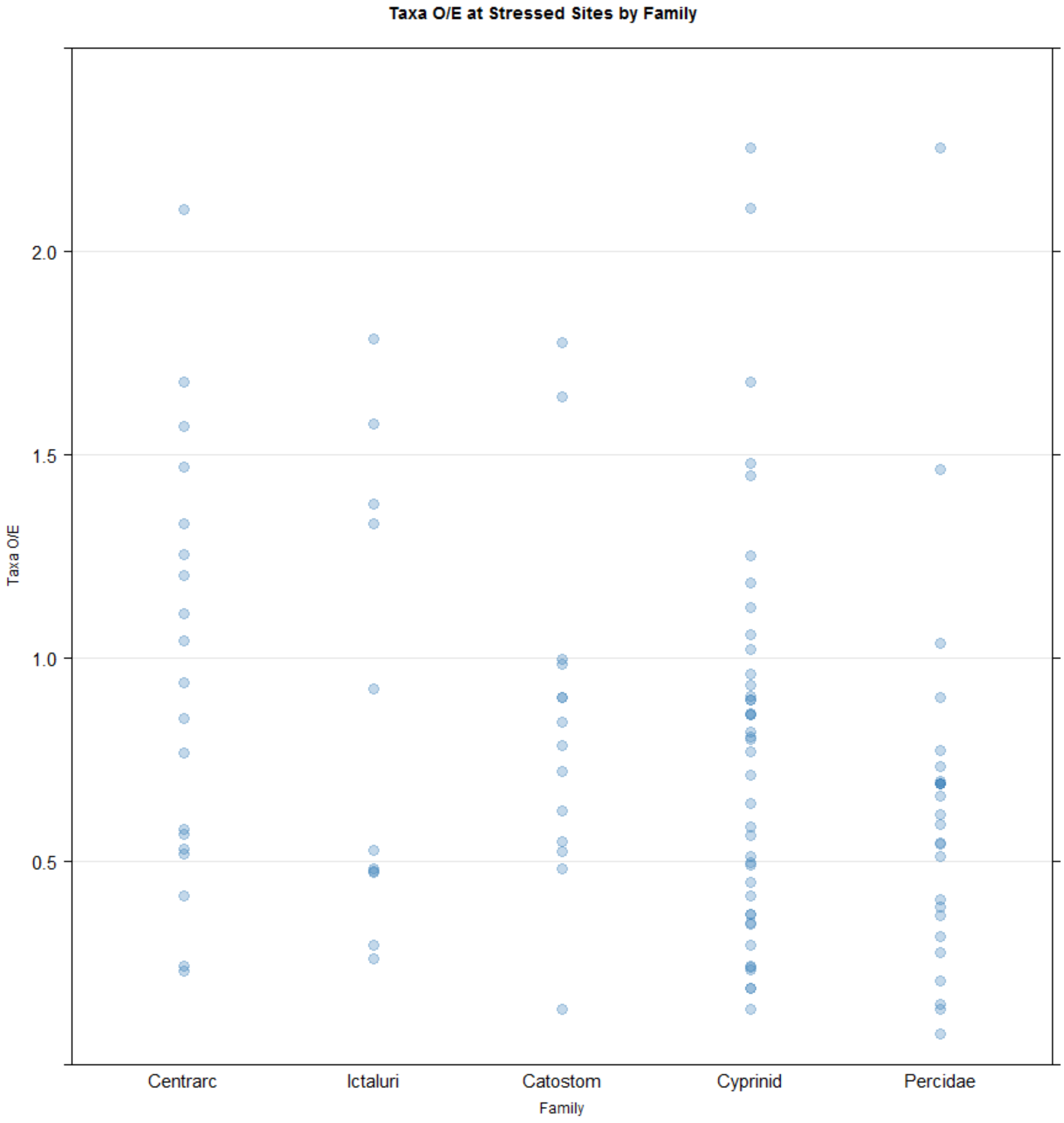


Figure 39. Mean taxa O/E scores, displayed by the five most species-rich families, indicated taxa which occurred more or less frequently at stressed sites.

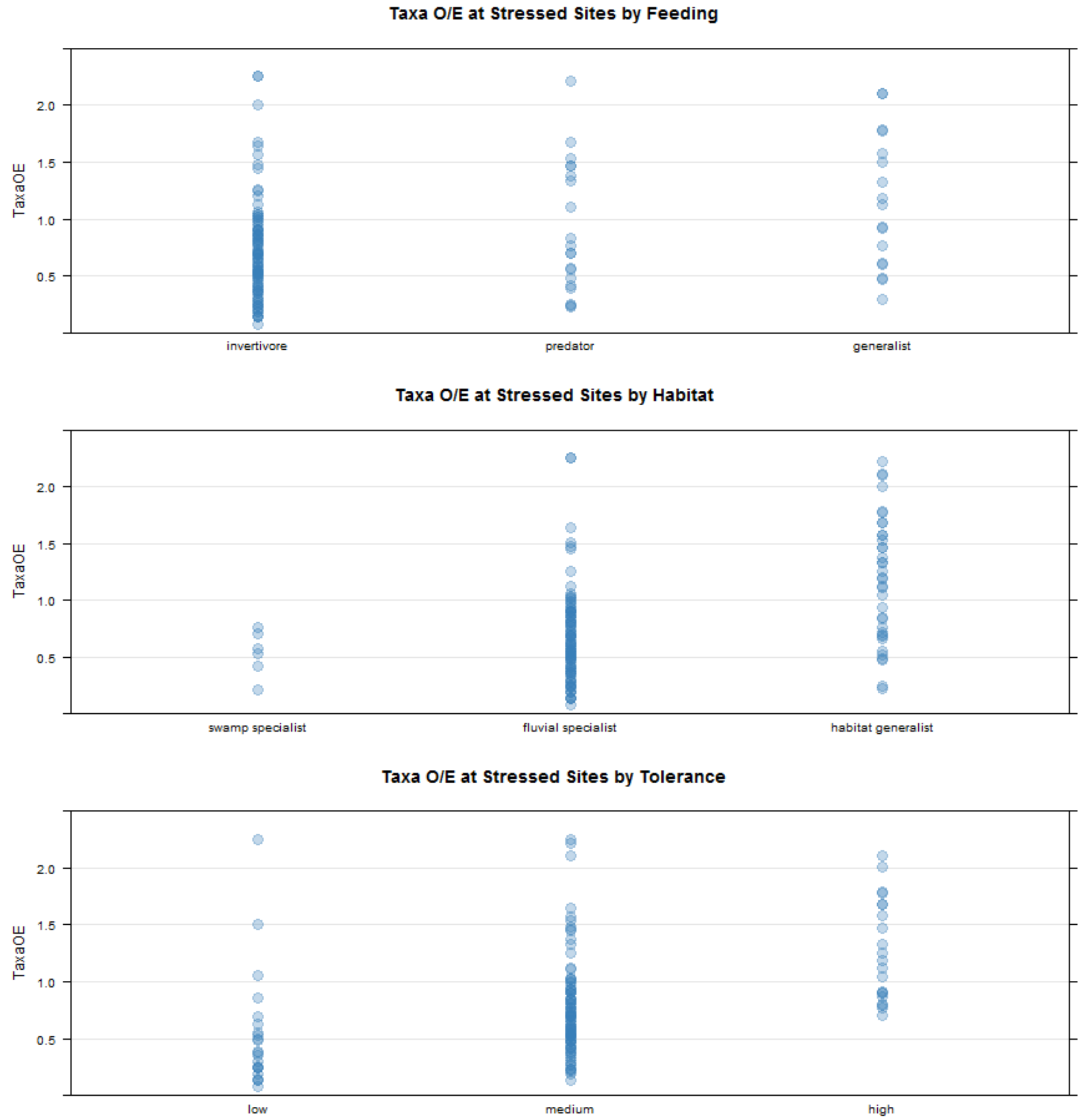


Figure 40. Strip plots of taxa O/E by feeding, habitat, and tolerance groups. Species defined as feeding generalists increased at stressed sites (1.15 average taxa O/E) while invertivores decreased (0.72). Increases in habitat generalist (1.22), are contrasted with decreases in fluvial and swamp specialists (0.67 and 0.54, respectively). The *a priori* designations of tolerance responded predictably with mean O/E taxa scores increasing notably for high tolerance taxa (1.33), decreasing slightly for medium (0.76), and decreasing dramatically for low tolerance taxa (0.53).

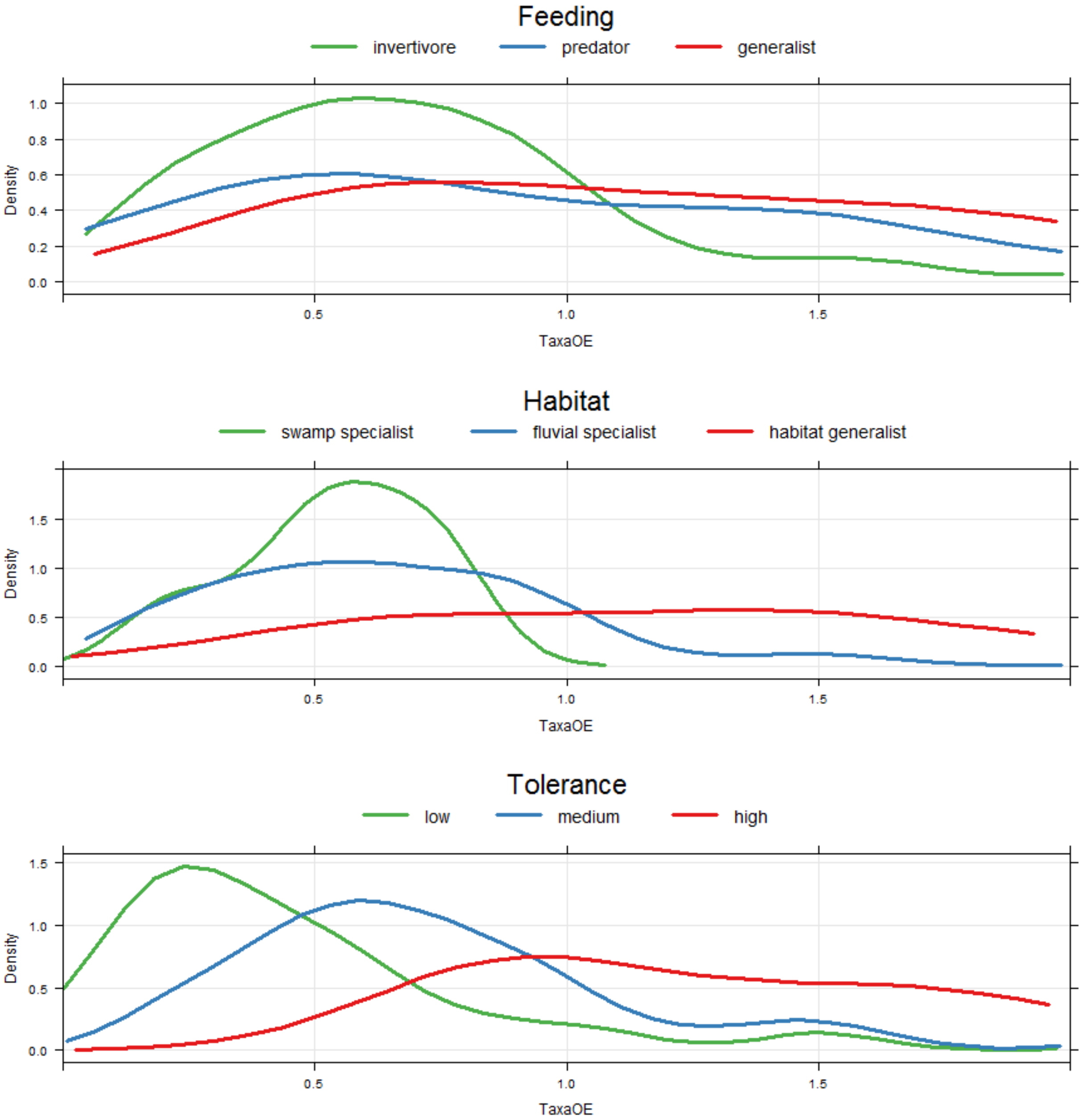


Figure 41. Density plots of taxa O/E by feeding, habitat, and tolerance groups show notable declines in invertivores, swamp and fluvial specialists, and increases in feeding and habitat generalists.

CHAPTER 4

DISCUSSION

This project has demonstrated the feasibility of developing an O/E assessment metric based on samples of fish assemblages in wadeable streams in a geologically and biologically diverse region. After a literature search on the topic, this appears to be the first O/E model that uses a Random Forest model (O/E-RF) to predict fish distributions for bioassessment. The O/E-RF approach is a particularly good complement to the IBI approach. The IBI uses *a priori* assumptions based on ecological theory to define biogeographical regions and select candidate metrics. In the O/E-RF framework, those decisions are largely made *a posteriori* in a data exploration approach via statistical clustering techniques and a machine learning algorithm. Both methods address the same question from slightly different approaches, and therefore provide two relatively independent lines of evidence.

The best predictors of sample assignment to a group were mean annual temperature, elevation, geographic coordinates, and slope. These variables are understood to be important in determining the distribution of organism due to their impact on the metabolism, particularly for fish which are poikilothermic. Classification error was generally low, with the exception of the Southeastern Plains and Southern Coastal Plains Ecoregions. This result may relate to a lack of narrow endemic fishes in these ecoregions, particularly in comparison to other regions in Georgia. The restriction of fish species to particular basins likely aided in the clustering of community types in these basins, whereas the clustering below the fall line was less well defined, and this uncertainty likely propagated through the RF modeling and calculation of O/E scores.

The multivariate cluster analysis grouped samples largely according to ecoregion and basin, providing an intuitive understanding of natural variability in Georgia's stream fish communities. Comparisons of O/E scores between sites considered "least impacted" and those categorized as "stressed" supported the hypothesis that loss of expected species richness could be a sensitive metric of human disturbance, with the exception of the basins and ecoregions dropped from the model.

The RF model was a useful statistical approach for predicting stream fish occurrences in Georgia, particularly the multiple variable importance measures it produced and its ability to handle complex interactions between variables. The majority of published O/E bioassessments used discriminant function models, but that approach was not pursued here because some comparisons indicate instability in the stepwise model selection procedures (Van Sickle et al., 2006). Although care was taken in selection of tuning parameters in the RF model, the response of community groups to some environmental variables may have been indicative of overfitting, or may have reflected the narrow range of data (just for wadeable streams) used to train the model. Predictive ability of the model could potentially be improved with additional environmental variables, such as additional measures of habitat connectivity between reaches, or leveraging the habitat and water quality data measured in the field. There is also opportunity for modeling refinements, such as the use of a more mature CART based modeling approach via the alternate "gbm" package in R, based on related applications that have observed superior performance by boosted regression trees (J R Leathwick, Elith, Francis, Hastie, & Taylor, 2006).

The clustering of fish community types also informs our understanding of the biogeography of Georgia's fishes. Georgia's stream fish community groups, to a first approximation, appear to be a product of ecoregion and basin factors. Georgia's northern ecoregions (Piedmont, Blue Ridge, and Ridge and Valley) are bisected from the southern ecoregions (Southern Coastal Plain and Southeastern Plains) by the Atlantic Seaboard Fall Line. Areas south of this geographic feature are hydrogeomorphically distinct, characterized by low-gradient, meandering rivers, broad floodplains with prolonged flooding, bottomland hardwoods, substantial sediment (and adsorbed contaminant) retention, and extensive hydrologic alteration from surface and groundwater withdrawals (Hupp, 2000). In these areas, the biogeographic groups, defined in the cluster analysis and predicted for each site by the RF model, aligned closest to ecoregions. Therefore, presumably the community groups were more strongly influenced by factors used to delineate these regions (e.g., soils and land cover) which reflect habitat constraints (e.g., the underlying geology and its impact on water chemistry). Conversely, in the northern plateau and mountainous areas above the fall line, groups aligned more closely with basins and subbasins, potentially reflecting constraints on dispersal in steep topography or long isolation of populations resulting in basin endemics.

This research demonstrated the advantage and constraints of applying the O/E assessment approach to stream fishes and presented an additional metric for monitoring and assessing Georgia's streams. Model results in the southern part of the state seem to have been limited by the relatively low fish diversity in the region – leading to greater variability in scores from stochastic effects in species occurrence and detection. Additionally, the complex groundwater dynamics from the karst topography in the region were likely not accounted for in the environmental predictor variables, particularly the temperature and flow variables which were based on surface precipitation and runoff models. The impact designations also could not account for some presumed stressors in the region, such as reductions in baseflow from groundwater pumping and small impoundments. Furthermore, the extensive agricultural land use in the area may have limited the availability of reference sites, and more generally, the IBI program has had less intensive sampling in the area, limiting data available for training and validating the model.

Based on the analysis of a simulated bioassessment dataset, stochastic variation in scores, even at least impacted sites, clearly can obscure the ability of the O/E method to discriminate among levels of impact when the number of possible taxa is low (e.g., <10). The method was designed for application to species-rich faunal groups, and application to low numbers of taxa reduces discriminatory power. Although occurrence data for over two-hundred fish taxa from Georgia’s streams were available for this analysis, this diversity is still about one-third of the macroinvertebrate diversity (over six-hundred taxa) used in the development of the initial RIVPACS modeling framework (J F Wright et al., 1996). Furthermore, the distribution of fish taxa may be more geographically restricted, since some macroinvertebrate taxa have life stages able to disperse across catchment boundaries, whereas fish are generally limited to distribution via the stream network (with the exceptions, perhaps, of migration during flooding or via human introductions). Ecologists using invertebrate O/E models have also explored the biodiversity and taxonomic resolution necessary to build effective O/E models, and have concluded that higher taxonomic resolution (e.g., genera, species, or subspecies) may be necessary in areas where families have a high amount of adaptive radiation (Hawkins, Norris, Wright, Sutcliffe, & Furse, 2000). Other analyses of bioassessment modeling have investigated additional possible conditions that could lead to systematic biases in O/E scores, depending on the number of reference sites used to build the model, the distribution of those sites, the distribution of estimated capture probabilities of the taxa in the reference sites, and the capture probability value used to screen taxa for inclusion in the model (Yuan, 2006).

The impairment threshold used in this analysis (O/E scores of 0.8 or less) was derived from similar O/E modeling for fishes, macroinvertebrates, and diatoms (Carlisle et al., 2008), however these values are tentative, and additional analysis would be required to determine threshold relevant to bioassessment in Georgia. The selection of impairment thresholds for regulatory purposes requires additional expert and stakeholder involvement and is beyond the scope of this work. One common approach to calibrating bioassessment indices is the use of an expert elicitation process where biological criteria scores are correlated with desired biological conditions such as the presence of endemic, sensitive, and rare taxa, ecosystem function, and waterbody connectivity (Davies & Jackson, 2006).

The observation that taxa O/E scores at stressed sites increased the most for centrachids, and decreased the most for percids, cyprinids, and catostomids, confirms a similar result made in a larger regional model of southeastern U.S. stream fishes (Meador & Carlisle, 2009). The successful reproduction of their finding, using over a decade of samples from a large independent data set, provides a strong additional line of evidence demonstrating the sensitivity of stream fish families, such as fluvial percids, to human disturbance.

These results also points to several potential next steps for research. Additional analyses of functional traits should be pursued to examine which taxa are extirpated more easily as the biotic community becomes increasingly disturbed (as measured by O/E scores). The analysis of increasers and decreasers provided initial indications of some taxa, family, and functional group sensitivities. This line of questioning could be extended by modeling species occurrences as a function of taxa traits in combination with O/E scores. These empirical results, indicating which fish guilds respond most to disturbances, could help inform a more mechanistic understanding of what life history traits, pathways, and anthropogenic stressors impact Georgia's stream fish populations.

While the refinement of the state's multimetric IBI is still ongoing, the results of this O/E modeling could support a number of tentative recommendations for the Stream Team. First, prior to the classification of stream types or selection of metrics for IBIs, it may be useful to complete exploratory data analyses, such as clustering fish community types and examining modelled species distributions. Whereas defining IBI site classifications and metrics likely begins with a literature search, these analyses require few assumptions regarding the ecology of fishes, and may elucidate patterns in fish distributions or responses to disturbance. The results of the exploratory analyses can then be incorporated into IBI development if they appear meaningful, rejected if they conflict with our understanding of fish ecology, or researched further. Second, the O/E metric constructed in this research can provide a measure of stream health to be analyzed independently, or as an additional line of evidence, in areas where the model performed strongest. A stream fish IBI approach may perform the best in screening sites in Georgia, since the metrics are purposely chosen on the basis of their specific response in disturbed versus undisturbed sites. The IBI, therefore, may be an ideal tool for identifying reaches with undiagnosed chemical or physical stressors. The O/E approach may have particular utility after a site has been identified as impacted, in order to determine the underlying cause. The species distribution model, built via the O/E methodology or another methods, provides an expectation of which fish fauna are predicted to be present in a stream. The absence of these taxa or functional groups can be used to infer what physiological or behavioral traits are being selected against at the site, potential pathways of the impact, and what the ultimate cause may be. Finally, the inability of the O/E approach to more explicitly incorporate rare taxa into the metric highlights the need for further improvements in methodology. Although bioassessments generally ask whether a particular site is disturbed for the purposes of restoring and maintaining water quality, other management objectives specifically entail conserving a region’s native biota. The monitoring data collected for bioassessments can answer not only, “what proportion of sites are relatively undisturbed?” but additionally, “are species disappearing, and at what rate?” Future refinements in stream bioassessment will need to broaden the focus from the health of fish populations in a given reach, to the viability of fish populations regionally.

REFERENCES

Andreasen, J. K., O’Neill, R. V, Noss, R., & Slosser, N. C. (2001). Considerations for the development of a terrestrial index of ecological integrity. *Ecological Indicators*, *1*(1), 21–35.

Barbour, M., Gerritsen, J., Snyder, B., & Stribling, J. (1999). *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers*. Washington, DC.

Beitinger, T. L. (1990). Behavioral reactions for the assessment of stress in fishes. *Journal of Great Lakes Research*, *16*(4), 495–528.

Borisko, J. P., Kilgour, B. W., & Stanfield, L. W. (2007). An Evaluation of Rapid Bioassessment Protocols for Stream Benthic Invertebrates in Southern Ontario, Canada. *Water Quality Research Journal of Canada*, *42*(3), 184–193. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=asf&AN=501371949&site=ehost-live

Borja, Á., Franco, J., Valencia, V., Bald, J., Muxika, I., Jesús Belzunce, M., & Solaun, O. (2004). Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. *Marine Pollution Bulletin*, *48*(3-4), 209–218. http://doi.org/10.1016/j.marpolbul.2003.12.001

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. http://doi.org/10.2307/1942268

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. http://doi.org/10.1023/A:1010933404324

Brown, M. T., & Vivas, M. B. (2005). Landscape development intensity index. *Environmental Monitoring and Assessment*, *101*(1-3), 289–309. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15736887

Buisson, L., Blanc, L., & Grenouillet, G. (2008). Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*, *17*(2), 244–257. http://doi.org/10.1111/j.1600-0633.2007.00276.x

Cao, Y., Larsen, D. P., & Thorne, R. S.-J. (2001). Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society*, *20*(1), 144–153. http://doi.org/10.2307/1468195

Cao, Y., Williams, D. D., & Williams, N. E. (1998). How important are rare species in aquatic community ecology and bioassessment ? *Limnology and Oceanography*, *43*(7), 1403–1409.

Carlisle, D. M., Falcone, J., & Meador, M. R. (2009). Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. *Environmental Monitoring and Assessment*, *151*(1-4), 143–60. http://doi.org/10.1007/s10661-008-0256-z

Carlisle, D. M., Hawkins, C. P., Meador, M. R., Potapova, M., & Falcone, J. (2008). Biological assessments of Appalachian streams based on predictive models for fish , macroinvertebrate , and diatom assemblages. *Journal of the North American Benthological Society*, *27*(November 2007), 16–37. http://doi.org/10.1899/06-081.1

Chessman, B. C., & Royal, M. J. (2004). Bioassessment without reference sites: use of environmental filters to predict natural assemblages of river macroinvertebrates. *Journal of the North American Benthological Society*, *23*(3), 599–615. http://doi.org/10.1899/0887-3593(2004)023<0599:BWRSUO>2.0.CO;2

Clarke, R. T., Furse, M. T., Wright, J. F., & Moss, D. (1996). Derivation of a biological quality index for river sites: Comparison of the observed with the expected fauna. *Journal of Applied Statistics*, *23*(2-3), 311–332. http://doi.org/10.1080/02664769624279

Clarke, R. T., Wright, J. F., & Furse, M. T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, *160*(3), 219–233. http://doi.org/10.1016/S0304-3800(02)00255-7

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. http://doi.org/10.1890/07-0539.1

Dale, V. H., & Beyeler, S. C. (2001). Challenges in the development and use of ecological indicators. *Ecological Indicators*, *1*(1), 3–10. http://doi.org/10.1016/S1470-160X(01)00003-6

Davies, & Jackson. (2006). The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications*, *16*(4), 1251–1266. Retrieved from http://www.esajournals.org/doi/abs/10.1890/1051-0761(2006)016%5B1251:TBCGAD%5D2.0.CO%3B2

Davies, P. E., Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Development of a national river bioassessment system (AUSRIVAS) in Australia. In *Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997.* (pp. 113–124). Freshwater Biological Association (FBA).

Davis, W. S., & Simon, T. P. (1995). *Biological assessment and criteria: tools for water resource planning and decision making*. CRC Press.

Dewald, T., McKay, L., Bondelid, T., Johnston, C., Moore, R., & Rea, A. (2012). NHDPlus Version 2: User Guide. Retrieved from http://www.horizon-systems.com/nhdplus/

Dickens, C. W. S., & Graham, P. M. (2002). The South African Scoring System (SASS) version 5 rapid bioassessment method for rivers. *African Journal of Aquatic Science*, *27*(1), 1–10.

ESRI. (2014). ArcGIS Desktop 10.2. Redlands, CA: Environmental Systems Research Institute.

European Community. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, L327 Official Journal of the European Parliament 1–82 (2000). http://doi.org/10.1039/ap9842100196

Flebbe, P. a., Roghair, L. D., & Bruggink, J. L. (2006). Spatial Modeling to Project Southern Appalachian Trout Distribution in a Warmer Climate. *Transactions of the American Fisheries Society*, *135*(5), 1371–1382. http://doi.org/10.1577/T05-217.1

Freeman, M. C., & Crawford, M. (1988). Fish Assemblage Stability in a Southern Appalachian Stream. *Canadian Journal of Fisheries and Aquatic Science*, *45*, 1949–1958. Retrieved from http://cwt33.ecology.uga.edu/publications/978.pdf

Fry, J. A., Xian, G., Jin, S., Dewitz, J. A., Homer, C. G., LIMIN, Y., … Wickham, J. D. (2011). Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, *77*(9), 858–864.

GAEPD. (2003). Georgia Department of Natural Resources Basins (DNR Basins). Retrieved October 15, 2015, from http://epd.georgia.gov/dnr-basins

Gara, B. D., & Micacchion, M. (2010). *Assessment of wetland mitigation projects in Ohio. Volume 2: Developing a GIS-based tool to optimize vernal pool wetland mitigation site selection. Ohio EPA Technical Report WET/2010-1B.* Columbus, OH.

GAWRD. (2005). *Standard Operating Procedures for Conducting Biomonitoring on Fish Communities in Wadeable Streams in Georgia*. Social Circle, GA. Retrieved from http://www.georgiawildlife.com/sites/default/files/uploads/wildlife/fishing/pdfs/streamsurvey/SOP_Part1.pdf

GAWRD. (2015). Protected Fishes of Georgia. Retrieved January 1, 2015, from http://georgiawildlife.com/sites/default/files/uploads/wildlife/nongame/text/html/protected_species/fishes.html

Giattina, J., & Garton, R. (1983). A Review of the preference-avoidance response of fishes to aquatic contaminants. *Residue Reviews*, *87*, 43–90. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4612-5479-9_2

Gorman, O. T., & Karr, J. R. (1978). Habitat structure and stream Fish communities. *Ecology*, *59*(3), 507–515. http://doi.org/10.2307/1936581

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., & Stevenson, R. J. (2000). Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society*, *19*(3), 541–556.

Hawkins, C. P., Norris, R. H., Hogue, J. N., & Feminella, J. W. (2000). Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*, *10*(5), 1456–1477. http://doi.org/10.1890/1051-0761(2000)010[1456:DAEOPM]2.0.CO;2

Hawkins, C. P., Norris, R. H., Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Effects of taxonomic resolution and use of subsets of the fauna on the performance of RIVPACS-type models. In *Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997.* (pp. 217–228). Freshwater Biological Association (FBA).

Hawkins, C. P., Olson, J. R., & Hill, R. a. (2010). The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society*, *29*(1), 312–343. http://doi.org/10.1899/09-092.1

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., … Pont, D. (2010). The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. *Science of The Total Environment*, *408*(19), 4007–4019. http://doi.org/10.1016/j.scitotenv.2010.05.031

Hitt, N. P., & Angermeier, P. L. (2008). River-Stream Connectivity Affects Fish Bioassessment Performance. *Environmental Management*, *42*(1), 132–150. http://doi.org/10.1007/s00267-008-9115-5

Hitt, N. P., & Angermeier, P. L. (2011). Fish community and bioassessment responses to stream network position. *Journal of the North American Benthological Society*, *30*(1), 296–309. http://doi.org/10.1899/09-155.1

Holland-Bartels, L. E., & Dewey, M. R. (1997). The influence of seine capture efficiency on fish abundance estimates in the upper Mississippi River. *Journal of Freshwater Ecology*, *12*(1), 101–111.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., … Wickham, J. (2007). Completion of the 2001 national land cover database for the counterminous United States. *Photogrammetric Engineering and Remote Sensing*, *73*(4), 337.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., … Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, *81*(5), 345–354.

Hughes, D. L., Gore, J., Brossett, M. P., & Olson, J. R. (2009). *Rapid bioassessment of stream health*. CRC Press.

Hupp, C. (2000). Hydrology, Geomorphology and Vegetation of Coastal Plain Rivers in the South-Eastern USA. *Hydrological Processes*, *14*(16-17), 2991–3010. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/1099-1085(200011/12)14:16/17<2991::AID-HYP131>3.0.CO;2-H/abstract

Jackson, C. R., & Pringle, C. M. (2010). Ecological Benefits of Reduced Hydrologic Connectivity in Intensively Developed Landscapes. *BioScience*, *60*(1), 37–46. http://doi.org/10.1525/bio.2010.60.1.8

Joy, M. K., & Death, R. G. (2002). Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology*, *47*(11), 2261–2275. http://doi.org/10.1046/j.1365-2427.2002.00954.x

Karr, J. R. (1981). Assessment of biotic integrity using fish communities. *Fisheries*, *6*(6), 21–27. http://doi.org/10.1577/1548-8446(1981)006

Karr, J. R. (1991). Biological integrity : A long-neglected aspect of water resource management. *Ecological Applications*, *1*(1), 66–84.

Karr, J. R., & Chu, E. W. (1997). *Biological Monitoring and Assessment: Using Multimetric Indexes Effectively (EPA 235-R97-001)*. Seattle, WA.

Karr, J. R., Fausch, K. D., Angermeier, P. L., Yant, P. R., & Schlosser, I. J. (1986). *Assessing Biological Integrity in Running Waters A Method and Its Rationale*. Champaign: Illinois Natural History Survey Special Publication.

Kennard, M. J., Pusey, B. J., Arthington, a. H., Harch, B. D., & Mackay, S. J. (2006). Development and application of a predictive model of freshwater fish assemblage composition to evaluate river health in eastern Australia. *Hydrobiologia*, *572*(1), 33–57. http://doi.org/10.1007/s10750-005-0993-8

Klerks, P. L., & Weis, J. S. (1987). Genetic adaptation to heavy metals in aquatic organisms: a review. *Environmental Pollution*, *45*(3), 173–205.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand : an analysis using boosted regression trees. *Marine Ecology Progress Series*, *321*, 267–281.

Leathwick, J. R., Rowe, D., Richardson, J., Elith, J., & Hastie, T. (2005). Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, *50*(12), 2034–2052. http://doi.org/10.1111/j.1365-2427.2005.01448.x

Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. In *2000 Annual Meeting of the Society for Academic Emergency Medicine* (p. 14). San Francisco, CA. http://doi.org/10.1.1.95.4103

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(December), 18–22. http://doi.org/10.1177/154405910408300516

Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151. http://doi.org/10.1111/j.1466-8238.2007.00358.x

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. http://doi.org/10.1002/widm.8

Maloney, K. O., Weller, D. E., Russell, M. J., & Hothorn, T. (2009). Classifying the biological condition of small streams: an example using benthic macroinvertebrates. *Journal of the North American Benthological Society*, *28*(4), 869–884. http://doi.org/10.1899/08-142.1

Marchant, R. (2002). Do rare species have any place in multivariate analysis for bioassessment? *Journal of the North American Benthological Society*, *21*(2), 311–313. http://doi.org/10.2307/1468417

Marcinek, P., & Lanford, P. (2013). GA DNR Stream Survey Team's State of the Fish. In *Proceedings of the 2013 Georgia Water Resources Conference, April 10-11, 2013, Athens, Georgia*. Athens, GA. Retrieved from https://smartech.gatech.edu/handle/1853/51481

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59–69. http://doi.org/10.1111/j.1472-4642.2008.00491.x

Master, L. L., Flack, S. R., & Stein, B. A. (1998). *Rivers of life: critical watersheds for protecting freshwater biodiversity*. (L. L. Master, S. R. Flack, & B. A. Stein, Eds.). Retrieved from http://www.natureserve.org/library/riversoflife.pdf

Meador, M. R., & Carlisle, D. M. (2009). Predictive Models for Fish Assemblages in Eastern U.S. Streams: Implications for Assessing Biodiversity. *Transactions of the American Fisheries Society*, *138*(4), 725–740. http://doi.org/10.1577/T08-132.1

Meador, M. R., Coles, J. F., & Zappia, H. (2005). Fish Assemblage Responses to Urban Intensity Gradients in Contrasting Metropolitan Areas: Birmingham, Alabama and Boston, Massachusetts. *American Fisheries Society Symposium*, *47*, 409–423. Retrieved from http://water.usgs.gov/nawqa/ecology/pubs/MeadorOthers2005_AFS_FishChap.pdf

Moore, K. M. S., & Gregory, S. V. (1988). Response of young-of-the-year cutthroat trout to manipulation of habitat structure in a small stream. *Transactions of the American Fisheries Society*, *117*(2), 162–170.

Moss, D., Wright, J. F., Furse, M. T., & Clarke, R. T. (1999). A comparison of alternative techniques for prediction of the fauna of running-water sites in Great Britain. *Freshwater Biology*. http://doi.org/10.1046/j.1365-2427.1999.00376.x

Oberdorff, T., & Pont, D. (2001). A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology*, *46*(3), 399–415. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2427.2001.00669.x/full

Olden, J. (2003). A species-specific approach to modeling biological communities and its potential for conservation. *Conservation Biology*, *17*(3), 854–863.

Olden, J., & Jackson, D. (2002). A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, 1976–1995. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2427.2002.00945.x/full

Omernik, J. (1987). Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*. Retrieved from http://www.tandfonline.com/doi/abs/10.1111/j.1467-8306.1987.tb00149.x

Osborne, L. L., Kohler, S. L., Bayley, P. B., Day, D. M., Bertrand, W. A., Wiley, M. J., & Sauer, R. (1992). Influence of stream location in a drainage network on the index of biotic integrity. *Transactions of the American Fisheries Society*, *121*(5), 635–643.

Palmer, M. A., & Febria, C. M. (2012). The Heartbeat of Ecosystems. *Science*, *336*(6087), 1393–4. http://doi.org/10.1126/science.1223250

Parsley, M. J., Palmer, D. E., & Burkhardt, R. W. (1989). Variation in capture efficiency of a beach seine for small fishes. *North American Journal of Fisheries Management*, *9*(2), 239–244.

Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, *216*(3-4), 316–322. http://doi.org/10.1016/j.ecolmodel.2008.05.006

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*(2), 391–409. http://doi.org/10.2307/1468026

Pont, D., Hughes, R. M., Whittier, T. R., & Schmutz, S. (2009). A Predictive Index of Biotic Integrity Model for Aquatic-Vertebrate Assemblages of Western U.S. Streams. *Transactions of the American Fisheries Society*, *138*(2), 292–305. http://doi.org/10.1577/T07-277.1

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., … Schmutz, S. (2006). Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology*, *43*(1), 70–80. http://doi.org/10.1111/j.1365-2664.2005.01126.x

Pont, D., Hugueny, B., & Oberdorff, T. (2005). Modelling habitat requirement of European fishes: do species have similar responses to local and regional environmental constraints? *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(1), 163–173. http://doi.org/10.1139/f04-183

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, *9*(2), 181–199.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. Retrieved from http://www.r-project.org

Rehn, A. C., Mazor, R. D., & Ode, P. R. (2015). *The California Stream Condition Index (CSCI): A New Statewide Biological Scoring Tool for Assessing the Health of Freshwater Streams*. Retrieved from http://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/csci_tech_memo.pdf

Reynoldson, T. B., Norris, R. H., Resh, V. H., Day, K. E., & Rosenberg, D. M. (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*(4), 833–852.

Roset, N., Grenouillet, G., Goffaux, D., Pont, D., & Kestemont, P. (2007). A review of existing fish assemblage indicators and methodologies. *Fisheries Management and Ecology*, *14*(6), 393–405. http://doi.org/10.1111/j.1365-2400.2007.00589.x

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*(5), 696–704. http://doi.org/10.1577/1548-8659

Schlosser, I. J. (1985). Flow regime, juvenile abundance, and the assemblage structure of stream fishes. *Ecology*, 1484–1490.

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294*. Retrieved from http://pubs.usgs.gov/wsp/wsp2294/pdf/wsp_2294.pdf

Seegert, G. (2000). The development, use, and misuse of biocriteria with an emphasis on the index of biotic integrity. *Environmental Science & Policy*, *3*, 51–58. http://doi.org/10.1016/S1462-9011(00)00027-7

Stein, B. A. (2002). *States of the Union: Ranking America's Biodiversity*. *NatureServe*. Arlington, VA. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:States+of+the+Union+:+Ranking+America's+Biodiversity#0

Stoddard, J., & Larsen, D. (2006). Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications*, *16*(4), 1267–1276. Retrieved from http://www.esajournals.org/doi/abs/10.1890/1051-0761(2006)016%5B1267:SEFTEC%5D2.0.CO%3B2

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. http://doi.org/10.1186/1471-2105-8-25

U.S.C. Clean Water Act of 1972 (2002). USA. Retrieved from http://www.epw.senate.gov/water.pdf

Van Sickle, J., Hawkins, C. P., Larsen, D. P., & Herlihy, A. T. (2005). A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society*. http://doi.org/10.1899/0887-3593(2005)024<0178:ANMFTE>2.0.CO;2

Van Sickle, J., Huff, D. D., & Hawkins, C. P. (2006). Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology*, *51*(2), 359–372. http://doi.org/10.1111/j.1365-2427.2005.01487.x

Van Sickle, J., Larsen, D. P., & Hawkins, C. P. (2007). Exclusion of rare taxa affects performance of the O / E index in bioassessments. *Journal of the North American Benthological Society*, *26*(2), 319–331. http://doi.org/10.1899/0887-3593(2007)26[319:EORTAP]2.0.CO;2

Waite, I. R. (2014). Agricultural disturbance response models for invertebrate and algal metrics from streams at two spatial scales within the U.S. *Hydrobiologia*, *726*(1), 285–303. http://doi.org/10.1007/s10750-013-1774-4

Waite, I. R., Kennen, J. G., May, J. T., Brown, L. R., Cuffney, T. F., Jones, K. a, & Orlando, J. L. (2012). Comparison of Stream Invertebrate Response Models for Bioassessment Metrics. *Journal of the American Water Resources Association*, *48*(3), 570–583. http://doi.org/10.1111/j.1752-1688.2011.00632.x

Warren, M. L., Angermeier, P. L., Burr, B. M., & Haag, W. R. (1997). Decline of a diverse fish fauna: patterns of imperilment and protection in the Southeastern United States. In *Aquatic Fauna in Peril: The Southeastern Perspective* (pp. 105–164).

Warren, M. L., Burr, B. M., Walsh, S. J., Bart, H. L., Cashner, R. C., Etnier, D. a., … Starnes, W. C. (2000). Diversity, Distribution, and Conservation Status of the Native Freshwater Fishes of the Southern United States. *Fisheries*, *25*(March 2015), 7–31. http://doi.org/10.1577/1548-8446(2000)025<0007:DDACSO>2.0.CO;2

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*(2), 260–267. http://doi.org/10.1111/j.2041-210X.2011.00170.x

Wright, J. F., Blackburn, J. H., Gunn, R. J. M., Furse, M. T., Armitage, P. D., Winder, J. M., … Moss, D. (1996). Macroinvertebrate frequency data for the RIVPACS III sites in Great Britain and their use in conservation evaluation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *6*(3), 141–167.

Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*. (J. Wright, D. Sutcliffe, & M. Furse, Eds.). Ambleside, Cumbria, UK: Freshwater Biological Association. Retrieved from http://core.kmi.open.ac.uk/download/pdf/11020772.pdf

Yoder, C. O., & Rankin, E. T. (1998). The role of biological indicators in a state water quality management process. *Environmental Monitoring and Assessment*, *51*(1-2), 61–88.

Yuan, L. L. (2006). Theoretical predictions of observed to expected ratios in RIVPACS-type predictive model assessments of stream biological condition. *Journal of the North American Benthological Society*, *25*(4), 841–850. http://doi.org/10.1899/0887-3593(2006)025[0841:TPOOTE]2.0.CO;2

Zuellig, R. E., Carlisle, D. M., Meador, M. R., & Potapova, M. (2012). Variance partitioning of stream diatom, fish, and invertebrate indicators of biological condition. *Freshwater Science, 31*(1), 182–190. http://doi.org/10.1899/11-040.1

APPENDIX A

SUMMARY OF MOST WIDELY CITED BIOASSESSMENT INDICES

Table A.1. Summary of the most widely cited publications describing the development or evaluation of bioassessment indices. The terms "Index of Biotic Integrity" and "RIVPACS" were queried in Google Scholar. Relevant publications, which were cited at least 50 and 10 times, respectively for the two search terms, were reviewed, up to the first 100 results. Relevant publications included those which describe the development and evaluation of bioassessment models, excluding summaries, reviews, or description of evaluation methods that were not specific to a given assemblage or location. Where there were multiple publications regarding the same bioassessment index development, the most highly cited publication was retained. “Other” assemblages includes periphyton, phytoplankton, zooplankton, macrophytes, or aquatic amphibians. Table is sorted in descending order based on number of citations.

	Multimetric	**Multivariate**	**Stream**	**River**	**Lake**	**Wetland**	**Estuary**	**Fish**	**Macroinvert.**	**Other**	
n =	**51**	**14**	**46**	**11**	**6**	**3**	**5**	**44**	**20**	**9**	
Country, Location	**Methods**		**Waterbodies**					**Assemblagse**			**Reference**
USA, Tennessee Valley	1		1						1		Kerans, B. L., & Karr, J. R. (1994). A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological applications*, *4*(4), 768-785.
USA, Midwest	1		1					1			Fausch, K. D., Karr, J. R., & Yant, P. R. (1984). Regional application of an index of biotic integrity based on stream fish communities. *Transactions of the American Fisheries Society*, *113*(1), 39-55.
USA, Chesapeake Bay	1						1		1		Weisberg, S. B., Ranasinghe, J. A., Dauer, D. M., Schaffner, L. C., Diaz, R. J., & Frithsen, J. B. (1997). An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries*, *20*(1), 149-158.
Canada, Ontario	1		1					1			Steedman, R. J. (1988). Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, *45*(3), 492-501.
Canada, Great Lakes		1			1				1		Reynoldson, T. B., Bailey, R. C., Day, K. E., & Norris, R. H. (1995). Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state.*Australian journal of ecology*, *20*(1), 198-219.

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
USA, Oregon, Colorado, New England & Appalachians	1		1					1			Miller, D. L., Hughes, R. M., Karr, J. R., Leonard, P. M., Moyle, P. B., Schrader, L. H., ... & Orth, D. J. (1988). Regional applications of an index of biotic integrity for use in water resource management. *Fisheries*, *13*(5), 12-20.
France, Seine Basin	1			1				1			Oberdorff, T., & Hughes, R. M. (1992). Modification of an index of biotic integrity based on fish assemblages to characterize rivers of the Seine Basin, France. *Hydrobiologia*, *228*(2), 117-130.
USA, Mid-Appalachians	1		1					1			Hill, B. H., Herlihy, A. T., Kaufmann, P. R., Stevenson, R. J., McCormick, F. H., & Johnson, C. B. (2000). Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society*, *19*(1), 50-67.
United Kingdom, All		1	1						1		Wright, J. F., Furse, M. T., & Armitage, P. D. (1993). RIVPACS-a technique for evaluating the biological quality of rivers in the UK. *European Water Pollution Control*, *3*, 15-15.
USA, Wisconsin	1		1					1			Lyons, J., Wang, L., & Simonson, T. D. (1996). Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management*, *16*(2), 241-256.
France, All		1	1	1				1			Oberdorff, T., Pont, D., Hugueny, B., & Chessel, D. (2001). A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology*, *46*(3), 399-415.
USA, Illinois, Ohio & West Virginia	1		1					1			Angermeier, P. L., & Karr, J. R. (1986). Applying an index of biotic integrity based on stream-fish communities: considerations in sampling and interpretation. *North American Journal of Fisheries Management*, *6*(3), 418-429.

Country, Location	**Multimetric**	**Multivariate**	**Stream**	**River**	**Lake**	**Wetland**	**Estuary**	**Fish**	**Macroinvert.**	**Other**	**Reference**
USA, Wisconsin	1		1					1			Lyons, J. (1992). Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. *General technical report NC (USA)*.
USA, Mid-Atlantic Highlands	1		1					1			McCormick, F. H., Hughes, R. M., Kaufmann, P. R., Peck, D. V., Stoddard, J. L., & Herlihy, A. T. (2001). Development of an index of biotic integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society*,*130*(5), 857-877.
USA, Massachusetts	1						1	1			Deegan, L. A., Finn, J. T., Ayvazian, S. G., Ryder-Kieffer, C. A., & Buonaccorsi, J. (1997). Development and validation of an estuarine biotic integrity index. *Estuaries*, *20*(3), 601-617.
Mexico, West Central	1		1	1				1			Lyons, J., Navarro-Pérez, S., Cochran, P. A., Santana, E. C., & Guzmán-Arroyo, M. (1995). Index of Biotic Integrity Based on Fish Assemblages for the Conservation of Streams and Rivers in West-Central Mexico. *Conservation Biology*, *9*(3), 569-584.
USA, West Virginia	1		1					1			Leonard, P. M., & Orth, D. J. (1986). Application and testing of an index of biotic integrity in small, coolwater streams. *Transactions of the American Fisheries Society*, *115*(3), 401-414.
USA, Southeast	1						1		1		Van Dolah, R. F., Hyland, J. L., Holland, A. F., Rosen, J. S., & Snoots, T. R. (1999). A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. *Marine Environmental Research*, *48*(4), 269-283.
Australia, All		1	1						1		Davies, P. E., Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Development of a national river bioassessment system (AUSRIVAS) in Australia. In *Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997*. (pp. 113-124). Freshwater Biological Association (FBA).

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
Australia, New South Wales	1		1	1				1			Harris, J. H., & Silveira, R. (1999). Large-scale assessments of river health using an Index of Biotic Integrity with low-diversity fish communities.*Freshwater Biology*, *41*(2), 235-252.
USA, Mid-Atlantic Highlands	1		1						1		Klemm, D. J., Blocksom, K. A., Fulk, F. A., Herlihy, A. T., Hughes, R. M., Kaufmann, P. R., ... & Davis, W. S. (2003). Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management*, *31*(5), 0656-0669.
USA, Great Lakes	1					1		1	1	1	Wilcox, D. A., Meeker, J. E., Hudson, P. L., Armitage, B. J., Black, M. G., & Uzarski, D. G. (2002). Hydrologic variability and the application of index of biotic integrity metrics to wetlands: a Great Lakes evaluation. *Wetlands*, *22*(3), 588-615.
USA, Illinois & Indiana	1		1					1			Karr, J. R., Yant, P. R., Fausch, K. D., & Schlosser, I. J. (1987). Spatial and temporal variability of the index of biotic integrity in three midwestern streams.*Transactions of the American Fisheries Society*, *116*(1), 1-11.
USA, Maryland	1		1					1			Stribling, JB (1998) Development of a benthic index of biotic integrity for Maryland streams. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division.
USA, Colorado	1		1					1	1	1	Griffith, M. B., Hill, B. H., McCormick, F. H., Kaufmann, P. R., Herlihy, A. T., & Selle, A. R. (2005). Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. *Ecological Indicators*, *5*(2), 117-136.
Belgium, Flanders	1		1	1				1			Belpaire, C., Smolders, R., Auweele, I. V., Ercken, D., Breine, J., Van Thuyne, G., & Ollevier, F. (2000). An Index of Biotic Integrity characterizing fish populations and the ecological quality of Flandrian water bodies. *Hydrobiologia*,*434*(1-3), 17-33.

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
USA, Wisconsin	1			1				1			Lyons, J., Piette, R. R., & Niermeyer, K. W. (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*(6), 1077-1094.
USA, Oregon & Washington		1	1						1		Ostermiller, J. D., & Hawkins, C. P. (2004). Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models.*Journal of the North American Benthological Society*, *23*(2), 363-382.
USA, Interior Plateau	1		1							1	Wang, Y. K., Stevenson, R. J., & Metzmeier, L. (2005). Development and evaluation of a diatom-based Index of Biotic Integrity for the Interior Plateau Ecoregion, USA. *Journal of the North American Benthological Society*, *24*(4), 990-1008.
USA, Colorado	1		1					1			Bramblett, R. G., & Fausch, K. D. (1991). Variable fish communities and the index of biotic integrity in a western Great Plains river. *Transactions of the American Fisheries Society*, *120*(6), 752-769.
USA, Georgia	1		1					1			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*(5), 1118-1133.
USA, Great Lakes	1					1			1		Burton, T. M., Uzarski, D. G., Gathman, J. P., Genet, J. A., Keas, B. E., & Stricker, C. A. (1999). Development of a preliminary invertebrate index of biotic integrity for Lake Huron coastal wetlands. *Wetlands*, *19*(4), 869-882.
USA, Great Lakes	1					1		1			Uzarski, D. G., Burton, T. M., Cooper, M. J., Ingram, J. W., & Timmermans, S. T. (2005). Fish habitat use within and across wetland classes in coastal wetlands of the five Great Lakes: development of a fish-based index of biotic integrity. *Journal of Great Lakes Research*, *31*, 171-187.

Country, Location	**Multimetric**	**Multivariate**	**Stream**	**River**	**Lake**	**Wetland**	**Estuary**	**Fish**	**Macroinvert.**	**Other**	**Reference**
USA, Northeastern	1		1	1						1	Hill, B. H., Herlihy, A. T., Kaufmann, P. R., DeCelles, S. J., & Vander Borgh, M. A. (2003). Assessment of streams of the eastern United States using a periphyton index of biotic integrity. *Ecological Indicators*, *2*(4), 325-338.
New Zealand, All		1	1					1			Joy, M. K., & Death, R. G. (2002). Predictive modeling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology*, *47*(11), 2261-2275.
Brazil, South	1		1					1			Bozzetti, M., & Schulz, U. H. (2004). An index of biotic integrity based on fish assemblages for subtropical streams in southern Brazil. *Hydrobiologia*, *529*(1-3), 133-144.
Republic of Guinea, Konkoure River	1			1				1			Hugueny, B., Camara, S., Samoura, B., & Magassouba, M. (1996). Applying an index of biotic integrity based on fish assemblages in a West African river.*Hydrobiologia*, *331*(1-3), 71-78.
Mexico, West Central	1		1						1		Weigel, B. M., Henne, L. J., & Martínez-Rivera, L. M. (2002). Macroinvertebrate-based index of biotic integrity for protection of streams in west-central Mexico. *Journal of the North American Benthological Society*,*21*(4), 686-700.
USA, Mississippi	1		1					1			Shields Jr, F. D., Knight, S. S., & Cooper, C. M. (1995). Use of the index of biotic integrity to assess physical habitat degradation in warmwater streams.*Hydrobiologia*, *312*(3), 191-208.
USA, Appalachian		1	1					1	1	1	Carlisle, D. M., Hawkins, C. P., Meador, M. R., Potapova, M., & Falcone, J. (2008). Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society*, *27*(1), 16-37.
Cameroon, Lower Ntem River	1		1					1			Toham, A. K., & Teugels, G. G. (1999). First data on an index of biotic integrity (IBI) based on fish assemblages for the assessment of the impact of deforestation in a tropical West African river system. *Hydrobiologia*, *397*, 29-38.

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
Mexico, Central	1				1			1			Lyons, J., Gutierrez-Hernandez, A., Diaz-Pardo, E., Soto-Galera, E., Medina-Nava, M., & Pineda-Lopez, R. (2000). Development of a preliminary index of biotic integrity (IBI) based on fish assemblages to assess ecosystem condition in the lakes of central Mexico. *Hydrobiologia*, *418*(1), 57-72.
USA, Mid-Atlantic	1						1		1		Llansó, R. J., Scott, L. C., Hyland, J. L., Dauer, D. M., Russell, D. E., & Kutz, F. W. (2002). An estuarine benthic index of biotic integrity for the mid-Atlantic region of the United States. II. Index development. *Estuaries*, *25*(6), 1231-1242.
USA, Minnesota	1				1			1			Drake, M. T., & Pereira, D. L. (2002). Development of a fish-based index of biotic integrity for small inland lakes in central Minnesota. *North American Journal of Fisheries Management*, *22*(4), 1105-1123.
USA, Maryland	1		1					1			Roth, N., Southerland, M., Chaillou, J., Klauda, R., Kazyak, P., Stranko, S., ... & Morgan II, R. (1998). Maryland biological stream survey: development of a fish index of biotic integrity. In *Monitoring Ecological Condition at Regional Scales* (pp. 89-106). Springer Netherlands.
New Zealand, All	1		1					1			Joy, M. K., & Death, R. G. (2004). Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. *Environmental Management*, *34*(3), 415-428.
USA, Northwest Great Plains	1		1					1			Bramblett, R. G., Johnson, T. R., Zale, A. V., & Heggem, D. G. (2005). Development and evaluation of a fish assemblage index of biotic integrity for northwestern Great Plains streams. *Transactions of the American Fisheries Society*, *134*(3), 624-640.

Country, Location	**Multimetric**	**Multivariate**	**Stream**	**River**	**Lake**	**Wetland**	**Estuary**	**Fish**	**Macroinvert.**	**Other**	**Reference**
USA, Western	1		1					1		1	Pont, D., Hughes, R. M., Whittier, T. R., & Schmutz, S. (2009). A predictive index of biotic integrity model for aquatic-vertebrate assemblages of western US streams. *Transactions of the american Fisheries Society*, *138*(2), 292-305.
Brazil, Rio Paraiba do Sul	1			1				1			Araujo, F. G., Fichberg, I., Pinto, B. C. T., & Peixoto, M. G. (2003). A preliminary index of biotic integrity for monitoring the condition of the Rio Paraíba do Sul, southeast Brazil. *Environmental management*, *32*(4), 516-526.
Australia, Eastern		1	1	1				1			Kennard, M. J., Pusey, B. J., Arthington, A. H., Harch, B. D., & Mackay, S. J. (2006). Development and application of a predictive model of freshwater fish assemblage composition to evaluate river health in eastern Australia.*Hydrobiologia*, *572*(1), 33-57.
USA, Midwest	1		1					1			Mundahl, N. D., & Simon, T. P. (1999). Development and application of an index of biotic integrity for coldwater streams of the upper Midwestern United States. *Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida*, 383-415.
USA, Florida	1				1			1			Schulz, E. J., Hoyer, M. V., & Canfield Jr, D. E. (1999). An index of biotic integrity: a test with limnological and fish data from sixty Florida lakes.*Transactions of the American Fisheries Society*, *128*(4), 564-577.
China, Upper Yangtze River	1			1				1			Zhu, D., & Chang, J. (2008). Annual variations of biotic integrity in the upper Yangtze River using an adapted index of biotic integrity (IBI). *Ecological Indicators*, *8*(5), 564-572.
USA, Chesapeake Bay	1						1			1	Lacouture, R. V., Johnson, J. M., Buchanan, C., & Marshall, H. G. (2006). Phytoplankton index of biotic integrity for Chesapeake Bay and its tidal tributaries. *Estuaries and Coasts*, *29*(4), 598-616.

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
Romania, All	1		1					1			Angermeier, P. L., & Davideanu, G. (2004). Using Fish Communities to Assess Streams in Romania: Initial Development of an Index of Biotic Integrity*.*Hydrobiologia*, *511*(1-3), 65-78.
China, Anhui Province	1		1						1		Wang, B. X., Yang, L. F., Hu, B. J., & SHAN, L. N. (2005). A preliminary study on the assessment of stream ecosystem health in south of Anhui Province using benthic-index of biotic integrity. *Acta Ecologica Sinica*, *25*(6), 1481-1490.
USA, Great Lakes	1				1					2	Kane, D. D., Gordon, S. I., Munawar, M., Charlton, M. N., & Culver, D. A. (2009). The Planktonic Index of Biotic Integrity (P-IBI): an approach for assessing lake ecosystem health. *Ecological Indicators*, *9*(6), 1234-1247.
USA, Texas	1		1					1			Linam, G. W., Kleinsasser, L. J., & Mayes, K. B. (2009). Regionalization of the index of biotic integrity for Texas streams.
USA, Wisconsin	1				1			1			Jennings, M. J., Lyons, J., Emmons, E. E., Hatzenbeler, G. R., Bozek, M. A., Simonson, T. D., ... & Fago, D. (1999). Toward the development of an index of biotic integrity for inland lakes in Wisconsin. *Assessing the sustainability and biological integrity of water resource quality using fish communities*, 541-562.
Great Britain, All; Australia, All; USA, California		1	1						1		Hawkins, C. P., Norris, R. H., Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Effects of taxonomic resolution and use of subsets of the fauna on the performance of RIVPACS-type models. In *Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997.* (pp. 217-228). Freshwater Biological Association (FBA).

Country, Location	Multimetric	Multivariate	Stream	River	Lake	Wetland	Estuary	Fish	Macroinvert.	Other	Reference
Great Britain, All; Sweden, All; Czech Republic, All		1	1						1		Davy-Bowker, J., Clarke, R. T., Johnson, R. K., Kokes, J., Murphy, J. F., & Zahradkova, S. (2006). A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates. In *The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods* (pp. 91-105). Springer Netherlands.
Spain, All		1	1						1		Poquet, J. M., Alba-Tercedor, J., Puntí, T., del Mar Sánchez-Montoya, M., Robles, S., Alvarez, M., ... & Prat, N. (2009). The MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. *Hydrobiologia*, *623*(1), 153-171.
USA, Wyoming		1	1						1		Hargett, E. G., ZumBerge, J. R., Hawkins, C. P., & Olson, J. R. (2007). Development of a RIVPACS-type predictive model for bioassessment of wadeable streams in Wyoming. *Ecological Indicators*, *7*(4), 807-826.
USA, Oregon		1	1						1		Hubler, S. (2008). PREDATOR: Development and use of RIVPACS-type macroinvertebrate models to assess the biotic condition of wadeable Oregon streams. *Unpublished report prepared by the Oregon Department of Environmental Quality, Watershed Assessment Section*.
USA, Eastern		1	1					1			Meador, M. R., & Carlisle, D. M. (2009). Predictive models for fish assemblages in eastern US streams: implications for assessing biodiversity.*Transactions of the American Fisheries Society*, *138*(4), 725-740.

APPENDIX B

TAXA OCCURRENCE BY BASIN AND ECOREGION

Table B.1. Percentages of all samples (including reference, stressed, validation, and no designation) that taxa were observed in, by ecoregion and basin. Dash "-" represents that taxa was not observed (0%) and tile "~" signifies observed in a marginal number (<0.5%) of samples within that ecoregion or basin. Cells are shaded with increasing gray opacity at 25, 50, and 75% thresholds.

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Aca_pom %	5	-	-	-	23	12	19	-	-	-	-	-	3	1	3	18	4	33	14	-	-
Amb_ari %	4	10	3	13	-	-	-	60	-	1	18	10	-	-	-	-	-	-	-	26	-
Amb_rup %	-	11	-	26	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	49
Ame_bru %	11	6	36	1	3	6	6	-	-	32	10	12	-	38	30	25	26	-	-	3	1
Ame_cat %	-	-	1	-	-	-	-	-	-	-	-	-	-	1	2	-	1	-	-	-	-
Ame_mel %	-	~	1	6	-	-	-	-	-	~	3	-	-	-	-	-	-	-	-	-	3
Ame_nat %	38	2	29	23	31	27	38	100	40	37	12	35	28	22	27	21	33	67	48	26	15
Ame_neb %	5	1	7	2	12	9	31	-	-	4	2	4	8	7	9	13	5	33	7	5	1
Ame_pla %	3	~	3	-	-	-	6	-	-	1	-	-	-	4	4	4	8	-	-	-	-
Ame_ser %	~	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-
Ame_spp %	~	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-
Ami_cal %	7	-	1	-	29	33	25	-	20	1	-	4	-	2	3	20	1	-	17	5	-
Ang_ros %	27	-	6	-	18	18	38	100	20	3	-	4	-	28	15	52	7	33	9	-	-
Aph_say %	79	-	9	-	59	58	69	100	60	16	-	59	85	16	49	75	24	100	69	-	-
Apl_gru %	-	-	~	5	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-
Cam_spp %	-	76	44	96	-	-	-	-	-	56	91	17	-	16	2	-	8	-	-	95	93
Cat_com %	-	~	3	22	-	-	-	-	-	9	-	-	-	~	-	-	-	-	-	-	27
Cen_mac %	11	-	1	-	53	45	50	-	40	-	-	-	8	4	6	41	5	67	28	-	-
Cho_cor %	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
Chr_ten %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Cli_fun %	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	17

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Cot_spp %	-	84	18	80	-	-	-	-	-	21	81	-	-	-	-	-	14	-	-	77	88
Cyp_cae %	-	~	-	-	-	-	-	-	-	-	~	-	-	-	-	-	-	-	-	-	-
Cyp_cam %	2	-	5	-	-	-	-	-	-	-	-	-	-	9	13	9	-	-	-	-	-
Cyp_cat %	-	31	7	7	-	-	-	-	-	-	43	-	-	-	-	-	-	-	-	21	-
Cyp_can %	-	-	~	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-
Cyp_gal %	-	3	-	11	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	17
Cyp_gib %	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	85	-
Cyp_lee %	2	-	-	-	-	-	-	-	-	-	-	-	-	1	2	4	-	-	-	-	-
Cyp_lut %	-	-	2	2	-	-	-	-	-	6	1	-	-	-	-	-	-	-	-	-	-
Cyp_niv %	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-
Cyp_hyb %	-	-	-	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Cyp_spi %	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
Cyp_tri %	-	32	3	23	-	-	-	-	-	-	47	-	-	-	-	-	-	-	-	-	-
Cyp_ven %	23	3	13	21	-	-	-	100	-	25	23	50	28	1	-	-	-	-	2	13	-
Cyp_xae %	-	-	4	-	-	-	-	-	-	-	-	-	-	6	10	-	-	-	-	-	-
Cyp_car %	-	1	1	5	-	-	-	-	-	1	2	-	-	~	3	-	1	-	-	-	3
Dor_cep %	~	~	3	7	-	-	-	-	-	3	4	1	3	1	2	-	2	-	-	3	2
Ela_eve %	1	-	-	-	7	3	-	-	-	-	-	-	-	-	-	4	2	-	5	-	-
Ela_zon %	5	-	-	-	2	6	13	-	-	-	-	1	3	3	2	5	-	-	-	-	-
Enn_glo %	1	-	-	-	21	24	13	-	-	-	-	-	-	~	-	9	2	-	5	-	-
Enn_obe %	1	-	-	-	27	9	13	-	-	-	-	-	-	-	1	16	1	33	16	-	-
Eri_ins %	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
Eri_obl %	22	-	18	-	14	6	38	-	-	14	-	12	-	25	41	39	16	-	7	-	-
Eri_spp %	-	-	~	-	1	-	-	-	-	-	-	-	-	-	1	2	1	-	-	-	-
Eri_suc %	9	-	~	-	17	9	6	-	20	1	-	4	8	1	-	23	4	67	19	-	-

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Eso_ame %	82	-	11	-	92	94	94	80	80	19	-	58	85	23	43	91	23	100	91	-	-
Eso_nig %	24	-	7	1	12	15	25	20	-	7	1	21	8	16	14	25	10	33	10	-	-
Eth_ble %	-	12	-	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	40
Eth_bre %	-	4	1	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-
Eth_cae %	-	-	-	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18
Eth_chl %	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Eth_chu %	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33	-
Eth_coo %	-	11	~	52	-	-	-	-	-	-	42	-	-	-	-	-	-	-	-	-	-
Eth_dur %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Eth_edw %	14	-	-	-	-	-	-	60	-	7	-	18	23	-	-	-	-	-	-	-	-
Eth_eto %	-	3	1	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-
Eth_fri %	~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
Eth_fus %	3	-	~	-	7	12	6	20	-	-	-	1	10	1	1	2	1	-	3	-	-
Eth_hop %	13	-	13	-	-	-	6	-	-	-	-	-	-	25	36	27	14	-	-	-	-
Eth_ins %	-	1	17	-	-	-	-	-	-	-	-	-	-	26	34	5	8	-	-	-	-
Eth_jes %	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17
Eth_jor %	-	3	~	9	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-	-
Eth_olm %	14	-	7	-	1	-	13	-	-	-	-	-	-	9	11	39	29	-	-	-	-
Eth_par %	1	-	~	-	-	-	-	-	-	~	-	-	-	~	1	-	-	-	-	-	-
Eth_ruf %	-	11	-	25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	48
Eth_sco %	-	14	5	-	-	-	-	-	-	-	22	-	-	-	-	-	-	-	-	-	-
Eth_ser %	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-
Eth_sti %	-	5	3	25	-	-	-	-	-	-	22	-	-	-	-	-	-	-	-	28	-
Eth_swa %	13	-	2	-	-	-	-	40	-	3	-	34	21	-	-	-	-	-	-	-	-
Eth_tal %	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	85	-

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Eth_ten %	-	-	-	34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	40
Eth_tri %	-	~	-	3	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-
Eth_zon %	-	4	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
Fun_cat %	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11
Fun_chr %	1	-	-	-	2	-	-	-	-	-	-	1	3	-	-	-	1	-	3	-	-
Fun_esc %	1	-	-	-	-	-	-	-	-	~	-	1	-	-	-	-	-	-	-	-	-
Fun_lin %	4	-	-	-	3	3	6	-	-	-	-	1	-	2	2	5	2	-	3	-	-
Fun_oli %	6	-	~	29	-	-	-	-	-	10	4	-	-	-	-	-	-	-	-	3	27
Fun_ste %	-	11	10	29	-	-	-	-	-	14	38	-	-	-	-	-	-	-	-	-	-
Gam_spp %	52	1	28	51	50	45	56	100	60	29	25	41	62	25	41	66	33	-	47	21	25
Hem_fla %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Het_for %	~	-	-	-	3	6	-	-	-	-	-	1	-	-	-	2	-	-	-	-	-
Hyb_spp %	19	3	38	19	-	-	-	20	-	24	2	52	5	22	42	27	39	-	-	72	23
Hyp_eto %	-	47	29	52	-	-	-	-	-	39	91	-	-	-	-	-	-	-	-	97	-
Hyp_nig %	1	32	11	32	-	-	-	-	-	-	-	-	-	-	13	-	50	-	-	-	83
Ich_cas %	-	6	~	1	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	6
Ich_gag %	23	16	17	15	-	-	-	100	-	31	33	48	10	-	-	-	-	-	-	69	-
Ich_gre %	-	17	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	29
Ict_fur %	-	-	-	1	-	-	-	-	-	-	~	-	-	-	-	-	-	-	-	-	-
Ict_pun %	3	2	5	6	1	-	-	-	-	3	6	5	8	2	5	5	1	-	2	13	1
Lab_sic %	30	-	6	1	11	15	38	20	20	11	-	27	38	10	13	21	3	-	22	-	1
Lam_aep %	-	-	-	1	-	-	-	-	-	-	~	-	-	-	-	-	-	-	-	-	-
Lep_ocu %	1	-	~	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-
Lep_oss %	1	-	1	-	2	3	-	-	-	-	-	-	-	1	2	7	1	-	3	-	-
Lep_pla %	1	-	~	-	1	-	-	-	-	-	-	-	-	~	-	4	-	-	2	-	-

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Lei_spp %	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
Lep_aur %	85	55	88	96	31	27	63	100	40	77	83	85	95	89	84	73	72	67	62	100	70
Lep_cya %	22	30	40	96	-	-	6	-	-	46	61	19	51	20	22	9	30	-	33	67	62
Lep_gul %	66	1	32	51	52	61	81	40	60	35	22	54	77	38	41	71	22	67	76	36	26
Lep_mac %	72	51	78	92	44	61	56	100	60	70	80	76	90	72	66	70	50	33	69	100	69
Lep_mar %	55	-	1	-	42	42	81	80	40	3	-	23	51	17	26	61	14	33	47	-	-
Lep_meg %	2	5	5	54	-	-	-	-	-	6	33	-	-	7	-	2	3	-	-	3	9
Lep_mic %	21	3	15	51	6	9	19	-	60	15	24	22	31	10	14	18	9	-	14	23	28
Lep_pun %	82	2	14	61	38	42	63	100	60	35	31	79	90	24	27	66	14	100	55	3	17
Lep_spp %	9	5	9	41	8	6	6	-	-	11	18	6	31	5	5	5	5	-	17	3	28
Let_app %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Lux_chr %	-	2	6	82	-	-	-	-	-	-	29	-	-	-	-	-	-	-	-	97	43
Lux_coc %	-	18	-	17	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	46
Lux_zon %	4	12	16	-	-	-	-	100	-	48	9	1	-	-	-	-	2	-	-	15	1
Lyt_atr %	1	-	2	-	-	-	-	-	-	6	-	3	-	-	-	-	-	-	-	-	-
Lyt_bel %	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	-
Lyt_fas %	-	-	-	21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24
Lyt_lir %	-	~	-	17	-	-	-	-	-	-	11	-	-	-	-	-	-	-	-	-	-
Mic_cat %	2	-	2	-	-	-	-	-	-	2	-	10	-	-	-	-	-	-	-	-	-
Mic_dol %	-	3	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
Mic_hen %	-	5	3	9	-	-	-	-	-	-	14	-	-	-	1	-	-	-	-	10	-
Mic_pun %	1	6	4	5	-	-	-	-	-	11	-	-	-	2	-	2	-	-	-	-	17
Mic_sal %	49	23	43	56	18	18	31	100	40	46	37	60	62	32	37	41	32	33	40	54	46
Mic_spp %	-	42	28	58	-	-	-	-	-	16	78	-	-	13	14	9	11	-	-	67	17
Min_mel %	31	2	16	24	1	-	13	100	-	23	15	49	13	15	14	23	11	-	5	10	9

Taxa	Ecoregion					Basin															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Mor_chr %	-	-	~	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-
Mor_sax %	~	-	-	-	-	-	-	-	-	~	-	-	-	-	-	-	-	-	-	-	-
Mor_spp %	-	-	~	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-
Mox_car %	-	~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Mox_col %	2	~	4	-	-	-	-	-	-	-	-	-	-	5	6	9	5	-	-	-	-
Mox_duq %	-	31	7	48	-	-	-	-	-	-	47	-	-	-	-	-	-	-	-	36	36
Mox_ery %	-	13	1	43	-	-	-	-	-	-	26	-	-	-	-	-	-	-	-	3	25
Mox_lac %	4	~	8	-	-	-	-	-	-	22	-	15	-	-	-	-	-	-	-	-	-
Mox_poe %	-	5	5	12	-	-	-	-	-	-	14	-	-	-	-	-	-	-	-	59	-
Mox_rup %	1	5	30	-	-	-	-	-	-	7	-	-	-	53	39	-	26	-	-	-	-
Mox_spa %	5	-	3	-	-	-	-	-	-	10	-	14	-	-	-	-	-	-	-	-	-
Mox_spb %	-	-	2	-	-	-	-	-	-	-	-	-	-	~	2	11	4	-	-	-	-
Mug_cep %	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
Noc_spp %	12	42	81	10	-	-	-	-	-	54	20	36	-	78	66	25	78	-	-	95	40
Not_cry %	19	3	11	8	30	24	31	20	60	12	6	10	10	15	17	36	12	33	16	5	7
Not_amp %	16	-	21	-	-	-	-	80	-	35	-	43	-	28	3	-	-	-	-	-	-
Not_asp %	-	~	-	3	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-
Not_bai %	-	-	1	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-
Not_cha %	4	-	-	-	-	-	6	-	-	-	-	1	-	2	1	11	-	-	-	-	-
Not_chr %	-	8	2	35	-	-	-	-	-	-	32	-	-	-	-	-	-	-	-	-	-
Not_cum %	26	-	3	-	4	-	6	-	-	-	-	4	36	13	20	46	12	-	-	-	-
Not_har %	7	-	-	-	-	-	-	-	-	5	-	9	-	2	-	-	-	-	-	-	-
Not_hud %	2	~	19	-	-	-	-	-	-	4	-	10	-	23	30	11	13	-	-	-	-
Not_hyp %	-	-	4	-	-	-	-	-	-	8	-	6	-	-	-	-	-	-	-	-	-
Not_leu %	-	16	~	4	-	-	-	-	-	1	-	-	-	-	-	-	2	-	-	-	28

Taxa	**Ecoregion**					**Basin**															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Not_lon %	17	-	11	-	-	-	-	-	-	34	1	41	-	2	-	-	-	-	-	-	-
Not_lut %	13	13	60	-	-	-	-	-	-	16	11	38	-	74	65	25	72	-	-	-	2
Not_mac %	6	-	-	-	6	12	-	-	-	1	-	3	-	~	2	13	1	-	5	-	-
Not_pet %	17	-	3	-	4	9	13	20	20	2	-	7	-	14	13	18	6	-	-	-	-
Not_sce %	-	-	~	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-
Not_spp %	~	~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	1
Not_spe %	-	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	18
Not_sti %	-	11	5	16	-	-	-	-	-	-	27	-	-	-	-	-	-	-	-	21	-
Not_tel %	-	~	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
Not_tex %	39	-	8	-	-	-	-	100	20	22	-	65	77	2	2	-	-	-	31	-	-
Not_vol %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Not_xae %	-	25	5	39	-	-	-	-	-	-	56	-	-	-	-	-	-	-	-	-	-
Not_fun %	-	-	8	-	-	-	-	-	-	10	-	-	-	-	-	-	-	-	-	72	-
Not_gyr %	7	-	5	-	11	12	19	-	-	~	-	1	8	1	17	27	7	-	5	-	-
Not_ins %	1	-	12	-	-	-	-	-	-	-	-	-	-	11	15	2	29	-	-	-	-
Not_lep %	43	11	13	4	3	6	6	100	-	23	15	53	62	11	6	27	9	-	28	21	-
Onc_myk %	-	52	2	7	-	-	-	-	-	12	15	-	-	-	-	-	18	-	-	-	40
Ops_emi %	14	-	1	-	1	3	-	40	-	3	-	22	38	1	-	5	-	-	2	-	-
Per_fla %	1	8	6	-	-	-	-	-	-	10	3	1	-	1	6	-	3	-	-	10	7
Per_aur %	-	2	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Per_cap %	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
Per_evi %	-	10	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17
Per_kat %	-	19	3	26	-	-	-	-	-	-	37	-	-	-	-	-	-	-	-	5	-
Per_kus %	-	1	~	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Per_nig %	60	23	54	41	6	-	19	100	-	73	61	82	62	41	33	39	36	-	38	15	-

Taxa	**Ecoregion**					**Basin**															
	SEP	BRM	PDM	RGV	SCP	SAT	ALT	APA	AUC	CHT	COO	FLI	OCH	OCM	OCO	OGE	SAV	STM	SUW	TAL	TEN
Per_pal %	-	25	5	1	-	-	-	-	-	-	29	-	-	-	-	-	-	-	-	23	-
Per_sci %	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
Per_smi %	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	62	-
Pet_spp %	-	2	-	11	-	-	-	-	-	-	~	-	-	-	-	-	-	-	-	-	15
Phe_cat %	-	7	2	5	-	-	-	-	-	-	10	-	-	-	-	-	-	-	-	21	-
Phe_cra %	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Phe_ura %	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Pim_spp %	-	-	2	15	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	23	16
Pom_ann %	-	-	~	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
Pom_nig %	7	3	12	24	6	12	6	-	20	12	12	12	5	9	11	11	6	-	-	15	12
Pte_spp %	38	-	~	-	3	6	6	100	-	15	-	49	18	6	6	7	2	-	9	-	-
Pyl_oli %	1	~	1	1	-	-	-	-	-	~	1	1	-	2	1	-	-	-	-	-	-
Rhi_spp %	-	20	-	32	-	-	-	-	-	-	7	-	-	-	-	-	7	-	-	-	52
Sal_tru %	-	25	~	-	-	-	-	-	-	5	4	-	-	-	-	-	14	-	-	-	16
Sal_fon %	-	3	-	-	-	-	-	-	-	-	~	-	-	-	-	-	3	-	-	-	1
Sem_spp %	11	77	59	69	-	-	-	40	-	55	83	29	3	23	43	9	54	-	-	85	75
Tri_mac %	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
Umb_pyg %	~	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-

APPENDIX C

ABBREVIATIONS FOR TAXA, BASINS, AND ECOREGIONS

Table C.1. Abbreviations of Georgia's Basins and Ecoregions.

Acronym	Description
ALT	Altamaha Basin
APA	Apalachicola Basin
AUC	Aucilla Basin
CHT	Chattahoochee Basin
COO	Coosa Basin
FLI	Flint Basin
OCH	Ochlockonee Basin
OCM	Ocmulgee Basin
OCO	Oconee Basin
OGE	Ogeechee Basin
SAT	Satilla Basin
SAV	Savannah Basin
STM	St. Marys Basin
SUW	Suwanee Basin
TAL	Tallapoosa Basin
TEN	Tennessee Basin
BRM	Blue Ridge Ecoregion
PDM	Piedmont Ecoregion
RGV	Ridge and Valley Ecoregion
SCP	Southern Coastal Plain
SEP	Southeastern Plains Ecoregion

Table C.2. Abbreviations of stream fish taxa observed in GADNR Stream Team samples.

Abbreviation	**Scientific Name**	**Common Name**
Aca_pom	*Acantharchus pomotis*	Mud sunfish
Amb_ari	*Ambloplites ariommus*	Shadow bass
Amb_rup	*Ambloplites rupestris*	Rock bass
Ame_bru	*Ameiurus brunneus*	Snail bullhead
Ame_cat	*Ameiurus catus*	White catfish
Ame_mel	*Ameiurus melas*	Black bullhead
Ame_nat	*Ameiurus natalis*	Yellow bullhead
Ame_neb	*Ameiurus nebulosus*	Brown bullhead
Ame_pla	*Ameiurus platycephalus*	Flat bullhead
Ame_ser	*Ameiurus serracanthus*	Spotted bullhead
Ame_spp	*Ameiurus spp*	Ameiurus species
Ami_cal	*Amia calva*	Bowfin
Ang_ros	*Anguilla rostrata*	American eel
Aph_say	*Aphredoderus sayanus*	Pirate perch
Apl_gru	*Aplodinotus grunniens*	Freshwater drum
Cam_spp	*Campostoma anomalum*	Central stoneroller
Cam_spp	*Campostoma oligolepis*	Largescale stoneroller
Cam_spp	*Campostoma pauciradii*	Bluefin stoneroller
Cat_com	*Catostomus commersonii*	White sucker
Cen_mac	*Centrarchus macropterus*	Flier
Cho_cor	*Chologaster cornuta*	Swampfish
Chr_ten	*Chrosomus tennesseensis*	Tennessee dace

Abbreviation	**Scientific Name**	**Common Name**
Cli_fun	*Clinostomus funduloides*	Rosyside dace
Cot_spp	*Cottus bairdi*	Mottled sculpin
Cot_spp	*Cottus carolinae*	Banded sculpin
Cot_spp	*Cottus chattahoochae*	Chattahoochee sculpin
Cot_spp	*Cottus spp*	Cottus species
Cot_spp	*Cottus tallapoosae*	Tallapoosa sculpin
Cyp_cae	*Cyprinella caerulea*	Blue shiner
Cyp_cam	*Cyprinella callisema*	Ocmulgee shiner
Cyp_cat	*Cyprinella callistia*	Alabama shiner
Cyp_can	*Cyprinella callitaenia*	Bluestripe shiner
Cyp_gal	*Cyprinella galactura*	Whitetail shiner
Cyp_gib	*Cyprinella gibbsi*	Tallapoosa shiner
Cyp_lee	*Cyprinella leedsi*	Bannerfin shiner
Cyp_lut	*Cyprinella lutrensis*	Red Shiner
Cyp_niv	*Cyprinella nivea*	Whitefin shiner
Cyp_hyb	*Cyprinella hybrid*	Cyprinella hybrid
Cyp_spi	*Cyprinella spiloptera*	Spotfin shiner
Cyp_tri	*Cyprinella trichroistia*	Tricolor shiner
Cyp_ven	*Cyprinella venusta*	Blacktail shiner
Cyp_xae	*Cyprinella xaenura*	Altamaha shiner
Cyp_car	*Cyprinus carpio*	Common carp
Dor_cep	*Dorosoma cepedianum*	Gizzard shad
Ela_eve	*Elassoma evergladei*	Everglades pygmy sunfish
Ela_zon	*Elassoma zonatum*	Banded pygmy sunfish
Enn_glo	*Enneacanthus gloriosus*	Bluespotted sunfish

Abbreviation	**Scientific Name**	**Common Name**
Enn_obe	*Enneacanthus obesus*	Banded sunfish
Eri_ins	*Erimystax insignis*	Blotched chub
Eri_obl	*Erimyzon oblongus*	Creek chubsucker
Eri_spp	*Erimyzon spp*	Erimyzon species
Eri_suc	*Erimyzon sucetta*	Lake chubsucker
Eso_ame	*Esox americanus*	Redfin pickerel
Eso_nig	*Esox niger*	Chain pickerel
Eth_ble	*Etheostoma blennioides*	Greenside darter
Eth_bre	*Etheostoma brevirostrum*	Holiday darter
Eth_cae	*Etheostoma caeruleum*	Rainbow darter
Eth_chl	*Etheostoma chlorobranchium*	Greenfin darter
Eth_chu	*Etheostoma chuckwachatte*	Lipstick darter
Eth_coo	*Etheostoma coosae*	Coosa darter
Eth_dur	*Etheostoma duryi*	Black darter
Eth_edw	*Etheostoma edwini*	Brown darter
Eth_eto	*Etheostoma etowahae*	Etowah darter
Eth_fri	*Etheostoma fricksium*	Savannah darter
Eth_fus	*Etheostoma fusiforme*	Swamp darter
Eth_hop	*Etheostoma hopkinsi*	Christmas darter
Eth_ins	*Etheostoma inscriptum*	Turquoise darter
Eth_jes	*Etheostoma jessiae*	Blueside darter
Eth_jor	*Etheostoma jordani*	Greenbreast darter
Eth_olm	*Etheostoma olmstedi*	Tessellated darter
Eth_par	*Etheostoma parvipinne*	Goldstripe darter
Eth_ruf	*Etheostoma rufilineatum*	Redline darter

Abbreviation	**Scientific Name**	**Common Name**
Eth_sco	*Etheostoma scotti*	Cherokee darter
Eth_ser	*Etheostoma serrifer*	Sawcheek darter
Eth_sti	*Etheostoma stigmaeum*	Speckled darter
Eth_swa	*Etheostoma swaini*	Gulf darter
Eth_tal	*Etheostoma tallapoosae*	Tallapoosa darter
Eth_ten	*Etheostoma tennesseense*	Tennessee darter
Eth_tri	*Etheostoma trisella*	Trispot darter
Eth_zon	*Etheostoma zonale*	Banded darter
Fun_cat	*Fundulus catenatus*	Northern studfish
Fun_chr	*Fundulus chrysotus*	Golden topminnow
Fun_esc	*Fundulus escambiae*	Russetfin topminnow
Fun_lin	*Fundulus lineolatus*	Lined topminnow
Fun_oli	*Fundulus olivaceus*	Blackspotted topminnow
Fun_ste	*Fundulus stellifer*	Southern studfish
Gam_spp	*Gambusia affinis*	Western mosquitofish
Gam_spp	*Gambusia holbrooki*	Eastern mosquitofish
Gam_spp	*Gambusia spp*	Moquitofish species
Hem_fla	*Hemitremia flammea*	Flame chub
Het_for	*Heterandria formosa*	Least killifish
Hyb_spp	*Hybognathus regius*	Eastern silvery minnow
Hyb_spp	*Hybopsis amblops*	Bigeye chub
Hyb_spp	*Hybopsis lineapunctata*	Lined chub
Hyb_spp	*Hybopsis rubrifrons*	Rosyface chub
Hyb_spp	*Hybopsis winchelli*	Clear Chub
Hyp_eto	*Hypentelium etowanum*	Alabama hogsucker

Abbreviation	**Scientific Name**	**Common Name**
Hyp_nig	*Hypentelium nigricans*	Northern hogsucker
Ich_cas	*Ichthyomyzon castaneus*	Chestnut lamprey
Ich_gag	*Ichthyomyzon gagei*	Southern brook lamprey
Ich_gre	*Ichthyomyzon greeleyi*	Mountain brook lamprey
Ict_fur	*Ictalurus furcatus*	Blue catfish
Ict_pun	*Ictalurus punctatus*	Channel catfish
Lab_sic	*Labidesthes sicculus*	Brook silverside
Lam_aep	*Lampetra aepyptera*	Least brook lamprey
Lep_ocu	*Lepisosteus oculatus*	Spotted gar
Lep_oss	*Lepisosteus osseus*	Longnose gar
Lep_pla	*Lepisosteus platyrhincus*	Florida gar
Lei_spp	*Lepisosteus spp*	Lepisosteus species
Lep_aur	*Lepomis auritus*	Redbreast sunfish
Lep_cya	*Lepomis cyanellus*	Green sunfish
Lep_gul	*Lepomis gulosus*	Warmouth
Lep_mac	*Lepomis macrochirus*	Bluegill
Lep_mar	*Lepomis marginatus*	Dollar sunfish
Lep_meg	*Lepomis megalotis*	Longear sunfish
Lep_mic	*Lepomis microlophus*	Redear sunfish
Lep_pun	*Lepomis punctatus*	Spotted sunfish
Lep_spp	*Lepomis spp*	Lepomis species
Let_app	*Lethenteron appendix*	American brook lamprey
Lux_chr	*Luxilus chrysocephalus*	Striped shiner
Lux_coc	*Luxilus coccogenis*	Warpaint shiner
Lux_zon	*Luxilus zonistius*	Bandfin shiner

Abbreviation	**Scientific Name**	**Common Name**
Lyt_atr	*Lythrurus atrapiculus*	Blacktip shiner
Lyt_bel	*Lythrurus bellus*	Pretty shiner
Lyt_fas	*Lythrurus fasciolaris*	Scarlet shiner
Lyt_lir	*Lythrurus lirus*	Mountain shiner
Mic_cat	*Micropterus cataractae*	Shoal bass
Mic_spp	*Micropterus chattahoochae*	Redeye bass Chattahoochae
Mic_spp	*Micropterus coosae*	Redeye bass
Mic_dol	*Micropterus dolomieu*	Smallmouth bass
Mic_hen	*Micropterus henshalli*	Alabama bass
Mic_pun	*Micropterus punctulatus*	Spotted bass
Mic_sal	*Micropterus salmoides*	Largemouth bass
Mic_spp	*Micropterus sp Altamaha bass*	Altamaha Black Bass
Mic_spp	*Micropterus sp Bartrams bass*	Bartrams Black Bass
Mic_spp	*Micropterus tallapoosae*	Tallapoosa Black Bass
Min_mel	*Minytrema melanops*	Spotted sucker
Mor_chr	*Morone chrysops*	White bass
Mor_sax	*Morone saxatilis*	Striped bass
Mor_spp	*Morone spp*	Morone species
Mox_car	*Moxostoma carinatum*	River redhorse
Mox_col	*Moxostoma collapsum*	Notchlip redhorse
Mox_duq	*Moxostoma duquesnei*	Black redhorse
Mox_ery	*Moxostoma erythrurum*	Golden redhorse
Mox_lac	*Moxostoma lachneri*	Greater jumprock
Mox_poe	*Moxostoma poecilurum*	Blacktail redhorse
Mox_rup	*Moxostoma rupiscartes*	Striped jumprock

Abbreviation	**Scientific Name**	**Common Name**
Mox_spa	*Moxostoma sp Apalachicola redhorse*	Apalachicola redhorse
Mox_spb	*Moxostoma sp Brassy Jumprock*	Brassy jumprock
Mug_cep	*Mugil cepahlus*	Flathead grey mullet
Noc_spp	*Nocomis leptocephalus*	Bluehead chub
Noc_spp	*Nocomis micropogon*	River chub
Not_cry	*Notemigonus crysoleucas*	Golden shiner
Not_amp	*Notropis amplamala*	Longjaw minnow
Not_asp	*Notropis asperifrons*	Burrhead shiner
Not_bai	*Notropis baileyi*	Rough shiner
Not_cha	*Notropis chalybaeus*	Ironcolor shiner
Not_chr	*Notropis chrosomus*	Rainbow shiner
Not_cum	*Notropis cummingsae*	Dusky shiner
Not_har	*Notropis harperi*	Redeye chub
Not_hud	*Notropis hudsonius*	Spottail shiner
Not_hyp	*Notropis hypsilepis*	Highscale shiner
Not_leu	*Notropis leuciodus*	Tennessee shiner
Not_lon	*Notropis longirostris*	Longnose shiner
Not_lut	*Notropis lutipinnis*	Yellowfin shiner
Not_mac	*Notropis maculatus*	Taillight shiner
Not_pet	*Notropis petersoni*	Coastal shiner
Not_sce	*Notropis scepticus*	Sandbar shiner
Not_spp	*Notropis spp*	Notropis species
Not_spe	*Notropis spectrunculus*	Mirror shiner
Not_sti	*Notropis stilbius*	Silverstripe shiner
Not_tel	*Notropis telescopus*	Telescope shiner

Abbreviation	**Scientific Name**	**Common Name**
Not_tex	*Notropis texanus*	Weed shiner
Not_vol	*Notropis volucellus*	Mimic shiner
Not_xae	*Notropis xaenocephalus*	Coosa shiner
Not_fun	*Noturus funebris*	Black madtom
Not_gyr	*Noturus gyrinus*	Tadpole madtom
Not_ins	*Noturus insignis*	Margined madtom
Not_lep	*Noturus leptacanthus*	Speckled madtom
Onc_myk	*Oncorhynchus mykiss*	Rainbow trout
Ops_emi	*Opsopoeodus emiliae*	Pugnose minnow
Per_fla	*Perca flavescens*	Yellow perch
Per_aur	*Percina aurolineata*	Goldline darter
Per_cap	*Percina caprodes*	Logperch
Per_evi	*Percina evides*	Gilt darter
Per_kat	*Percina kathae*	Mobile logperch
Per_kus	*Percina kusha*	Bridled darter
Per_nig	*Percina nigrofasciata*	Blackbanded darter
Per_pal	*Percina palmaris*	Bronze darter
Per_sci	*Percina sciera*	Dusky darter
Per_smi	*Percina smithvanizi*	Muscadine darter
Pet_spp	*Petromyzontidae spp*	Petromyzontidae species
Phe_cat	*Phenacobius catostomus*	Riffle minnow
Phe_cra	*Phenacobius crassilabrum*	Fatlips minnow
Phe_ura	*Phenacobius uranops*	Stargazing minnow
Pim_spp	*Pimephales notatus*	Bluntnose minnow
Pim_spp	*Pimephales vigilax*	Bullhead minnow

Abbreviation	Scientific Name	Common Name
Pom_ann	*Pomoxis annularis*	White crappie
Pom_nig	*Pomoxis nigromaculatus*	Black crappie
Pte_spp	*Pteronotropis euryzonus*	Broadstripe shiner
Pte_spp	*Pteronotropis grandipinnis*	Apalachee shiner
Pte_spp	*Pteronotropis metallicus*	Metallic Shiner
Pte_spp	*Pteronotropis stonei*	Lowland shiner
Pyl_oli	*Pylodictis olivaris*	Flathead catfish
Rhi_spp	*Rhinichthys atratulus*	Blacknose dace
Rhi_spp	*Rhinichthys cataractae*	Longnose dace
Sal_tru	*Salmo trutta*	Brown trout
Sal_fon	*Salvelinus fontinalis*	Brook trout
Sem_spp	*Semotilus atromaculatus*	Creek chub
Sem_spp	*Semotilus thoreauianus*	Dixie chub
Tri_mac	*Trinectes maculatus*	Hogchoker
Umb_pyg	*Umbra pygmaea*	Eastern mudminnow

APPENDIX D

MEAN TAXA RESPONSES IN DISTURBED SITES

Table D.1. Mean taxa O/E at stressed sites indicates taxa which were found more or less often than expected in disturbed areas. These “increasers” taxa are associated with disturbed areas whereas “decreasers” are less likely to be present there (*sensu* Meador & Carlisle, 2009). Median of mean taxa O/E is 0.70. The traits were synthesized by the GADNR Stream Team.

Taxa O/E	**Family**	**Genus**	**Species**	**Feeding**	**Habitat**	**Tolerance**
2.60	Centrarchidae	*Micropterus*	*punctulatus*	predator	habitat generalist	high
2.23	Cyprinidae	*Notropis*	*spectrunculus*	invertivore	fluvial specialist	medium
2.23	Percidae	*Percina*	*evides*	invertivore	fluvial specialist	low
2.22	Sciaenidae	*Aplodinotus*	*grunniens*	predator	habitat generalist	medium
2.11	Cyprinidae	*Notemigonus*	*crysoleucas*	generalist	habitat generalist	high
2.10	Centrarchidae	*Pomoxis*	*nigromaculatus*	generalist	habitat generalist	medium
2.01	Poeciliidae	*Gambusia*	*spp.*	invertivore	habitat generalist	high
1.79	Ictaluridae	*Ameiurus*	*melas*	generalist	habitat generalist	high
1.79	Catostomidae	*Catostomus*	*commersonii*	generalist	habitat generalist	high
1.69	Cyprinidae	*Pimephales*	*spp.*	invertivore	habitat generalist	high
1.68	Centrarchidae	*Micropterus*	*salmoides*	predator	habitat generalist	high
1.65	Catostomidae	*Moxostoma*	*poecilurum*	invertivore	fluvial specialist	medium
1.58	Ictaluridae	*Ameiurus*	*natalis*	generalist	habitat generalist	high
1.57	Centrarchidae	*Lepomis*	*microlophus*	invertivore	habitat generalist	medium
1.54	Lepisosteidae	*Lepisosteus*	*platyrhincus*	predator	habitat generalist	medium
1.49	Petromyzontidae	*Ichthyomyzon*	*greeleyi*	generalist	fluvial specialist	low
1.48	Cyprinidae	*Notropis*	*baileyi*	invertivore	fluvial specialist	medium
1.47	Centrarchidae	*Lepomis*	*cyanellus*	predator	habitat generalist	high
1.47	Percidae	*Perca*	*flavescens*	predator	habitat generalist	medium
1.43	Cyprinidae	*Clinostomus*	*funduloides*	invertivore	fluvial specialist	medium
1.38	Ictaluridae	*Ictalurus*	*punctatus*	predator	habitat generalist	medium
1.33	Centrarchidae	*Lepomis*	*gulosus*	predator	habitat generalist	medium
1.33	Ictaluridae	*Ameiurus*	*nebulosus*	generalist	habitat generalist	high
1.26	Centrarchidae	*Lepomis*	*macrochirus*	invertivore	habitat generalist	high
1.25	Cyprinidae	*Notropis*	*leuciodus*	invertivore	fluvial specialist	medium
1.21	Centrarchidae	*Lepomis*	*spp.*	invertivore	habitat generalist	unknown

Taxa O/E	Family	Genus	Species	Feeding	Habitat	Tolerance
1.19	Cyprinidae	*Cyprinus*	*carpio*	generalist	habitat generalist	high
1.13	Cyprinidae	*Cyprinella*	*venusta*	invertivore	fluvial specialist	medium
1.12	Clupeidae	*Dorosoma*	*cepedianum*	generalist	habitat generalist	high
1.11	Centrarchidae	*Micropterus*	*henshalli*	predator	habitat generalist	medium
1.05	Cyprinidae	*Cyprinella*	*xaenura*	invertivore	fluvial specialist	low
1.04	Centrarchidae	*Lepomis*	*auritus*	invertivore	habitat generalist	high
1.04	Percidae	*Percina*	*nigrofasciata*	invertivore	fluvial specialist	medium
1.03	Fundulidae	*Fundulus*	*olivaceus*	invertivore	fluvial specialist	medium
1.03	Cyprinidae	*Notropis*	*maculatus*	invertivore	fluvial specialist	medium
1.00	Catostomidae	*Minytrema*	*melanops*	invertivore	fluvial specialist	medium
0.99	Catostomidae	*Moxostoma*	*erythrurum*	invertivore	fluvial specialist	medium
0.96	Cyprinidae	*Notropis*	*petersoni*	invertivore	fluvial specialist	medium
0.94	Centrarchidae	*Lepomis*	*megalotis*	invertivore	habitat generalist	medium
0.93	Cyprinidae	*Nocomis*	*spp.*	generalist	fluvial specialist	medium
0.92	Ictaluridae	*Ameiurus*	*brunneus*	generalist	fluvial specialist	medium
0.91	Percidae	*Etheostoma*	*scotti*	invertivore	fluvial specialist	medium
0.90	Cyprinidae	*Notropis*	*lutipinnis*	invertivore	fluvial specialist	high
0.90	Catostomidae	*Hypentelium*	*nigricans*	invertivore	fluvial specialist	high
0.90	Cyprinidae	*Notropis*	*stilbius*	invertivore	fluvial specialist	medium
0.90	Cyprinidae	*Campostoma*	*spp.*	herbivore	fluvial specialist	high
0.90	Catostomidae	*Moxostoma*	*rupiscartes*	invertivore	fluvial specialist	medium
0.87	Cyprinidae	*Notropis*	*longirostris*	invertivore	fluvial specialist	high
0.86	Cyprinidae	*Phenacobius*	*catostomus*	invertivore	fluvial specialist	low
0.86	Cyprinidae	*Cyprinella*	*callisema*	invertivore	fluvial specialist	medium
0.85	Centrarchidae	*Lepomis*	*punctatus*	invertivore	habitat generalist	medium
0.84	Catostomidae	*Erimyzon*	*oblongus*	invertivore	fluvial specialist	medium
0.84	Esocidae	*Esox*	*americanus*	predator	habitat generalist	medium
0.82	Fundulidae	*Fundulus*	*stellifer*	invertivore	fluvial specialist	medium

Taxa O/E	Family	Genus	Species	Feeding	Habitat	Tolerance
0.82	Cyprinidae	*Luxilus*	*coccogenis*	invertivore	fluvial specialist	medium
0.81	Cyprinidae	*Luxilus*	*chrysocephalus*	invertivore	fluvial specialist	high
0.80	Cyprinidae	*Notropis*	*texanus*	invertivore	fluvial specialist	medium
0.79	Catostomidae	*Hypentelium*	*etowanum*	invertivore	fluvial specialist	high
0.78	Percidae	*Etheostoma*	*coosae*	invertivore	fluvial specialist	medium
0.77	Cyprinidae	*Semotilus*	*spp*	generalist	fluvial specialist	high
0.77	Centrarchidae	*Enneacanthus*	*gloriosus*	invertivore	swamp specialist	medium
0.77	Lepisosteidae	*Lepisosteus*	*oculatus*	predator	habitat generalist	medium
0.74	Petromyzontidae	*Ichthyomyzon*	*castaneus*	parasitic	fluvial specialist	medium
0.73	Percidae	*Etheostoma*	*fusiforme*	invertivore	fluvial specialist	medium
0.72	Catostomidae	*Erimyzon*	*sucetta*	invertivore	habitat generalist	medium
0.71	Cyprinidae	*Notropis*	*hudsonius*	invertivore	fluvial specialist	medium
0.71	Aphredoderidae	*Aphredoderus*	*sayanus*	invertivore	swamp specialist	high
0.70	Amiidae	*Amia*	*calva*	predator	habitat generalist	medium
0.70	Percidae	*Etheostoma*	*edwini*	invertivore	fluvial specialist	low
0.70	Percidae	*Etheostoma*	*duryi*	invertivore	fluvial specialist	medium
0.70	Percidae	*Etheostoma*	*zonale*	invertivore	fluvial specialist	medium
0.70	Percidae	*Percina*	*caprodes*	invertivore	habitat generalist	medium
0.70	Percidae	*Percina*	*sciera*	invertivore	fluvial specialist	medium
0.69	Salmonidae	*Salmo*	*trutta*	predator	fluvial specialist	medium
0.66	Atherinopsidae	*Labidesthes*	*sicculus*	invertivore	habitat generalist	medium
0.66	Percidae	*Percina*	*kathae*	invertivore	fluvial specialist	medium
0.65	Cyprinidae	*Pteronotropis*	*spp.*	invertivore	fluvial specialist	medium
0.62	Catostomidae	*Moxostoma*	*collapsum*	invertivore	fluvial specialist	low
0.62	Percidae	*Etheostoma*	*stigmaeum*	invertivore	fluvial specialist	medium
0.61	Petromyzontidae	*Ichthyomyzon*	*gagei*	generalist	fluvial specialist	medium
0.60	Cottidae	*Cottus*	*spp.*	generalist	fluvial specialist	medium
0.60	Percidae	*Etheostoma*	*tennesseense*	invertivore	fluvial specialist	medium

Taxa O/E	Family	Genus	Species	Feeding	Habitat	Tolerance
0.59	Cyprinidae	*Luxilus*	*zonistius*	invertivore	fluvial specialist	medium
0.58	Centrarchidae	*Acantharchus*	*pomotis*	invertivore	swamp specialist	medium
0.57	Centrarchidae	*Ambloplites*	*rupestris*	predator	fluvial specialist	medium
0.56	Cyprinidae	*Hybopsis*	*spp.*	invertivore	fluvial specialist	medium
0.56	Esocidae	*Esox*	*niger*	predator	habitat generalist	medium
0.55	Catostomidae	*Moxostoma*	*lachneri*	invertivore	fluvial specialist	low
0.55	Percidae	*Etheostoma*	*swaini*	invertivore	fluvial specialist	medium
0.54	Percidae	*Etheostoma*	*blennioides*	invertivore	fluvial specialist	medium
0.53	Centrarchidae	*Centrarchus*	*macropterus*	invertivore	swamp specialist	medium
0.53	Catostomidae	*Moxostoma*	*duquesnei*	invertivore	fluvial specialist	low
0.53	Ictaluridae	*Noturus*	*insignis*	invertivore	fluvial specialist	medium
0.52	Centrarchidae	*Lepomis*	*marginatus*	invertivore	habitat generalist	medium
0.51	Percidae	*Etheostoma*	*rufilineatum*	invertivore	fluvial specialist	medium
0.51	Cyprinidae	*Rhinichthys*	*spp.*	invertivore	fluvial specialist	medium
0.50	Cyprinidae	*Lythrurus*	*atrapiculus*	invertivore	fluvial specialist	low
0.49	Cyprinidae	*Notropis*	*xaenocephalus*	invertivore	fluvial specialist	medium
0.48	Catostomidae	*Moxostoma*	*sp. brassy jumprock*	invertivore	fluvial specialist	low
0.48	Ictaluridae	*Pylodictis*	*olivaris*	predator	habitat generalist	medium
0.48	Ictaluridae	*Ameiurus*	*platycephalus*	generalist	habitat generalist	medium
0.47	Ictaluridae	*Noturus*	*leptacanthus*	invertivore	fluvial specialist	medium
0.47	Anguillidae	*Anguilla*	*rostrata*	generalist	fluvial specialist	medium
0.45	Cyprinidae	*Cyprinella*	*callistia*	invertivore	fluvial specialist	medium
0.43	Fundulidae	*Fundulus*	*lineolatus*	invertivore	swamp specialist	medium
0.42	Centrarchidae	*Micropterus*	*spp.*	predator	fluvial specialist	medium
0.42	Cyprinidae	*Notropis*	*amplamala*	invertivore	fluvial specialist	medium
0.41	Percidae	*Etheostoma*	*hopkinsi*	invertivore	fluvial specialist	medium
0.39	Percidae	*Etheostoma*	*olmstedi*	invertivore	fluvial specialist	medium
0.39	Salmonidae	*Oncorhynchus*	*mykiss*	predator	fluvial specialist	low

Taxa O/E	Family	Genus	Species	Feeding	Habitat	Tolerance
0.37	Cyprinidae	*Cyprinella*	*gibbsi*	invertivore	fluvial specialist	medium
0.37	Cyprinidae	*Lythrurus*	*bellus*	invertivore	fluvial specialist	low
0.37	Percidae	*Etheostoma*	*inscriptum*	invertivore	fluvial specialist	medium
0.35	Cyprinidae	*Notropis*	*chrosomus*	invertivore	fluvial specialist	low
0.35	Cyprinidae	*Notropis*	*cummingsae*	invertivore	fluvial specialist	medium
0.32	Percidae	*Percina*	*smithvanizi*	invertivore	fluvial specialist	medium
0.30	Petromyzontidae	*Petromyzontidae*	*spp.*	generalist	no designation	unknown
0.30	Cyprinidae	*Lythrurus*	*lirus*	invertivore	fluvial specialist	low
0.30	Ictaluridae	*Noturus*	*funebris*	invertivore	fluvial specialist	medium
0.28	Percidae	*Etheostoma*	*tallapoosae*	invertivore	fluvial specialist	medium
0.26	Ictaluridae	*Noturus*	*gyrinus*	invertivore	fluvial specialist	medium
0.25	Salmonidae	*Salvelinus*	*fontinalis*	predator	fluvial specialist	low
0.24	Cyprinidae	*Opsopoeodus*	*emiliae*	invertivore	habitat generalist	low
0.24	Centrarchidae	*Ambloplites*	*ariommus*	predator	fluvial specialist	low
0.24	Cyprinidae	*Cyprinella*	*trichroistia*	invertivore	fluvial specialist	low
0.23	Cyprinidae	*Cyprinella*	*nivea*	invertivore	fluvial specialist	medium
0.23	Centrarchidae	*Micropterus*	*dolomieu*	predator	habitat generalist	medium
0.22	Elassomatidae	*Elassoma*	*zonatum*	invertivore	swamp specialist	medium
0.21	Percidae	*Etheostoma*	*caeruleum*	invertivore	fluvial specialist	medium
0.19	Cyprinidae	*Lythrurus*	*fasciolaris*	invertivore	fluvial specialist	medium
0.19	Cyprinidae	*Notropis*	*hypsilepis*	invertivore	fluvial specialist	low
0.15	Percidae	*Etheostoma*	*etowahae*	invertivore	fluvial specialist	low
0.14	Cyprinidae	*Cyprinella*	*galactura*	invertivore	fluvial specialist	medium
0.14	Percidae	*Percina*	*palmaris*	invertivore	fluvial specialist	low
0.14	Catostomidae	*Moxostoma*	*sp. Apalachicola redhorse*	invertivore	fluvial specialist	low
0.08	Percidae	*Etheostoma*	*jessiae*	invertivore	fluvial specialist	low