AN EVALUATION OF THE RELATIVE INFLUENCE OF SPATIAL, STATISTICAL, AND BIOLOGICAL FACTORS ON THE ACCURACY OF STREAM FISH SPECIES PRESENCE MODELS

by

JOHN CARROLL RUIZ

Under the direction of James T. Peterson

ABSTRACT

Models relating fish species presence to landscape and stream features are increasingly being used by natural resource managers. The accuracy of these models directly influences the ability to make sound stream management decisions. To evaluate the effect of biotic and abiotic factors on model accuracy, I fit parametric (logistic regression) and nonparametric (k-nearest neighbor) models of species presence at two spatial scales using watershed and stream reach characteristics. I then evaluated the influence of model type, spatial scale, and species-specific traits, on the cross validation errors of the 4 best-fitting scale-specific, parametric and non parametric models. I found that errors were highly variable within a species and error rates were related to model type, spatial scale, and species-specific traits. Findings from this study suggest that species specific traits had a large effect along with the effects of scale and that resource managers should consider these effects when modeling presence or absence of fish species.

INDEX WORDS: Model accuracy, Logistic regression, Nearest neighbor, Prediction error, Spatial scale, Species traits, Species presence/absence

AN EVALUATION OF THE RELATIVE INFLUENCE OF SPATIAL, STATISTICAL, AND BIOLOGICAL FACTORS ON THE ACCURACY OF STREAM FISH SPECIES PRESENCE MODELS

by

JOHN CARROLL RUIZ

B.S., University of Georgia, 2001

A Thesis Submitted to the Graduate Faculty of

The University of Georgia

in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2005

AN EVALUATION OF THE RELATIVE INFLUENCE OF SPATIAL, STATISTICAL, AND BIOLOGICAL FACTORS ON THE ACCURACY OF STREAM FISH SPECIES PRESENCE MODELS

by

JOHN CARROLL RUIZ

Major Professor:	James T. Peterson
Committee:	Robert J. Cooper
	Mary C. Freeman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2005

DEDICATION

This thesis is dedicated to my daughter, Malina Pearl Ruiz, who was born while I was pursuing my degree. She is a constant reminder of dedication and happiness to whom I am grateful.

AKNOWLEDGMENTS

Many people deserve thanks for their efforts and guidance with this project. My major professor, Dr. James T. Peterson, who throughout this project eagerly provided help with statistical analyses, writing, and encouragement. My committee members, Dr. Robert J. Cooper and Dr. Mary C. Freeman, also provided valuable insight to the project. I would not have been able to complete my field research without the help of fellow graduate students, Nolan Banish, Shane Hawthorne, Jeremy McCargo, Dale McPhearson, and Colin Shea. Additionally, Anne Wimberly deserves thanks for guidance in GIS maps and figures.

Finally, I extend thanks to my friends and family for providing encouragement while I pursued this degree. I especially thank my wife, Natasha, for her love and devotion. Their support permitted me to complete this degree and for that I am grateful.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The southeastern United States has the highest diversity and number of endemic fishes in North America (Burr and Mayden 1992; Warren et al. 1997) and the highest proportion of fish fauna threatened with extinction (Minckley and Decon 1991; Warren and Burr 1994; Richter et al. 1997). The decline of native fish species in the southern United States has been attributed to habitat degradation and fragmentation (Angermeier 1995; Warren et al. 1997). Southeastern fishes are further threatened by continued physical habitat alteration in the form of stream channelization, impoundment, sedimentation, and flow modification (Walsh et al. 1995; Etnier 1997; Burkhead et al. 1997). A national risk assessment classified aquatic ecosystems in seven southeastern states (Alabama, Florida, Georgia, North and South Carolina, Tennessee, and Virginia) as “extreme risk” due to increased land development, numbers of imperiled species (including fishes), and already endangered ecosystems (Noss and Peters 1995). The conservation of these species will require the knowledge of their distribution and habitat requirements. The Southeastern Fishes Council also has recognized the need for up-to-date information on the current distribution and status of southeastern fish populations (Warren et al. 2000). This need is based, in part, on a lack of information on the distribution and status of fishes and the U.S. Fish and Wildlife Service’s (USFWS) decision to discontinue the designation of category 2 species (species for which information indicates that listing is possibly appropriate, but there is no conclusive data) as candidates for listing under the Endangered Species Act of 1973 (Warren et al. 2000).

The status of fish populations in Georgia depends, in part, on the quality of water and aquatic habitats. Therefore, it is important to identify and protect existing areas of high quality habitat. The Flint River Basin, located in the western third of Georgia, and has a large diversity of fish fauna with 85 extant species representing 19 families (NAWQA 2000). Although the Flint River Basin has been transformed by human activity, much of its biodiversity is intact. All of the basins major tributaries are unimpounded and it has a free flowing stream reach of 201 kilometers that only 42 other rivers in the continental U.S. can claim (Behnke 1990). Within the Flint Basin are several protected tracts of land adjacent to major tributaries or the Flint River. These include the 7972 hectare Chickasawhatchee Wildlife Management Area (WMA), a major recharge zone for the Floridian Aquifer that contains the headwaters of Chickasawhatchee, Kiokee, and Spring Creeks and the 11735 hectare protected Jones Center which contains a proportion of Ichawaynochaway creek and borders the Flint River. The high diversity of fish fauna combined with relatively intact habitats and stream connectivity within the Flint River Basin make this an ideal site for the management of fishes with the Gap Analysis Program (GAP). GAP is a national program designed to identify areas of biodiversity and protect them from decline.

To conserve or restore fish in the Flint River Basin, fisheries professionals need to know the current status and distribution of lotic fish populations. This will require the development of accurate models of species distribution and population status. Towards this end, I will develop empirical models of fish distribution at multiple scales in the Flint River basin using existing land cover data and fish sampling data with the following objectives. I will examine the accuracy of models at two different scales and determine the best scale to model species distribution. I will relate species specific traits to scale specific estimates of model accuracy to determine which traits influence model accuracy. My study will provide resource managers with a set of guidelines to determine the best method to model species distribution and identify areas of large biodiversity that can be used in conservation planning in the state of Georgia.

CHAPTER 2

LITERATURE REVIEW

The character of individual streams and rivers reflects an integration of physical and biological processes occurring in the entire catchment, yet ecological studies and management of natural resources have traditionally occurred at the scale of a stream reach, forest stand, or vegetation plot (Johnson and Gage 1997). Recently, biologists and resource managers have begun to consider the importance of examining the relationship between taxa distribution and habitat at regional and National scales. The Gap Analysis Program (GAP) is an effort to address the conservation of biodiversity on a National scale by identifying gaps in areas managed exclusively or primarily for the long-term maintenance of populations of native species (Scott et al. 1993). This information is critical to conservation and management of fish species in the Flint Basin.

GAP began in 1987 and moved from a research and development program within the United States Geological Survey (USGS) to operational status in early 1998 (Jennings 2000). GAP's purpose is twofold: (1) to provide regional assessments of the current status and distribution of native vertebrate taxa and land cover types and (2) facilitate the application of this information to land management activities (Scott and Jennings 1998). GAP is based on Burley's concept: the identification of gaps in networks of conservation lands by identifying and classifying flora and fauna and the habitats that they use (Jennings 2000). This information then is used to provide resource managers with tools for examining the existing system of protected areas and to identify critical areas for protection and management (Jennings 2000). The method requires large amounts of data to assess the distribution and status of species and habitats relative to existing conservation areas (Jennings 2000).

To identify conservation gaps, the presence of species must be determined. This can be accomplished with empirical models (i.e. models that rely on observational data) relating species presence to landscape and local features. Because of the difficulty in collecting data over a large area, there is a limited number of such studies. However, recent advances in remote sensing and geographic information systems (GIS) have made this approach viable (Howell 2000). For example, broad-scale landscape variables have been used to predict the distribution of salmonids in the Columbia River Basin (Haas et al. 2000), and presence or absence of native warmwater fishes in Wyoming (Lipsey et al. *in review*). These studies were restricted to a single spatial scale. For example, (Lipsey et al. *in review*) modeled species presence distribution using stream reach characteristics such as, elevation, channel slope, and stream size. Ecological patterns, however, are often scale-dependent. That is, patterns such as fish-habitat relations may differ with changes in spatial scale (Wiens 1989). Hence, a single scale may be inadequate for modeling the distribution of all species.

The proper scale at which to model is important. Some fish species are resource generalists and may not respond to small scale habitat features thus; modeling them at a large scale may be more appropriate (Kolasa 1989). Other species may be specialists and have specific habitat needs that may respond to small scale features. For example, the Redeye chub *Notropis harperi* is found almost exclusively in groundwater dominated stream reaches (Mettee et al. 1996). Arbitrarily selecting a scale to impose on a community of organisms that operate on different scales may obfuscate species habitat relationships to varying degrees for different species. For example, modeling at a large scale may result in poorly predicting models of species occurrence because microhabitat features may be limiting species distribution (Lipsey et al. *in review*). The implications of this are great as ecosystem management requires an understanding of both large and small scale features (Levin 1992). To determine the proper scale, species-habitat relationships must be examined at multiple scales (Wiens 1989).

The conservation of biodiversity is of global importance. Species extinction is a worldwide problem and a major focus among conservationists because an estimated 1000 species are becoming extinct each year and the rate could reach 5000 per year in the near future (Scott et. al. 1987). GAP has been devised to identify gaps in biodiversity on a National scale by creating literature- based models to predict these gaps. However, managers should use empirical models to predict gaps or areas of high biodiversity with greater accuracy (Conroy and Noon 1996). Thus, this project will evaluate the factors influencing the accuracy of empirical models of species presence to provide resource managers with information on the most useful tools for the conservation of fish biodiversity.

CHAPTER 3

OBJECTIVES

My objectives were threefold: (1) to develop models of species presence at two scales; (2) relate model accuracy to scale, type of statistical model, and species specific-specific traits; and (3) develop guidelines for future modeling of fish species presence in warmwater streams.

CHAPTER 4

METHODS

Study site

The Flint River Basin is located in the western third of Georgia and extends from Atlanta to the Florida border (Figure 1). The Basin drains an area of 21911 km^2 and the mainstem of the Flint River is 561 river kilometers long (Couch et al. 1996). The major tributaries of the Flint Basin are Spring Creek, Muckalee, Kinchafoonee, and Ichawaynochaway Creeks. There are also 3 major reservoirs on the mainstem Flint River (Lake Worth, Blackshear, and Seminole) and numerous small impoundments located within the Flint Basin. The Flint River flows through 2 physiographic provinces, the Piedmont and the Coastal Plain. As the river flows through these provinces it is subject to their geomorphic characteristics. Gradients in the Piedmont are steeper and shoals or rapids are more common while gradients in the Coastal Plain are lower and have a lower frequency of shoals or rapids.

Land use in the Flint River Basin is primarily agriculture and forestry related with forestry practices dominating the Piedmont and row crop agriculture in the Coastal Plain (Peterson et al. 2003). Urban land use in the basin is concentrated around metropolitan areas such as Atlanta, Albany and Bainbridge.

Fish data

I identified and compiled existing fish sampling data from previous studies in the Flint River basin by the Georgia Department of Natural Resources (GADNR), USGS, and University of Georgia. I used fish sample data from 1980 to the present and defined this as “current time” because major land use characteristics during this time frame have not changed substantially over most of the Basin (Peterson et al. 2003). Fish samples previously collected by GADNR were collected using a standard protocol, a single pass with a backpack electrofisher in a site length 35 times the mean stream width and maximum length of 500-m (GADNR 2000). Methods used by USGS were similar, but site lengths were 6-10 times the mean stream width to ensure all habitat types were represented (NAWQA 2000).

Supplemental sample

I visually examined the spatial distribution of existing samples and identified locations in the Basin with little or no prior samples. Areas determined to have no prior samples then were sampled to minimize gaps in the spatial distribution of samples throughout the Basin. To ensure proper representation in areas missing samples, stream types were grouped into strata based on stream order and major tributary. Sampling sites were randomly selected from within each stratum.

Fish sampling

Fishes were sampled within the Basin at locations with no prior sampling data or to repeat samples in selected streams. I used 3 different types of electrofishing gear for which efficiency is known, and based on stream size following Peterson and Rabeni (1995). To represent all habitat types within small (< 6-m wetted width and maximum depth of ≤1-m) and medium (> 6-m wetted width and maximum depth of ≤ 1.5-m) streams, site lengths were 100-m and in larger streams (> 15-m wetted width and max depth of < 2.5-m) 150-m. In small streams, I used a Smith Root backpack electrofisher model LR-24 (Murphy et al. 1996). I placed 6 mm mesh blocknets at the upper and lower end of the site. Fish collections consisted of 3 passes, one up, one down, and one up and fish collected in the downstream blocknet. In medium streams, a barge electrofisher was used (Murphy et al. 1996). The barge electrofisher consists of 2 handheld anodes and one cathode on the bottom of the barge. Power was supplied by a Honda 3500 generator and output controlled with a Smith Root Model 6A shock-box. Again, 6 mm mesh blocknets were placed at the upper and lower end of the site and fish collection consisted of 3 passes, one up, one down, and one up and fish collected in the downstream blocknet. In large streams, a boat electrofisher was used (Murphy et al. 1996). Fish collection consisted of 6 passes, up middle, down river left, up river right, down middle, up river left, and down river right. All fish collected were identified and released. Species that could not be identified in the field were preserved in 10% formalin and brought back to the lab for identification.

Habitat variables

Land use was estimated from 1992 Landsat thematic mapper imagery provided by the UGA Natural Resources Spatial Analysis Laboratory (Peterson et al. 2003). Large-scale habitat measures (i.e. parent geology, vegetation cover) were determined by examining Global GIS land coverage maps at the 1:100,000 scale (Scott and Jennings 1998). Land coverage consisted of 15 categories: water, low intensity residential, high intensity residential, commercial / industrial / transportation, transitional, deciduous forest, evergreen forest, mixed forest, shrubland, orchards/vineyards, grasslands/herbaceous, pasture, row crops, woody wetlands and emergent/herbaceous wetlands. Elevation and slope were derived from the National Elevation Dataset (USGS National Elevation Dataset, 1999). The National Elevation Dataset is a 1:24,000 digital elevation model for the conterminous United States. The SLOPE command in Arc/Info Grid was used to calculate the slope for each collection site. Link magnitude is defined as the number of contributing tributaries upstream from the sampling site and is calculated using the method of Shreve (1967). Downstream link (D-link) is defined as the link magnitude of the stream reach immediately below the site (Osborne et al. 1992). Link and D-link were determined on 1:100,000 scale USGS maps of the Flint River Basin.

I evaluated the effect of spatial scale on the accuracy of species presence models and defined large scale as a subwatershed (USGS 12 digit Hydraulic Unit Code, HUC) that is approximately 7300 Ha in size and small-scale as a stream reach as defined by Frissell et al. (1986). I fit models at two scales, the subwatershed level and reach level, where reaches were nested within subwatersheds. To minimize the influence of incomplete detection on the accuracy of species presence models, I selected species that occurred in at least 20 sample sites and had detection probabilities ≥ 70 percent. Detection probabilities were estimated using repeat visit data from a subset of sample sites (Peterson et al. 2003).

Life History

Life history characteristics of fish species, such as seasonal habitat preference (i.e. riffle, run, pool), tolerance (i.e., to dissolved oxygen and temperature), and spawning behavior may affect how a species relates to the environment, which may affect model accuracy. Consequently I classified species based on 7 life history characteristics (Table 1) and related these to model accuracy for each species. The species characteristics are defined as: class, whether the species is a resource generalist or specialist; size of mature adult, where small was ≤100mm, medium was 101mm-200mm and large was >200mm; ontogenetic shift, species utilizes different habitats at different life stages; tolerance to anthropogenic change; endemism, species is endemic to Apalachicola basin; introduction, species has been introduced to basin; cosmopolitan, species is found commonly throughout basin. For each species, I determined the life history characteristics using Fishes of Alabama (Mettee et al.1996), and other published studies, (Nelson 1968, Mathur

1973, and Balon 1975). I also contacted 3 experts in the region Bud Freeman (UGA), Mary Freeman (USGS) and John Biagi (GADNR) for additional input.

Analyses

When modeling species occurrences, biological responses are typically estimated by assuming some predefined statistical distribution (e.g. logistic regression models assume binomial responses); hence, model accuracy is likely a function of how well the distribution approximates the biological response (Haas et al. 2000). To evaluate the influence of model type on model accuracy, I fit two types of models, parametric and non-parametric. To evaluate the effect of scale, each model type was fit at 2 separate scales, subwatershed and reach, resulting in 4 models per species.

Logistic regression is a parametric technique that is useful for modeling binary responses, such as species presence or absence (Agresti 1990). I fit logistic regression to the largescale subwatershed species presence. However, logistic regression cannot account for spatial dependence (i.e. spatial autocorrelation; Sokal and Rohlfs 1995) and can result in lower model accuracy. To account for spatial autocorrelation among reaches, I fit hierarchical logistic regression models to the reach level data with reaches nested within subwatersheds. Hierarchical models can account for spatial dependence and can model data at multiple spatial scales (Snijders and Bosker 1999). Both hierarchical and non-hierarchical logistic regression models were fit with SAS software (Proc NLMIX, SAS Institute 2001). I fit all possible subsets of models using local and landscape characteristics (Appencix A). The best fitting model for each species was determined via Akaike Information Criteria (AIC, Burnham and Anderson 1998;Tables 2 and 3).

In contrast to parametric techniques, non-parametric classifiers do not require distributional assumptions and may be more robust and accurate than parametric models (Haas et al. 2000). Thus, I fit non parametric nearest neighbor models to species presence data at both the subwatershed and reach scale. Nearest neighbor (KNN) models are nonparametric classifiers that use the response distribution of K nearest neighbors based on the values of predictor variables (Hand 1982). I fit all possible subsets of models using local and landscape characteristics (Appencix A). The best fitting model for each species and scale was determined (Tables 4 and 5) via a Monte Carlo hypothesis testing procedure (Hass et al. 2000), where predictors were accepted when $\alpha < 0.05$ level.

Estimation of model accuracy

I estimated the accuracy of the best fitting parametric and non parametric models by estimating the expected error rate. The expected error rate (EER) is the most relevant measure of model accuracy (Efron 1983). Ten-fold cross validation is the best estimator of the EER that provides a measure of overall predictive ability without excessive variance (Breiman and Spector 1992). This procedure is completed by omitting 10 percent of the observations and fitting the model with the remaining 90% observations. The probability of presence is estimated for the omitted observations, and species presence is estimated using a 0.5 cutoff (i.e., probability ≥ 0.5 estimated presence). This procedure is repeated for all observations.

For each of the 4 combinations of scale and model type, I identified the best fitting model for each species (determined using different techniques such as Monte Carlo Hypothesis testing and Akaike information criteria). Once the best fitting models were determined, I estimated three different types of error rates for each of the 4 best fitting models using ten-fold cross validation. The first type of error was overall error, which is the proportion of observations that were misclassified without regard to presence or absence of a species. Prediction errors, also called commission errors, were the proportion of category-specific observations the model predicted incorrectly. For example, the proportion of observations that a species was predicted to be present but was actually absent. Classification errors, also defined as omission error, were the proportion of category specific observations that were incorrectly classified as present or absent at a location. Category-wise prediction and classification errors then were added and hence, prediction and classification errors used in the analysis could have values greater than one. I examined the relationship between error rates, model scale, model type, and biological traits using hierarchical linear regression (Bryk and Raudenbush 1992). Here, the dependent variables in the regression were the cross-validation error rates. I fit 17 candidate models that represented factors that I believe influenced the ability of the model to predict species presence (Table 1). These factors include scale, model type, difference, which is the absolute value of difference between the number of known presences and absences, and 7 biological traits for each species. The relative fit of candidate models was determined using AIC. To determine relative plausibility of models, Akaike weights were calculated following Burnham and Anderson (2001). Weights range from 0 to 1 with the highest weight signifying the most plausible model. A confidence set of models was then determined as those models with weights within 10% of the best model weight following Thompson and Lee (2000). All inferences were based on the confidence set.

Prior to fitting the hierarchical linear regression error models, I partitioned variation in cross validation error rates among and within species with random effects Analysis of Variance (ANOVA). Hierarchical linear regression categorical predictors such as species characteristics, scale, and model type also were binary coded prior to modeling (Table 2). Additionally, goodness-of-fit for each candidate model was determined by examining residual and normal probability plots (Bryk and Raudenbush 1992).

CHAPTER 5

RESULTS

Fish samples were collected from 257 sites nested within 44 subwatersheds. There was a total of 47 species analyzed based on average detection probabilities that were greater than 70% (Table 3). Species were modeled with logistic regression and k-nearest neighbor techniques at 2 separate scales, largescale (subwatershed) and smallscale (reach) level, resulting in 4 models for each species. Because species such as bluegill, yellowfin shiner and spottail shiner occurred in every subwatershed, largescale models could not be fit with these species. The best fitting species presence models were relatively accurate with overall cross validation error rates averaging from 28% to 71% (Appendix B). In general, logistic (parametric) models were more accurate than KNN (nonparametric) models. The best fitting KNN models also tended to include fewer predictors than logistic regression models. Similarly, smallscale (reach) models included more predictors than the largescale (subwatershed level) models.

The random effects ANOVA indicated that there was greater variation in cross validation error rates within species than among species for all error measures (Figure 2). On average, variation in cross validation error rates within species was 5.9 times greater than among species. Variation in error rates within species also was greatest for overall error, 99.9%, and least for classification error, 65.0%.

Overall error

Four candidate models constituted the set of confidence models for overall error (Table 4). The best fitting model relating overall error to scale, model type, and species traits included smallscale, logit (logistic regression), difference, small size, large size, and a interaction between large size and smallscale (Table 5). This model was 1.01 times more likely than the next best approximating model, which contained the same variables as the best-fitting model with the exception of the small size by smallscale interaction. The confidence set included these two models and two additional models containing cosmopolitan and specialist species. Parameter estimates suggested that overall error rates were on average 0.134 lower for logistic and hierarchical logistic regression (Table 5). Overall error rates also were on average 0.3 greater for (reach) small scale models compared to (subwatershed) largescale models. The parameter estimates for the remaining variables in the candidate set, difference, small and large size, cosmopolitan, and specialist were relatively imprecise and the confidence intervals were very wide and contained zero. Hence, my inferences for the effect of these variables were limited.

Prediction error

An examination of the residuals from the global hierarchical linear model relating prediction error to model type, scale, and species specific traits indicated the fit was adequate. Five models represented the confidence set of models for prediction error (Table 6). The best-fitting model relating prediction error to scale, model type and species traits included smallscale, specialist, and a specialist by smallscale interaction and was 1.8 times more likely than the next best approximating model, which contained smallscale, cosmopolitan species trait and a cosmopolitan species trait by smallscale interaction (Table 6). The confidence set included these 2 models and 3 additional models that contained the difference and logistic regression parameters. Parameter estimates suggested that prediction error rates for generalist species were on average 0.277 greater for smallscale (reach) models compared to largescale (subwatershed) models, whereas prediction error rates were on average 0.199 lower for specialist species at small scales (Table 7). Parameter estimates for logistic regression, difference, and cosmopolitan species suggested a positive relationship with error rates but my inferences were limited because the estimates were imprecise (Table 7).

Classification error

Examination of the residual plot from the global hierarchical linear model relating classification error to model type, scale, and species specific traits suggested an adequate fit. The best-fitting model relating classification error to scale, model type, and species traits included smallscale, logistic regression, difference, specialist, and a specialist by smallscale interaction (Table 8). This model was 1.2 times more likely than the next best approximating model, which contained the same variables as the best fitting model but replaced specialists with cosmopolitan species. The confidence set included these 2 models and 3 additional models containing small size, large size, and high and low tolerance (Table 8). Classification error rates were on average 0.089 lower for logistic regression, whereas classification error rates were on average 0.111 greater for smallscale (reach) models (Table 9). Parameter estimates suggested that specialist species had lower classification error rates at large scales and that they became even lower at smallscales. Parameters estimates for logistic regression and difference suggested a negative relationship with classification error rates. However, my inferences were limited because the estimates were imprecise and spanned zero.

CHAPTER 6

DISCUSSION

The within-species variation in error rates represents the amount of variation due to factors such as type of statistical model and spatial scale, whereas the variation among species represents the amount of variation in error rates presumably due to biological differences among species. I found that variation in within species error rates was higher than among species across all error measures. This result suggests that factors such as type of model or spatial scale may have a greater effect on the accuracy of species presence models. Although the variation in error rates among species was lower, the hierarchical models suggest that species-specific traits also influenced model accuracy. Thus, developing guidelines for species presence modeling requires an understanding of those factors influencing accuracy both among and within-species.

Within species influences

Accuracy of species presence models depends, in part, on how well the statistical method can approximate the species response. In general, previous studies have found that model accuracy varied among species but that nonparametric models were more accurate than parametric models across species (Hass et al 2000; Olden and Jackson 2002). I also found that no single method was best for modeling all species at all scales. However, logistic regression models were on average more accurate than nonparametric KNN models across spatial scales. This apparent discrepancy may be due to the difference in complexity of the species responses modeled and the ability to account for spatial dependence. Hass et al. (2000) modeled three population status categories: strong, depressed, and absent, whereas I modeled two: presence and absence. Three category logit models are more complex than two category models because modeling requires simultaneously fitting two linear models rather than one (Agresti 1990). Hence, it may be more difficult to approximate a species response with two linear models, increasing model error. Additionally, I was able to account for spatial dependence among sample units at small scales using hierarchical logistic regression. Spatial dependence can affect model accuracy and increase error rates (Sokal and Rohlf 1995). The parametric and non parametric models that were used in these previously cited studies and the KNN models I used cannot account for spatial dependence which may have lowered model accuracy. Thus, I believe that parametric hierarchical models may be more accurate than nonparametric methods when spatial autocorrelation is present.

The accuracy of my species presence models was influenced by the spatial scale modeled, but the effect of scale differed among species. I assumed that species were present in sampling locations because the species could exist under the stream conditions (e.g., habitat availability) during sampling. The habitat conditions in individual stream reaches reflect an integration of physical and biological processes occurring in the entire catchment. Largescale habitat features I used to model species presence, including physiographic province, parent geology, and land use, represented those features influencing habitat forming processes, such as sediment delivery and the size, timing, and duration of streamflows (Frissell et al. 1986). The smallscale or local features I used to model species presence, including reach slope and link magnitude and represented those factors that modify the largescale processes to produce the local habitat conditions of the reach (Frissell et al. 1986). Micro-habitat characteristics in stream reaches tend to be more strongly related to local characteristics such as reach slope, whereas reach habitat features are more loosely associated with largescale (subwatershed) features

(Richards et al. 1996). Therefore, the small scale (reach) features are probably a better indicator in the availability of habitats at the scale that fish were sampled and largescale watershed features probably reflect average habitat conditions for streams in a watershed. For example, largescale processes, such as climate, determine thermal and flow regime thereby determining the available species pool (i.e. the number of species that can possibly be at a specific site). Small scale processes, such as gradient, further refine largescale processes, thereby determining the structure and availability of specific habitat. I believe that the effect of spatial scale on the accuracy of species presence models depends on how strongly species utilize local habitat features. For example, small scale (reach) attributes such as stream gradient positively influence current velocity so the likelihood for finding species that require high current velocity habitats might be higher in steeper reaches. In contrast, gradient may not be a useful predictor for a species that can inhabit a wide variety of high and low current habitats. Thus, I hypothesize that the effect of scale should depend on how strongly species are linked to local (macrohabitat) features.

Among species influences

Because of widely varying life-history strategies, individual species require a variety of environmental conditions from the micro- to the macrohabitat level (Kolasa 1989). I assumed that species were located in sample sites that fulfilled their life history requirements. Therefore, a single scale may be inadequate for modeling the distribution of all species. The available species pool is determined by largescale and smallscale environmental factors, hydrologic processes, and biology of the fish species (Ricklefs 1987). Largescale features such as climate, geology, and spatial location within a watershed provide a filter of environmental conditions that determine the available species pool (Poff 1997). For example, largescale features such as climate influence average stream flows and temperatures at the watershed scale. Therefore, species pool within the watershed will be those species that can persist in these conditions. Smallscale (reach) features such as elevation, link, and d-link magnitude define local environmental conditions further. Therefore, species that are related more strongly to specific microhabitats may be more strongly related to local smallscale features. For example, shoal bass require shoal habitats (Mettee et al. 1996), a relatively smallscale feature compared to a watershed. In general, I found that the species presence models for generalist species were more accurate at larger scales, which was consistent with previous studies (Smogor et al. 1995; Meyer 2004). I believe that this is because the life history requirements of generalist species are broad, thus enabling them to occupy a wider variety of habitats. That is, the distribution of generalist species is weakly related to smallscale (reach) factors, such as d-link. Consequently, generalist species are most likely better predicted at larger scales such as physiographic region because of shared features such as climate. Conversely, I found that smaller scales were better for predicting specialist species or species with specific habitat requirements (e.g., non-cosmopolitan species). I believe that this is due, in part, to the fact that the smallscale features I used are more closely related to the local habitat features that specialist species. Thus, I hypothesize that the life history requirements of a species determines the scale that is most appropriate for modeling its distribution and that specialist species are more likely to be modeled more accurately at smaller scales, whereas generalist species are more likely to be more accurate at larger scales.

False absence and presence

Another factor that can influence the accuracy of species presence models is false absences (i.e., the failure to record a species when in fact it is present) and false presences (i.e., recording a species present when it is in fact absent). To minimize the effect of false presences, a local expert (B. Freeman, University of Georgia; personal communication) checked species records and beyond that, I assumed that the presence of all fish species were recorded correctly. False absence can bias statistical estimates potentially increasing errors (Tyre et al. 2003). I attempted to minimize the influence of false absence by including species and sample units with detection probabilities ≥70%. However, false absences may have had an effect. The probability of detecting a species is a positive function of the number of samples collected in a sample unit (Tyre et al. 2003). I found that largescale (subwatershed) models were more accurate than smallscale (reach) models. Largescale models had multiple samples within subwatersheds, so detection probabilities were higher at the subwatershed level compared to the smallscale (reach) level. Thus, the lower errors for the largescale models were consistent with a possible false absence effect on model accuracy and average detection probabilities were approximately equal 88% for both generalist and specialist species. However, the better predicting models for specialist species at smallscales were not consistent with this effect. Therefore, I believe that false absences did not have a substantial influence on my conclusions.

Management recommendations

I found that, for relatively common fish species in the Flint River Basin, species-specific traits and spatial scale influenced the accuracy of species presence models. I also found that model accuracy was related to the type of statistical model. However, out of numerous statistical techniques I examined only two, logistic regression and k-nearest neighbor. Therefore, the patterns I observed may not be applicable to more rare species or different types of statistical techniques. For example, discriminant analysis, a parametric technique, was found to predict very poorly compared to logistic regression (Press and Wilson 1978) and probably predicts much more poorly than non parametric techniques such as k-nearest neighbor. However, I believe that many of my conclusions should be useful to other biologists modeling fish distributions in other systems.

It can be expensive to develop fish distribution models for numerous species over large geographical areas. The expense of these projects are often related to the amount of information and amount of detail required to model species distributions. For example, estimating stream link is relatively a time consuming task compared to estimating the amount of land use in a particular watershed because there are often hundreds of stream reaches per watershed and each of these reaches in the watershed upstream of the site must be counted. However, I found that species presence model error rates were related to spatial scales and that some species could be modeled using less expensive large scale data. By taking into account how scale and species traits affect the performance of the model, resource managers can more efficiently develop species distribution models that are also more accurate, saving time and money in a budget-constrained profession. Therefore, I recommend that the life histories of the species being modeled be determined prior to modeling species distribution to determine whether the species is a habitat generalist or specialist so they may be modeled at the appropriate scale.

Most aquatic conservation measures are usually implemented to protect threatened and endangered species. These species are typically resource specialists that require specific habitat features (Warren et al. 2000). I found that species whose life histories did not require them to use local smallscale features and could use a wide variety of habitats had lower error rates when modeled at larger scales. Conversely, species whose life histories required them to use specific local smallscale features had lower error rates at small scales. This was consistent with previous studies that suggested modeling at larger scales was best for species that utilize a wide variety of habitat types (Smogor et al. 1995). Consequently, I suggest that rare and threatened species be modeled at smaller spatial scales.

Accuracy of modeling species presence can vary depending on how well the statistical method can approximate the species response. Previous studies have found that model accuracy varied among species but that nonparametric models were more accurate than parametric models (Hass et al 2000; Olden and Jackson 2002). I also found that no single method was best for modeling all species at all scales. However, logistic regression models were on average more accurate than nonparametric KNN models across scales. This apparent discrepancy may be due to the difference in complexity of the species responses modeled and the ability to account for spatial dependence. Non parametric models are also known as black box classifiers; therefore, interpretations of relationships are difficult to interpret and the ability to estimate uncertainty is often poor. Conversely, results from parametric models are easily interpreted and the ability to estimate uncertainty good. Thus, I recommend that biologists give greater emphasis to estimating uncertainty and examine cross validation errors to determine the proper statistical model type rather than relying on a single model type.

LITERATURE CITED

Agresti, A. 1990. Categorical Data Analysis. John Wiley & Sons, New York.

Angermeier, P. L. 1995. Ecological attributes of extinction-prone species: loss of freshwater fishes of Virginia. Conservation Biology. [Conserv. Biol.] Vol. 9 pp. 143-148.

Balon, E. K. 1975. Reproductive guilds of fishes: A proposal and definition. Journal of Fisheries Resources Board Canada 32: 821-864.

Benke, A. C. 1990. A perspective on America's vanishing streams. Journal of North American Benthological Society. 9(1): 77-88.

Breiman, L., P. Spector. 1992. Submodel selection and evaluation in regression. The X-random case. International Statistical Review. 60(3):291-319.

Bryk, A. S. and S. W. Raudenbush. 1992. Hierarchical linear models: applications and data analysis methods. Sage, Thousand Oaks, California.

Burkhead, N. M., S. J. Walsh, B. J. Freeman, and J. D. Williams. 1997. Status and restoration of the Etowah River, an imperiled southern Appalachian ecosystem. Pages 375-444 *in* G.W. Benz and D.E. Collins, eds. Aquatic fauna in peril: the Southeastern perspective. Special Publ. 1, Southeast Aquatic Research Institute, Lenz Design and Communications, Decatur, GA.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.

Burr, B. M., and R. L. Mayden. 1992. Phylogenetics and North American freshwater fishes. Pages 18-75 *in* R.L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford University Press, Stanford, CA.

Conroy, M. J., and B. R. Noon. 1996. Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. Ecological Applications 6(3):763-767.

Couch, C. A., E. H. Hopkins, and P. S. Hardy. 1996. Influences of Environmental Settings on Aquatic Ecosystems in the Apalachicola-Chattahoochee-Flint River Basin U.S. Geological Survey Water-Resources Investigations Report 95-4278

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association., 783: 316-331.

Etnier, D. A. 1997. Jeopardized southeastern freshwater fishes: a search for causes. Pages 88-104 *in* G.W. Benz and D.E. Collins, eds. Aquatic fauna in peril: the Southeastern perspective. Special Publ. 1, Southeast Aquatic Research Institute, Lenz Design and Communications, Decatur, GA.

Frissell, C. A., W. J. Liss, C. E. Warren and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental Management Vol:10:2 pp. 199-214.

GADNR (Georgia Department of Natural Resources) 2000. Standard Operating Procedures for Conducting Biomonitoring on Fish Communities in the Piedmont Ecoregion of Georgia.

Hand, D. J. 1982. Kernel discriminant analysis. Research Studies Press, New York.

Hass, T. C., J. T. Peterson, and D. C. Lee. 2000. An evaluation of parametric and nonparametric models for predicting fish distribution and status with broadscale landscape information. Unpublished draft for Ecological Modeling.

Howell, J. E. 2000. Modeling Distribution of Neotropical Migrant Songbirds on a State-Wide Basis. Masters Thesis Proposal, University of Georgia, Warnell School of Forest Resources, Athens, GA.

Jennings, M. D. 2000. Gap analysis: concepts, methods and recent results. Landscape Ecology. Vol. 15. pp. 5-20.

Johnson, L. B. and S. H. Gage. 1997. Landscape approaches to the analysis of aquatic ecosystems. Freshwater Biology. Vol 37. pp. 113-132.

Kolasa, J. 1989. Ecological systems in hierarchical perspective: breaks in community structure and other consequences. Ecology. Vol. 70:1 pp. 36-47.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943-1967.

Lipsey, T. B. and W. A. Hubert. i*n review*. Assessment of Logistic Modeling to Predict Presence and Absence of Native Warmwater Fishes in Wyoming Streams. Unpublished draft.

Mettee, R. R., P. E. O'Neil, and J. M. Pierson. 1996. Fishes of Alabama and the Mobile Basin. Oxmoor House, Inc., Birmingham, AL.

Minckley, W. L., and J. E. Deacon, eds. 1991. Battle against extinction. University of Arizona Press. Tucson, AZ.

Murphy, B. R. and D. W. Willis. 1996. Fisheries Techniques. American Fisheries Society. Bethesda, Maryland.

NAWQA (National Water-Quality Assessment Program) 2000. Description of the ACF River Basin Study Area. http://ga.water.usgs.gov/nawqa/main.description.html

Nelson, J. S. 1968. Life history of the brook silverside, Labidesthes sicculus, in Crooked Lake, Indiana. Transactions of the American Fisheries Society 97: 293-29.

Noss, R. F. and R. L. Peters. 1995. Endangered ecosystems: a status report on America's vanishing habitat and wildlife. Defenders of Wildlife, Washington, DC.

Olden, J. D. and D. A. Jackson. 2002. A comparison of statistical approaches for modeling fish species distributions. Freshwater Biology. 47:1976-1995.

Osborne, L. L., and M. J. Wiley. 1992. Influence of tributary spatial position on the structure of warmwater fish communities. Canadian Journal of fisheries and Aquatic Sciences. 49:671-681.

Peterson, J. T., and C. F. Rabeni. 1995. Optimizing sampling effort for sampling warmwater stream fish communities. North American Journal of Fisheries Management 16:738-746.

Peterson, J. T., B. J. Freman, E. Kramer, E. Irwin, M.C. Freeman, C. Strait, A. Wimberly, and J.C. Ruiz. 2003. Aquatic Gap: Regional analysis of biodiversity in the ACT/ACF Basins.

Poff, N. L. 1997. Landscape filters and species traits: toward mechanistic understanding and prediction in stream ecology. Journal of the North American Benthological Society 16:391-408

Press, J. and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73: 699-705.

Richards, C., B. Johnson, and G. E. Host. 1996. Landscape influences on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Science 53(suppl. I):295-311.

Richter, B. D., D. P. Braun, M. A. Mendelson, and L. L. Master. 1997. Threats to imperiled freshwater fauna. Conservation Biology. Vol. 11:5 pp.1081-1093.

Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. Science. 235:167-171.

SAS Institute Inc. 2001. version 8.2. SAS Institute Inc., Cary North Carolina. USA.

Scott, M. J. and M. D. Jennings. 1998. Large-Area Mapping of Biodiversity. Annals of the Missouri Botanical Garden. Vol. 85:1 pp. 34-47.

Scott, M. J., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'erchia, T.C. Edwards Jr., J. Ulliman, and R.G. Wright. 1993. Gap Analysis: A Geographic Approach to Protection of Geological diversity. Wildlife Mongraphs: 123:1-41.

Shreve, R. L. (1967) Infinite topologically random channel networks. Journal of Geology, 75: 178-186.

Smogor R. A., P. L. Angermeier, and C. K. Gaylord. 1995. Distribution and abundance of American Eels in Virginia streams: Tests of null models across spatial scales. Transactions of the American Fisheries Society. 124:789-803.

Snijders, T., and R. Bosker. 1999. Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, California, USA.

Sokal, R. R. and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. Freeman, New York.

Thompson, W. L. and D. C. Lee. 2000. Modeling relationships between landscape-level attributes and snorkel counts of Chinook salmon and steelhead parr in Idaho. Canadian Journal of Fisheries and Aquatic Sciences 57:1834-1842.

Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. Ecological Applications. Vol. 13:6 pp 1790-1801.

Walsh, S. W., N. M. Burkhead, and J. D. Williams. 1995. Southeastern freshwater fishes. Pages 177-147 in E.T. LaRoe, ed. Our living resources. A report to the nation on the distribution, abundance, and health of U.S. plants, animals, and ecosystems. U.S. Department of the Interior, National Biological Service, Washington, DC.

Warren, M. L. Jr., and B. M. Burr. 1994. Status of freshwater fishes of the United States: overview of an imperiled fauna. Fisheries. Vol. 19:1 pp.6-18.

Warren, M. L., P. L. Angermeier, B. M. Burr, and W. R. Hagg. 1997. Decline of a diverse fish fauna: patterns of imperilment and protection in the southeastern United States. Pages 105-164 *in* G.W. Benz and D.E. Collins, eds. Aquatic fauna in peril: the Southeastern perspective. Special Publ. 1, Southeast Aquatic Research Institute, Lenz Design and Communications, Decatur, GA.

Warren, M. L., B. M. Brooks, S. J. Walsh, H. L. Bart, R. C. Cashner, D. A. Etnier, B. J. Freeman, B. R. Kuhajda, R. L. Mayden, H. W. Robinson, S. T. Ross, and W. C. Starnes. 2000. Diversity, Distribution, and Conservation Status of the Native Freshwater Fishes of the Southern United States. Fisheries 25(10):7-31

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology. 3:385-397.

Table 1. Hypotheses and correspond candidate models the factors believed to influence the accuracy of species presence models.

Hypothesis	*Model*
1) Model error differs with model type, difference between observed presence and absence, scale, and species traits. The effect of scale will vary with species trait.	
a) Cosmopolitan, highly tolerant, class (generalist) species will have higher error rates than other species and the error rates will be even higher at small scales	smallscale logit diff class cosmo high class*smallscale cosmo*smallscale high*smallscale
b) Class (generalist) species will have higher error rates than other species and the error rates will be even higher at small scales.	smallscale logit diff class class*smallscale
c) Low tolerance and highly tolerant species will have higher error than other species and the error rates will be even higher for highly tolerant species at small scales.	smallscale logit diff low high high*smallscale
d) Small and large body size species will have higher error than others species and error rates will be even higher for larger species at smaller scales.	smallscale logit diff sm lg lg*smallscale
d) Small and large body size species will have higher error than others species and error rates will be even higher for smaller species at smaller scales.	smallscale logit diff sm lg sm*smallscale
f) Cosmopolitan, species will have higher error rates than other species and the error rates will be even greater at small scales	smallscale logit diff cosmo cosmo*smallscale

Table 1. continued

2) Model error differs with difference between observed presence and absence, scale, and species traits. The effect of scale will vary with species trait.	
a) Cosmopolitan, highly tolerant, class (generalist) species will have higher error rates than other species and the error rates will be greater at small scales	smallscale diff class cosmo high class*smallscale cosmo*smallscale high*smallscale
b) Class (generalist) species will have higher error rates than other species and the error rates will be even greater at small scales	smallscale diff class class*smallscale
c) Low tolerance and highly tolerant species will have higher error than other species and the error rates will be even higher for highly tolerant species at small scales.	smallscale diff low high high*smallscale
d) Small and large body size species will have higher error than others species and error rates will be even higher for larger species at smaller scales.	smallscale diff sm lg lg*smallscale
e) Small and large body size species will have higher error than others species and error rates will be even higher for smaller species at smaller scales.	smallscale diff sm lg sm*smallscale
f) Cosmopolitan, species will have higher error rates than other species and the error rates will be even greater at small scales	smallscale diff cosmo cosmo*smallscale

Table 1. continued

3) Model error differs with scale and species traits. The effect of scale will vary with species trait.	
a) Cosmopolitan, highly tolerant, class (generalist) species will have higher error rates than other species and the error rates will be even higher at small scales	smallscale class cosmo high class*smallscale cosmo*smallscale high*smallscale
b) Class (generalist) species will have higher (generalist) error rates than other species and the error rates will be even higher at small scales	smallscale class class*smallscale
c) Low tolerance and highly tolerant species will have higher error than other species and the error rates will be even higher for highly tolerant species at small scales.	smallscale low high high*smallscale
d) Small and large body size species will have higher error than others species and error rates will be even higher for larger species at smaller scales.	smallscale sm lg sm*smallscale
e) Cosmopolitan, species will have higher error rates than other species and the error rates will be even greater at small scales	smallscale cosmo cosmo*smallscale
f) Small and large body size species will have higher error than others species and error rates will be even higher for larger species at smaller scales.	smallscale sm lg lg*smallscale

Table 2. Species characteristics, their states and binomial values

Characteristic	**States**	**Value**	
Specialization	Specialist	1	
	Generalist	0	
Ontogenetic shift	Yes	1	
	No	0	
Endemic	Yes	1	
	No	0	
Introduced	Yes	1	
	No	0	
Cosmopolitan	Yes	1	
	No	0	
Tolerance	Low	1	
	High	0	
		Small	**Large**
Size of adult	Small	1	0
	Medium	0	0
	Large	0	1

Table 3. The number of sample sites detected, the mean, minimum (min) and maximum (max) detection probabilities, and traits of the species used in the analysis of accuracy of species presence models. The species traits are: class, resource generalists (G) and specialists(S); adult body size, large (L), medium (M), and small (S); and tolerance to anthropogenic alteration, high (H), moderate (M) and low (L).

			Detection probabilities			Species traits			
Common name	Scientific name	Sites present	Mean	Min.	Max.	Class	Cosmopolitan	Adult size	Tolerance
Southern brook lamprey	*Icthyomyzon gagei*	60	0.905	0.698	1.00	G	1	L	M
Redfin pickerel	*Esox americanus*	101	0.905	0.698	1.00	G	1	L	M
Chain pickerel	*Esox niger*	63	0.905	0.698	1.00	G	1	L	H
Bluestripe shiner*	*Cyprinella callitaenia*	29	0.898	0.759	1.00	G	1	S	M
Blacktail shiner	*Cyprinella venusta*	129	0.898	0.759	1.00	G	1	M	H
Silverjaw minnow	*Ericymba buccata*	32	0.898	0.759	1.00	G	0	L	L
Clear chub	*Hybopsis winchelli*	100	0.898	0.759	1.00	G	1	S	M
Bluehead chub	*Nocomis leptocephalus*	42	0.898	0.759	1.00	G	1	L	M
Golden shiner[I]	*Notemigonus crysoleucas*	71	0.898	0.759	1.00	G	1	L	H
Redeye chub	*Notropis harperi*	40	0.898	0.759	1.00	S	0	S	L
Spottail shiner	*Notropis hudsonius*	72	0.898	0.759	1.00	G	1	S	H
Highscale shiner	*Notropis hypsilepis*	29	0.898	0.759	1.00	G	1	S	H
Longnose shiner	*Notropis longirostris*	85	0.898	0.759	1.00	G	1	S	H
Yellowfin shiner	*Notropis lutipinnis*	35	0.898	0.759	1.00	G	1	S	H

* Endemic species, [I] species are stocked or introduced, [O] Species exhibits ontogenetic shift

Table 3. continued.

			Detection probabilities			Species traits			
Common name	Scientific name	Sites present	Mean	Min.	Max.	Class	Cosmopolitan	Adult size	Tolerance
Coastal shiner	*Notropis petersoni*	42	0.898	0.759	1.00	G	1	S	H
Weed shiner	*Notropis texanus*	151	0.898	0.759	1.00	G	1	S	H
Pugnose minnow	*Opsepuedus emilae*	54	0.898	0.759	1.00	G	1	S	H
Sailfin shiner	*Pternotropis hypselopterus*	128	0.898	0.759	1.00	G	1	S	H
Spotted sucker	*Minytrema melanops*	89	0.868	0.696	1.00	G	1	L	H
Greyfin redhorse	*Moxostoma sp.cf. poecilurum*	28	0.868	0.696	1.00	G	1	L	H
Greater jumprock*	*Scartomyzon lachneri*	39	0.868	0.696	1.00	G	1	L	H
Snail bullhead	*Ameirurs brunneus*	36	0.867	0.696	1.00	G	0	L	L
Yellow bullhead	*Ameirurs natalis*	57	0.867	0.696	1.00	G	1	L	H
Brown bullhead	*Ameirurs nebulosus*	21	0.867	0.696	1.00	G	1	L	H
Cannel catfish[I]	*Ictalurus punctatus*	33	0.867	0.696	1.00	G	1	L	H
Speckeld madtom	*Noturus leptacanthus*	22	0.867	0.696	1.00	G	1	S	H
Pirate pearch	*Aphredoderus sayanus*	108	0.898	0.759	1.00	G	1	M	H
Eastern starhead topminnow	*Fundulus escambiae*	21	0.905	0.784	1.00	G	1	S	H
Mosquitofish	*Gambusia affinis*	110	0.905	0.784	1.00	G	1	S	H

---

* Endemic species, [I] species are stocked or introduced, [O] Species exhibits ontogenetic shift

Table 3. continued

			Detection probabilities			Species traits			
Common name	Scientific name	Sites present	Mean	Min.	Max.	Class	Cosmopolitan	Adult size	Tolerance
Brook silverside	*Labidesthes sicculus*	86	0.905	0.784	1.00	G	1	M	H
Shadow bass	*Ambloplites ariommus*	27	0.877	0.711	1.00	G	1	L	M
Banded pygmy sunfish	*Elassoma zonatum*	27	0.877	0.711	1.00	G	0	S	M
Redbreast sunfish[I]	*Lepomis auritus*	185	0.877	0.711	1.00	G	1	L	M
Green sunfish[I]	*Lepomis cyanellus*	73	0.877	0.711	1.00	G	1	S	M
Warmouth	*Lepomis gulosus*	116	0.877	0.711	1.00	G	1	L	M
Bluegill	*Lepomis macrochirus*		0.877	0.711	1.00	G	1	L	M
Dollar sunfish	*Lepomis marginatus*	44	0.877	0.711	1.00	G	1	M	H
Redear sunfish[I]	*Lepomis microlophus*	65	0.877	0.711	1.00	G	1	L	M
Spotted sunfish	*Lepomis punctatus*	152	0.877	0.711	1.00	G	1	L	H
Largemouth bass[IO]	*Micropterus salmoides*	116	0.901	0.771	1.00	G	1	L	M
Shoal bass*[O]	*Micropterus cataractae*	102	0.901	0.771	1.00	S	1	S	H
Black crappie[I]	*Pomoxis nigromaculatus*	37	0.877	0.711	1.00	G	1	L	M
Brown darter	*Etheostoma edwini*	54	0.866	0.704	1.00	G	0	S	M
Swamp darter	*Etheostoma fusiforme*	35	0.866	0.704	1.00	G	1	S	H
Gulf darter	*Etheostoma swaini*	44	0.866	0.704	1.00	G	1	L	H
Halloween darter*	*Percina sp.*	27	0.866	0.704	1.00	S	0	M	L
Blackbanded darter	*Percina nigrofasciata*	174	0.866	0.704	1.00	G	1	M	H

* Endemic species, [I] species are stocked or introduced, [O] Species exhibits ontogenetic shift

Table 4. Predictor variables, AICc, DAICc, and Akaike weights (*w*) for the set of candidate models (i*)* relating fish species traits, statistical model type and scale to overall error rates.

Overall error candidate models	AICc	DAICc	*wi*
Smallscale, logit, diff, small,large, large*smallscale	-223.3	0.0	0.403
Smallscale, logit, diff, small, large, small*smallscale	-223.3	0.0	0.386
Smallscale, logit, diff, specialist, specialist*smallscale	-220.7	2.7	0.105
Smallscale, logit, diff, cosmopolitan, cosmopolitan*smallscale	-220.2	3.1	0.084
Smallscale, logit, diff, low tolerance, high tolerance, high tolerance*smallscale	-217.3	6.1	0.000
Smallscale, logit, diff, specialist, cosmopolitan, high tolerance, specialist*smallscale cosmopolitan*smallscale	-213.5	9.9	0.000
Smallscale, diff, small, large, large*smallscale	-171.1	52.2	0.000
Smallscale, diff, small, large, small*smallscale	-171.1	52.3	0.000
Smallscale, diff, specialist, specialist*smallscale	-168.3	55.1	0.000
Smallscale, diff, cosmopolitan, cosmopolitan*smallscale	-167.8	55.5	0.000
Smallscale, diff, low tolerance, high tolerance, high tolerance*smallscale	-164.8	58.5	0.000
Smallscale, specialist, specialist*smallscale	-138.6	84.8	0.000
Smallscale, cosmopolitan, cosmopolitan*smallscale	-138.5	84.9	0.000
Smallscale, small, large, large*smallscale	-133.8	89.5	0.000
Smallscale, small, large, small*smallscale	-133.8	89.6	0.000
Smallscale, specialist, cosmopolitan, high tolerance, specialist*smallscale cosmopolitan*smallscale	-133.7	89.7	0.000
Smallscale, low tolerance, high tolerance, high tolerance*smallscale	-130.1	93.3	0.000

Table 5. Estimates of fixed and random effects for confidence set of models using overall error rates to relate fish species traits, statistical model type and scale.

Model: Parameter estimate	Estimate (standard error)		90% confidence interval Lower	Upper
Smallscale, logit, diff, small, large, large*smallscale:				
Fixed effects				
Intercept	0.267	(0.040)	0.201	0.334
Smallscale	0.312	(0.030)	0.261	0.362
Logit	-0.134	(0.016)	-0.162	-0.107
Difference	-0.002	(0.000)	-0.002	-0.002
Small	0.065	(0.042)	-0.005	0.136
Large	0.103	(0.045)	0.028	0.177
Smallscale*large	0.010	(0.033)	-0.045	0.064
Random effects				
Species	0.005	(0.002)	0.003	0.011
Residual	0.012	(0.002)	0.010	0.015
Smallscale, logit, diff, small, large, small*smallscale:				
Fixed effects				
Intercept	0.265	(0.039)	0.199	0.331
Smallscale	0.316	(0.030)	0.266	0.365
Logit	-0.134	(0.016)	-0.162	-0.107
Difference	-0.002	(0.001)	-0.002	-0.002
Small	0.066	(0.046)	-0.010	0.141
Large	0.107	(0.042)	0.038	0.177
Smallscale*small	-0.001	(0.033)	-0.056	0.054
Random effects				
Species	0.005	(0.002)	0.003	0.011
Residual	0.012	(0.002)	0.010	0.015

Table 5. Continued

Model:	Estimate	(standard error)	90% confidence interval Lower	Upper
Parameter estimate				
Smallscale, logit, diff, specialist, specialist*smallscale:				
Fixed effects				
Intercept	0.339	(0.019)	0.307	0.371
Smallscale	0.315	(0.027)	0.270	0.361
Logit	-0.134	(0.016)	-0.161	-0.107
Difference	-0.002	(0.000)	-0.002	-0.002
Specialist	0.013	(0.066)	-0.096	0.123
Specialist*smallscale	-0.083	(0.067)	-0.195	0.029
Random effects				
Species	0.006	(0.002)	0.004	0.012
Residual	0.012	(0.002)	0.010	0.015
Smallscale, logit, diff, cosmopolitan, cosmopolitan*smallscale:				
Fixed effects				
Intercept	0.377	(0.046)	0.299	0.454
Smallscale	0.265	(0.058)	0.170	0.361
Logit	-0.134	(0.016)	-0.162	-0.107
Difference	-0.002	(0.000)	-0.002	-0.002
Cosmopolitan	-0.042	(0.049)	-0.123	0.038
Cosmopolitan*smallscale	0.053	(0.051)	-0.031	0.137
Random effects				
Species	0.006	(0.002)	0.004	0.012
Residual	0.012	(0.002)	0.010	0.015

Table 6. Predictor variables, AICc, DAICc, and Akaike weights (w) for the set of candidate models (i) relating fish species traits, statistical model type and scale to prediction error rates.

Prediction error candidate models	AICc	DAICc	*wi*
Smallscale, specialist, specialist*smallscale	4.5	0.0	0.334
Smallscale, cosmopolitan, cosmopolitan*smallscale	5.7	1.2	0.182
Smallscale, diff, specialist, specialist*smallscale	6.2	1.7	0.143
Smallscale, diff, cosmopolitan, cosmopolitan*smallscale	7.4	2.9	0.078
Smallscale, logit, diff, specialist, specialist*smallscale	8.1	3.6	0.054
Smallscale, low tolerance, high tolerance, high tolerance*smallscale	9.1	4.7	0.033
Smallscale, small, large, small*smallscale	9.2	4.8	0.032
Smallscale, small, large, large*smallscale	9.4	4.9	0.029
Smallscale, logit, diff, cosmopolitan, cosmopolitan*smallscale	9.4	4.9	0.029
Smallscale, diff, small, large, small*smallscale	10.0	5.5	0.022
Smallscale, diff, small, large, large*smallscale	10.2	5.7	0.019
Smallscale, diff, low tolerance, high tolerance, high tolerance*smallscale	10.4	5.9	0.018
Smallscale, logit, diff, small, large, small*smallscale	12.0	7.5	0.008
Smallscale, logit, diff, small, large, large*smallscale	12.2	7.7	0.007
Smallscale, logit, diff, low tolerance, high tolerance, high tolerance*smallscale	12.4	7.9	0.007
Smallscale, specialist, cosmopolitan, high tolerance, specialist*smallscale, cosmopolitan*smallscale	13.1	8.6	0.004
Smallscale, logit, diff, specialist, cosmopolitan, high tolerance, specialist*smallscale, cosmopolitan*smallcale	17.0	12.5	0.001

Table 7. Estimates of fixed and random effects for confidence set of models using prediction error rates to relate fish species traits, statistical model type and scale.

Model:	Estimate		90% confidence interval	
Parameter estimate	(standard error)		Lower	Upper
Smallscale, specialist, specialist*smallscale:				
Fixed effects				
Intercept	0.479	(0.029)	0.431	0.527
Smallscale	0.271	(0.033)	0.216	0.326
Specialist	0.006	(0.112)	-0.180	0.192
Specialist*smallscale	-0.199	(0.130)	-0.414	0.016
Fixed effects				
Species	0.012	(0.005)	0.006	0.029
Residual	0.047	(0.006)	0.039	0.058
Smallscale, cosmopolitan, cosmopolitan*smallscale:				
Fixed effects				
Intercept	0.494	(0.077)	0.365	0.623
Smallscale	0.149	(0.089)	0.002	0.296
Cosmopolitan	-0.017	(0.083)	-0.154	0.120
Smallscale*cosmopolitan	0.125	(0.095)	-0.033	0.283
Fixed effects				
Species	0.012	(0.005)	0.007	0.029
Residual	0.047	(0.006)	0.039	0.059
Smallscale, diff, specialist, specialist*smallscale:				
Fixed effects				
Intercept	0.485	(0.030)	0.434	0.536
Smallscale	0.300	(0.052)	0.213	0.387
Difference	0.001	(<0.001)	-0.001	0.000
Specialist	0.005	(0.113)	-0.183	0.192
Specialist*smallscale	-0.180	(0.131)	-0.398	0.037
Random effects				
Species	0.013	(0.006)	0.007	0.031
Residual	0.046	(0.006)	0.038	0.058

Table 7. Continued

Model:	Estimate		90% confidence interval	
Parameter estimate	(standard error)		Lower	Upper
Smallscale, diff, cosmopolitan, cosmopolitan*smallscale:				
Fixed effects				
Intercept	0.498	(0.078)	0.367	0.629
Smallscale	0.197	(0.111)	0.013	0.381
Difference	0.000	(<0.001)	-0.001	0.000
Cosmopolitan	-0.015	(0.083)	-0.153	0.124
Smallscale*cosmopolitan	0.106	(0.099)	-0.057	0.269
Random effects				
Species	0.013	(0.006)	0.007	0.031
Residual	0.047	(0.006)	0.039	0.058
Smallscale, logit, diff, specialist, specialist*smallscale:				
Fixed effects				
Intercept	0.477	(0.034)	0.420	0.534
Smallscale	0.300	(0.052)	0.214	0.387
Logit	0.015	(0.032)	-0.037	0.068
Difference	0.000	(<0.001)	-0.001	0.000
Specialist	0.005	(0.113)	-0.183	0.192
Specialist*smallscale	-0.180	(0.131)	-0.398	0.037
Random effects				
Species	0.013	(0.006)	0.007	0.031
Residual	0.046	(0.006)	0.038	0.057

Table 8. Predictor variables, AICc, DAICc, and Akaike weights (w) for the set of candidate models (i) relating fish species traits, statistical model type and scale to classification error rates.

Class error candidate models	AICc	DAICc	*wi*
Smallscale, logit, diff, specialist, specialist*smallscale	8.1	0.0	0.312
Smallscale, logit, diff, cosmopolitan, cosmopolitan*smallscale	8.6	0.5	0.243
Smallscale, logit, diff, small, large, small*smallscale	9.4	1.3	0.163
Smallscale, logit, diff, small, large, large*smallscale	9.7	1.6	0.141
Smallscale, logit, diff, low tolerance, high tolerance, high tolerance*smallscale	12.1	3.9	0.045
Smallscale, logit, diff, specialist, cosmopolitan, high tolerance, specialist*smallscale cosmopolitan*smallscale	13.6	5.6	0.019
Smallscale, diff, specialist, specialist*smallscale	14.1	6.0	0.016
Smallscale, specialist, specialist*smallscale	14.5	6.4	0.013
Smallscale, diff, cosmopolitan, cosmopolitan*smallscale	14.55	6.5	0.012
Smallscale, cosmopolitan, cosmopolitan*smallscale	14.65	6.6	0.012
Smallscale, diff, small, large, small*smallscale	15.25	7.1	0.009
Smallscale, diff, small, large, large*smallscale	15.5	7.4	0.008
Smallscale, diff, low tolerance, high tolerance, high tolerance*smallscale	17.9	9.8	0.002
Smallscale, small, large, small*smallscale	18.2	10.1	0.002
Smallscale, small, large, large*smallscale	18.3	10.2	0.002
Smallscale, low tolerance, high tolerance, high tolerance*smallscale	18.9	10.8	0.001
Smallscale, specialist, cosmopolitan, high tolerance, specialist*smallscale, cosmopolitan*smallscale	19.7	11.6	0.001

Table 9. Estimates of fixed and random effects for confidence set of models using classification error rates to relate fish species traits, statistical model type and scale.

Model:	Estimate		90% confidence interval	
Parameter estimate	(standard error)		Lower	Upper
Smallscale, logit, diff, specialist, specialist*smallscale:				
Fixed effects				
Intercept	0.760	(0.035)	0.701	0.819
Smallscale	0.111	(0.052)	0.026	0.197
Logit	-0.089	(0.031)	-0.140	-0.038
Difference	-0.001	(0.000)	-0.001	0.000
Specialist	-0.193	(0.120)	-0.391	0.005
Specialist*smallscale	-0.138	(0.128)	-0.349	0.074
Random effects				
Intercept				
Species	0.018	(0.006)	0.011	0.036
Residual	0.044	(0.005)	0.036	0.054
Smallscale, logit, diff, cosmopolitan, cosmopolitan*smallscale:				
Fixed effects				
Intercept	0.609	(0.083)	0.469	0.749
Smallscale	0.036	(0.109)	-0.144	0.216
Logit	-0.089	(0.031)	-0.141	-0.038
Difference	-0.001	(0.000)	-0.001	0.000
Cosmopolitan	0.158	(0.088)	0.013	0.304
Cosmopolitan*smallscale	0.073	(0.096)	-0.086	0.231
Random effects				
Intercept				
Species	0.018	(0.006)	0.011	0.036
Residual	0.044	(0.005)	0.036	0.054

Table 9. Continued

Model: Parameter estimate	Estimate	(standard error)	90% confidence interval Lower	Upper
Smallscale, logit, diff, small, large, small*smallscale:				
Fixed effects				
Intercept	0.727	(-0.072)	0.606	0.849
Smallscale	0.116	(0.057)	0.021	0.210
Logit	-0.089	(0.031)	-0.141	-0.038
Difference	-0.001	(0.000)	-0.002	0.000
Small	-0.060	(0.084)	-0.198	0.079
Large	0.110	(0.077)	-0.018	0.237
Small*smallscale	0.036	(0.063)	-0.069	0.140
Random effects				
Intercept				
Species	0.016	(0.006)	0.010	0.034
Residual	0.044	(0.005)	0.037	0.055
Smallscale, logit, diff, small, large, large*smallscale				
Fixed effects				
Intercept	0.718	(0.073)	0.596	0.840
Smallscale	0.133	(0.057)	0.038	0.228
Logit	-0.089	(0.031)	-0.141	-0.038
Difference	-0.001	(0.000)	-0.001	0.000
Small	-0.042	(0.078)	-0.171	0.087
Large	0.114	(0.083)	-0.023	0.251
Large*smallscale	-0.010	(0.063)	-0.113	0.094
Random effects				
Intercept				
Species	0.016	(0.006)	0.010	0.034
Residual	0.044	(0.005)	0.037	0.055

Table 9. Continued

Model: Parameter estimate	Estimate	(standard error)	90% confidence interval Lower	Upper
Smallscale, logit, diff, low tolerance, high tolerance, high tolerance*smallscale:				
Fixed effects				
Intercept	0.811	(0.055)	0.719	0.903
Smallscale	0.073	(0.067)	-0.037	0.184
Logit	-0.089	(0.031)	-0.141	-0.038
Difference	-0.001	(0.000)	-0.001	0.000
Low tolerance	-0.220	(0.099)	-0.383	-0.056
High tolerance	-0.071	(0.065)	-0.179	0.036
High tolerance*smallscale	0.058	(0.063)	-0.047	0.163
Random effects				
Intercept				
Species	0.019	(0.006)	0.012	0.037
Residual	0.044	(0.005)	0.036	0.054

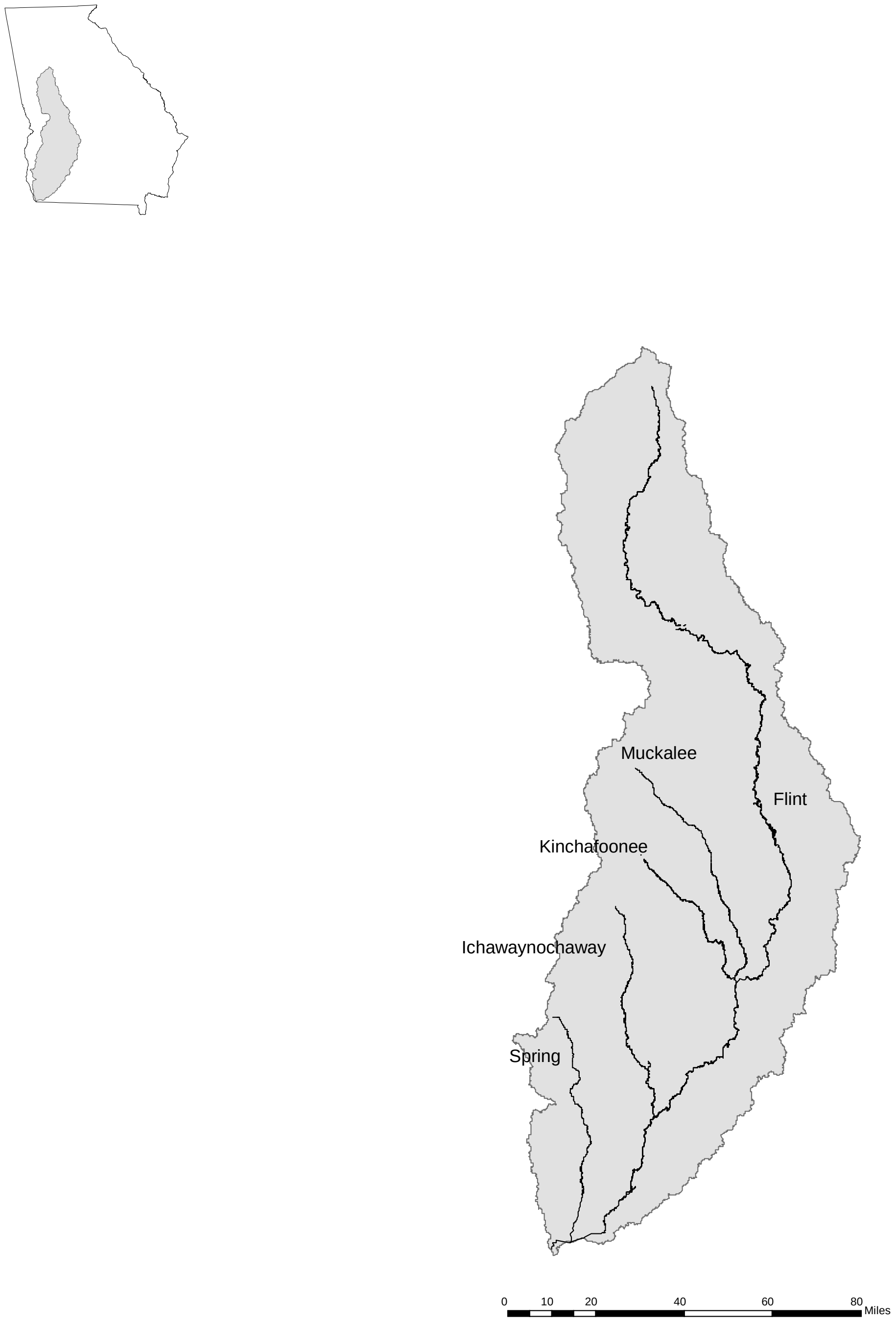

Figure 1.1 Location of Flint River Basin and Major Tributaries.

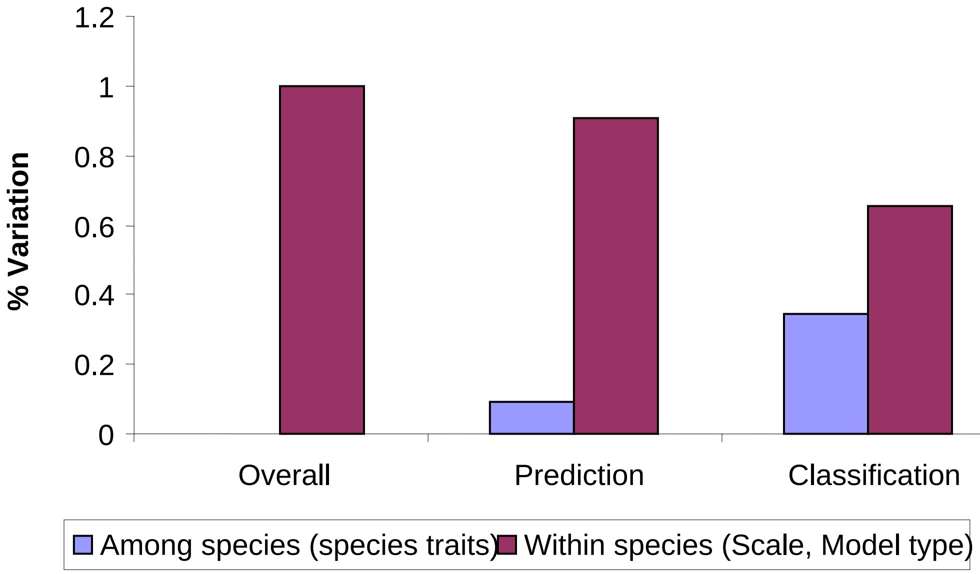

Figure 1.2 Proportion of variation between species trait, scale, and model type.

APPENDIX A

Appendix A. Sample reach and subwatershed characteristics used to fit species presence models.

Characteristic
Sample reach
Site elevation (m)
Stream gradient (%)
Stream aspect (degrees)
Stream link magnitude
Link magnitude of nearest downstream reach
Isolation (% of sampled stream reaches)
Subwatershed
Drainage density (km/km^2)
Road density (km/km^2)
Percent of adjacent subwatersheds occupied (varied among species)
Physiographic providence composition (% of subwatershed)
Piedmont province
Parent geology composition (% of subwatershed)
Alluvium
Amphibolite
Gneiss
Granite
Limestone
Ultramafic
Sand
Land use/ land type composition (% of subwatershed)
Forested
Grassland
Rowcrop agriculture
Urban
Wetland

APPENDIX B

Appendix B. Model errors for individual species by scales and statistical method. Prediction and classification errors are the sum of category-wise error rates.

		Logit		KNN	
Species	**Error**	**Reach**	**Watershed**	**Reach**	**Watershed**
Southern brook lamprey	Overall	0.236	0.142	0.295	0.378
	Prediction	0.665	0.288	0.589	0.831
	Classification	0.676	0.314	0.426	0.880
Redfin pickerel	Overall	0.394	0.068	0.552	0.049
	Prediction	0.739	0.909	0.707	0.103
	Classification	0.736	0.929	0.918	0.909
Chain pickerel	Overall	0.297	0.301	0.685	0.215
	Prediction	1.020	0.835	0.776	0.529
	Classification	1.009	0.885	0.916	0.642
Bluestripe shiner	Overall	0.098	0.222	0.226	0.727
	Prediction	0.485	0.550	0.707	0.727
	Classification	0.531	0.505	0.379	1.000
Blacktail shiner	Overall	0.386	0.130	0.441	0.091
	Prediction	0.730	0.635	0.656	0.091
	Classification	0.774	0.390	0.885	1.000
Silverjaw minnow	Overall	0.292	0.109	0.257	0.211
	Prediction	0.701	0.205	0.512	0.385
	Classification	0.673	0.229	0.395	0.448
Clear chub	Overall	0.444	0.115	0.615	0.136
	Prediction	0.898	0.670	1.588	0.136
	Classification	0.893	0.698	1.010	1.000
Bluehead chub	Overall	0.220	0.103	0.242	0.325
	Prediction	0.779	0.291	0.606	0.565
	Classification	0.758	0.223	0.309	0.433
Golden shiner	Overall	0.300	0.153	0.630	0.227
	Prediction	0.799	0.388	0.815	0.227
	Classification	0.845	0.551	0.901	1.000
Redeye chub	Overall	0.394	0.068	0.552	0.049
	Prediction	0.739	0.909	0.707	0.103
	Classification	0.736	0.929	0.918	0.909

Appendix B. continued

Spottail shiner	Overall	0.296	0.028	0.175	0.270
	Prediction	0.709	0.120	0.681	0.571
	Classification	0.682	0.036	0.228	0.344
Highscale shiner	Overall	0.134	0.059	0.321	0.078
	Prediction	0.722	0.102	0.778	0.212
	Classification	0.796	0.196	0.464	0.160
Longnose shiner	Overall	0.357	0.143	0.264	0.295
	Prediction	0.803	0.342	0.539	0.295
	Classification	0.800	0.351	0.480	1.000
Yellowfin shiner	Overall	0.263	0.073	0.263	0.701
	Prediction	0.693	0.170	0.693	1.108
	Classification	0.390	0.199	0.390	1.008
Costal shiner	Overall	0.156	0.305	0.304	0.545
	Prediction	0.575	0.613	0.686	0.545
	Classification	0.624	0.621	0.435	1.000
Weed shiner	Overall	0.307	0.115	0.359	0.068
	Prediction	0.582	0.789	0.664	0.720
	Classification	0.701	0.478	0.843	0.833
Pugnose minnow	Overall	0.222	0.280	0.597	0.345
	Prediction	0.803	0.677	0.766	1.208
	Classification	1.008	0.689	0.772	1.055
Sailfin shiner	Overall	0.212	0.017	0.176	0.085
	Prediction	0.438	0.030	0.319	0.164
	Classification	0.418	0.051	0.298	0.250
Spotted sucker	Overall	0.409	0.141	0.654	0.114
	Prediction	0.839	0.667	1.156	0.114
	Classification	0.823	1.031	1.008	1.000
Greyfin redhorse	Overall	0.148	0.240	0.144	0.530
	Prediction	0.907	0.495	0.646	0.583
	Classification	0.945	0.500	0.476	0.898
Greater jumprock	Overall	0.145	0.199	0.303	0.526
	Prediction	0.549	0.427	0.716	0.794
	Classification	0.461	0.485	0.464	0.850

Appendix B. continued

Snail bullhead	Overall	0.168	0.281	0.420	0.636
	Prediction	0.695	0.607	0.787	0.636
	Classification	0.691	0.635	0.561	1.000
Yellow bullhead	Overall	0.282	0.227	0.632	0.318
	Prediction	0.922	0.522	0.854	0.318
	Classification	0.943	0.562	0.872	1.000
Brown bullhead	Overall	0.101	0.338	0.627	0.681
	Prediction	1.076	0.770	0.905	0.681
	Classification	1.019	0.762	0.735	0.998
Cannel catfish	Overall	0.170	0.285	0.332	0.382
	Prediction	0.974	0.565	0.764	0.727
	Classification	0.987	0.570	0.484	0.764
Speckled madtom	Overall	0.443	0.115	0.498	0.124
	Prediction	0.682	0.534	0.943	0.565
	Classification	0.883	0.347	0.993	1.012
Pirate perch	Overall	0.331	0.165	0.412	0.091
	Prediction	0.640	0.852	0.553	0.091
	Classification	0.636	0.800	0.718	1.000
Eastern starhead topminnow	Overall	0.145	0.173	0.265	0.536
	Prediction	1.045	0.410	0.771	0.845
	Classification	1.036	0.424	0.313	0.853
Mosquitofish	Overall	0.445	0.068	0.564	0.068
	Prediction	0.838	0.068	0.950	0.068
	Classification	0.847	1.000	0.994	1.000
Brook silverside	Overall	0.309	0.175	0.425	0.159
	Prediction	0.703	0.748	0.773	0.990
	Classification	0.725	0.764	0.756	1.026
Shadow bass	Overall	0.150	0.445	0.652	0.613
	Prediction	1.096	0.947	0.896	1.463
	Classification	1.046	0.949	0.789	1.054
Banded pygmy sunfish	Overall	0.126	0.210	0.392	0.353
	Prediction	0.676	0.468	0.801	0.585
	Classification	0.690	0.456	0.471	0.575

Appendix B. continued

Redbreast sunfish	Overall	0.250	0.036	0.280	0.049
	Prediction	0.417	0.399	0.280	0.081
	Classification	0.875	0.061	1.000	0.957
Green sunfish	Overall	0.187	0.231	0.218	0.455
	Prediction	0.895	0.465	0.675	0.455
	Classification	0.968	0.465	0.556	1.000
Warmouth	Overall	0.433	0.159	0.549	0.114
	Prediction	0.742	0.765	0.549	0.114
	Classification	0.812	0.729	1.000	1.000
Dollar sunfish	Overall	0.224	0.228	0.421	0.353
	Prediction	0.833	0.462	0.767	0.374
	Classification	0.857	0.495	0.594	0.864
Redear sunfish	Overall	0.321	0.195	0.604	0.273
	Prediction	0.936	0.465	0.848	0.273
	Classification	0.950	0.581	0.874	1.000
Spotted sunfish	Overall	0.409	0.055	0.409	0.047
	Prediction	0.409	0.246	0.409	0.096
	Classification	1.000	1.010	1.000	1.001
Largemouth bass	Overall	0.543	0.068	0.549	0.068
	Prediction	1.031	0.068	0.549	0.068
	Classification	1.014	1.000	1.000	1.000
Shoal bass	Overall	0.116	0.200	0.204	0.415
	Prediction	0.530	0.441	0.656	0.704
	Classification	0.588	0.458	0.344	0.702
Black crappie	Overall	0.142	0.269	0.476	0.523
	Prediction	0.594	0.537	0.861	0.523
	Classification	0.728	0.544	0.723	1.000
Brown darter	Overall	0.277	0.202	0.341	0.176
	Prediction	0.794	0.415	0.624	0.308
	Classification	0.769	0.481	0.439	0.452

Appendix B. continued

Swamp darter	Overall	0.158	0.335	0.648	0.307
	Prediction	0.857	0.670	0.851	0.491
	Classification	0.943	0.670	0.779	0.614
Gulf darter	Overall	0.323	0.153	0.589	0.136
	Prediction	0.825	0.744	0.767	0.136
	Classification	0.845	0.865	0.850	1.000
Halloween darter	Overall	0.096	0.072	0.080	0.220
	Prediction	0.498	0.229	0.433	0.660
	Classification	0.290	0.334	0.184	0.465
Blackbanded darter	Overall	0.305	0.029	0.325	0.045
	Prediction	0.620	0.373	0.972	0.091
	Classification	0.917	0.030	0.997	0.904