# Michigan Stream Fish: Distribution Models, Future Predictions, and Urban Impacts

by

Paul J. Steen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Natural Resources and Environment)
in The University of Michigan
2008

Doctoral Committee:

Professor Michael J. Wiley, chair
Professor George W. Kling II
Adjunct Professor Paul W. Seelbach
Jeffrey Standford Schaeffer, US Geological Survey

To Kellie

Your daily support
and your continual sacrifices
allowed me to reach this goal.

And to Dad and Mom

Our walks through the woods
taught me to love God's creation
although I didn't realize it at the time.

# Acknowledgements

Many people contributed to this project, and this dissertation would not have been written without their assistance. My heart-felt thanks go to the members of my committee: Thank you Mike Wiley, Paul Seelbach, George Kling, and Jeff Schaeffer. Your help and ideas were essential in turning my scattered thoughts into cohesive arguments and interesting science. Thank you, especially, Mike: your constant guidance was invaluable for my time spent in graduate school.

The great people at the US Geological Survey, Great Lakes Science Center provided funding and an office to me for many years. Thank you, Dora Reader, for making all of it possible. Thank you, Jeff Schaeffer, Jaci Savino, and Leon Carl for providing me with such generous funding. To everyone else involved in the GAP project: Thank you, Jana Stewart, Jim McKenna, John Lyons, Scott Nelson, Steve Aichele, Ed Bissell, Limei Zhang, and Allain Rasolofoson. Your work in building habitat databases and GIS data is the backbone of this dissertation.

The scientists of the Michigan Department of Natural Resources, Fisheries Division were a constant resource, a sounding board, and the source of much of my data. Thank you, Troy Zorn, Paul Seelbach, Li Wang, and Kevin Wehrly. A great thank you goes to Arthur Cooper: You were ever faithful in answering my numerous email questions about the GIS data.

Finally, I wish to thank my family. My Dad, Mom, Father-in-law, Mother-in-law, and siblings and their spouses were all a source of encouragement and inspiration. Of course, the greatest thanks of all go to my wife, Kellie. Thank you, Kellie! This belongs to you as much as it does to me. Your encouragement helped me get something done every day, even on the days when I felt like I couldn’t do it. Thank you for believing in me.

Chapter 2 has been previously published. This article is Contribution 1309 of the USGS Great Lakes Science Center:

> Steen, P. J., D. R. Passino-Reader, and M. J. Wiley. 2006. Modeling brook trout presence and absence from landscape variables using four different analytical techniques. Pages 513-531 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Chapter 3 has been accepted for publication by the Transactions of the American Fisheries Society and is currently in press. This article is Contribution 1448 of the USGS Great Lakes Science Center.

# Table of Contents

Chapter

# List of Tables

## List of Figures

List of Appendices

# Abstract

To identify abiotic requirements necessary to maintain growth, survival, and reproduction of species, researchers often use models to search for patterns between species occurrences and environmental characteristics of sampled locations. Models are used to test hypotheses about processes that are important to organisms, and used to make species distributions and abundance predictions for management application.

Several analytical methodologies were assessed for modeling associations between fish species and habitat characteristics. Brook trout presence/absence models were created using different techniques: multiple regression, logistic regression, neural networks, and classification trees. Results showed that all methods could be successful provided underlying assumptions were met. This analysis indicated classification trees were a technique uniquely suited to the creation of large numbers of interpretable models.

Classification tree methods and landscape-scale habitat variables were used to create and validate presence/absence models and relative abundance models for Michigan stream fish. Ninety-three presence/absence models were on average 72% correct and 46 relative abundance models were on average 76% correct when tested against independent data. Water temperature and catchment area were determined to be dominant constraints on fish distributions.

Classification trees were applied to land-use alteration and climate change scenarios to understand how fish communities of the Muskegon River system (Michigan, USA) would be structured through the year 2100. Models predicted cold-water species reduction due to water temperature warming, and walleye loss due to increased urban development. Warm-water species were expected to have substantial range increases. Classification trees were used to explore how managers might manipulate predictor variables to maximize probability of species presence.

Urban development has been shown to have strong negative impacts on fish community quality. However, effort is needed into understanding why streams with similar urban levels have fish communities of significantly different quality. Univariate tests and covariance structure analysis were used to investigate how natural and anthropogenic features are related to variance of fish biotic integrity in urban streams. Urbanized streams with more natural land-cover, more point source discharges, better water quality, and that are adjacent to non-urbanized streams supported higher quality fish communities.

# Chapter 1

# Introduction

Understanding biotic and abiotic processes that determine where and why species are able to exist is one of the fundamental aims of ecology. It permeates all aspects of ecology: academic ecologists examine mechanisms of species coexistence and competition, while fisheries managers manipulate stream habitat to optimize managed species reproduction. Conservation biology, the branch of ecology dedicated to the preservation of biotic diversity, is particularly focused on understanding the mechanisms that regulate species distributions. Knowledge of the requirements necessary for growth, survival, and reproduction of species is required for wise conservation planning.

In practical application, however, the biotic and abiotic requirements of a species are incompletely known. Environmental complexity and species interactions make it difficult to learn the exact constraints on a population. To identify abiotic requirements, researchers often use statistical models to search for patterns between species occurrences or abundances and the environmental characteristics of sampled locations. These models serve two important purposes: (1) they are used to formulate and test hypotheses about the factors and processes that are important to organisms, and (2) they can used to make predictions of species distributions and abundances for use in management decisions.

Creating fish distribution models from habitat variables using regression analysis has a long history and strengths and weaknesses of this approach are well understood by researchers (Fausch et al. 1988). Neural network models and classification trees, however, are fairly new methods to the ecological field. Studies directly comparing these newer techniques with more traditional approaches are relatively rare, but when reported, have typically found that the new techniques are able to predict more accurately than simple linear modeling (Lek et al. 1996; Franklin 1998; Vayssieres et al. 2000). However, careful comparisons of performance of these approaches for modeling fish distributions at a large geographic extent, such as the extent of the state of Michigan, have not been previously reported (but see Mastrorillo et al. 1997; Olden and Jackson 2001,2002 for smaller scale analyses). Likewise, a direct comparison of neural net and classification tree approaches for fishes has not been previously reported. In Chapter 2, I compared the accuracy of Michigan brook trout distribution models created with regression, neural networks, and classification tree approaches.

Fish habitat requirements have often been modeled using site or local-scale environmental variables (Fausch et al. 1988). Habitat variables measured at this scale are useful to managers because small scale habitat can be manipulated (Fausch et al. 1988; Vaughan and Ormerod 2003). However, fish species are clearly influenced by processes that operate on larger spatial scales and slower temporal scales than that those measured at the local-scale (Richards et al. 1996; Leftwich et al. 1997; Rathert 1999; Allan 2004). For example, stream temperatures are critical to fish and influenced by a combination of local and landscape-scale processes (Wehrly et al. 2003; Wehrly et al. 2006). Also, the hydrologic flow regime of a stream is crucial to fish communities and is driven by factors measured at a catchment scale (Poff et al. 1997). While other fish distribution models based partially on landscape-scale data have been created for Michigan (Zorn 2003), a new database containing extensive fish samples and literally hundreds of habitat variables was recently developed (Brendan et al. 2006). A new regional modeling effort, exploring this new data source, is just beginning but promises to provide new insights into the importance of landscape-scale habitat variables on fish (Riseng et al. 2006; Wang et al. 2007). In Chapter 3, I used this data source to build 93 fish presence/absence models and 46 relative abundance models and used the models to predict fish distributions and identify fish requirements.

Fish distribution models can have an important role in conservation planning and management decisions. For example, models can be used to anticipate future changes in fish in order to give managers time to alter management practices and to identify areas with a high risk of habitat degradation. Several studies have examined how future temperature changes are likely to reduce cold-water fish distribution (Meisner 1990; Eaton and Scheller 1996; Flebbe 1996; Jager et al. 1999; Stefan et al. 2001) and increase the abundance of warm-water fish such as bass (McCauley and Kilgour 1990; Magnuson et al. 1990; King et al. 1999). Additionally, future land-use/cover shifts could have a large impact on fish community health as studies have consistently shown these to be related (Scott et al. 1986; Weaver and Garman 1994; Hall et al. 1999; Wang et al. 2001; Tabit and Johnson 2002; Snyder et al. 2003; Wang et al. 2003; Zimmerman et al. 2003; Miltner et al 2004; Barker et al. 2006; Riseng 2006). In chapter 4, I created fish distribution predictions based on potential future changes in both temperature and land-use/cover and associated them with a GIS to show how the changes will affect specific streams. Such model applications have not been developed for Michigan prior to this study.

The quality of fish communities seems to be correlated with the amount of urbanized land in a stream's watershed (Wang et al. 2001; Wang and Kanehl 2003; Wang et al. 2003; Walsh 2004; Carter and Fend 2005; Fitzpatrick et al. 2005; Kennen et al. 2005; Limburg et al. 2005; Walters et al. 2005). In the United States, over 130,000 kilometers of streams and rivers have already been affected by urbanization, and land-use change projections predict that developed area is going to increase by 80% in the next 25 years (Paul and Meyer 2001; Pijanowski et al. 2001; Alig et al. 2004; Walsh et al. 2005). As stopping the spread of urbanization altogether is not a politically feasible or necessarily desirable goal, it is important to understand how to aid the establishment or maintenance of high quality fish communities despite the amount of urban disturbance. While previous studies have concentrated on the relationship between urbanization and fish integrity (Weaver and Garman 1994; Kemp and Spotila 1997; Tabit and Johnson 2002; Walters et al. 2003; Morgan and Cushman 2005), I was unable to find studies that gave a quantitative analysis of the variance around this relationship. In Chapter 5, I conducted a study to account for this variance and discussed important management implications for fish communities in urban streams.

The overall objectives of my dissertation are:

1) To evaluate the methodology for four different presence/absence modeling techniques using data from fish sampling and enduring landscape habitat variables for rivers across the state of Michigan (Chapter 2).

2) To use the selected methodology to build distribution and abundance models for all common Michigan stream fish and to assess, describe, and understand the model patterns and relationships (Chapter 3).

3) To use the models to probe some key issues in fish community conservation; explore potential future fish distributions in the Muskegon river system given changes in land-use and temperature (Chapter 4).

4) To understand more fully the relationship between urbanization, fish community integrity, and the variation that occurs around this relationship, and to use this information to understand how urban streams can possess healthy fish communities (Chapter 5).

# References

Alig, R. J., J. D. Kline, and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. Landscape and Urban Planning 69:219-234.

Allan, J. D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35:257-284.

Barker, L. S., G. K. Felton, and E. Russek-Cohen. 2006. Use of Maryland biological stream survey data to determine effects of agricultural riparian buffers on measures of biological stream health. Environmental Monitoring and Asessment 117:1-19.

Brendan, T. O., R. D. Clark, A. R. Cooper, P. W. Seelbach, L. Wang, S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49-74 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape Influences on Stream Habitat and Biological Assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland

Carter, J. L., and S. V. Fend. 2005. Setting limits: The development and use of factor-ceiling distributions for an urban assessment using macroinvertebrates. Pages 179-191 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Eaton, J. G., and R. M. Scheller. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. Limnology and Oceanography 41.

Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream fish from habitat variables: 1950-1985. General Technical Report PNW-GTR-213, United States Department of Agriculture, Forest Services, Pacific Northwest Research Station, Portland, OR.

Fitzpatrich, F. A., M. W. Diebel, A. H. Mitchell, T. L. Arnold, M. A. Lutz, and K. D. Richards. 2005. Effects of urbanization on the geomorphology, habitat, hydrology, and fish index of biotic integrity of streams in the Chicago, Illinois, and Wisconsin. Pages 87-116 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Flebbe, P. A., L. D. Roghair, and J. L. Bruggink. 2006. Spatial modeling to project southern Appalachian trout distribution in a warmer climate. Transactions of the American Fisheries Society 135:1371-1382.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9:733-748.

Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture, urbanization, and climate on water quality in the northern Great Plains. Limnology and oceanography 44:739-756.

Jager, H. I., W. Van Winkle, and B. D. Holcomb. 1999. Would hydrologic climate changes in Sierra Nevada streams influence trout persistence? Transactions of the American Fisheries Society 128:222-240.

Kemp, S. J., and J. R. Spotila. 1997. Effects of urbanization on brown trout Salmo trutta, other fish and macroinvertebrates in Valley Creek, Valley Forge, Pennsylvania. The American Midland Naturalist 138:55-68.

Kennen, J. G., M. Chang, and B. H. Tracy. 2005. Effects of landscape change on fish asseblage structure in a rapidly growing metropolitan area in North Carolina, USA. Pages 39-52 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

King, J. R., B. J. Shuter, and A. P. Zimmerman. 1999. Empirical links between thermal habitat, fish growth, and climate change. Transactions of the American Fisheries Society 128:656-665.

Leftwich, K. N., P. L. Angermeier, and C. A. Dolloff. 1997. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Transactions of the American Fisheries Society 126:725-734.

Lek, S., M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga, and S. Aulagnier. 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling 90:39-52.

Limburg, K. E., K. M. Stainbrook, J. D. Erickson, and J. M. Gowdy. 2005. Urbanization consequences: Case studies in the Hudson River watershed. Pages 23-37 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Magnuson, J. J., J. D. Meisner, and D. K. Hill. 1990. Potential changes in the thermal habitat of Great Lakes fish after global climate warming. Transactions of the American Fisheries Society 119:254-264.

Mastrorillo, S., S. Lek, F. Dauba, and A. Belaud. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshwater Biology 38:237-246.

McCauley, R. W., and D. M. Kilgour. 1990. Effect of air temperature on growth of largemouth bass in North America. Transactions of the American Fisheries Society 119:276-281.

Meisner, J. D. 1990. Potential loss of thermal habitat for brook trout, due to climatic warming, in two southern ontario streams. Transactions of the American Fisheries Society 119:282-291.

Miltner, R. J., D. White, and C. Yoder. 2004. The biotic integrity of streams in urban and suburbanizing landscapes. Landscape and Urban Planning 69:87-100.

Morgan, R. P., and S. E. Cushman. 2005. Urbanization effects on stream fish assemblages in Maryland, USA. Journal of the North American Benthological Society 24:643-655.

Olden, J. D., and D. A. Jackson. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130:878-897.

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

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. The Annual Review of Ecology and Systematics 32:333-365.

Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. Journal of the North American Benthological Society 16:391-409.

Pijanowski, B. C., D. G. Brown, B. A. Shellito, and G. A. Manik. 2001. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems 26:553-575.

Rathert, D., D. White, J. C. Sifneos, and R. M. Hughes. 1999. Environmental correlates of species richness for native freshwater fish in Oregon. Journal of Biogeography 26:257-273

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influence on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Sciences 53:295-311.

Riseng, C. M., M. J. Wiley, R. J. Stevenson, T. G. Zorn, and P. W. Seelbach. 2006. Comparison of coarse versus fine scale sampling on statistical modeling of landscape effects and assessment of fish assemblages of the Muskegon River, Michigan. Pages 555-575 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors.

Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Scott, J. B., C. R. Stewart, and Q. J. Stober. 1986. Effect of urban development on fish population dynamics in Kelsey Creek, Washington. Transactions of the American Fisheries Society 115:555-567.

Snyder, C. D., J. A. Young, R. Villella, and D. P. Lemarie. 2003. Influences of upland and riparian land use patterns on stream biotic integrity. Landscape Ecology 18:647-664.

Stefan, H. G., X. Fang, and J. G. Eaton. 2001. Simulated fish habitat changes in North America lakes in response to projected climated warming. Transactions of the American Fisheries Society 130:459-477.

Tabit, C. R., and G. M. Johnson. 2002. Influence of urbanization on the distribution of fishes in a southeastern upper Piedmont drainage. Southeastern Naturalist 1:253-268.

Vaughan, I. P., and S. J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology 17:1601-1611.

Vayssieres, M., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science 11:679-694.

Walsh, C. J. 2004. Protection of in-stream biota from urban impacts: minimise catchment imperviousness or improve drainage density? Marine and Freshwater Research 55:317-326.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. The Journal of the North American Benthological Society 24:706-723.

Walters, D. M., D. S. Leigh, and A. B. Bearden. 2003. Urbanization, sedimentation, and the homogenization of fish assemblages in the Etowah River Basin, USA. Hydrobiologia 494:5-10.

Walters, D. M., M. C. Freeman, D. S. Leigh, B. J. Freeman, and C. M. Pringle. 2005. Urbanization effects on fishes and habitat quality in a Southern Piedmont river basin. Pages 69-85 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization of stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 2001:255-266.

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

Wang, L., J. Lyons, P. W. Rasmussen, and P. W. Seelbach. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Canadian Journal of Fisheries and Aquatic Sciences 60:491-505.

Wang,L., T. Brenden, P. W. Seelbach, A. Cooper, D. Allan, R. Clark, Jr., and M. Wiley. 2007. Landscape based identification of human disturbance gradients and references for Michigan streams. Environmental Monitoring and Assessment. Online-First.

Weaver, L. A., and G. C. Garman. 1994. Urbanization of a watershed and historical changes in a stream fish assemblage. Transactions of the American Fisheries Society 123:162-172.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2003. Classifying regional variation in thermal regime based on stream fish community patterns. Transactions of the American Fisheries Society 132:18-37.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2006. Influence of landscape features on summer water temperatures in Lower Michigan streams. Pages 113-127 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Zimmerman, J. K. H., B. Vondracek, and J. Westra. 2003. Agricultural land use effects on sediment loading and fish assemblages in two Minnesota (USA) watersheds. Environmental Management 32:93-105.

Zorn, T. G. 2003. Fishes of lower Michigan: distribution patterns, abundance models, and causal relationships. Doctoral Dissertation. University of Michigan, Ann Arbor.

# Chapter 2

# Modeling Brook Trout Presence and Absence from Landscape Variables Using Four Different Analytical Techniques

## Abstract

As a part of the Great Lakes Regional Aquatic Gap Analysis Project, I evaluated methodologies for modeling associations between fish species and habitat characteristics at a landscape scale. To do this, I created brook trout *Salvelinus fontinalis* presence and absence models based on four different techniques: multiple linear regression, logistic regression, neural networks, and classification trees. The models were tested in two ways: by application to an independent validation database and cross-validation using the training data, and by visual comparison of statewide distribution maps with historically recorded occurrences from the Michigan Fish Atlas. Although differences in the accuracy of our models were slight, the logistic regression model predicted with the least error, followed by multiple regression, then classification trees, then the neural networks. These models will provide natural resource managers a way to identify habitats requiring protection for the conservation of fish species.

## Introduction

It is necessary to have knowledge of the habitats required to maintain the growth, survival, and reproduction of freshwater fish species and populations in order to have wise conservation planning and decision making. In practical application, however, habitat requirements are often incompletely known. Therefore, biologists commonly use data on a fish's habitat selection, based on field observations of species occurrence or densities (Rosenfeld 2003). Given data on habitat characteristics and observed fish distributions, correlative habitat associations can be used to predict the occurrence or densities of fish in locations where samples have not been collected. These predictions are useful for identifying habitat units important to target species but vulnerable to alteration and degradation by humans, and lacking protective status. Such habitats represent "gaps" in conservation strategy.

The goal of the U. S. Geological Survey, Gap Analysis Program (GAP) is to "keep common species common" by identifying those species not adequately represented in existing conservation areas (Scott et al. 1993). In the past decade, gap analyses have been performed in terrestrial systems across the U.S., and in the mid-1990s an aquatic gap pilot began in Missouri. In 2001 GAP funded the first regional aquatic gap analysis in the eight Great Lakes states: Minnesota, Wisconsin, Illinois, Indiana, Michigan, Ohio, Pennsylvania, and New York. The goals of this project, called the Great Lakes Regional Aquatic Gap Analysis, are 1) to evaluate biological diversity of Great Lakes aquatic habitats and identify gaps in the distribution and protection of these species and their habitats, and 2) to use an integrated approach in which common methods and protocols are established and results are comparable across the Great Lakes landscape (Myers et al. 2002; Morrison et al. 2003).

A critical step in achieving the aquatic gap project goals is to predict patterns of species occurrence from regional habitat data. Earlier aquatic gap projects used several different methods to model empirical associations of fish species presence and absence with habitat characteristics: 1) classification and regression trees (CART) analysis (Sowa 1999), 2) multiple linear regression (Sowa and Rabeni 1995; Sowa 1999), 3) and genetic algorithms (Alex Covert, USGS Ohio Water Science Center, personal communication). In this study, I build on the earlier aquatic gap projects by assessing several methods (multiple linear regression, logistic regression, neural networks, and classification trees) and comparing their predictive abilities.

Predicting fish distributions from habitat variables using regression analysis has a long history in ecological applications and is well understood by researchers (Fausch et al. 1988). Neural networks and classification trees, however, are fairly new methods to the ecological field. This study uses these techniques but does not go into detail in describing how they work; this has been done well in other papers both for neural networks (Rumelhart et al. 1986; Mastrorillo et al. 1997; Boddy and Morris 1999; Lek and Guegan 1999; Olden and Jackson 2001) and for classification trees (Breiman et al. 1984; Bell 1999; De'ath and Fabricius 2000; De'ath 2002).

Studies directly comparing these newer techniques with more traditional approaches are relatively rare, but where they exist have typically found that the new techniques are able to predict more accurately than simple linear modeling (Lek et al 1006; Franklin 1998; Vayssieres et al. 2000). However, careful comparisons of performance of these approaches for modeling fish distributions at the large geographic extent contemplated in the Great Lakes Aquatic Gap Program have not been previously reported (but see Mastrorillo et al. 1997; Olden and Jackson 2001,2002 for smaller scale analyses). Likewise, comparison of neural net and classification tree approaches for fishes have not been previously reported.

The main goal of this study was to evaluate the methodology for four different presence/absence modeling techniques (multiple regression, logistic regression, neural networks, classification trees) using data from fish sampling and enduring landscape habitat variables for rivers across the state of Michigan. This model comparison will aid us in selecting the approach, or approaches, to be used to produce fish distribution maps for the Great Lakes Regional Aquatic Gap Analysis. For this study, I analyzed the streams of Michigan for the presence and absence of brook trout *Salvelinus fontinalis*, a popular sport fish whose basic habitat requirements are well known (Smith 1985).

## Methods

### Developing the database

The Great Lakes Regional Aquatic Gap Analysis, in collaboration with the Michigan Department of Natural Resources (MDNR), has established a high-resolution, GIS-linked database with characteristics of Michigan's rivers. This database provided the environmental variables that served as the independent predictors for the models. The database is referenced to a group of ArcGIS line coverages (ESRI 2002), in which each river is broken down to confluence-to-confluence reaches, and each reach contains information for a wide variety of landscape-scale environmental variables, such as air temperature, soil permeability, land-cover, and geology (S. Aichele, USGS, personal communication)(Table 2.1, Figure 2.1). The line coverages are based on the USGS National Hydrography Dataset at the 1:100,000 scale (NHD 2007).

Many variables are measured at four different scales (Figure 2.2). Riparian variables refer to the land 60 meters on each side of the stream reach of interest. Network variables refer to the 60-meter river buffer, plus the 60-meter river buffer of every stream reach upstream from the reach of interest. Sub-watershed variables refer to the catchment lateral to the stream reach, and watershed variables include the reach's catchment and the catchment of all the reaches upstream.

I obtained spatially referenced fish assemblage samples from the Fisheries Division of the MDNR and extracted data for brook trout occurrences from this "training" dataset. Fish were collected by tow-barge, backpack and boat electrofishing, rotenone, seines, trap nets, and fyke nets (Merna 1988). While the amount of data available from the MDNR was extensive, I limited our analysis to samples collected in 1980-2002, and strived for even spatial coverage across the state (Figure 2.1A). To test the models, I used an independent data set from the Michigan Rivers Inventory (MRI) project (Seelbach and Wiley 1997), where fish were collected primarily by tow-barge and backpack electrofishing and rotenone in 1980-2002 (Figure 2.1B).

Sampling points were associated to the stream reaches in a single table in which rows represented stream reaches and columns contained values for the habitat variables and a record of brook trout presence (E. Bissell, USGS, Water Resource Division, Lansing, Michigan, Personal Communication).

I deleted replicate samples so that every reach was represented by only one observation. When different samples for the same reach disagreed on brook trout presence, I kept the observation where the fish was present. After eliminating replicates, I had 901 observations for the MDNR data and 635 for the MRI data. Hereafter, observations marked as "present" are called "presence reaches" and observations marked as "absent" are called "absence reaches".

As with many presence and absence databases, the number of absence reaches was much greater than the number of presence reaches (Zorn 2003). To prevent the models from weighting more towards absence prediction than presence, I randomly selected a subset of absence reaches equal to the number of presence reaches. To do this, I first divided the 682 absence reaches of the MDNR data into three groups by Shreve link number. Next, I randomly selected 104 or 105 absence reaches from each size group so that the total number of absence reaches (314) equaled the total number of presence reaches. These 628 MDNR observations were used as the training data for formulating the models. The absence segments that were not selected were discarded, but I kept all 635 reaches of the MRI data to validate the models formulated by the MDNR data.

Stratification by Shreve link number was necessary to ensure that the whole range of stream sizes available were included in the modeling. Had I merely used simple random sampling to select the 314 absence reaches, the sampling would have been biased towards smaller rivers due to the high ratio of small to large streams in the database. The end result would have been models only applicable to small streams.

### Multiple Regression and Logistic Regression Modeling

For the multiple regression model, when necessary I transformed each variable with one of three transformations (logarithmic, square root, or arcsine) in order to meet the assumption of a linear relationship between the independent and dependent variable (Zar 1999) (Table 2.1). Upon creation of the final multiple regression model, partial regression plots were created for each of the model's predictors to test the assumption of linearity. These plots show the effect of a predictor on the response variable with the effect of all of the other predictors removed (Faraway 2005). For both the multiple regression and logistic regression, each variable was standardized (mean 0, standard deviation 1) to allow for easier comparison of the relative importance of the regression coefficients (Faraway 2005).

Using a stepwise selection technique similar to that of Zorn (2003), I ran multiple linear regression and logistic regression models in SPSS 12.0 for Windows (SPSS 2003). For both regression models, I manually entered a habitat variable into the equation. If the variable was significant ($p < 0.10$), I left it in the equation and added another variable. If the new variable was not significant ($p \geq 0.10$), I removed it from the model. If the new variable was significant and caused the original variable to become insignificant, I removed the original variable if its removal caused the model's adjusted $R^2$ to increase. I repeated this procedure until every environmental variable had been given a chance to enter the model. While SPSS can carry out this procedure automatically, I performed it manually in order to dictate the order that variables entered the model (Table 2.1). I placed variables assumed to have a more direct effect on fish presence and absence higher in the list, and variables with indirect or unknown effects lower in the list. This procedure gives variables assumed to be more important to fish a greater chance of being included in the model (Zorn 2003).

As the next step, I examined the $\beta$-value (regression coefficient) and the 90% confidence interval of the $\beta$-value for each variable included in the model. If the $\beta$-value had a confidence interval large enough that I was not confident in its predictive abilities, I would remove the variable from the model. If I believed a variable to be important but it was not included in the model, I would add this variable and recalculate the model. If the variable was significant ($p < 0.10$) I allowed the variable to stay in the model. By adding these steps to the end of the regression modeling process, I added subjective decision making based on our ecological knowledge of the fish to an otherwise objective routine.

To check if the constant variance and independence assumptions of the error terms of the multiple regression model held true, I created a diagnostic plot of the residuals versus fitted values. Patterns in this plot indicate assumption violations (Faraway 2005). A Q-Q plot of the residuals was created to check error normality; a straight line in this plot indicates normality (Faraway 2005).

After creating the final regression models, I applied the model to the MRI test data and also performed n-fold (leave-one-out) cross validation on the training data as a secondary test of the model's performance. Reaches with a predicted value $\geq 0.5$ were considered to have trout present; those with values $< 0.5$ were considered as lacking trout.

## Neural Network Modeling

In using a neural network, it is possible to include all of the available data in the network and get a solution that effectively predicts the dependent variable (Lek et al. 1996). However, many of our environmental variables are probably not related to brook trout presence, and including them in the model will increase computation times and cloud our ability to understand the relationship between the fish and more important variables (Olden and Jackson 2002). Therefore, in order to build a model that can both predict and provide some explanatory value, I needed to reduce the initial 46 habitat variables into a more manageable number.

Olden and Jackson (2002) have developed a randomization approach for both pruning variables and understanding how variables contribute to a neural network. To use this approach with our data, I first developed several neural networks in a feed-forward, back propagation procedure using the training data and the computer software program Neuralyst 1.4 (Rumelhart et al. 1986; Shih 1995; Boddy and Morris 1999). The networks were constructed with three layers of 46 neurons, 46 neurons, and 1 neuron, respectively, and were trained for 1000 epochs (Shih 1995). However, each of these networks had different random initial weights and so performed differently in how well they predicted the test data.

From these networks, I selected the network that had the greatest percentage of correct predictions for the test data (our "optimized" network) and calculated the sum of the input layer-hidden layer weights and the hidden layer-output layer weights for each input variable (Figure 2.3). Essentially, the contribution of each variable to the network depends on the magnitude and direction of the sum of these weights. The greater this sum (either negatively or positively), the more impact that variable has on the final solution produced by the network (Olden and Jackson 2002).

The sum of the weights alone does not fully inform us of how important a variable is, because at this point I do not know if the sum is significantly different than random. To determine significance, I randomly rearranged the brook trout presence and absence values among the different observations, and then constructed a new neural network with the same parameters and initial weights that were used in our optimized network. This network was allowed to train for 1000 epochs, and then as earlier I computed the sum of the input layer-hidden layer weights and the hidden layer-output layer weights for each habitat variable. Since I rearranged the presence/absence values among the observations, if a variable is important to brook trout the new network will not produce a sum of weights that is of greater magnitude than the sum of weights from the original network.

However, there is a slight chance that by rearranging the values, I actually made a variable more important to brook trout presence. Therefore, I created a new network many times (in this case, 1000 times), each time randomly rearranging the presence/absence values, and each time calculating the sum of the weights for each variable. The statistical significance of each habitat variable was the proportion of the values (including the original sum) that were more extreme than the observed sum (Olden and Jackson 2002). For example, if only 9 of the 1000 sums are greater than our observed sum, then the probability of a type I error and statistical significance of the variable was (9+1)/1000= .01. When a variable had a low p-value, I inferred that it played an important role in the formation of the neural network. This significance test was similar to the significance test of a regression coefficient: I was testing the null hypothesis that a variable does not have an effect, and at low probabilities the null hypothesis was rejected and I concluded that the variable did have an effect.

After the 1000 iterations, seven variables had a significance level less than 0.1. These variables were considered to be the most important of the original 46 habitat variables and were used to construct a new neural network. This neural network was created with the training data in a manner similar to our original network. I applied this new network to the MRI test data to determine its predictive abilities and applied the network to the training data through a n-fold cross-validation procedure as a secondary test of the model’s performance. Reaches with a predicted value $\geq$ 0.5 or greater were considered to have trout present; trout were presumed absent at values < 0.5.

## Classification Tree Modeling

I used CART 5.0 to train the training data in a classification tree (Steinbery and Colla 1997). Since the program selects the variables that result in the best training of the data, I did not have to enter variables individually or prune variables as I did for the regression and neural network techniques. CART produced a series of trees in which different predictor variables were used as binary splits. As CART created the trees, it also tested the training data in a n-fold cross-validation procedure. The tree that I selected to use as our predictive model was the one that resulted in the highest agreement between the predicted presence and absence classification and the known presence and absence classification for the cross-validation. After selecting the optimal tree, I ran the MRI test data through the tree as a measure of the model’s validity. Unlike the other modeling methods, the classification tree did not predict a value for each reach; rather, the model directly classified a reach as either “present” or “absent.”

## Model Application

The models were used to predict brook trout presence and absence for all of the stream reaches in Michigan. I used these predictions to create statewide distribution maps in ArcMap 8.3 (ESRI 2002). The predictions were entered into a spreadsheet and then joined to the stream layer by a number unique to each reach. The background layers for these maps came from the Michigan Geographic Data Library (MGDL 2007) and the stream layer was developed by the Great Lakes GAP Project and the MDNR, as mentioned above.

After building the predictive maps, I visually compared these with the Michigan Fish Atlas 2003, v.1.1, 2nd edition (MGDL 2007). The Michigan Fish Atlas is a point shapefile that contains 2468 georeferenced brook trout samples collected from 1847 to the present. As these samples came from locations where brook trout are known to live, a successful predictive map will have similar distribution patterns to the Fish Atlas.

## Results

### Multiple Regression

Eight variables were included in the multiple regression model (Adjusted $R^2$ = 0.436, Table 2.2). Of these, the most influential was the July mean air temperature (W_JULY_MN, $\beta = -0.184$). Other important variables included stream size (CHAN_LINK, $\beta = -0.133$), and the percentage of forest land-cover in the watershed (W_FOR, $\beta = 0.091$).

The model was applied to the MRI test data and predicted 86% of the presence reaches correctly and 76% of the absence reaches correctly (Table 2.3). When the training data were used in a cross-validation test, 81% of the presence reaches were predicted correctly and 80% of the absence reaches were predicted correctly.

### Logistic Regression

Nine variables were included in the logistic regression model (Cox and Snell $R^2$ = 0.480, Table 2.4). Most of the same variables that were significant in the multiple regression were also significant in the logistic regression. The most influential variable in both models was July mean air temperature (W_JULY_MN, $\beta = -1.3614$). Stream size (CHAN_LINK, $\beta = -1.2475$) and percentage of forest land-cover in the watershed (W_FOR, $\beta = 0.4737$) were again important.

The logistic regression model was applied to the MRI test data and predicted 87% of the presence reaches correctly and 75% of the absence reaches correctly (Table 2.3). In cross-validation tests, 80% of the presence reaches were predicted correctly and 81% of the absence reaches were predicted correctly.

### Neural Network

I tested two different sizes of neural networks: a full model, which used all 46 variables, and a pruned model, which used the seven most important variables from the full model. The full model predicted the MRI validation data set well: 85.7% of the presence reaches were predicted correctly and 71.8% of the absence reaches were predicted correctly (Table 2.3). The cross-validation of the full model predicted the presence and absence reaches correctly 75.4% and 77.7%, respectively.

By using the randomization procedure, I identified the seven most important variables from this model (Table 2.5). These included the July mean air temperature, which had a negative relationship with brook trout presence, and the Darcy value of the riparian zone, which had a positive relationship with brook trout presence. Also important were channel gradient, and wetland and open/field land-covers.

The simplified model, which contained these seven predictors, was also applied to the MRI test data and the training data were cross-validated (Table 2.3). The pruned model predicted absence reaches better than the full neural network but presence reaches were predicted worse. For the MRI data, 81.3% of the presence reaches were predicted correctly and 77.4% of the absence reaches were predicted correctly (Table 2.3). The cross-validation of the pruned model predicted the presence and absence reaches correctly 68.8% and 84.4%, respectively.

### Classification Tree

The CART program produced several trees of differing sizes, and the tree selected as the final model was the one that best predicted the cross-validation data (Figure 2.4). The first split in the tree was made by July mean air temperature (W_JULY_MN), and similarly to the regression models, other important variables included stream size (CHAN_LINK), and percent of forest land-cover in the watershed (W_FOR). One unique property of classification trees is that the model can use the same variable more than once; at the bottom of the tree the model uses the W_JULY_MN variable again to make another split.

I applied the MRI test data to the tree to validate the model. The tree predicted 84.3% of the presence reaches and 77.7% of the absence reaches correctly. (Table 2.3). Cross-validation of the training data predicted 75.2% of the presence reaches and 78.3% of the absence reaches correctly.

### Predicted Distribution Maps and Model Comparisons

The Michigan Fish Atlas shows that brook trout has historically been found throughout Michigan’s Upper Peninsula and northern Lower Peninsula (Figure 2.5A). While the populations are not as dense as in the north, brook trout is also found in southwest Michigan, as well. This species generally does not live in southeast Michigan, although the Fish Atlas does record a few scattered populations there.

The distribution maps produced from our four models were able to repeat this general pattern (Figures 2.5B-2.5F). Each map predicted brook trout to be widespread in the Upper Peninsula, with an occasional stream marked as absent. In the Lower Peninsula, the models predicted brook trout throughout the north and along the west coast, and generally predicted absence in southeast Michigan. The map created by the logistic regression model appeared to be most accurate when compared to the Fish Atlas; all of the maps are examined more closely below.

## Discussion

### Ecological Significance

An advantage of performing this modeling exercise on brook trout is that habitat requirements of this fish are well known, so it is relatively easy to determine if our models are consistent with ecological knowledge of the fish. In general, I would expect brook trout to prosper in small to medium size streams with plenty of groundwater flow, which provides cold water and a stable environment (Smith 1985; Zorn et al. 2002). Wehrly et al. (2003) reported that brook trout in Lower Michigan are restricted to streams with mean July water temperatures < 21 °C. Since streams typically grow warmer as they grow larger (Wehrly et al. 1997), I should expect to find brook trout more consistently in smaller, headwater streams rather than in larger rivers (Smith 1985; Zorn et al. 2002). In terms of land-cover, I expect that streams favorable to brook trout would have minimal thermal pollution. Therefore I expect that streams with riparian zones and catchments high in forests, and low in urban and agricultural land-uses, would be more likely to provide suitable habitat for the fish.

## Regression Models

In the multiple regression model, six of the variables have β value signs that are consistent with our knowledge of brook trout ecology (Table 2.2), and in the logistic regression model, eight variables have consistent signs (Table 2.4). For example, the lower the July mean air temperature (W_JULY_MN) and the smaller the stream (CHAN_LINK), the more likely that brook trout is predicted present. If the stream has a high slope in the catchment (W_SLOPE) and high soil permeability in the riparian zone (RT_PERM), two variables that increase groundwater flow, I would also expect a greater chance of brook trout presence. All of these variables are more likely found in headwater streams, so it is reasonable that the further away the reach is from the Great Lakes (DOWNLENGTH), the more likely it is to contain brook trout.

Prior to examining the regression coefficients, I was unsure of how two of the significant variables would affect brook trout presence: percent of land-cover containing wetlands and percent of land-cover containing open water. The models predicted that wetlands and open water would negatively influence presence. This seems reasonable as these land-covers could result in surface water warming, which contributes to poor brook trout habitat (Wehrly et al. 2003).

Hindering our interpretation of the significant regression variables is the problem of multicollinearity. Due to existing correlations between the variables, I can only interpret the effect of a predictor on the response when I also consider the effect of every other variable on the response at the same time (Faraway 2005). Understandably, with close to ten predictors in our models, this is a very difficult task. The result is that I can really only generalize about an individual predictor; I can say a predictor seems to have a certain effect but cannot say the exact magnitude of that effect. However, multicollinearity does not affect the accuracy of our predictions; it only affects the explanatory value of the model.

The multiple regression and logistic regression models fit the MRI test data well, especially in regards to predicting presence, which both models correctly predicted over 85% of the time (Table 2.3). The cross-validation of each model was also quite successful, predicting the reaches correctly about 80% of the time. In order to determine which model predicted better overall, I added the percent correct prediction for the MRI test data and cross-validation to produce a value that I called “Performance” (Table 2.3). Using this value, it appears that logistic regression predicted the data slightly better than the multiple regression (324.1 > 321.6).

The statewide mapped predictions seem generally reasonable for both models, both predicting brook trout occurrences not only in the north but also scattered in the southern interlobate and glacial outwash regions of Western Michigan where in fact isolated populations do occur (Figure 2.5B,C). However, the logistic regression map followed the Fish Atlas patterns more closely. In the Fish Atlas, brook trout is uniformly distributed in the Upper Peninsula, except for a few areas (i.e. the Manistique, the Cedar, and the Escanaba watersheds) in which there are “holes”. The map produced by logistic regression properly models both the Manistique hole and the Cedar-Escanaba hole. In the Lower Peninsula, the Fish Atlas shows that brook trout are not found in the streams of the Pigeon and Birch watersheds, which are located in the area of land protruding into southern Lake Huron (this area is called the "tip of the thumb" due to the generally accepted idea that the Lower Peninsula looks like a mitten). The logistic regression model accurately predicts absence in this area; the multiple regression largely predicts absence but does predict presence in several streams.

In addition, the logistic regression model predicted a band of brook trout presence in southeast Michigan that runs in a southwest-northeast direction. Part of this band can be seen in the Fish Atlas (Figure 2.5A). This band lies on a glacial interlobate formation, and has higher slopes, higher soil permeability, and faster groundwater flow than the flat lake plain geology of the rest of southeast Michigan (Bent 1971). It is reasonable to expect that the streams of this area have the proper habitat to support brook trout.

### Multiple Regression Assumptions

In general, multiple regression works best when the response variable is continuous, not dichotomous or categorical (Zar 1999). The multiple regression model predicted quite well, but the question must be asked if it is acceptable to use this model with presence-absence data.

I checked the assumptions of the models with diagnostic plots. The partial regression plots demonstrated that transformations of the variables helped improve the linear relationship between the predictors and the response, but did not achieve perfect linearity, resulting in a model that has less fit than a perfect linear model. In addition, diagnostics on the residuals showed that while the error terms met the assumption of linearity, they violated the assumption of constant variance and independence. As a result, probability based assessments of the model parameterization and goodness of fit are in question. While this model was interesting as an intellectual exercise, the result of these problems is that I would not use a model of this type for critical management decisions. Logistic regression, which was developed for dichotomous responses and which has much less stringent assumptions, is certainly the presence/absence regression model of choice.

### Neural Network Models

Both the full neural network and the pruned neural network did a good job in predicting the test data and accurately cross-validating the training data. The variables identified as significant in the full model were, in general, consistent with our ecological knowledge of brook trout. Air temperature had a negative relationship with brook trout presence; the regression models predicted the same relationship (Table 2.5). In addition, the Darcy variable and channel gradient had a positive relationship with brook trout presence, which is also similar to the regression models. The effect of the land-cover variables on brook trout was not clear. Both open/field land-cover and wetland land-cover variables were significant in the model, but each of these variables were included in two different scales that had opposite relationships to brook trout. For example, WT_OPEN had a positive relationship to presence, but W_OPEN had a negative relationship to presence (Table 2.5). As these two variables are positively correlated (r = 0.66), I would expect them to have a similar effect, but our analysis showed that they did not.

Interestingly, the pruned network predicted the test data and the cross-validation of the training data just as well as the full network (performance value: 311.9 to 310.6), demonstrating the validity of the randomization approach for network pruning (Table 2.3). However, examination of the predictive maps produced by the models showed that the full network map was more realistic than the pruned network map (Figure 2.5D,E). The pruned network captures well the Manistique hole and the Cedar-Escanaba hole in the Upper Peninsula and the glacial interlobate band in the southeast Lower Peninsula, but it predicts presence in the thumb and draws an unexpected straight line of predicted presences running west to east across the Lower Peninsula. After some investigation into this line, it appeared that the line marks a change in the stream reaches' air temperature values. A similar problem with the air temperature variable was seen in the classification tree model and is discussed more in the next section.

The predictive abilities of the pruned network seemed reliable until they were applied to the whole state and viewed geographically. For this particular study, I decided it would be best to run the randomization procedure to determine variable significance, but use the full model in making the predictions. Other fish may be able to be modeled reasonably through the pruned network; each case should be evaluated individually.

### Classification Tree Model

The classification tree model included several variables that were also in the regression models (i.e. W_JULY_MN, CHAN_LINK, W_FOR). The tree splits the data with these variables in a way that is consistent with our ecological knowledge of brook trout, and this model predicted the test data and cross-validation of the training data almost as well as the regression models. The performance value of this model (315.5) is lower than the regression models and higher than the neural network models, but all of these differences are actually quite small (Table 2.3).

The statewide predictive map is generally reasonable for the classification tree model; it follows the general pattern shown in the Michigan Fish Atlas and in the regression models (Figure 2.5F). The model does not correctly capture the Manistique hole and the Cedar-Escanaba hole, as did the logistic regression model. In addition, several streams at the tip of the thumb are predicted present, but the Fish Atlas does not record brook trout being found in this area. This problem is most likely the result of the importance placed on the July mean air temperature variable; the tip of the thumb is far enough north to have a lower air temperature than the rest of the thumb, and lower air temperatures cause the fish to be predicted present. This problem actually represents an important problem with all of the models using air temperature.

Studies have shown that water temperature, not air temperature, is the most important habitat variable for fish (Wehrly et al. 2003; Zorn 2003). In this study, water temperature data were not available, so I used air temperature only. It is a curious coincidence that the air temperature the classification model chose as a maximum cutoff for brook trout is the same mean water temperature (21 °C) reported by Wehrly et al. (2003) as the upper limit for Michigan brook trout. Water temperatures and air temperatures are in fact not predictably related in Michigan due to the spatially variable contribution of ground water to stream channels (Wiley et al. 1997; Baker et al. 2003; Wehrly et al. 2003). However, in general, northern Michigan has lower air temperatures than southern Michigan, and due to Michigan’s glacial history, northern Michigan tends to have higher soil permeability, greater groundwater flows, and thus colder water temperatures than southern Michigan (Bent 1971). This coincidence results in all of the models being able to capture the north-south distribution gradient when using air temperature, even though the driving factor is water temperature. In the southern part of the state, the models seem to struggle with presence predictions since the air temperature and water temperature are not as closely related as they are in the northern half of Michigan. As mentioned above, this problem becomes especially evident in the predictive map of the pruned neural network.

In order to more thoroughly check this north-south distribution problem, I examined the residuals of the MRI data. I divided the data into two parts by presence and absence and constructed linear regressions of the residuals of the MRI test data on the latitude of the observations (Figure 2.6). Regressions for all of the models showed that presence reaches in the northern Lower Peninsula tend to have lower residuals than presence stream reaches in the southern Lower Peninsula. This indicates that the models were able to more accurately predict presence reaches in the north than in the south. Absence reaches followed the opposite pattern: residuals in the southern Lower Peninsula tend to be smaller than residuals in the northern Lower Peninsula, which indicates the models can more accurately predict absences in the southern Lower Peninsula than the north. In future studies, this problem will be addressed as water temperature data are obtained and built into the models. I expect that the inclusion of water temperature will remove this inconsistency in the models between the north and the south, as the water temperature value does not rely fully on latitude.

## Error in Databases

The fish samples used in this study were obtained by several methods. Different sampling methods have different catch efficiencies depending on the fish species and stream in question, and at some sampling points the method may not have been optimal for sampling brook trout. Consequently brook trout may be found at sites where I marked them as absent. It is difficult to say that a fish does not reside in a particular location; it may be that I simply have not found them. However, this issue may be of greater consequence for a different fish; trout are susceptible to all the sampling methods and so were probably collected with little error. Patton et al. (2000) and Cao et al. (2005) found that electrofishing and seining yielded 97% comparability in species captured from Wyoming streams.

In addition, the data were collected over a period of 22 years. Over this time, sampling sites where fish were present may now be absent, and vice-versa, due to normal fish movement and changing habitat conditions. Land-cover data and stream line segment locations are likewise dated. Land-cover was based on air photos taken in 1978. However, I believe these data are generally representative of conditions in the past 25 years, which is what I examined. In using data of this type, I can avoid the naturally occurring year-to-year variation.

Another less obvious source of error in our evaluation is the unintended bias in our training data itself. Since it represents a random sample of the larger database, it is likely to under-represent the extremes of the larger distribution. The models are likely to favor methods that are best at predicting sites near the average (center of the distribution) and be biased against methods that provide a better fit to the tails of the distribution relative to the center.  This may be the reason why the classification tree was able to slightly out perform the neural network.

## Model Comparison

In this study, all four models did well at predicting the distribution of brook trout throughout the state of Michigan.  When applied to test data and when cross-validated with the training data, the order for accuracy was as follows: logistic regression, multiple regression, classification tree, and neural network.  The predictive map produced by the logistic regression model also was the closest to the Michigan Fish Atlas patterns. However, the relative differences in the model predictions were quite small.  Therefore, if I was to select the model that I deemed to be most useful, the choice might need to be based on criteria other than this predictive success.

As discussed above, multiple linear regression is not ideal for predicting dichotomous response variables (Zar 1999).  Statistical assumptions about the error terms are never met with dichotomous data.  However, logistic regression was developed for modeling with presence/absence data and has much less stringent assumptions than multiple regression.  Logistic regression is also quite familiar to most ecologists, is widely used in the literature, and is included in most statistical packages.

Classification trees and neural networks, on the other hand, are distribution-free, nonlinear modeling procedures, and therefore especially of interest to ecologists, who often encounter messy data and non-linear responses.  However, these methods are fairly new and unfamiliar to many researchers, and while the software is available, it will need to be purchased separately from a standard statistical package and may be expensive.

Given that a researcher has software for both of these modeling types on hand, classification trees will probably be preferred due to their easy setup and clear explanatory value.  It is very logical and easy to follow classification trees and understand how and why the trees make the classification decisions.  In contrast a neural network can be quite confusing to the beginner and requires many more steps and more time to get any type of explanatory value.  For these reasons, I prefer classification trees to neural networks in habitat analysis.

Table 2.1. Environmental variables included in the models. "Order" refers to the order in which the variables were entered into the regressions. Variables transformed for the regression models are marked.

| Order | Variable Name | Unit | Code | Transformation |
|---|---|---|---|---|
| | *Air Temperature Variable* | | | |
| 1 | Watershed July Mean Air Temperature | °C | W_JULY_MN | None |
| | *Channel Geometry/Position* | | | |
| 2 | Shreve Stream Order | None | CHAN_LINK | Log |
| 3 | Lake Immediately Downstream | Binary | DLAKE | Square Root |
| 4 | Distance Downstream to Great Lake | Meters | DOWNLENGTH | None |
| | *Flow / Hydrologic Variables* | | | |
| 5 | Channel Slope | % | CHAN_GRAD | Log |
| 6 | Channel Sinuosity | None | CHAN_SINU | None |
| 7 | Riparian Mean Darcy Value | None | R_DARCY | None |
| 8 | Riparian Mean Slope | % | R_SLOPE | Square Root |
| 9 | Riparian Mean Soil Permeability | Inches/hour | R_PERM | Square Root |
| 10 | Riparian Trace Mean Darcy Value | None | RT_DARCY | None |
| 11 | Riparian Trace Mean Slope | % | RT_SLOPE | Square Root |
| 12 | Riparian Trace Mean Soil Permeability | Inches/hour | RT_PERM | Square Root |
| 13 | Watershed Mean Darcy Value | None | W_DARCY | None |
| 14 | Watershed Mean Slope | % | W_SLOPE | Square Root |
| 15 | Watershed Mean Soil Permeability | Inches/hour | W_PERM | Square Root |
| 16 | Watershed Trace Mean Darcy Value | None | WT_DARCY | None |
| 17 | Watershed Trace Mean Slope | % | WT_SLOPE | Square Root |
| 18 | Watershed Trace Mean Soil Permeability | None | WT_PERM | Square Root |
| 19 | Mean Precipitation in Watershed | mm | W_PRECIP | None |
| | *Surficial Geology in Watershed Trace* | | | |
| 20 | Coarse Soil Texture in Watershed | % | COARSE | None |
| 21 | Fine Soil Texture in Watershed | % | FINE | None |
| 22 | Medium Soil Texture in Watershed | % | MEDIUM | None |
| | *Percent Landuse* | | | |
| 23 | Riparian Urban | % | R_URB_P | Log |
| 24 | Riparian Agriculture | % | R_AGR_P | ArcSine |
| 25 | Riparian Wetland | % | R_WET_P | ArcSine |
| 26 | Riparian Forest | % | R_FOR_P | ArcSine |
| 27 | Riparian Open/ Fields | % | R_OPEN_P | Log |
| 28 | Riparian Open Water | % | R_WAT_P | Log |
| 29 | Riparian Trace Urban | % | RT_URB_P | Log |
| 30 | Riparian Trace Agriculture | % | RT_AGR_P | ArcSine |
| 31 | Riparian Trace Wetland | % | RT_WET_P | ArcSine |
| 32 | Riparian Trace Forest | % | RT_FOR_P | ArcSine |

Table 2.1, continued.

| | | | | |
|---|---|---|---|---|
| 33 | Riparian Trace Open/ Fields | % | RT_OPEN_P | Log |
| 34 | Riparian Trace  Open Water | % | RT_WAT_P | Log |
| 35 | Watershed Urban | % | W_URB_P | Log |
| 36 | Watershed Agriculture | % | W_AGR_P | ArcSine |
| 37 | Watershed Wetland | % | W_WET_P | ArcSine |
| 38 | Watershed Forest | % | W_FOR_P | ArcSine |
| 39 | Watershed Open / Fields | % | W_OPEN_P | Log |
| 40 | Watershed Open Water | % | W_WAT_P | Log |
| 41 | Watershed Trace Urban | % | WT_URB_P | Log |
| 42 | Watershed Trace Agriculture | % | WT_AGR_P | ArcSine |
| 43 | Watershed Trace Wetland | % | WT_WET_P | ArcSine |
| 44 | Watershed Trace Forest | % | WT_FOR_P | ArcSine |
| 45 | Watershed Trace Open/ Fields | % | WT_OPEN_P | Log |
| 46 | Watershed Trace Open Water | % | WT_WAT_P | Log |

Table 2.2. Variables significant in the multiple linear regression, their β values (regression coefficients), standard error of the β, and significance. Variables were entered into the model in a stepwise procedure and were included in the final model if found significant ($p < 0.10$).

| Variable | β | Standard Error | Significance |
|---|---|---|---|
| (Intercept) | 0.5 | | |
| W_JULY_MEAN | -0.184 | 0.023 | <0.001 |
| LINK | -0.133 | 0.016 | <0.001 |
| W_FOR | 0.091 | 0.022 | <0.001 |
| R_PERM | 0.061 | 0.017 | <0.001 |
| WT_WET | -0.058 | 0.018 | 0.001 |
| R_DARCY | 0.055 | 0.016 | <0.001 |
| RT_WAT | -0.042 | 0.016 | 0.01 |
| WT_OPEN | 0.036 | 0.017 | 0.03 |

Table 2.3. Percentage of correct predictions for the MRI test data and the training data upon cross-validation. "Performance" is the sum of the correct predictions for a model and gives a measure of the model's relative predictive ability. (MLR- multiple linear regression; LR- logistic regression; CART - classification tree; NN(46)- neural network with 46 habitat variables; NN(7)- pruned neural network with 7 habitat variables).

| | MRI | | X-Validation | | Performance |
|---|---|---|---|---|---|
| | Presence | Absence | Presence | Absence | |
| MLR | 85.7 | 75.7 | 80.6 | 79.6 | 321.6 |
| LR | 87.1 | 75.9 | 79.9 | 81.2 | 324.1 |
| CART | 84.3 | 77.7 | 75.2 | 78.3 | 315.5 |
| NN (46) | 85.7 | 71.8 | 75.4 | 77.7 | 310.6 |
| NN (7) | 81.3 | 77.4 | 68.8 | 84.4 | 311.9 |

Table 2.4. Variables significant in the logistic regression, their β values, standard error of the β values, and significance. Variables were entered into the model in a stepwise procedure and were included in the final model if found significant ($p < 0.10$).

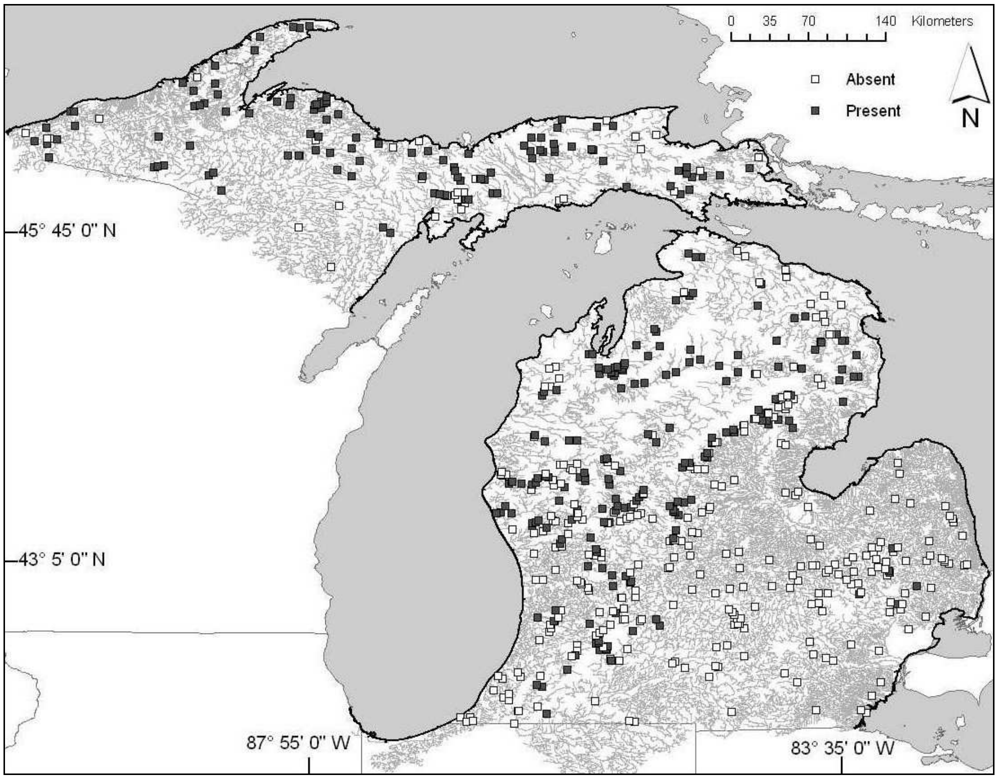
| Variable | β | Std. Error | Significance |
|---|---|---|---|
| (Intercept) | -0.0726 | | |
| W_JULY_MEAN | -1.3614 | 0.186 | <0.001 |
| CHAN_LINK | -1.2475 | 0.153 | <0.001 |
| W_FOR | 0.4737 | 0.179 | 0.008 |
| WT_WET | -0.4224 | 0.14 | 0.002 |
| R_PERM | 0.4139 | 0.127 | 0.001 |
| W_SLOPE | 0.3823 | 0.153 | 0.013 |
| R_DARCY | 0.3615 | 0.133 | 0.007 |
| DOWNLENGTH | 0.3486 | 0.131 | 0.008 |
| WT_OPEN | 0.2317 | 0.133 | 0.082 |

Table 2.5. Variables determined to be significant after a neural network randomization procedure, their relationship to brook trout presence, and associated p-values. These variables were then used to construct a pruned neural network.

| Variable | Relationship | P-value |
|---|---|---|
| WT_OPEN | + | 0.016 |
| W_JULY_MN | - | 0.017 |
| CHAN_GRAD | + | 0.021 |
| RT_WET | + | 0.044 |
| R_DARCY | + | 0.053 |
| W_OPEN | - | 0.057 |
| WT_WET | - | 0.087 |

Figure 2.1. Michigan streams reaches containing habitat variables overlaid by A) MDNR sampling points for brook trout presence and absence (training data) and B) MRI sampling points for brook trout presence and absence (testing data).

A)



B)

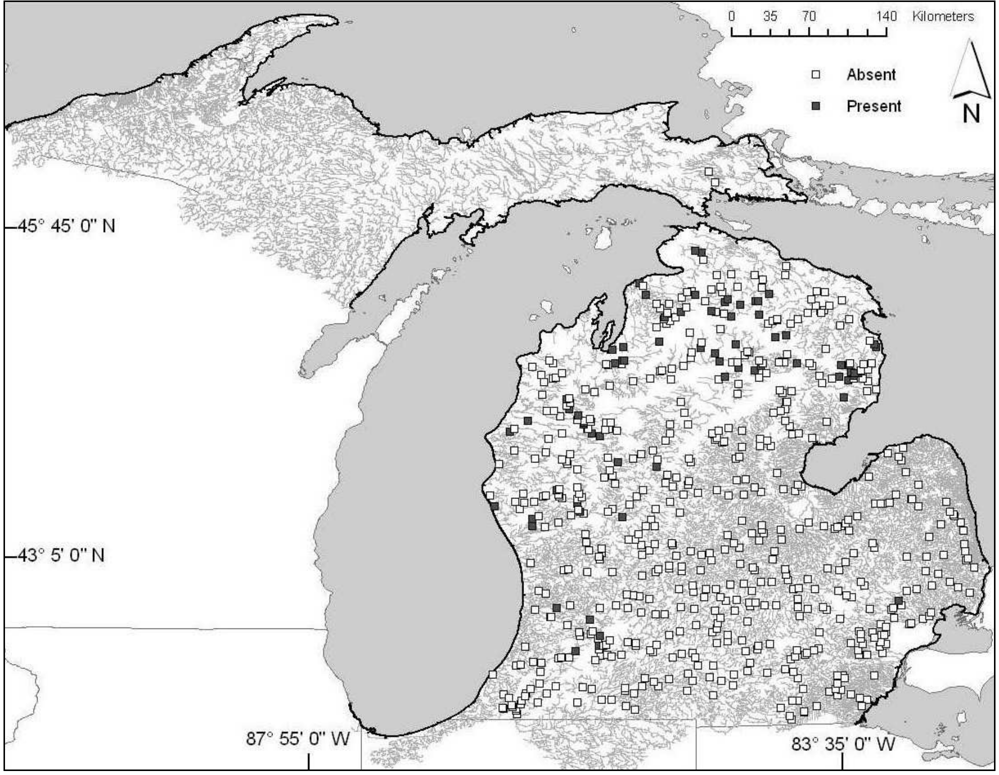


Figure 2.2. Variables are measured on four scales for each individual stream reach: A) Riparian, B) Network, C) Sub-watershed, D) Watershed.  See text for details.

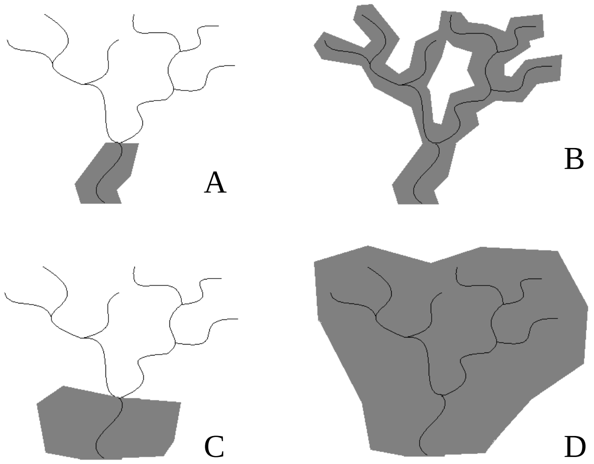
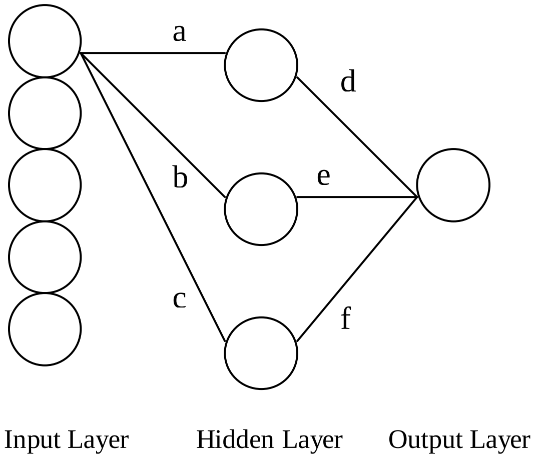


Figure 2.3. In order to prune the neural network, the sum of the input layer-hidden layer weights and the hidden layer-output layer weights is calculated for each input variable. In this example, I calculate this sum for one variable in a neural network with 3 hidden neurons in the hidden layer.



Input Layer-Hidden Layer Weights:
a, b, and c

Hidden Layer-Output Layer Weights:
d, e, and f

Sum of weights for variable of interest:
$a + b + c + d + e + f$

Figure 2.4. Classification tree created by CART that had highest correct percentage of predictions for the test data. An observation is tested one at a time, starting with the top of the tree and working down, following the splitting rules until it is classified as present or absent.

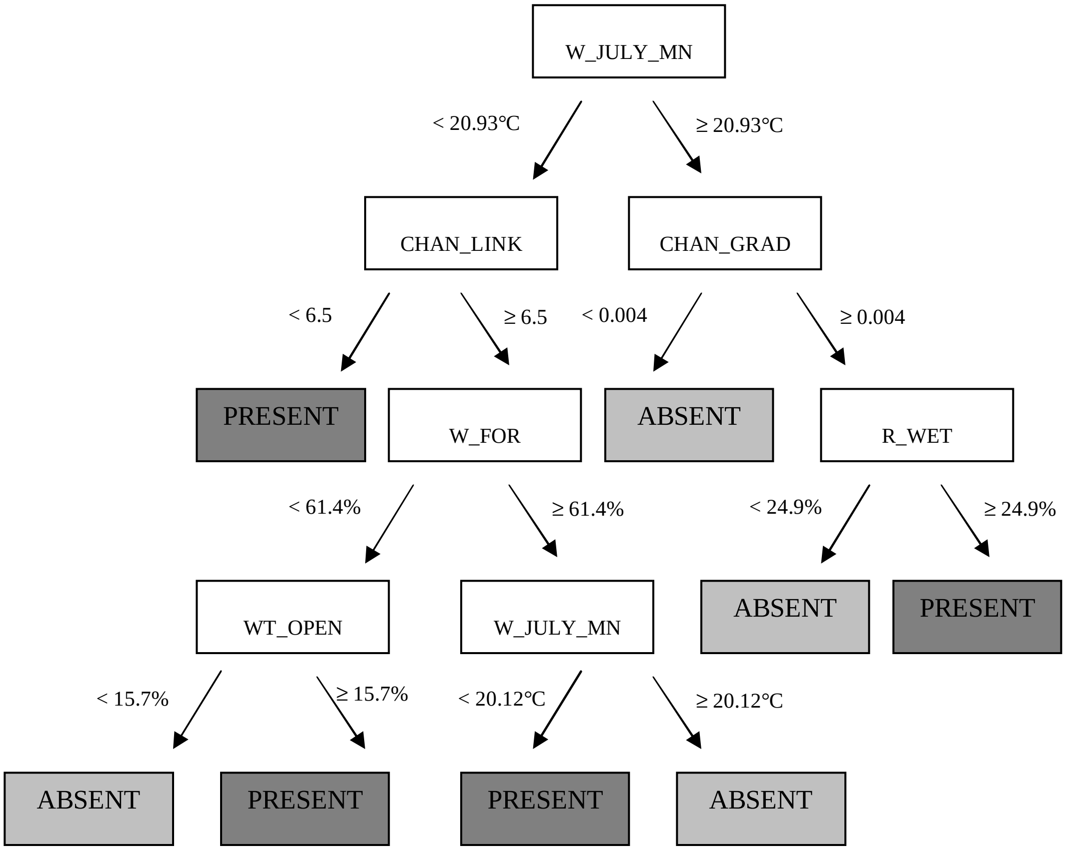


Figure 2.5. A) Brook trout samples in the Michigan Fish Atlas, and B-F) predicted brook trout distributions in Michigan using B) multiple linear regression, C) logistic regression, D) neural network with all variables, E) neural network with seven variables, F) classification tree. A black stream indicates predicted presence, and a light gray stream indicates predicted absence.

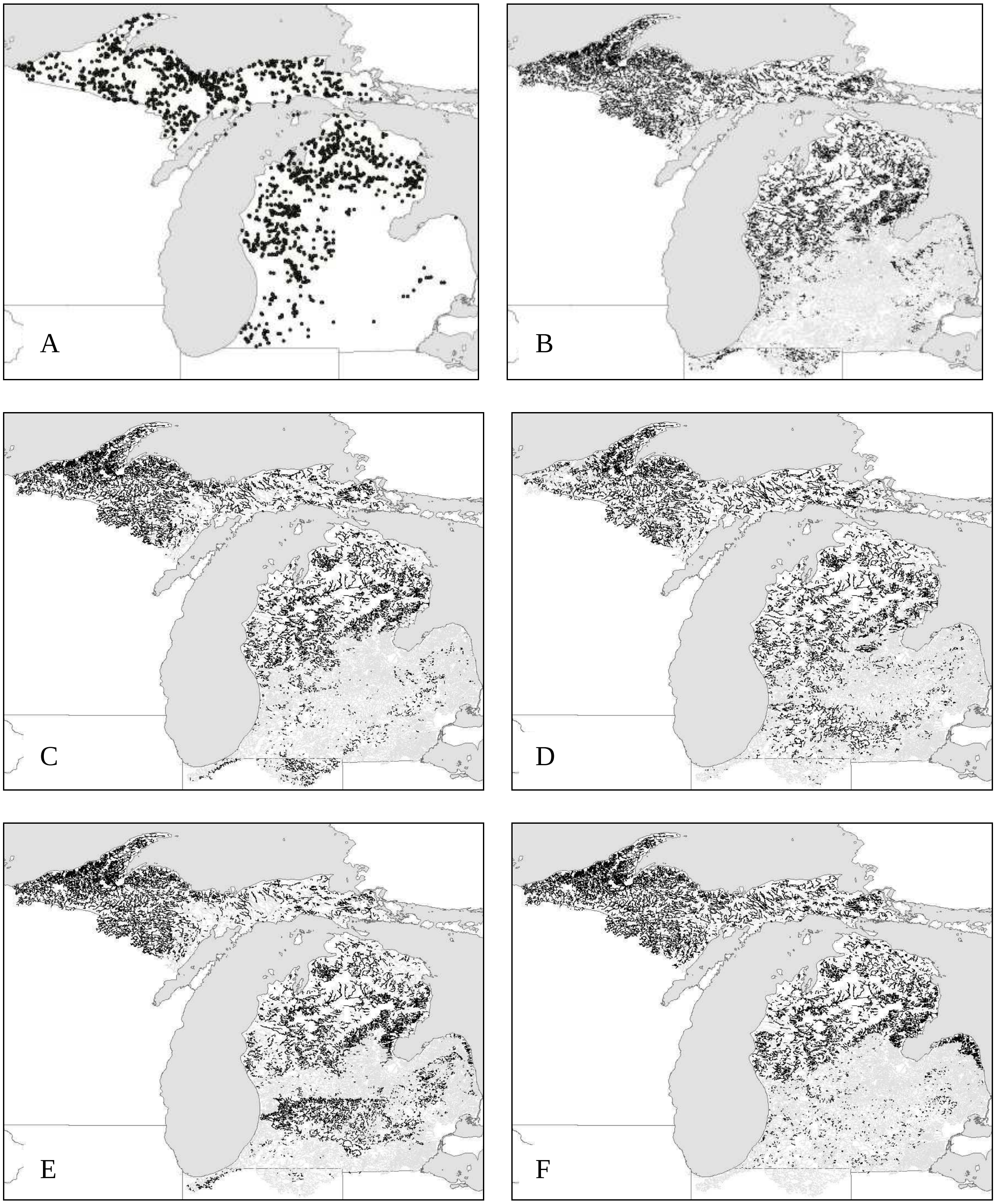
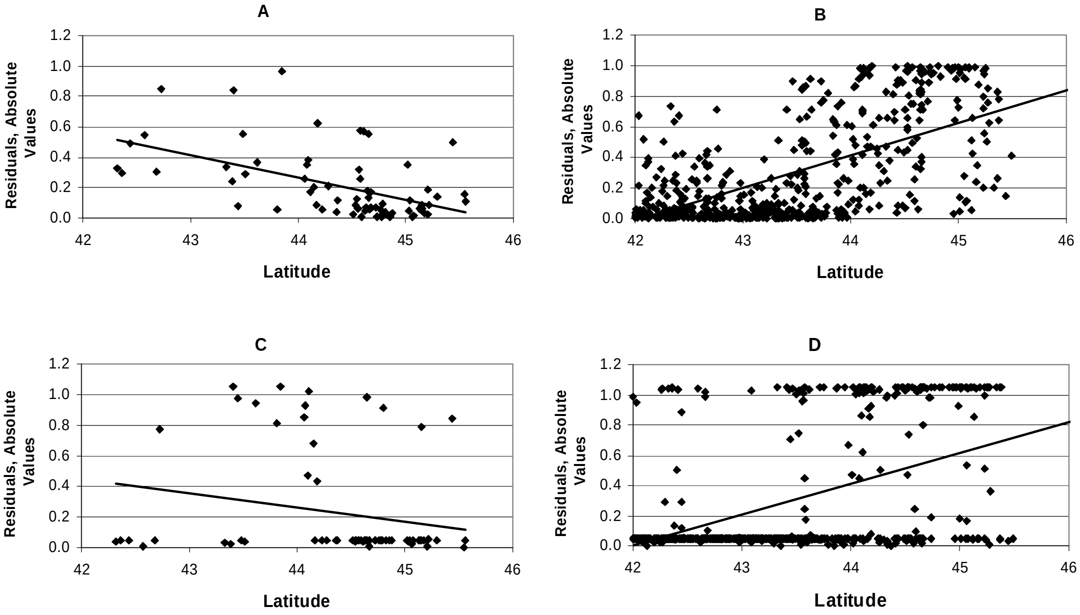


Figure 2.6. Scatterplots of the absolute values of residuals for the MRI test data regressed against the latitude of the sampling point for the A) Presence samples in the logistic regression model ($R^2$= 0.27, F= 25.4, p-value <0.000); B) Absence samples in the logistic regression model ($R^2$= 0.41 , F= 402.4, p-value <0.000 ); C) Presence samples in the full neural network model ($R^2$= 0.04 , F= 3.1, p-value = 0.084); D) Absence samples in the full neural network model ($R^2$= 0.25 , F= 187.4 , p-value <0.000). Graphs for the other models are not shown; all of the models displayed the same patterns demonstrated by the graphs shown here.



# References

Baker, M. E., M. J. Wiley, and P. W. Seelbach. 2003. A GIS model of subsurface water potential for aquatic resource inventory, assessment, and environmental management. Journal of Environmental Management 32:706-719.

Bell, J. F. 1999. Tree-based methods. Pages 89-105 *in* A. H. Fielding, editor. Machine learning methods for ecological applications. Kluwer Academic Publishers, Boston, Massachusetts.

Bent, P. C. 1971. Influence of surface glacial deposits on streamflow characteristics. U.S. Geological Survey, Water Resources Division, Open-file Report, Lansing, Michigan.

Boddy, L., and C. W. Morris. 1999. Artificial neural networks for pattern recognition. Pages 38-87 *in* A. H. Fielding, editor. Machine Learning Methods for Ecological Applications. Kluwer Academic Publishers, Boston, Massachusetts.

Breiman, L. F. J., R. Olshen, and C. Stone. 1984. Classification and regression trees. Chapman and Hall, New York.

Cao, Y., C.P. Hawkins, and A. W. Storey. 2005. A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis. Freshwater Biology 50: 1105-1115.

De'ath, G. 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. Ecology 83:1105-1117.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

ESRI. 2002. ArcGIS 8.3. 380 New York Street, Redlands, California.

Faraway, J. J. 2005. Linear Models with R. Chapman and Hall/CRC, New York.

Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream fish from habitat variables: 1950-1985. General Technical Report PNW-GTR-213, United States Department of Agriculture, Forest Services, Pacific Northwest Research Station, Portland, Oregon.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9:733-748.

Lek, S., M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga, and S. Aulagnier. 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling 90:39-52.

Lek, S., and J. F. Guegan. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling 120:65-73.

Mastrorillo, S., S. Lek, F. Dauba, and A. Belaud. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshwater Biology 38:237-246.

Merna, J. W., editor. 1988. Manual of fisheries survey methods. Michigan Department of Natural Resources, Fisheries Management Report 9, Ann Arbor.

MGDL (Michigan Geographic Data Library), 2007. Available: http://www.michigan.gov/cgi/ (December 2007).

Morrison, S. S., D. R. Passino-Reader, D. N. Myers, J. E. McKenna, J. Stewart, B. C. Scudder, K. Lohman, and S. Aichele. 2003. Great Lakes regional aquatic gap analysis- Preserving biodiversity in the Great Lakes basin. USGS GLSC Fact Sheet 2003-1. U.S. Geological Survey, Ann Arbor, Michigan.

Myers, D. N., J. E. McKenna, D. R. Passino-Reader, and J. Stewart. 2002. Great Lakes Aquatic GAP Project. USGS National GAP Analysis Program Bulletin 11:33-37.

NHD (National Hydrography Dataset), 2007. Available: http://nhd.usgs.gov/ (December 2007).

Olden, J. D., and D. A. Jackson. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130:878-897.

Olden, J. D., and D. A. Jackson. 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling 154:135-2002.

Patton, T. M., W. A. Hubert, F .J. Rahel, and K. G. Gerov. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. North American Journal of Fisheries Management 20: 394-398.

Rosenfeld, J. 2003. Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches. Transactions of the American Fisheries Society 132:953-968.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning representations by back-propagation errors. Nature (London) 323:533-536.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D. Erchia, T. C. Edwards, J. Ulliman, and G. Wright. 1993. Gap analysis: A geographic approach to protection of biodiversity. Wildlife Monographs 123:1-41.

Seelbach, P. W., and M. J. Wiley. 1997. Overview of the Michigan Rivers Inventory (MRI) project. Fisheries Technical Report 97-3, Michigan Department of Natural Resources, Ann Arbor.

Shih, Y. 1995. Neuralyst User Guide. Cheshire Engineering Corporation, Pasadena, California.

Smith, C. L. 1985. The Inland Fishes of New York State. The New York State Department of Environmental Conservation, Albany.

Sowa, S. P. 1999. Establishing initial conservation priorities., Report of the MoRAP, U.S. Geological Survey, Columbia, Missouri.

Sowa, S. P., and C. F. Rabeni. 1995. Regional evaluation of the relation of habitat to distribution and abundance of smallmouth bass and largemouth bass in Missouri streams. Transactions of the American Fisheries Society 124:240-251.

Steinberg, D., and P. Colla. 1997. CART-- Classification and Regression Trees. Salford Systems, San Diego, California.

SPSS. 2003. SPSS for Windows v12.0. Chicago, Illinois.

Vayssieres, M., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science 11:679-694.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 1997. Landscape-based models that predict July thermal characteristics of lower Michigan rivers. Fisheries Research Report 2037, Michigan Department of Natural Resources, Ann Arbor.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2003. Classifying regional variation in thermal regime based on stream fish community patterns. Transactions of the American Fisheries Society 132:18-37.

Wiley, M. J., S. L. Kohler, and P. W. Seelbach. 1997. Reconciling landscape and local v iews of aquatic communities: lessons from Michigan trout streams. Freshwater Biology 37:133-148.

Zar, J. H. 1999. Biostatistical Analysis, 4th edition. Prentice Hall, Upper Saddle River, New Jersey.

Zorn, T. G. 2003. Fishes of Lower Michigan: distribution patterns, abundance models, and causal relationships. Doctoral Dissertation. University of Michigan, Ann Arbor.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's Lower Peninsula. Transactions of the American Fisheries Society 131: 70-85

# Chapter 3

# Classification tree models for predicting distributions of Michigan stream fish from landscape variables

## Abstract

Traditionally, fish habitat requirements have been described from correlations between occurrence and site-scale environmental variables. However, recent studies have shown that studying landscape-scale processes improves our understanding of what drives species assemblages and distribution patterns across the landscape. In this study, my goal was to learn more about the constraints on the distribution of Michigan stream fish using landscape-scale habitat variables. I used classification trees and landscape-scale habitat variables to create and validate presence/absence models and relative abundance models for Michigan stream fishes. I developed 93 presence/absence models that were on average 72% correct when compared to independent data, and I developed 46 relative abundance models that were on average 76% correct when compared to independent data. The models were used to create statewide predictive distribution and abundance maps that can be used for a variety of conservation and scientific purposes.

## Introduction

Environmental complexity and species interactions make it difficult to learn the exact abiotic habitat constraints on a population. Researchers often use statistical models for this by searching for patterns between species occurrences or abundances and the environmental characteristics of sampled locations. These models serve two important purposes: they are used to formulate and test hypotheses about the factors and processes that are important to organisms, and they are sometimes used to make predictions of species distributions and abundances for use in management and conservation decisions.

Traditionally, fish habitat requirements have been described from site or local-scale environmental variables (Fausch et al. 1988). Habitat variables measured at this scale are useful to managers because small-scale habitat can be manipulated (Fausch et al. 1988; Vaughan and Ormerod 2003). Local-scale variables such as cover or substrate are measured on short river reaches and affect food, refuge habitat, spawning habitat, and ultimately fish abundance. Three well known modeling approaches, the U.S. Habitat Suitability Index (HSI), the River Invertebrate Prediction and Classification System (RIVPACS), and Australian Rivers Assessment Scheme, are based on local-scale environmental variables (Seelbach et al. 2002a). There are problems with modeling on a site-scale level; it is expensive, or in some cases impossible, to measure site attributes everywhere within a study region (Seelbach et al. 2002a). Beyond this practical concern, an important ecological tenet states that “different processes are likely to be important on different scales”(Levin 1992); researchers may be completely unaware of important large-scale processes that impact fish if they only use site-scale habitat data (Wiley et al. 1997; Fausch et al. 2002; Allan 2004).

In the past fifteen years, the advent of powerful geographic information system (GIS) tools has made it possible to study spatial variation in fish distributions and abundance from a larger, landscape perspective and to incorporate habitat attributes measured at larger spatial scales. GIS-based modeling uses a variety of large-scale map-based variables (e.g., geology and climate), which influence an aquatic system's hydrological and thermal characteristics (Wiley et al. 1997). Modeling at this scale often uses land-use patterns as well, because they influence amounts and rates at which sediment, pollutants, and water are delivered to the system (Schlosser 1991).

Fish species are clearly influenced by processes that operate on larger spatial scales and slower temporal scale than those measured at the local-scale (Richards et al. 1996; Leftwich et al. 1997; Rathert 1999; Allan 2004). While fish are responding mechanistically to what is happening in their immediate surroundings, those local-scale factors are directly caused by the larger landscape. For example, while stream temperature is measured at a specific location, it is controlled by a combination of local and landscape-scale processes (Wehrly et al. 2003; Wehrly et al. 2006). Also, the hydrologic flow regime of a stream is crucial to fish communities and is driven by factors operating at a catchment scale (Poff et al. 1997).

Models based on landscape-scale processes are becoming more common. Wiley et al. (1997) produced trout population density models using only landscape-scale variables, while Zorn et al. (1998, 2003) used catchment area and low-flow yield as key variables in predicting fish assemblages in Michigan. Zorn et al. (2003, 2004) also used landscape-scale variables with multiple linear regression to predict fish assemblages. Close associations have also been recognized between fish assemblages and hydrologic variability, watershed size, gradient, and percent forest cover (Poff and Allan 1995; Maret et al. 1997).

In addition to providing understanding into processes that drive the fish distributions, there are many other reasons to develop models that study the relationship of landscape-scale environmental variation and fish populations. Such models provide insight to how aquatic ecological systems function, predict potential population sites, and identify areas for population restoration (Fausch et al 1988; Maret et al. 1997; Wiley et al. 1997; Olden 2001; Olden and Jackson 2002). This is especially important for Michigan stream fish communities. Michigan possesses a diverse array of streams ranging from nationally renowned trout fisheries to diverse warm- and cool-water communities that support recreational angling for a variety of game species. In addition, maintaining the diversity of non-game stream fishes is an important conservation goal. Both fisheries managers and non-game biologists need further understanding of the processes that regulate stream fish communities within the state; however, broad scale knowledge of Michigan stream communities has been hindered because although historical fish data are plentiful, a relatively small percentage of stream reaches have been sampled.

In this study, my goal was to learn more about large-scale factors that influence the distribution of Michigan stream fish. To do this, I used landscape-scale habitat variables and three sources of data on Michigan fish distributions to create and validate models that predicted presence/absence (PA) and relative abundance (RA) of Michigan fishes.

Specific objectives were as follows:

1. To build classification tree fish models for Michigan stream fish.
2. To assess each model for validity using an independent dataset.

3. To describe the general structure and behavior of the models.
4. To understand patterns in model error and to understand model limitations.
5. To use the models to describe relationships between fish communities and landscape-scale habitat variables.

## Methods

### Data Description- Habitat Variables

Data for predictor variables used in this study were obtained through the combined efforts of the Great Lakes Aquatic GAP Project (GLGAP; GLSC 2007) and the Classification and Impairment Assessment of Upper Midwest Rivers (CIAUMR; Brenden et al. 2006; UM 2007). These groups have established a high-resolution, GIS-linked database containing characteristics of Michigan’s rivers. The database was referenced to a group of ArcGIS line coverages (ESRI 2002), in which each river was divided into inter-confluence reaches. Line coverages were based on the USGS National Hydrography Dataset (NHD, 2007) at the 1:100,000 scale, but were updated to provide more accurate representation of Michigan rivers (Brenden et al. 2006). There are 31,817 Michigan stream reaches (86,983 kilometers of stream length) included in the database, and the database contained information on wide variety of landscape-scale environmental variables for each stream reach, including soil permeability, land cover, stream position, bedrock and surficial geology, modeled water temperature, climate data, modeled exceedence flows, and modeled phosphorus (Brenden et al. 2006).

The database contained approximately 320 variables for each stream reach; I chose to combine some and remove others to end up with a list of 23 variables that I hypothesized to have the most direct mechanistic relationships to fish distributions (Table 3.1). Reducing the number of predictors was essential to reduce collinearity between model variables, improve model interpretability, and reduce probability of spurious correlations. Not all correlated variables were removed; for example, it was important to leave in the different types of land-use and land-cover as these variables are important for managers as examples of landscape-scale variables that can be manipulated. Choosing these variables was a key step in the modeling process, and the decision was based on past work on Michigan fish (Zorn 2003) as well as preliminary classification trees in which I included all possible variables. The variables that I retained and their importance to fish are discussed in the next several paragraphs.

Water temperature has important effects on growth and survival of fish and affects dissolved oxygen levels (Smale and Rabeni 1995; Wehrly et al. 2003; Bailey and Alanara 2006; Rand et al. 2006; Wehrly et al. 2006). Since water temperature data were not available for every stream reach, a temperature model was developed to make predictions of mean July stream temperature (Li Wang, Michigan Department of Natural Resources, personal communication). In addition to water temperature, I also used mean annual air temperature, which is a reasonable approximation of ground water temperature and thus water temperature during base-flow conditions.

Of the different types of land-use data available, I used percent of forest, wetlands, agriculture, and urban on two scales: a 60 meter (30 meters to each side of the stream) riparian network stream buffer for the stream reach of interest and all streams upstream, and the total catchment area ($km^2$) of the stream reach. The riparian area of a stream is an important indicator of erosion control, pollution filtering capacity, shading, and woody debris potential, while land-use of the entire catchment area of a stream has important effects on water chemistry and stream hydrology (Wang et al. 1997; Synder et al. 2003; Wang et al. 2003).

Surficial geology has impacts on water chemistry and hydrology (Bent 1971). I obtained surficial geology data from 1:250,000 scale maps. I calculated the sum of the coarse-textured geological areas (outwash, coarse textured end moraine and till, lacustrine sand and gravel, dune sand) for the watershed of each stream reach and divided by the watershed area to produce the percent of coarse surficial geology in the watershed. This was also done with fine-textured surficial geology (fine textured till, fine-texture end-moraine, and lacustrine clay and silt).

Several habitat variables were built from GIS-obtained information to serve as surrogates for site-scale habitat features that are important in shaping fish communities (Table 3.1). Ninety percent exceedence flow yield (exceedence flow/catchment area) served as a replacement for velocity at baseflow and indicates the relative contribution of groundwater, while specific stream power at 90% exceedence flow (10*90% ex. flow * gradient / catchment area) can indicate a stream's substrate, with a high power stream able to scour fine sediment from the channel bed. Ten percent exceedence flow is a measure of a stream's peak flow that can limit recruitment and abundance of the population, and specific stream power at 10% exceedence flow is a measure of the stream's maximum erosive force and sediment transport capability. All flow estimates were standardized as "yields" by dividing values by catchment area.

Phosphorus is an essential nutrient that can limit productivity in aquatic systems (Vanni 1987; Vanni et al. 1997; Zorn et al. 2003). Since total phosphorus measurements were not available for every Michigan stream reach, I predicted it using a multiple regression equation based off of 1985-1992 Michigan Rivers Inventory (MRI) phosphorus measurements and the other variables in Table 3.1 [ln(Total Phosphorus) = -6.996 + (% Agriculture in watershed* 1.497) + (ln (Stream power at 90% exceedence flow)* -0.222) + (10% exceedence flow yield* 59.977)], n = 172, p <.001, adjusted $R^2$ = 0.54) (Seelbach and Wiley 1997).

There were several measured connectivity variables that take advantage of the stream connection properties inherent to the NHD (Brenden et al. 2006). Variables built from these analyses include distance from the stream to the receiving Great Lake and distance from the stream to upstream and downstream lakes and ponds. Streams reaches disconnected from the Great Lakes by dams or waterfalls were noted. It is expected that these variables will be important to lake fish species that migrate into streams for parts of their life cycle (e.g., Chinook salmon), or fish that live in both lakes and rivers (e.g., most centrarchids). Also, the variable LINKDCATCH was created to measure the distance from the stream reach of interest to the closest downstream stream reach that has a 10% greater catchment area than that of the stream of interest (Osborne and Wiley 1992). This distance might prove useful for explaining occurrences of large river fish in small tributaries, or small stream fish in nearby larger rivers.

### Data Description- Fish Data

I used three fish databases to create and validate the models. The Michigan Rivers Inventory (MRI) dataset contains quantitative fish samples obtained through electroshocking and rotenone sampling. The samples in this dataset were obtained during the 1980s and 1990s and cover the geographic extent of Michigan, but do have a bias towards small to medium streams that can be waded (Seelbach and Wiley 1997). I compiled fish counts from the years 1980-2002 from the Fish Collection System (FCS) of the MDNR Fisheries Division. These records were collected with a wide variety of catch techniques, including electroshocking, rotenone, and seining. Given the poor catch efficiency of seining methods, for sites that were seined I only recorded the presence of fish caught and did not consider missing fish as "absent". I also used the Michigan Fish Atlas, created by the University of Michigan's Museum of Zoology (Bailey et al. 2000). This database has occurrence records of Michigan fish going back to the mid-19th century. However, for this study I only used data from collections made during 1980-2000, in order to match the time frame of the MRI and FCS data. These records were also collected with a wide variety of catch techniques and provide good spatial coverage of the state.

For all three datasets, I deleted replicate samples so that a stream reach was represented by only one sampling effort. When different samples for the same reach disagreed on a species presence or abundance, I kept the observation where the fish was present or in higher abundance. This assumed that the stream reach has the potential to hold the higher amount of fish, and the lower fish count was a result of disturbance unrelated to the habitat factors.

### Classification trees

Classification trees are created through a data partitioning technique; a value of a variable is used to split the data into two subsets that are as pure as possible for the response (Breiman et al. 1984, Bell 1999). Each subset is then split repeatedly until all data within the subset are classified into a single class, or until a pre-determined stopping point is reached. Each of these subsets is called a node, and the final subsets (those which are not split) are called terminal nodes. Subsets are connected through variables and splitting rules in such a way as to create an inverted tree diagram, which can be used graphically to display the model's decision rules (Figure 3.1). To make a prediction, an unclassified observation is dropped into the top of the tree and follows the splitting rules until it reaches a terminal node. The predicted value for the observation is the value at its terminal node. Bell (1999) gives a thorough and understandable explanation of classification trees.

Classification tree is an empirical modeling technique that can deal with strongly non-linear, high order relationships, missing values, different data types (continuous, ratings, categorical); it can predict as well or better than traditional approaches and the graphical output is easy to interpret (Breiman et al. 1984, Bell 1999, Olden and Jackson 2002). For these reasons, the use of trees in ecological studies has increased dramatically in the past five years (De'ath and Fabricius 2000; Vayssieres et al. 2000; De'ath 2002; Taverna et al. 2005; Holland et. 2005; Baker et al. 2006; Steen et al. 2006; Usio et al. 2006).

In a previous study, I modeled brook trout with several different analytical techniques and determined that a classification tree method was successful in modeling with landscape-scale data (Steen et al. 2006, Chapter 2). In this study, I decided to use classification trees to develop the models for all common species of Michigan stream fish.

### Presence/Absence modeling procedure

I created a species-specific PA classification tree model for each of the 93 fish species that had more than 30 occurrences in the training dataset (Table 3.2). I used the MRI dataset as training data and the FCS dataset as testing data. I selected the MRI set as the training dataset because it had higher sample sizes for most of the non-game fishes than did the FCS dataset. For 11 species, either the number of occurrences in the FCS data was low (less than 3 occurrences) or the identifications of the fish were suspect. For these species, I withheld 20% of the MRI data from training to serve as a test dataset (Table 3.3). I used the Fish Atlas data as a supplemental training database; if the MRI data did not contain at least thirty species occurrences, I added Fish Atlas data to the MRI data for model training purposes.

The training data for a species, having been pruned down through the procedures above, were entered into CART 5.0 (Steinberg and Colla 1997). This program produced a series of differently sized classification trees, each with different misclassification rate for both the training data and an independent data set created from a cross-validation of the training data. Next, I selected the tree that minimized error in both the training data and cross-validation. If a tree was greater than 7 terminal nodes but had a lower error rate than a smaller tree, I selected the smaller tree despite it having a higher error rate. I felt that as trees started growing past 7 terminal nodes, the interpretation of the tree grew difficult and would start to contain more spurious variable splits. This decision represents the desire to have trees that are accurate, yet easy to interpret. Certainly this is not an objective decision and reflects my judgment and preference.

Using this tree as a starting point, I determined if the variable splits in the tree could possess ecological meaning. Splits that lacked ecological meaning were those in which the tree created a split at an unreasonable value; for example, the most common spurious split was a percent land-use split of less than 1 percent. Since it was unlikely that these values had any significance to the fish, I removed these variables from the analysis and recreated the tree in order to develop a better model. If there were no spurious variable splits, I accepted the tree as the final PA model.

The FCS test dataset was applied to the final model to get a benchmark of the model's accuracy by predicting the percentage of observations predicted correctly. In addition, I calculated the true skill statistic (TSS) for the FCS data. TSS and its predecessor, Cohen's kappa, are relatively new ways to measure the accuracy of presence/absence models, and address the problem reported by Fielding and Bell (1997) of inflated accuracy ratings for rare species. TSS is a presence/absence assessment score that accounts for errors and success as a result of random guessing, and ranges from -1 to +1, where +1 indicate perfect prediction and values of -1 to 0 indicate a model that is worse than random (Allouche 2006). However, the majority of the discussion of this paper relies on the percentage accuracy rating rather than TSS, as percent accuracy is more intuitive than is TSS and creates results more interesting and easier to understand. In addition, the results indicated that TSS consistently underestimated the value of models for which there was a large discrepancy between number of present and absent observations.

### Presence/Absence model error

I identified sites from the FCS testing dataset that had misclassified fish predictions; in other words, sites where predicted presence/absence did not match the observation. These types of errors are usually described with the terms false positive (predicted present when observed absent) and false negative (predicted absent when observed present). For example, when a FCS sampling site has 10 false positive errors, this means that 10 fish species were predicted to be present in the stream, but were not found.

I examined the correlation matrix of the numbers of false positive and false negative errors made at a site and the habitat values for the stream reach where the sampling site was located. This was done in order to determine whether there are any patterns between model error and the habitat variables; such patterns can indicate if streams with particular habitat tend to have more or less accurate models. To prevent the models that performed poorly from interfering with these results, I only looked at PA models with a TSS greater than zero and at least 60% accuracy (in both absence and presence) when compared against the test dataset.

### Relative Abundance model procedure

For the RA models, I selected MRI data obtained from two-pass electroshocking depletion samples and converted the fish counts to estimated catch per hectare. The FCS dataset and Fish Atlas dataset were not used in RA modeling.

I built the RA models on an individual species basis. For each species with greater than or equal to 30 occurrences in the MRI data, I divided fish density estimates into three logarithmic-scale categories (low: 1-10; medium: 11-100; high: > 100 fish per hectare). I also tried dividing density estimates into categories by equal interval and by natural breaks. However, the models performed the same or worse using these category breaks, so I decided to use the logarithmic-scale out of simplicity; each fish species had the same abundance categories when using the logarithmic scale.

To build the trees, I followed the same steps used in the PA models, except that I used three density categories instead of presence/absence categories. Since the only density data available were from the MRI dataset I withheld 20% of the MRI sample for model validation. Several fish had greater than 30 samples, but too few fish in a category to use a 20% hold-out sample for a test dataset (e.g., 2 observations in the low category, 4 in medium, and 30 in high). In these cases, I used the 10-fold cross-validation procedure given by Steinberg and Colla (1997) to assess the model performance. In the cross-validation process, one tenth of the data is held back while the rest is used to create the tree, and error estimates are made for the withheld data. This is repeated until all the data has been withheld and tested, and the final testing accuracy is determined from the combination of all of the mini-test samples.

If a relative abundance model had an accuracy rating worse than guessing when compared to the test data or cross-validation (<33.3% for any category), I created a two-category classification tree for that species. For these models, I dropped the middle category so that the species was only predicted at a low and high relative abundance. This also involved dropping the training data that had been in the middle category (11-100 fish per hectare) and making the assumption that in the real world, no fish fall within this range. This resulted in models that were simpler and more removed from reality than the

3-category models, but I think this was necessary in order to build RA models with good accuracy levels for these species.

### Model analysis and predictions

For both model sets, I counted the number of times each variable occurred to indicate the most important variables for all of the fish. Then I more closely examined how the top five variables split in the trees to determine if there were any overall patterns caused by these variables. To prevent the models that performed poorly from interfering with these results, I only looked at PA models or 2-category RA models that had at least 60% accuracy (in any category: absence, presence, low, or high), when compared against the test dataset. For the PA models, I also required the model to have a TSS greater than zero in order to include the model in the analysis.

For every species, I applied the PA model to every stream reach in Michigan. For species with an abundance model, I applied the RA model to every stream that was predicted as present, and combined the two models to produce predictions with three or four categories: fish absence, low relative abundance, medium relative abundance (where available), and high relative abundance. The predictions were joined to the updated 1:100,000 NHD in a GIS to produce statewide distribution maps for each fish.

## Results

### Presence/Absence models

I developed PA models for 93 Michigan stream fish (Table 3.2). Despite the addition of the Fish Atlas data, I did not have enough data (less than 30 occurrences) to create PA models for 52 of the 145 fish species found in Michigan (Bailey and Smith 2002). However, while 18 of these fish are found in streams, 34 are primarily or exclusively lake species and the samples did not include lakes. The lake species are not included in Table 3.2.

Each PA model has two measurements of percent accuracy when compared to the testing data: percent correct of predicted presences and percent correct of predicted absences. The mean of these two scores gives us an accuracy measurement that is used to compare individual species models (henceforth, this measurement is referred to as the “average accuracy”).

For all 93 PA models combined, I predicted 72% of the test data observations correctly. Forty-four percent of the PA models had an average accuracy of between 65% and 75%, including fish species such as rock bass, northern pike, smallmouth bass, and yellow perch (Table 3.3, Figure 3.2). Four models had predictions that were worse than simply guessing (<50% average accuracy: creek chubsucker, freshwater drum, eastern sand darter, blacknose shiner). However, 21% of the models had an average accuracy of greater than 80% (e.g., greenside darter, redfin shiner, and white perch). Fish species associated with big, slow rivers were modeled particularly well. Four redhorse species (black, greater, golden, silver) had an average accuracy greater than 88%, and two other redhorse (river, shorthead) had average accuracies of 74% and 77%. Channel catfish had an average accuracy of 90%, and common carp had an average accuracy of 80%. Although cold-water species were not modeled as accurately as redhorses, these fish models also did well; brook trout, slimy sculpin, mottled sculpin, Chinook salmon, and Coho salmon all had average accuracies of about 75%.

I recorded the frequency of each habitat variable included in PA models that had an average accuracy greater than 60% and a TSS greater than zero. The two variables that appeared most often were water temperature and catchment area, being in 45 and 44 of the 82 models, respectively (Table 3.4). Other frequently occurring variables included air temperature, predicted total phosphorus, and the 10% exceedence flow yield. All land-use variables included in the models occurred with approximately the same frequency, though land-use measured on the larger watershed scale occurred slightly more frequently (on average, in 14 of the 82 models) than land-use measured on the riparian scale (on average, in 11 of the 82 models).

I examined the PA models to see if there were any patterns associated with the variable splits of the five most frequently occurring variables. Patterns in the variable splits would indicate if these important variables have a consistent effect on the fish. The pattern was quite clear for water temperature; in 39 of the 45 models containing water temperature, an increase in water temperature resulted in fish presence. Not surprisingly, cold-water species were associated with 5 of the other 6 models. Brook trout, brown trout, rainbow trout, mottled sculpin, and slimy sculpin were predicted absent when the temperature was on average above 19.9°C. An increase in temperature resulted in fish absence for pirate perch as well, but the split value for water temperature in this model was quite high (23°C), so this fish should not be grouped with the others. Models of cool-water species (e.g., muskellunge, brook stickleback, brassy minnow) did not have consistent water temperature patterns.

An increase of catchment area resulted in a prediction of presence in 39 of the 44 models containing catchment area, and an increase of phosphorus resulted in a presence prediction for 18 of the 24 models containing phosphorus. The results for air temperature and 10% exceedence flow yield were ambiguous as neither presence nor absence predictions were dominant when the variable value increased.

I looked at the correlation matrix between the number of errors (absolute number, not a percentage) made at a site in the testing data and the habitat variables for the stream. For false negative errors, the highest correlation was rather small (10% exceedence flow yield : r = 0.17). However, the number of false positive errors made at a site was correlated with several habitat variables. The strongest correlation was between number of false positive errors and water temperature (r = 0.66), indicating that as stream water temperature increased, more species are predicted to be in streams where they were not observed. Similarly, catchment area (r = 0.35) and agriculture (RT_AGR : r = 0.43, WT_AGR: r = 0.50) are also positively correlated with the number of false positive errors at a site. On the other hand, percent of forest in the riparian zone (r = -0.58) and watershed (r = -0.57) is negatively correlated to number of false positive errors, indicating that as percent forest increases, fewer errors are made in a stream.

## Relative Abundance models

I created 46 RA models, 10 models with three abundance levels, 36 models having two abundance levels. I did not have enough data to create models for 47 of the species that I had created PA models for. Similar to the PA models, I predicted some species very well (e.g., brook stickleback, pumpkinseed),but was unable to model other species much more accurately than simply guessing (e.g., rainbow darter, rosyface shiner) (Tables 3.5 and 3.6, Figure 3.2). Overall, though, the accuracy of the RA models exceeded expectations, especially for that of the two-category models. The average three-category model predicted low abundances correctly 71.8% of the time, medium abundances 58.5% of the time, and high abundances 79.4% of the time (Table 3.5). On average, the two-level model predicted low abundances 80.2% of the time and high abundances 76.9% of the time (Table 3.6).

I recorded the number of times that each habitat variable occurred in the more accurate RA models (all 3-level models, and > 60% accuracy for both % low and % high in the 2-level models) (Table 3.4). Catchment area was the most important (41.9% of models), followed by predicted total phosphorus (32.6%) and percentage of coarse surficial geology in the watershed (27.9%). While water temperature and air temperature were in about 50% and 30% of the presence/absence models, they are only in 8 (18.6%) and 9 (20.9%) of the 43 RA models, respectively. Interestingly, both gradient and downstream link have moved from the bottom of the presence/absence list to near the top of the RA list (Table 3.4).

I looked for patterns in the relative abundance trees by examining the splits of the most frequent variables. While the effect of catchment area and gradient were ambiguous, a decrease of the downstream link variable (LINKDCATCH) resulted in a greater abundance in 9 of the 10 RA models it appeared in, and an increase total predicted phosphorus increased abundance in 12 of the 14 RA models it appeared in. Also, an increase of the value of coarse surficial geology resulted in a lower abundance in 10 of 12 RA models, and an increase of 90% exceedence flow yield resulted in a lower abundance for 11 of 11 RA models.

### Distribution maps

Using the predictions generated from the models, I created either presence/absence or absence/abundance statewide distribution maps. I give an example of a map that combines the presence/absence model and relative abundance model to classify each Michigan stream as absent, low, or high in rock bass (Table 3.3). In this example, I can see that rock bass is predicted to be found in low densities throughout the larger rivers of the Upper Peninsula and northern Lower Peninsula. The highest density of rock bass is predicted to be in the south-central portion of the Lower Peninsula, throughout the upper portions of the Saginaw, Grand, Kalamazoo, and St. Joseph watersheds. These predictions were tested against both presence/absence independent data and a 20% hold-out sample from the abundance training data (Table 3.3).

All species maps are available upon request to the author or at the website http://www-personal.umich.edu/~psteen/. Also available are interactive maps that run in the free downloadable program ArcReader (www.esri.com/arcreader). This program allows a user to query specific streams in the GIS to obtain observed fish and predicted fish information as well as the habitat variables used in the models.

## Discussion

I created presence/absence models for 93 fish species typically found in Michigan streams, and developed relative abundance models for 46 of these 93 species. About 7 of every 10 predictions were accurate for the PA models, about 6 of every 10 predictions were accurate for the 3-category RA models, and about 8 of every 10 predictions were accurate for the 2-category RA models. This suggests that landscape scale factors alone can be used to predict overall occurrence and abundance of most fish species in Michigan Rivers when site-specific data are not available.

Optimally, I would be able to create models based on both landscape-scale and local-scale variables (Wiley et al. 1997). Habitat conditions at the site scale (e.g., channel morphology, substrate and cover conditions, etc.) can have very strong effects on localized fish abundance patterns in streams. Since many landscape-scale variables impact local-scale mechanisms, I indirectly model some aspects of the local-scale control. However, without direct measurement of local-scale variables I was unable to capture all of the variation that occurs around these variables. Also, since the fish were measured with a single sample, it was impossible to detect how temporal variation could change the species presence and abundance (Wiley et al. 1997). Additionally, research has shown that biological variables such as competition are important to species occurrence and abundance (Larson and Moore 1985; Flecker and Townsend 1994; Stoks and McPeek 2003). For these reasons, I would not expect model accuracies much higher than obtained with this model set, and errors in the predictions were expected.

However, using local-scale variables to build models like those in this study would be impossible; obtaining small-scale data on a scale as large as the state of Michigan would require prohibitive amounts of time and money. Given that research in landscape ecology has indicated that large-scale variables may be as or more important to fish than small-scale variables, and often correlate strongly with the small-scale variables, I feel using large-scale variables was justified and was the best approach that could be used to meet the goals of the study (Schlosser 1991; Wiley et al. 1997; Fausch et al. 2002; Allan 2004).

### Presence/Absence model summary

With about 70% prediction accuracy against a test dataset, the PA models performed very well, overall. Large river fish such as redhorses and channel catfish were modeled very well, indicating that these species may be fully reliant on large-scale processes to determine their distribution. Similarly, cold-water species were predicted very well. Centrarchids were typically modeled with moderate accuracy (approximately 65-75%), indicating that landscape-scale habitat and characteristics were important, but there are other factors in determining their distribution that I was not able to detect with these models. For example, it is likely that including temporal variation in fish populations would increase model accuracy.

However, there certainly was variation in model accuracy between different species, with some models barely better or worse than guessing. There are a variety of ways to explain why some fish were modeled poorly. Misidentification of fish during the data collection phase could have played a role in poor model performance, as some of the less accurate models are built on fish species that are difficult to identify quickly in the field. Three lamprey species were probably predicted poorly due to this reason (silver lamprey, northern brook lamprey, American brook lamprey). The stream habitat data was perhaps not causally linked to the distribution of lake species that are found in rivers, resulting in poor prediction of certain lake species (burbot, freshwater drum). Some species were found virtually everywhere, and so the models were not able to distinguish between presence and absence streams (white sucker, blacknose dace). Unfortunately, many rare fish were predicted poorly as well (blacknose shiner, creek chubsucker, eastern sand darter); these are fish that were historically widespread but due to pollution and siltation now have a much narrower distribution (Trautman 1981; Roberts et al. 2005). The predictive models of these rare species were inaccurate as to where the fish currently live but may perhaps indicate where the fish has the potential to live.

Zorn (1998, 2003) used low-flow yield (as an index of water temperature) and catchment area as primary ordination axes in separating clusters of fish assemblages, and explained that these two variables can reliably be used to determine what fish may reside in a particular stream section. Unsurprisingly, the two most important variables in the PA models were also water temperature and catchment area. Numerous other studies have found water temperature to be key in the classification of fish (Fausch et al. 1998; Matthews and Robison 1988; Lyons 1992; Hinz and Wiley 1997; Zorn et al. 2002; Wehrly et al. 2003; Steen et al. 2006), and there is also a long history of studies on how a stream changes depending on its position in the catchment (Hawkes 1975; Vannote et al. 1980; Wiley et al. 1990; Smith and Kraft 2005).

Many of the GIS-based habitat variables served as surrogates for site-scale habitat. These variables require a conceptual leap from site-based to landscape-based modeling and their importance in the models emphasizes the linkages between the two scales of data. Catchment area is one such variable; it is a measure of the amount of land draining to the stream, and therefore is used as a convenient way of indicating a stream's approximate discharge, width, depth, and gradient (Vannote 1980). These stream characteristics are highly correlated with site-scale habitat values, such as velocity, channel substrate, and dissolved nutrients (Vannote 1980; Wiley et al. 1990; Rahel and Hubert 1991; Lyons 1996). In the models, more fish seem to prefer streams with larger catchment areas, indicating that larger streams with low gradient, high discharge, and warm summer water temperatures tended to favor the greatest number of Michigan fish species. Larger streams also have greater habitat complexity, providing space for a variety of fish species with different habitat requirements. The importance of catchment area has also been seen in previous fish classification and ordination work (Zorn et al. 2002).

Stream yield and specific power variables are GIS-derived surrogates for stream discharge, stream velocity, substrate, erosive force, and sediment transport capability. On average, these variables were contained within about 18% of the models; so while they are not integral to every model, they still have important effects. For example, the models predicted correctly that black crappie, bowfin, northern pike, and black bullhead will tend to be absent in streams with high stream power, indicating a preference for low velocity, lentic conditions. Bluegill was found to be present in streams with a low 10% yield; the species avoids streams with high peak flows. Slimy sculpin tended to be absent from streams with a low 90% yield, showing a tendency for groundwater driven streams with consistent flow rather than flashy, runoff driven streams.

The connectivity variables (e.g., distance from Great Lake, pond, or larger river) were included in only in about 10% of the models; however, these variables were very important in the modeling of several species. In the Coho and Chinook salmon models, the first split in the classification tree was the variable describing the distance from the closest Great Lake. Both models indicate that either species of salmon are very unlikely to be found more than 122 kilometers from a Great Lake. Removing this variable from either model resulted in predictions that were only slightly better than guessing; therefore, this variable was integral for in successful prediction.

The variable measuring the distance from the stream to Great Lake also indicated whether a stream was disconnected from the Great Lakes due to a dam or waterfall. While this aspect of the variable was unexpectedly not important in the Coho and Chinook models, it was important in the rainbow trout model. The rainbow trout model reported that it was unlikely, though not impossible, for rainbow trout to be found in a stream above a dam or waterfall. This result was entirely logical given the life history of the migrating steelhead. (I should note that no distinction was made between steelhead and resident rainbow trout in the model development phase due to uncertainty in the sampling database.)

The distance from a pond or lake and distance from a large river were also key variables for several species. For example, largemouth bass, smallmouth bass, and yellow perch were more likely to be found within 20 km, 8 km, and 6 km of a pond or lake respectively. The bowfin model predicted the species to be found within 150 meters of the confluence of the stream of interest and a river that has a 10% greater catchment area. This variable was also important for brown bullhead (21 km) and longnose sucker (23 km). Once again, it was entirely logical that the models have included these variables, as these fish were good examples of species that are found in lakes or slow-moving backwaters but also live in stream environments.

## Presence/Absence Error Analysis

In PA models, there are possible error types: false negative and false positive error. In my PA models, false positive errors occur more frequently than false negative errors by a ratio of 8:1. False negative errors are typically seen as more severe than false positive errors (McKenna et al 2006); a false negative error is more likely to be caused by an error in the model rather than a fish that was missed in the sampling.  In addition, false negative errors have a severe impact on conservation work based from models: if a rare species is predicted to be absent from a set of streams in which it actually exists, those streams may not be given the level of protection needed to conserve the species.

When distribution models are used for conservation work, false positive errors tend to be a "safe" error; if I do not know whether a fish in a stream or not, it is safer to assume the fish is present.  A false positive error does not necessarily indicate a flaw in the model; if a species was not observed in the field, the sampling effort may have been insufficient, the fish may not have been identified correctly, or the fish has the potential to live in the stream but simply is not there (McKenna et al. 2006).

False positive errors may also have been caused by quality discrepancies between the training and testing data. Overall, I had a higher degree of confidence in the fish-identification accuracy and catch efficiency of the MRI training data than the FCS test data.  As a result, the FCS test data probably had a higher proportion of fish that were improperly identified and a higher proportion of errors due to fish that were not caught but should have been.  When the test data were predicted by the models, the end result of this discrepancy was a higher number of false presences errors.  In other words, the model said the fish should have been there, and perhaps it was, but the FCS data was not accurate enough to show this.  Therefore the number of false presences in the test data may be inflated and underestimate the accuracy of the models, especially for hard to identify species.

To check this hypothesis, I compared the average false presence error rate for game fish, which are easily identified (brook and brown trout, smallmouth bass, largemouth bass, Chinook and Coho salmon, walleye, and yellow perch) against the average false presence error rate for cyprinids modeled in this study (chubs, dace, and minnow), which are typically harder to identify. The average false presence error for gamefish was 19.2%, and the false presence error for the cyprinids was 27.2%. The difference between the two is not as large as I had anticipated (independent t-test: t= -1.5, df= 26, p-value =.16), so it is likely that this hypothesis cannot fully explain the abundance of false presence predictions. However, it is possible that the discrepancy between the datasets can account for some of the false presence errors that occur.

I noticed that several of the habitat predictor variables were correlated to the number of false positive errors made at a stream reach. Water temperature was most strongly correlated to false positive errors; as temperature increases, the models tend to overestimate the number of species in the stream. Since warm-water streams have a higher diversity of species, it is likely that sampling efforts missed species in these streams, which would cause false positive errors in the test data. Another cause of these errors may be the bias introduced into the models through the disproportionate amount of cold-water stream samples compared to warm water stream samples in the training data; predictions made on cold-water are more accurate since they are more similar to the data used to make the models.

## Relative Abundance model summary

When using abundance categories in modeling, determining where to place the boundaries of the categories is a difficult problem and usually results in inaccurate models when predicting observations that are not clearly one category or another. Due to this, I was only able to create 10 species models that had test data accuracies better than simply guessing (every abundance category ≥ 33.3%). In order to develop RA models for the other species, I decided to create models in which the middle category was removed so that there was clear distinction between the high and low categories.

Of the 44 RA models created, 10 had three categories of predicted abundance (low, medium, high), while 36 had two categories of predicted abundance (low, high). Interestingly, not only did the two-category models perform well, but also they were typically more accurate than the PA models when compared to the test data. This implied that there might be greater stream habitat differences between low/high abundance streams than there were between presence/absence streams. For example, a stream may be considered "present" with one fish in it, and another stream with 1000 fish is also considered "present". The classification tree will have difficulty in distinguishing between the marginal stream with one fish and a true absent stream, resulting in misclassified observations in the PA model. On the other hand, when the stream with one fish is classified as "low", and the stream with 1000 fish is classified as "high", the classification tree is able to separate them with greater accuracy, since there are greater habitat differences between these streams than between a marginal stream and an "absent" stream.

While most of the common species in Michigan were modeled for relative abundance; I should note that because of the low number of species modeled for RA, these results do not apply to all Michigan fish. Water temperature was an unimportant variable for most of the RA models; according to these model sets, it was more important for determining presence/absence of a species than for determining how many of the fish are in the stream. Zorn (2003) observed the same phenomenon with temperature when developing landscape-based multiple regression models. Gradient, coarse surficial geology, and 90% exceedence flow are more important in the RA models than in the PA models. An increase in these variables tended to result in a decrease of abundance of several species (e.g., black bullhead, bluntnose minnow, largemouth bass, white sucker, yellow perch) that prefer streams with low slope and more variable flows. Given that these flow characteristics were correlated with water temperature, their importance may explain the apparent unimportance of water temperature. Water temperature may not have been included in the abundance classification trees because the variation in the data was already captured.

In the PA models, probability of presence increased with increasing total predicted phosphorus, and similarly, the RA models show that abundance increased with increasing phosphorus. This is a logical result (though its frequency in the models may be somewhat surprising), as phosphorus can cause a bottom-up effect, increasing productivity in every trophic level (Vanni 1987; Vanni et al. 1997). Though not seen in these models because Michigan streams tend to have low phosphorus levels, high phosphorus levels cause eutrophication and anoxic conditions, which would effectively destroy a fish population. For this reason, this general pattern in phosphorus cannot be extrapolated beyond the phosphorus range in the data.

### Other general model limitations

Overall, these models do a fine job in prediction, but the models have limitations that should be recognized. Users of these models should be aware of these issues, and if similar models are constructed in the future, researchers should try to address these problems in order to minimize model error.

Data quality is always an issue when dealing with large datasets. Brenden et al. (2006) addressed specific limitations in the NHD and quality of the GIS-derived environmental variables. In short, some of these variables were obtained from low resolution maps (e.g., surficial geology, 1:250,000 scale) and will not have the accuracy I desire when operating on a NHD with a resolution of 1:100,000. In my models, coarse surficial geology occurred relatively often (18.3% of PA models, 27.9% of RA models), and it is possible that the scaling issue increased model error slightly.

I used several habitat variables that were built from models and then predicted across the state in order to produce a value for each stream reach (e.g., water temperature, total phosphorus, flow variables). Since these models contain error, it is logical to expect that the error will trickle down to the fish models, decreasing model accuracy. This problem is also known as “propagation of error”. As these habitat models are improved in the future, I hypothesize that the fish predictions will become more accurate.

The fish data were of good quality overall, but the fish were sampled over a long period of time, by different people and for different purposes, so it is impossible to determine which samples were poorly counted or implemented. The samplers may have misidentified or failed to catch some fish, particularly those that are hard to identify, rare, or small. Training a model on flawed data can confound the training process and produce a model that is inaccurate for the species, especially if the predictor variables are correlated with the likelihood of failing to detect a species in a survey. While this issue is indeed a problem, to minimize this error I included as many absence sites as possible in the training data for each species. By pooling absence sites I have replicate information on the probability of absence as indexed by the data. If a fish could potentially be missed at any particular site, it was my intention to include several sites with the same type of habitat for which the fish would not be missed. This process may not produce absolute truth for every site, but the overall distribution should be correct. The errors in the training data are reflected in the accuracy measurements; the models are be perfect but should be good enough for the use for which they are intended.

A major problem throughout this study has been the difficulty in developing statewide abundance predictions. I tried several methods (regression, regression trees, classification trees with different category boundaries), none of which performed to my satisfaction. In the final product, I was only able to produce accurate models by dropping out data points so that a clear distinction could be made between high and low abundance streams. While this procedure did produce models that were accurate in determining high and low abundance, dropping data is not to be taken lightly. However, given the options of having no relative abundance models at all, or having models with some problems but indeed providing predictions, I feel that the right decision was made as these models have a place in a management or conservation context.

Table 3.1. List of habitat and land-use stressor variables used in the creation of the PA and RA models for Michigan stream fishes. The descriptive statistics summarize the entire Michigan stream population as per the GAP/CIAUMR database.

| Variable Code | Variable Description | Unit | Min | Max | Mean | Std.D |
|---|---|---|---|---|---|---|
| *Temperature* | | | | | | |
| WATER_TEMP | Water temperature, predicted July mean | Celsius | 12.3 | 26.2 | 19.5 | 3.0 |
| WT_MAAT | Mean annual air temperature | Celsius | 3.7 | 9.8 | 7.3 | 1.7 |
| *Position in Catchment* | | | | | | |
| CATCHAREA | Area of the watershed | km$^2$ | 0.72 | 14103.5 | 721.0 | 1680.6 |
| *Connectivity* | | | | | | |
| UP_POND | Distance upstream to closest pond >=5 acres | meters | 0 | 57566.4 | 8948.0 | 10580.0 |
| DOWN_POND | Distance downstream to closest pond >=10 acres or Great Lake | meters | 0 | 195470.1 | 29732.2 | 35989.0 |
| LINKDCATCH | Distance from downstream reach with 10% >= catchment area than target reach | meters | 0 | 58851.0 | 2871.0 | 7115.2 |
| DOWN_LENGTH | Distance to Great Lake from downstream end of reach | meters | 0 | 130093.1 | 31886.8 | 31417.6 |
| *Geology/Hydrologic* | | | | | | |
| WT_FINE | Fine-grain surficial geology - percentage of watershed | % | 0 | 1 | 0.11 | 0.22 |
| WT_COARSE | Coarse-grain surficial geology- percentage of watershed | % | 0 | 1 | 0.65 | 0.36 |
| TEN_YIELD | 10% exceedence flow yield | cms/km$^2$* | 0.0075 | 0.0416 | 0.0186 | 0.0037 |
| NINETY_YIELD | 90% exceedence flow yield | cms/km$^2$ | 0.0001 | 0.0264 | 0.0039 | 0.0031 |
| GRADIENT | Channel gradient | unitless | 0 | 0.0288 | 0.0026 | 0.0038 |
| TEN_POWER | High flow-based specific power | cms/km$^2$ | 0 | 0.0073 | 0.0005 | 0.0008 |
| NINETY_POWER | Summer flow-based specific power | cms/km$^2$ | 0 | 0.0021 | 0.0001 | 0.0002 |
| *Land-use* | | | | | | |
| WT_FOREST | Forest Land cover - percentage of watershed | % | 0.02 | 0.95 | 0.41 | 0.24 |
| WT_WETLAND | Wetland land cover- percentage of watershed | % | 0 | 0.56 | 0.15 | 0.08 |
| WT_AGR | Agricultural land-use- percentage of watershed | % | 0 | 0.95 | 0.28 | 0.25 |
| WT_URBAN | Urban land-use- percentage of watershed | % | 0 | 0.64 | 0.05 | 0.07 |
| RT_FOREST | Forest land cover- percentage of riparian network | % | 0.02 | 0.90 | 0.28 | 0.16 |
| RT_WETLAND | Wetland land cover- percentage of riparian network | % | 0.01 | 0.94 | 0.37 | 0.17 |
| Variable Code | Variable Description | Unit | Min | Max | Mean | Std.D |
| RT_URBAN | Urban land-use- percentage of riparian network | % | 0 | 0.56 | 0.04 | 0.06 |
| *Water Quality* | | | | | | |
| TOTAL_P__PPM | Total phosphorus, predicted | ppm | 0.01 | 0.25 | 0.05 | 0.04 |

* cms = cubic meters per second

Table 3.2. List of which Michigan fish species were modeled for presence/absence (PA) and relative abundance (RA). Numbers in these columns refer to the number of species occurrences in the training data (No asterisk- MRI data, Asterisk- MRI and MI Fish Atlas Data). Species that did not have enough data to be modeled are not listed.

| Family | Scientific Name | Common Name | P/A | RA |
|---|---|---|---|---|
| Amiidae | *Amia calva* | Bowfin | 77* | |
| Aphredoderidae | *Aphredoderus sayanus* | Pirate Perch | 32 | 24 |
| Atherinidae | *Labidesthes sicculus* | Brook Silverside | 58* | |
| Catostomidae | *Carpiodes cyprinus* | Quillback | 72* | |
| Catostomidae | *Catostomus catostomus* | Longnose Sucker | 41 | |
| Catostomidae | *Catostomus commersonii* | White Sucker | 375 | 277 |
| Catostomidae | *Erimyzon claviformis* | Creek Chubsucker | 39 | |
| Catostomidae | *Erimyzon sucetta* | Lake Chubsucker | 57* | |
| Catostomidae | *Hypentelium nigricans* | Northern Hog Sucker | 182 | 109 |
| Catostomidae | *Minytrema melanops* | Spotted Sucker | 67* | |
| Catostomidae | *Moxostoma anisurum* | Silver Redhorse | 31 | 34 |
| Catostomidae | *Moxostoma carinatum* | River Redhorse | 25* | |
| Catostomidae | *Moxostoma duquesnei* | Black Redhorse | 36 | |
| Catostomidae | *Moxostoma erythurum* | Golden Redhorse | 111 | 82 |
| Catostomidae | *Moxostoma macrolepidotum* | Shorthead Redhorse | 56 | 24 |
| Catostomidae | *Moxostoma valenciennesi* | Greater Redhorse | 35 | 38 |
| Centrarchidae | *Ambloplites rupestris* | Rockbass | 243 | 161 |
| Centrarchidae | *Lepomis cyanellus* | Green Sunfish | 200 | 128 |
| Centrarchidae | *Lepomis gibbosus* | Pumpkinseed | 197 | 124 |
| Centrarchidae | *Lepomis gulosus* | Warmouth | 97* | |
| Centrarchidae | *Lepomis humilis* | Orangespotted Sunfish | 61* | |
| Centrarchidae | *Lepomis macrochirus* | Bluegill | 284 | 99 |
| Centrarchidae | *Lepomis peltastes* | Longear Sunfish | 40 | |
| Centrarchidae | *Micropterus dolomieu* | Smallmouth Bass | 157 | 89 |
| Centrarchidae | *Micropterus salmoides* | Largemouth Bass | 180 | 96 |
| Centrarchidae | *Pomoxis annularis* | White Crappie | 29* | |
| Centrarchidae | *Pomoxis nigromaculatus* | Black Crappie | 85 | 110 |
| Cobitidae | *Misgurnus anguillicaudatus* | Oriental Weatherfish | 29* | |
| Cottidae | *Cottus bairdii* | Mottled Sculpin | 83 | 172 |
| Cottidae | *Cottus cognatus* | Slimy Sculpin | 60 | 61 |
| Cyprinidae | *Campostoma anomalum* | Central Stoneroller | 87 | 72 |
| Cyprinidae | *Clinostomus elongatus* | Redside Dace | 45* | |
| Cyprinidae | *Couesius plumbeus* | Lake Chub | 43* | |
| Cyprinidae | *Cyprinella spiloptera* | Spotfin Shiner | 68 | 39 |
| Cyprinidae | *Cyprinus carpio* | Common Carp | 150 | 76 |
| Cyprinidae | *Hybognathus hankinsoni* | Brassy Minnow | 77* | |
| Cyprinidae | *Luxilus chrysocephalus* | Striped Shiner | 71* | |
| Cyprinidae | *Luxilus cornutus* | Common Shiner | 263 | 203 |
| Cyprinidae | *Lythrurus umbratilis* | Redfin Shiner | 71* | 37 |
| Cyprinidae | *Margariscus margarita* | Northern Pearl Dace | 91 | |
| Cyprinidae | *Nocomis biguttatus* | Horneyhead Chub | 142 | 92 |
| Cyprinidae | *Nocomis micropogon* | River Chub | 41 | |

Table 3.2, continued.

| Family | Scientific Name | Common Name | P/A | RA |
|---|---|---|---|---|
| Sciaenidae | *Aplodinotus grunniens* | Freshwater Drum | 50* | |
| Salmonidae | *Oncorhynchus kisutch* | Coho Salmon | 37* | |
| Salmonidae | *Oncorhynchus mykiss* | Rainbow Trout | 128 | 109 |
| Salmonidae | *Oncorhynchus tshawytscha* | Chinook Salmon | 45* | |
| Salmonidae | *Salmo trutta* | Brown Trout | 196 | 159 |
| Salmonidae | *Salvelinus fontinalis* | Brook Trout | 186 | 165 |
| Umbridae | *Umbra limi* | Central Mudminnow | 259 | 179 |
| | | **Number of Species** | **93** | **46** |

Table 3.3. Sample size and % correct agreement between predicted presence/absence values and observed values in the test dataset, for each PA model. The list is sorted by the average between % present and % absent (average accuracy). The average accuracy does not consider differences in N between % present and % absent.

| Common Name | N Present | % Present | N Absent | % Absent | Average Accuracy |
|---|---|---|---|---|---|
| Black Redhorse | 12 | 91.7 | 788 | 94.9 | 93.3 |
| White Perch | 27 | 100.0 | 781 | 81.3 | 90.7 |
| Channel Catfish | 54 | 81.5 | 760 | 98.0 | 89.8 |
| Greenside Darter* | 8 | 100.0 | 72 | 79.2 | 89.6 |
| Greater Redhorse | 13 | 84.6 | 801 | 93.3 | 89.0 |
| Redfin Shiner | 21 | 95.2 | 803 | 82.6 | 88.9 |
| Golden Redhorse | 47 | 83.0 | 780 | 94.0 | 88.5 |
| Silver Redhorse | 11 | 81.8 | 802 | 94.3 | 88.1 |
| White Bass | 19 | 94.7 | 793 | 79.3 | 87.0 |
| Roseyface Shiner* | 15 | 100.0 | 84 | 71.4 | 85.7 |
| Lake Chub | 3 | 100.0 | 803 | 70.0 | 85.0 |
| Chinook Salmon | 60 | 88.3 | 786 | 80.2 | 84.3 |
| Spotfin Shiner | 49 | 75.5 | 781 | 92.8 | 84.2 |
| Mimic Shiner | 17 | 88.2 | 786 | 78.2 | 83.2 |
| Blackstripe Topminnow* | 12 | 91.7 | 104 | 74.0 | 82.8 |
| Walleye | 149 | 71.8 | 698 | 93.0 | 82.4 |
| Sea Lamprey | 4 | 100.0 | 801 | 64.7 | 82.3 |
| River Chub | 24 | 70.8 | 800 | 93.0 | 81.9 |
| Common Carp | 156 | 84.6 | 723 | 76.1 | 80.4 |
| Emerald Shiner | 24 | 70.8 | 796 | 89.7 | 80.3 |
| Tadpole Madtom | 22 | 72.7 | 802 | 87.4 | 80.1 |
| Sand Shiner | 22 | 72.7 | 785 | 86.6 | 79.7 |
| Black Crappie | 85 | 72.9 | 751 | 86.0 | 79.5 |
| Stonecat | 81 | 66.7 | 758 | 92.1 | 79.4 |
| Yellow Bullhead | 97 | 78.4 | 745 | 78.9 | 78.6 |
| Pirate Perch | 26 | 76.9 | 780 | 79.7 | 78.3 |
| Slimy Sculpin | 28 | 85.7 | 775 | 70.3 | 78.0 |
| Spotted Sucker | 12 | 91.7 | 801 | 63.8 | 77.7 |
| Brook Trout | 504 | 75.6 | 586 | 79.7 | 77.7 |
| Shorthead Redhorse | 30 | 63.3 | 781 | 90.0 | 76.7 |
| Mottled Sculpin* | 15 | 80.0 | 51 | 72.5 | 76.3 |
| White Crappie | 12 | 75.0 | 789 | 76.4 | 75.7 |
| Brook Silverside | 7 | 85.7 | 787 | 65.6 | 75.7 |
| Central Stoneroller | 105 | 73.3 | 731 | 77.2 | 75.2 |
| Muskellunge | 53 | 84.9 | 739 | 64.4 | 74.7 |
| Rockbass | 302 | 73.8 | 663 | 75.4 | 74.6 |
| Northern Pike | 251 | 61.8 | 667 | 87.4 | 74.6 |
| Coho Salmon | 75 | 72.0 | 763 | 76.0 | 74.0 |
| Longnose Sucker | 7 | 85.7 | 802 | 62.2 | 74.0 |
| River Redhorse | 3 | 66.7 | 788 | 81.2 | 74.0 |
| Fathead Minnow | 37 | 83.8 | 777 | 63.4 | 73.6 |
| Smallmouth Bass | 185 | 61.6 | 721 | 85.0 | 73.3 |

Table 3.3, continued.

| Common Name | N Present | % Present | N Absent | % Absent | Average Accuracy |
|---|---|---|---|---|---|
| Longnose Gar | 11 | 63.6 | 800 | 83.0 | 73.3 |
| Quillback | 180 | 61.1 | 794 | 84.9 | 73.0 |
| Chestnut Lamprey | 5 | 60.0 | 802 | 85.8 | 72.9 |
| Grass Pickerel | 101 | 66.3 | 694 | 78.7 | 72.5 |
| Northern Logperch | 104 | 63.5 | 746 | 80.6 | 72.1 |
| Longnose Dace | 134 | 67.2 | 717 | 76.7 | 72.0 |
| Brassy Minnow | 5 | 80.0 | 801 | 63.5 | 71.8 |
| Green Sunfish | 357 | 77.0 | 592 | 66.4 | 71.7 |
| Striped Shiner* | 18 | 61.1 | 101.0 | 81.8 | 71.5 |
| Yellow Perch | 221 | 61.9 | 650 | 80.2 | 71.0 |
| Northern Hog Sucker | 99 | 68.7 | 699 | 73.2 | 70.9 |
| Finescale Dace* | 10 | 60.0 | 104 | 81.7 | 70.9 |
| Largemouth Bass | 275 | 61.1 | 630 | 80.5 | 70.8 |
| Creek Chub | 401 | 75.1 | 398 | 64.6 | 69.8 |
| Bluntnose Minnow | 235 | 70.6 | 685 | 68.9 | 69.8 |
| Common Shiner | 353 | 68.3 | 621 | 71.0 | 69.7 |
| Brook Stickleback | 117 | 75.2 | 718 | 63.9 | 69.6 |
| Oriental Weatherfish* | 8 | 75.0 | 103 | 64.1 | 69.6 |
| Orangespotted Sunfish* | 15 | 66.7 | 106 | 70.8 | 68.7 |
| Rainbow Trout | 363 | 67.8 | 783 | 68.3 | 68.0 |
| Johnny Darter | 271 | 72.7 | 519 | 63.2 | 67.9 |
| Warmouth | 22 | 72.7 | 776 | 63.1 | 67.9 |
| Rainbow Darter | 98 | 60.2 | 693 | 75.6 | 67.9 |
| Black Bullhead | 78 | 65.4 | 762 | 70.1 | 67.8 |
| Pumpkinseed | 116 | 66.4 | 676 | 69.1 | 67.7 |
| Brown Trout | 711 | 70.0 | 531 | 65.3 | 67.7 |
| Hornyhead Chub | 137 | 73.7 | 737 | 61.3 | 67.5 |
| Iowa Darter | 10 | 70.0 | 800 | 62.3 | 66.1 |
| Brown Bullhead | 33 | 60.6 | 777 | 71.6 | 66.1 |
| Redside Dace | 5 | 60.0 | 803 | 71.9 | 65.9 |
| Northern Redbelly Dace | 46 | 69.6 | 763 | 61.9 | 65.7 |
| Burbot | 98 | 53.0 | 752 | 77.7 | 65.4 |
| Lake Chubsucker | 5 | 60.0 | 786 | 70.4 | 65.2 |
| Central Mudminnow | 481 | 69.0 | 514 | 61.1 | 65.1 |
| Blackside Darter | 259 | 60.2 | 669 | 69.7 | 65.0 |
| Golden Shiner | 18 | 61.1 | 775 | 68.1 | 64.6 |
| Bluegill | 284 | 60.2 | 641 | 68.6 | 64.4 |
| White Sucker | 761 | 66.8 | 379 | 60.7 | 63.7 |
| Least Darter | 5 | 60.0 | 785 | 64.1 | 62.0 |
| Bowfin | 24 | 62.5 | 782 | 61.5 | 62.0 |
| Silver Lamprey* | 10 | 60.0 | 90 | 63.3 | 61.7 |
| Banded Killifish | 14 | 71.4 | 105 | 51.4 | 61.4 |
| Longear Sunfish | 8 | 50.0 | 783 | 71.6 | 60.8 |
| Northern Pearl Dace | 16 | 62.5 | 795 | 52.6 | 57.5 |
| Western Blacknose Dace | 464 | 85.6 | 514 | 24.1 | 54.9 |
| Northern Brook Lamprey | 19 | 31.6 | 796 | 77.6 | 54.6 |

Table 3.3, continued.

| Common Name | N Present | % Present | N Absent | % Absent | Average Accuracy |
|---|---|---|---|---|---|
| American Brook Lamprey | 8 | 25.0 | 799 | 84.0 | 54.5 |
| Creek Chubsucker | 14 | 14.3 | 781 | 84.6 | 49.5 |
| Freshwater Drum | 33 | 36.4 | 781 | 62.5 | 49.5 |
| Eastern Sand Darter | 8 | 37.5 | 106 | 59.4 | 48.5 |
| Blacknose Shiner | 17 | 17.6 | 796 | 56.9 | 37.3 |

* Test data is a 20% holdout from the MRI training data

Table 3.4. The number of times a habitat variable is included in the A) 82 Michigan stream fish PA models with a presence and absence accuracy greater than 60%, B) and all 10 of the 3-category Michigan stream fish RA models and the 33 2-category Michigan stream fish RA models with a low and high accuracy greater than 60%.

A)

| *Variable Code* | *Number* | *Percentage* |
|---|---|---|
| WATER_TEMP | 45 | 54.9 |
| CATCHAREA | 44 | 53.7 |
| WT_MAAT | 26 | 31.7 |
| TOTAL_P_PPM | 24 | 29.3 |
| TEN_YIELD | 22 | 26.8 |
| WT_FOREST | 17 | 20.7 |
| WT_COARSE | 15 | 18.3 |
| UP_POND | 15 | 18.3 |
| TEN_POWER | 15 | 18.3 |
| NINETY_YIELD | 14 | 17.1 |
| RT_AGR | 13 | 15.9 |
| WT_WETLAND | 13 | 15.9 |
| WT_AGR | 13 | 15.9 |
| WT_URBAN | 12 | 14.6 |
| RT_FOREST | 12 | 14.6 |
| RT_WETLAND | 11 | 13.4 |
| NINETY_POWER | 10 | 12.2 |
| DOWN_POND | 8 | 9.8 |
| RT_URBAN | 8 | 9.8 |
| WT_FINE | 7 | 8.5 |
| GRADIENT | 7 | 8.5 |
| LINKDCATCH | 6 | 7.3 |
| DOWN_LENGTH | 6 | 7.3 |

B)

| *Variable Code* | *Number* | *Percentage* |
|---|---|---|
| CATCHAREA | 18 | 41.9 |
| TOTAL_P__PPM | 14 | 32.6 |
| WT_COARSE | 12 | 27.9 |
| NINETY_YIELD | 11 | 25.6 |
| LINKDCATCH | 10 | 23.3 |
| GRADIENT | 9 | 20.9 |
| WT_MAAT | 9 | 20.9 |
| WATER_TEMP | 8 | 18.6 |
| RT_AGR | 7 | 16.3 |
| WT_WETLAND | 7 | 16.3 |
| RT_WETLAND | 7 | 16.3 |
| TEN_YIELD | 6 | 14.0 |
| NINETY_POWER | 6 | 14.0 |
| RT_FOREST | 6 | 14.0 |
| UP_POND | 6 | 14.0 |
| DOWN_POND | 4 | 9.3 |
| WT_FINE | 4 | 9.3 |
| TEN_POWER | 4 | 9.3 |
| RT_URBAN | 4 | 9.3 |
| WT_FOREST | 3 | 7.0 |
| DOWN_LENGTH | 3 | 7.0 |
| WT_AGR | 2 | 4.7 |
| WT_URBAN | 1 | 2.3 |

Table 3.5. Sample size and % correct agreement between predicted RA category and observed values in the test dataset, for each 3-category RA model. The list is sorted by the average between % low, % medium, and % high. The average value does not consider differences in N between the three categories.

| Common Name | N Low | % Low | N Medium | % Medium | N High | % High | Average of Low, Medium, and High |
|---|---|---|---|---|---|---|---|
| Brook Stickleback | 6 | 66.6 | 5 | 100.0 | 5 | 80.0 | 82.2 |
| Northern Pike | 21 | 85.7 | 20 | 60.0 | 5 | 100.0 | 81.9 |
| Brown Bullhead* | 19 | 79.0 | 11 | 63.6 | 4 | 100.0 | 80.9 |
| Central Stoneroller | 8 | 87.5 | 5 | 60.0 | 5 | 80.0 | 75.8 |
| Longnose Dace | 9 | 77.8 | 3 | 66.7 | 5 | 60.0 | 68.2 |
| Black Crappie* | 68 | 66.2 | 47 | 55.3 | 5 | 80 | 67.2 |
| Greater Redhorse* | 15 | 53.3 | 20 | 35.0 | 3 | 100.0 | 62.8 |
| Tadpole Madtom* | 9 | 66.7 | 19 | 52.6 | 26 | 68.8 | 62.7 |
| Redfin Shiner* | 12 | 75.0 | 21 | 33.3 | 4 | 75.0 | 61.1 |
| Silver Redhorse* | 20 | 60.0 | 12 | 58.3 | 2 | 50.0 | 56.1 |

*Species was tested using a cross-validation procedure rather than 20% of the original data (Steinberg and Colla 1997).

Table 3.6. Sample size and % correct agreement between predicted RA category and observed values in the test dataset, for each 2-category RA model. The list is sorted by the average between % low and % high. The average value does not consider differences in N between the two categories.

| Common Name | N Low | % Low | N High | % High | Average of Low and High |
|---|---|---|---|---|---|
| Channel Catfish | 4 | 100.0 | 3 | 100.0 | 100.0 |
| Golden Shiner | 6 | 100.0 | 2 | 100.0 | 100.0 |
| Pirate Perch | 2 | 100.0 | 4 | 100.0 | 100.0 |
| Common Carp | 10 | 80.0 | 9 | 100.0 | 90.0 |
| Pumpkinseed | 18 | 94.4 | 13 | 84.6 | 89.5 |
| Rockbass | 14 | 100.0 | 26 | 76.9 | 88.5 |
| Stonecat | 6 | 100.0 | 13 | 76.9 | 88.5 |
| Shorthead Redhorse | 4 | 75.0 | 3 | 100.0 | 87.5 |
| Slimy Sculpin | 8 | 87.5 | 7 | 85.7 | 86.6 |
| Bluntnose Minnow | 11 | 90.9 | 33 | 81.8 | 86.4 |
| Yellow Bullhead | 10 | 80.0 | 9 | 88.9 | 84.5 |
| Black Bullhead | 8 | 87.5 | 5 | 80.0 | 83.8 |
| Grass Pickerel | 5 | 100.0 | 3 | 66.7 | 83.3 |
| Golden Redhorse | 6 | 83.3 | 14 | 78.6 | 81.0 |
| Blackside Darter | 16 | 81.3 | 24 | 79.2 | 80.3 |
| Spotfin Shiner | 4 | 100.0 | 5 | 60.0 | 80.0 |
| Northern Hog Sucker | 11 | 90.9 | 16 | 68.8 | 79.8 |
| Green Sunfish | 14 | 78.6 | 18 | 77.8 | 78.2 |
| Largemouth Bass | 17 | 70.6 | 7 | 85.7 | 78.1 |
| Western Blacknose Dace | 19 | 89.5 | 17 | 64.7 | 77.1 |
| Bluegill | 15 | 73.3 | 10 | 80.0 | 76.7 |
| Hornyhead Chub | 9 | 66.7 | 14 | 85.7 | 76.2 |
| White Sucker | 32 | 75.0 | 37 | 75.7 | 75.4 |
| Rainbow Trout | 14 | 71.4 | 13 | 76.9 | 74.2 |
| Brook Trout | 17 | 64.7 | 24 | 83.3 | 74.0 |
| Smallmouth Bass | 12 | 75.0 | 10 | 70.0 | 72.5 |
| Mottled Sculpin | 24 | 75.0 | 19 | 68.4 | 71.7 |
| Yellow Perch | 12 | 66.7 | 4 | 75.0 | 70.8 |
| Central Mudminnow | 22 | 77.2 | 23 | 60.8 | 69.0 |
| Logperch | 10 | 80.0 | 7 | 57.1 | 68.6 |
| Johnny Darter | 21 | 71.4 | 32 | 65.6 | 68.5 |
| Common Shiner | 15 | 60.0 | 36 | 72.2 | 66.1 |
| Brown Trout | 19 | 63.2 | 21 | 66.7 | 64.9 |
| Creek Chub | 27 | 63.0 | 33 | 60.1 | 61.6 |
| Rainbow Darter | 13 | 53.8 | 11 | 63.3 | 58.6 |
| Roseyface Shiner | 5 | 60.0 | 8 | 50.0 | 55.0 |

Figure 3.1. Classification tree of the brown bullhead PA model. Variable descriptions are given in Table 3.1. An observation less than or equal to the split value is sent to the node to the left, otherwise, it goes to the right. The terminal node indicates the final classification of the observation. Terminal node 2 and 6 indicate how the classification tree deals with uneven sample sizes between presence and absence. These nodes are classified as "present" even though they have more "absent" observations, because the frequency of "present" observations is higher in the terminal node than it was in the mother node.

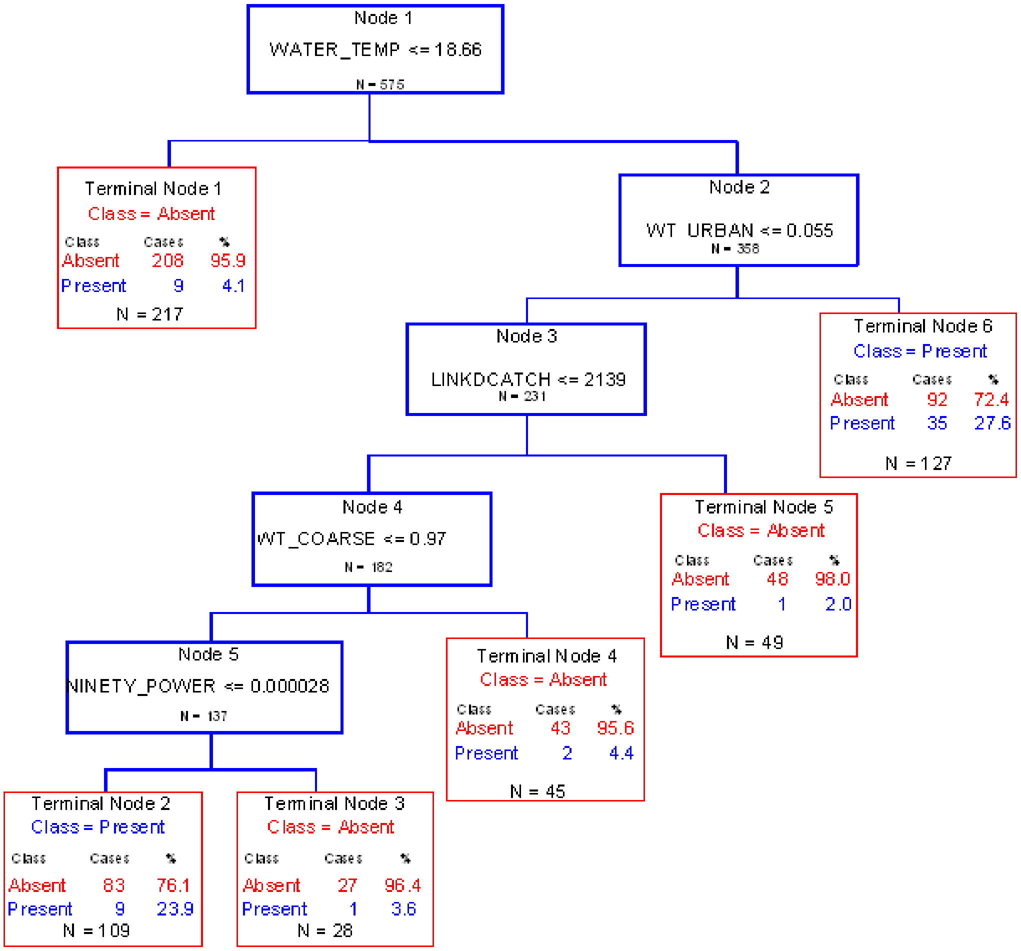


Figure 3.2. The percentage of Michigan stream fish models that fall within certain ranges of the average accuracy level for A) the 93 PA models, and B) the 46 RA models.

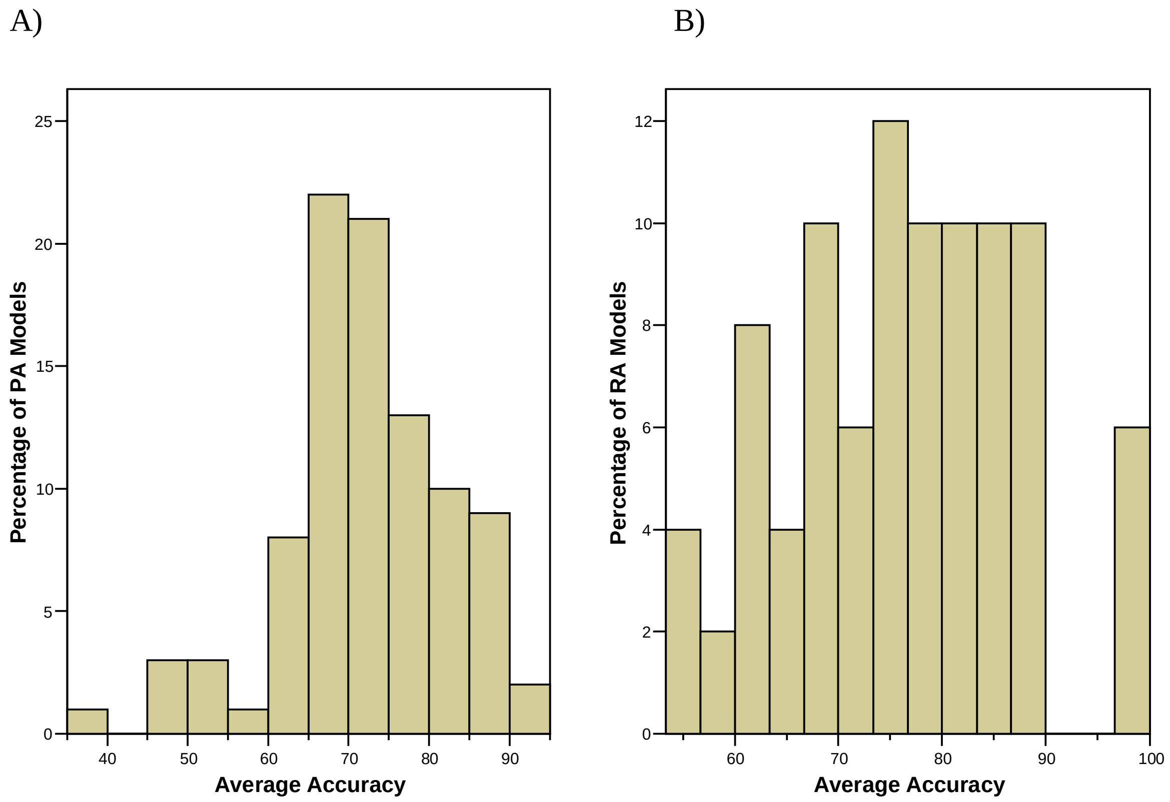
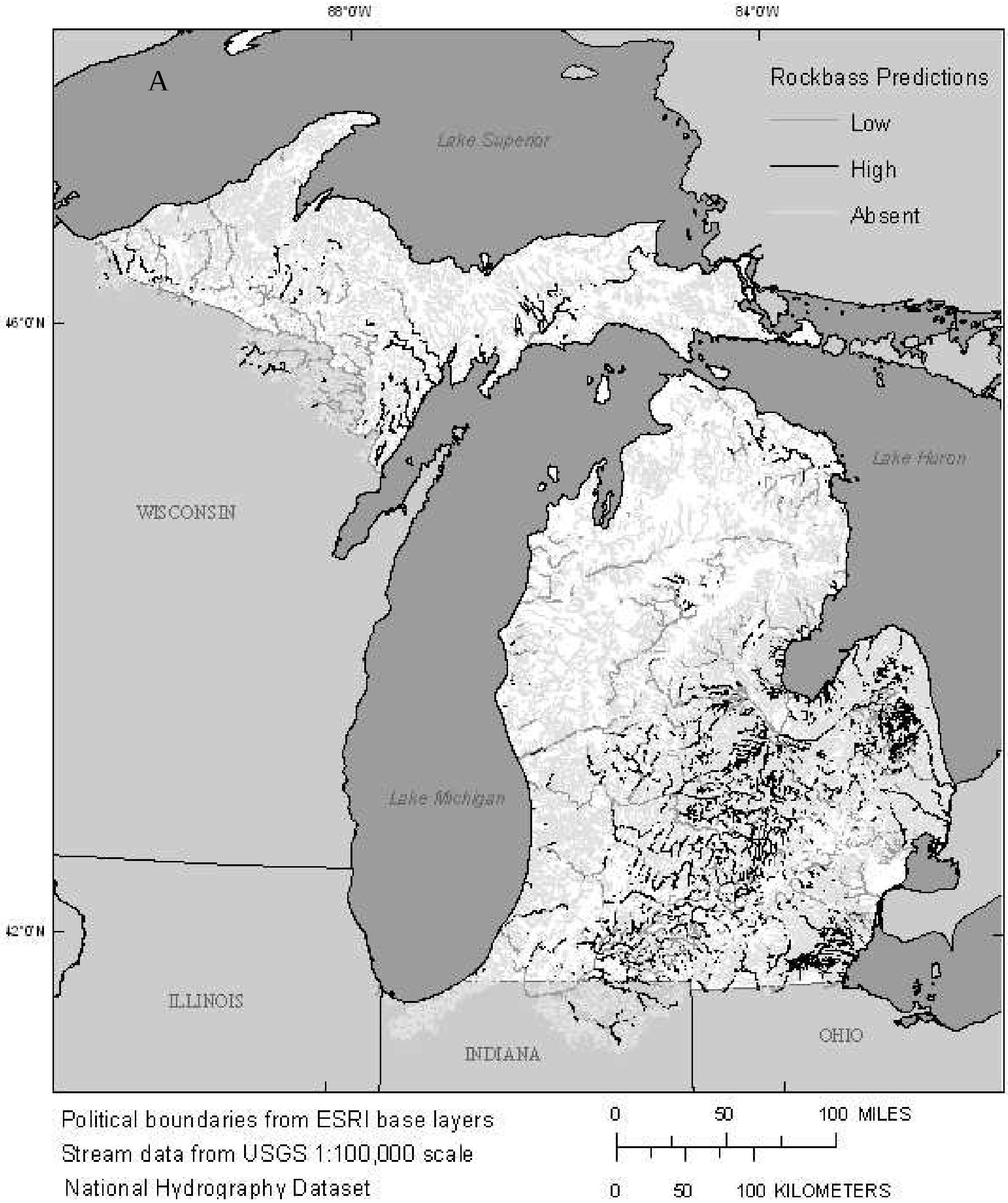
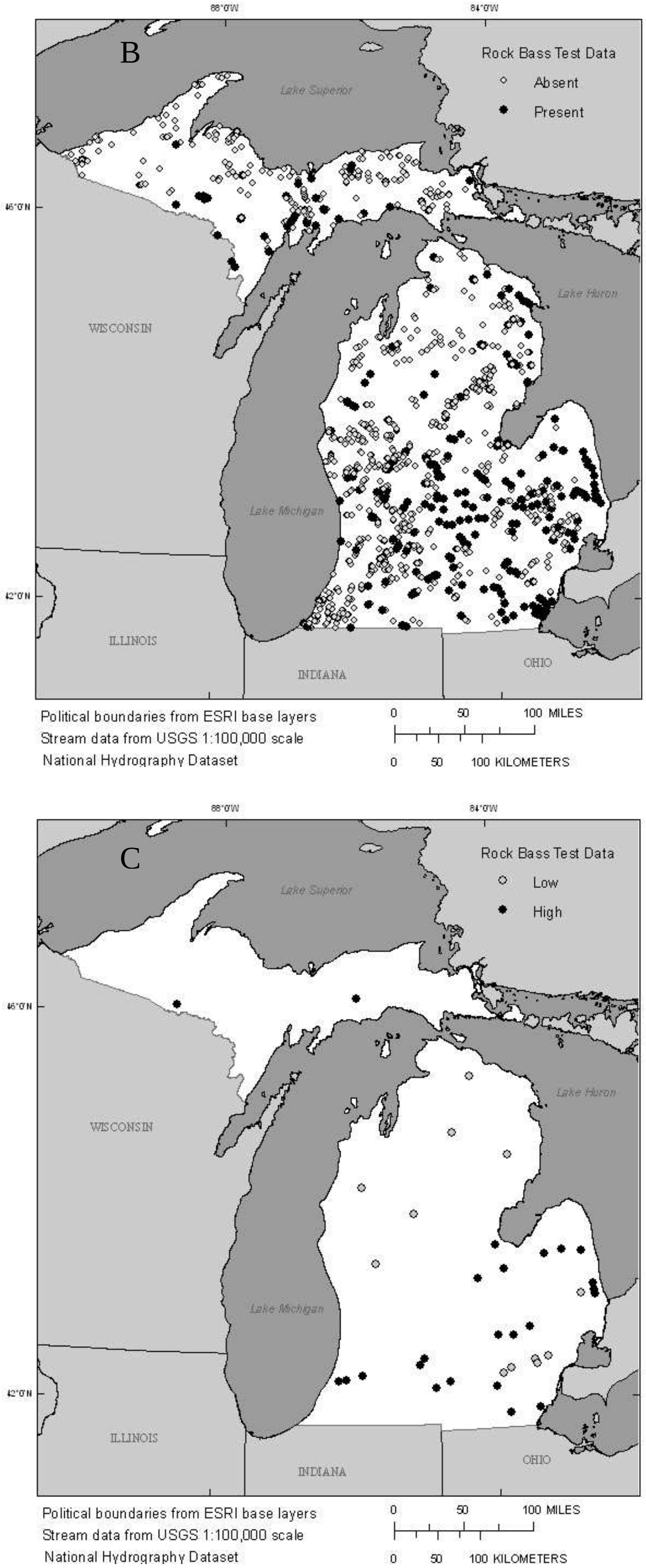


Figure 3.3. A) A rockbass distribution map that combines predictions from the PA model and RA model. B) Presence/absence data that was used to test this model: Presence 73.8% correct, Absence 75.4% correct. C) Abundance data that was used to test this model: Low 100% correct, High 76.9% correct. If PA model predicted a fish to be absent in a stream reach, the final prediction was “absent” regardless of abundance model result.



Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale
National Hydrography Dataset

Figure 3.3, continued.



Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale
National Hydrography Dataset

Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale
National Hydrography Dataset

# References

Allan, J. D. 2004. Landscapes and riverscapes: The influence of land-use on stream ecosystems. Annual Review of Ecology Evolution and Systematic 35:257-284.

Allouche, O, A. Tsoar, and R. Kadmon. Assessing the accuracy of species distribution models: prevalence, kappa, and the true skill statistic (TSS). Journal of Applied Ecology 43:1223-1232.

Bailey, J., and A. Alanara. 2006. Effect of feed portion size on growth of rainbow trout, Oncorhynchus mykiss (Walbaum), reared at different temperatures. Aquaculture 253:728-730.

Bailey, R. M., W. C. Latta, R. A. Simpson, and G. R. Smith. 2000. Distribution Maps of Michigan Fishes. University of Michigan, Museum of Zoology, Ann Arbor, Michigan.

Bailey, R.M and G.R. Smith. 2002. Names of Michigan Fishes. Michigan Department of Natural Resources, Fisheries Division.

Baker, C., R. Lawrence, C. Montague, and D. Patten. 2006. Mapping wetlands and riparian areas using Landsat ETM+ imagery and decision tree-based models. Wetlands 26:465-474.

Bell, J. F. 1999. Tree-based methods. Pages 89-105 *in* A. H. Fielding, editor. Machine learning methods for ecological applications. Kluwer Academic Publishers, Boston, Massachusetts.

Bent, P. C. 1971. Influence of surface glacial deposits on streamflow characteristics. U.S. Geological Survey, Water Resources Division, Open-file Report, Lansing, Michigan.

Breiman, L. F. J., R. Olshen, and C. Stone. 1984. Classification and regression trees. Chapman and Hall, New York.

Brenden, T. O., R. D. Clark, A. R. Cooper, P. W. Seelbach, L. Wang, S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49-74 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape Influences on Stream Habitat and Biological Assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

De'ath, G. 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. Ecology 83:1105-1117.

Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream fish from habitat variables: 1950-1985. United States Department of Agriculture, Forest Service, Pacific Northwest Research Station, General Technical Report PNW-GTR-213, Portland, Oregon.

Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: Bridging the gap between research and conservation of stream fish. BioScience 52:483-498.

Flecker, A. S., and C. R. Townsend. 1994. Community-wide consequences of trout introduction in New Zealand streams. Ecological Applications 4:798-807.

GLSC (Great Lake Science Center). 2007. The Great Lakes Aquatic GAP Project. Available: http://www.glsc.usgs.gov/GLGAP.htm (December 2007).

Hawkes, H. A. 1975. River zonation and classification. Pages 312-374 *in* B. A. Whitton, editor. River Ecology. University of California Press, Berkeley.

Hinz, L. C., and M. J. Wiley. 1997. Growth and production of juvenile trout in Michigan streams: Influence of temperature. Michigan Department of Natural Resources Fisheries Research Report 2041, Ann Arbor.

Holland, G. J., S. P. R. Greenstreet, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel *Ammodytes marinus* sediment habitat preferences in the marine environment. Marine Ecology- Progress Series 2005:269-282.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21-30.

Larson, G. L., and S. E. Moore. 1985. Encroachment of exotic rainbow trout intro stream populations of native brook trout in the southern Appalachian Mountains. Transactions of the American Fisheries Society 114: 195-203.

Leftwich, K. N., P. L. Angermeier, and C. A. Dolloff. 1997. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Transactions of the American Fisheries Society 126:725-734.

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

Lyons, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. United States Department of Agriculture, Forest Service, General Technical Report NC-149, St.Paul, Minnesota.

Lyons, J. 1996. Patterns in the species composition of fish assemblages among Wisconsin streams. Environmental Biology of Fishes 45: 329:341.

Maret, T. R., C. T. Robinson, and G. W. Minshall. 1997. Fish assemblages and environmental correlates in least-disturbed streams of the upper Snake River basin. Transactions of the American Fisheries Society 126:200-216.

Matthews, W. J., and H. R. Robison. 1988. The distribution of the fishes of Arkansas: a multivariate analysis. Copeia 1988:358-374.

McKenna, J. E., R. P. McDonald, C. Castiglione, S. S. Morrison, K. P. Kowalski, and D. Passino-Reader. 2006. A broadscale fish-habitat model development process: Genesee basin, New York. Pages 533-554 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

NHD (National Hydrography Dataset), 2007. Available: http://nhd.usgs.gov/ (December 2007).

Olden, J. D. 2001. A species-specific approach to modeling biological communities and its potential for conservation. Conservation Biology 17:854-863.

Olden, J. D., and D. A. Jackson. 2002. A comparison of statistical approaches for modelling 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.

Poff, N. L., and J. D. Allan. 1995. Functional organization of stream fish assemblages in relation to hydrological variability. Ecology 76:606-627.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The natural flow regime. BioScience 47:769-784.

Rahel, F. J., and W. A. Hubert. 1991. Fish assemblages and habitat gradients in a Rocky Mountain-Great Plains streams: biotic zonation and additive patterns of community change. Transactions of the American Fisheries Society 120:319-332.

Raleigh, R. F. 1982. Habitat suitability index models: Brook trout. U.S. Department of the Interior, Fish and Wildlife Service, FWS/OBS-82/10.24, Fort Collins, Colorado.

Rand, P. S., S. G. Hinch, J. Morrison, M. G. G. Foreman, M. J. MacNutt, J. S. MacDonald, M. C. Healey, A. P. Farrell, and D. A. Higgs. 2006. Effects of river discharge, temperature, and future climates on energetics and mortality of adult migrating Fraser River sockeye salmon. Transactions of the American Fisheries Society 135:655-667.

Rathert, D., D. White, J. C. Sifneos, and R. M. Hughes. 1999. Environmental correlates of species richness for native freshwater fish in Oregon. Journal of Biogeography 26:257-273.

Roberts, M. E., B. M. Burr, M. R. Whiles, and V. J. Santucci. 2005. Reproductive ecology and food habits of the blacknose shiner, Notropis heterolepis, in Northern Illinois. The American Midland Naturalist 155:70-83.

Seelbach, P. W., and M. J. Wiley. 1997. Overview of the Michigan Rivers Inventory (MRI) project. Michigan Department of Natural Resources, Fisheries Technical Report 97-3, Ann Arbor.

Seelbach, P. W., M. J. Wiley, P. Sorranno, and M. Bremigan. 2002a. Aquatic conservation planning: Using landscape maps to predict ecological reference conditions for specific water. Pages 454-477 *in* K. J. Gutzwiller, ed., editor. Applying Landscape Ecology in Biological Conservation. Springer, New York.

Schlosser, I. J. 1991. Stream fish ecology: a landscape perspective. BioScience 41:704-712.

Smale, M. A., and C. F. Rabeni. 1995. Hypoxia and hyperthermia tolerances of headwater stream fishes. Transactions of the American Fisheries Society 124:698-710

Smith, T. A., and C. E. Kraft. 2005. Stream fish assemblages in relation to landscape position and local habitat variables. Transactions of the American Fisheries Society 134:430-440.

Snyder, C. D., J. A. Young, R. Villella, and D. P. Lemarie. 2003. Influences of upland and riparian land use patterns on stream biotic integrity. Landscape Ecology 18:647-664.

Steen, P. J., D. Passino-Reader, and M. J. Wiley. 2006. Modeling brook trout presence and absence from landscape variables using four different analytical methods. Pages 513-531 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Steinberg, D., and P. Colla. 1997. CART-- Classification and Regression Trees. Salford Systems, San Diego, CA.

Stoks, R., and M. A. McPeek. 2003. Predators and life histories shape *Lestes* damselfly assemblages along a freshwater habitat gradient. Ecology 84:1576-1587.

Taverna, K., D. L. Urban, and R. I. McDonald. 2005. Modeling landscape vegetation pattern in response to historic land-use: a hypothesis-driven approach for the North Carolina Piedmont, USA. Landscape Ecology 20:689-702.

Trautman, M. B. 1981. The fishes of Ohio. Ohio State University Press, Columbus.

UM (University of Michigan). 2007. Ecological classification of rivers for environmental assessment. Available: http://sitemaker.umich.edu/riverclassproject (December 2007).

Usio, N., H. Nakajima, R. Kamiyama, I. Wakana, S. Hiruta, and T. N. 2006. Predicting the distribution of invasive crayfish (*Pacifastacus leniusculus*) in a Kusiro Moor marsh (Japan) using classification and regression trees. Ecological Research 21:271-277.

Vanni, M. J. 1987. Effects of nutrients and zooplankton size on the structure of a phytoplankton community. Ecology 68:624-635.

Vanni, M. J., C. D. Layne, and S. E. Arnott. 1997. "Top-down" trophic interactions in lakes: Effects of fish on nutrient dynamics. Ecology 78:1-20.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:131-137.

Vaughan, I. P., and S. J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology 17:1601-1611.

Vayssieres, M., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science 11:679-694.

Wang, L., J. Lyons, P. Kanehl, and R. Gatti. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22:6-12.

Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 2001:255-266.

Wang, L., J. Lyons, P. W. Rasmussen, and P. W. Seelbach. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Canadian Journal of Fisheries and Aquatic Sciences 60:491-505.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2003. Classifying regional variation in thermal regime based on stream fish community patterns. Transactions of the American Fisheries Society 132:18-37.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2006. Influence of landscape features on summer water temperatures in Lower Michigan streams. Pages 113-127 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Wiley, M. J., L. L. Osborne, and R. W. Larimore. 1990. Longitudinal structure of an agricultural prairie river system and its relationship to current stream ecosystem theory. Canadian Journal of Fisheries and Aquatic Sciences 47:373-384.

Wiley, M. J., S. L. Kohler, and P. W. Seelbach. 1997. Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. Freshwater Biology 37:133-148.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 1998. Patterns in the distributions of stream fishes in Michigan's lower peninsula. Michigan Department of Natural Resources, Fisheries Research Report 2035, Ann Arbor.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's lower peninsula. Transactions of the American Fisheries Society 131:70-85.

Zorn, T. G. 2003. Fishes of Lower Michigan: distribution patterns, abundance models, and causal relationships. PhD Dissertation, University of Michigan, Ann Arbor.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2004. Utility of species-specific, multiple linear regression models for predictions of fish assemblages in rivers of Michigan's Lower Peninsula. Michigan Department of Natural Resources, Fisheries Research Report 2072, Ann Arbor.

# Chapter 4

# Predicting past and future changes in Muskegon River watershed (Michigan, USA) game fish under land-use alteration and climate change scenarios

## Abstract

Future alterations in land-use and climate have the potential to cause substantial changes in the composition of stream fish communities. Stream fish predictive distribution models are an important tool to assess the probability of these changes causing species gain, loss, or extirpation. In this study, classification tree models predicting the probability of species presence were applied to the Muskegon Watershed (Michigan, USA). The models were applied to three potential future scenarios: 1) land-use change only, 2) land-use change and a 3 °C increase in air temperature by 2100, and 3) land-use change and a 5 °C increase in air temperature by 2100. The analysis indicated that the expected change in air and subsequent change in water temperatures resulted in the decline of cold-water fish in the Muskegon watershed by the end of the 21st century while warm-water species were predicted to significantly increase in range. Changes in land-use are expected to cause large changes in a few particular fish species such as walleye and Chinook salmon, but are not predicted to drive the overall changes in fish composition. Through interpretation of the classification tree models, managers can develop plans about how stream environmental conditions should be altered to maximize the probability of species residing in particular stream reaches.

## Introduction

The extirpation of Arctic grayling from Michigan streams, the reduction of lake trout in the Great Lakes, and loss of unique salmon sub-species in the western United States have something in common: they are the result of human disturbance in the environment. Over-fishing, pollution, dams, habitat degradation, and exotic species, among many other stressors, reduce native fish habitat and damage the integrity of an ecosystem's trophic structure. As a result of these anthropogenic impacts, we have seen fish community changes in the past and should expect more in the future.

It is important to be able to anticipate future impacts on fish communities. Knowledge of the loss of the grayling may have persuaded lawmakers to place more restrictions on 19th century logging practices, which contributed significantly to the species decline. Predicting likely future changes in fish communities can allow us to anticipate economic hardship in businesses dependent on sport fisheries, allow managers time to alter practices such as stocking patterns, catch limits, and fishing seasons, allow conservation groups to study and maintain areas with a high risk of habitat degradation, and create more public awareness of the importance of wise land-use management practices. Models of fish community changes can help us quantify the risk to fish populations, indicate what environmental conditions should be changed or maintained to obtain maximum fish potential, and inform our decision making processes.

Human use of land has constantly changed from the pre-settlement era to the present, and land-use shifts over the next century will undoubtedly continue to have effects on fauna of aquatic systems. It is anticipated that developed land in the US will increase by 79% over the next 25 years (Alig et al. 2004), which bodes poorly for fish communities because of the strong negative relationship between urban land, fish health, and fish biotic integrity (Scott et al. 1986; Weaver and Garman 1994; Wang et al. 2001; Tabit and Johnson 2002; Snyder et al. 2003; Wang et al. 2003a; Miltner et al. 2004). On the other hand, this increase of urban land will be mitigated somewhat by expected reduction of agricultural land (Pijanowski et al. 2001); agriculture is predicted to have negative effects on fish although these effects are not as strong as urban effects (Hall et al. 1999; Talmage et al. 2002; Wang et al. 2003b; Zimmerman et al. 2003; Barker et al. 2006). In addition; natural land-cover such as forests, which increase hydrologic stability and provide in-stream habitat, will likely continue to replace old agricultural fields (Roy et al. 2006; Barker et al. 2006; Meador and Goldstein 2003).

It is anticipated that climate change will also have major effects on the future of fish communities. Global warming will increase water temperatures, altered precipitation patterns and increased evaporation will result in changes in water quantity, and water quantity changes will affect water quality due to changes in concentration of the water’s chemical constituents (Regier and Meisner 1990, Schlinder 2001).

Several studies have predicted that water temperature increases due to climate change will result in reductions of brook trout, brown trout, and other coldwater species. Meisner (1990) predicted 40% trout reductions for southern Ontario streams, Flebbe (1996) predicted trout reductions between 53 and 97% in the southern Appalachian mountains, and Jager et al. (1999) predicted that climate change would restrict brown and rainbow trout from lower elevation to higher elevation Sierra Nevada streams. Eaton and Scheller (1996), using models based solely on thermal habitat, predicted that global warming of 4 °C would reduce cold and cool-water fish across the United States by 50%.

Across the US, increased water temperatures due to climate change are projected to reduce the number of lakes able to hold cold and cool-water fish communities by 45% and 30% respectively (Stefan et al. 2001).

On the other hand, higher water temperatures are predicted to increase growth of largemouth bass (McCauley and Kilgour 1990) and smallmouth bass (King et al. 1999). For the Great Lakes proper, Magnuson et al. (1990) reported that thermal habitat will increase for cold, cool, and warm water fish because most of the water is currently too cold to support even cold-water species much of the year. They predicted that the growing season length and range of depths with suitable temperatures would increase for all species.

In this study, I examined how game fish in the streams of the Muskegon River watershed (Michigan, USA) are expected to change in the 21st century with possible changes in land-use and stream temperature. The models and results were used to indicate how stream environmental conditions should be altered to maximize the probability of species residing in particular stream reaches.

This modeling effort represents a part of the Muskegon River Ecological Modeling System (MREMS), a modeling framework capable of predicting future and past states of the Muskegon River System and evaluating likely changes in hydrology, chemistry, and biology (Seelbach and Wiley 2005; Riseng et al. 2006). Risk analyses developed from MREMS are used to aid researchers and local stakeholders in monitoring and restoration activities.

Also, this study provides a justification for using classification tree models developed on present day habitat to predict future fish distributions. The models in this study were developed from a classification tree technique using data based on year 2001 habitat values. How do we know that models created on present day data can be applied to future scenarios? In order to justify using these models to predict in the future, it is necessary to provide a conceptual argument about how the patterns in the data identified by the classification tree algorithm can have real world meaning to the different fish species.

## Methods

### Study units

The Muskegon watershed, located in the western Lower Peninsula of Michigan, incorporates over 6,000 square kilometers of land, which is composed of a moderate mix of urban, agriculture, forest, and wetland land-use (O'Neal, 1997). The Muskegon River system, which is over 2,800 kilometers in length, drains into Lake Michigan. Water quality is good throughout the system, and stable flows due to permeable geology and high groundwater input provide for high quality cool and cold-water fisheries.

Five sections of the Muskegon system are particularly important to sport fisheries and were examined in greater detail throughout this study (Figure 4.1). The main branch of the Muskegon River from Muskegon Lake to Croton Dam is approximately 70 kilometers long, and supports populations of *Sander vitreus* walleye, *Micropterus dolomieu* smallmouth bass, *Esox Lucius* northern pike, *Oncorhynchus mykiss* steelhead (rainbow trout), *Salmon trutta* brown trout, and *Oncorhynchus tshawytscha* Chinook salmon (O'Neal, 1997, Hanchin, 2007). Hereafter, mention of the lower Muskegon River refers to this river section, while the Muskegon River system refers to all of the streams in the Muskegon watershed.

Flowing into the lower Muskegon River near Croton Dam is Bigelow Creek, a stretch of water only 18 kilometers long but known for extremely cold water and good *Salvelinus fontinalis* brook trout and brown trout populations as well as for providing important spawning grounds for steelhead and Chinook. Cedar Creek, a 38 kilometer cold-water stream that flows into Muskegon Lake, provides excellent habitat for brook and brown trout and also supports Chinook. Upstream from Croton Dam, the cold/cool water Middle Branch River (48 kilometers long) and Clam River (78 kilometers long) have good brook and brown trout populations. Portions of the Middle Branch River and Clam River have been designated as blue-ribbon trout streams by the Michigan Department of Natural Resources (Trout Unlimited 2008).

## Model Development and Application

Changes in the fish community of the Muskegon River system were predicted using classification tree fish distribution models that have been developed for the entire state of Michigan (Chapter 3). These models predicted fish species presence/absence on inter-confluence stream reaches based on the 1:100,000 National Hydrography Dataset (Brendan et al. 2006; NHD 2007). Using statewide fish and landscape-scale habitat data (Table 4.1), ninety-three presence/absence classification tree models were built for the most common Michigan river species. The models were compared to an independent data set to obtain a measure of model accuracy. In this study, only the models for 9 species of common game fish are used: brook trout, brown trout, Chinook salmon, *Oncorhynchus* *kisutch* Coho salmon, *Micropterus salmoides* largemouth bass, northern pike, rainbow trout, smallmouth bass, and walleye.

For each species, on every stream reach in the Muskegon River system, I estimated the frequency of species occurrence (FO) as the number of presence observations classified into a terminal node of the classification tree model compared to all observations classified into this terminal node. For example, across Michigan, stream reaches with a July mean water temperature greater than 19.3 °C contained brook trout in 9 of 276 (0.03) of the training observations and 29 of 205 (0.12) of the testing observations (combined together, 38 of 472 (0.07) of these observations contain brook trout) (Appendix A). I made the assumption that the rivers sampled in the training and testing data are representative of the streams in Michigan; therefore, I predicted that in the Muskegon system, stream reaches with a daily July mean water temperature greater than 19.3 °C contain brook trout at a frequency of 0.07. Using this measurement in our analysis built in realism beyond “present” and “absent” because brook trout may conceivably be found in warmer water temperatures. This 0.07 value represents one terminal node in the classification tree; stream reaches with different combinations of habitat values were classified into different terminal nodes that had different frequencies of occurrence.

For each game fish, I summarized the percent chance of occurrence for the entire Muskegon River system and for each study unit (Lower Muskegon, Cedar River, Bigelow Creek, Middle Branch River, Clam River). To do this, I took the average of the frequency of occurrence (FO) for the stream reaches composing each unit, weighted by stream length, and converted it to a percentage. For example, if the Cedar River was composed of three stream reaches with lengths of 10, 11, and 12 kilometers and frequencies of brown trout occurrence of 0.3, 0.5, and 0.8, respectively, then the average percent change of the occurrence per stream kilometer for this 33 kilometer stream reach would be (0.3*10)+(0.5*11)+(0.8*12)/ (10+11+12) = 0.55 or 55%. According to this interpretation, sampling any random kilometer in this 33-kilometer stream reach would result in a 55% chance of finding brown trout.

## Backcasting and Forecasting

To apply the fish models to both the past and present, changes were made in several of the predictive habitat variables on which the models are based (Table 4.1). Application of land transformation models (Pijanowski et al. 2001) to the Muskegon watershed produced estimates of urban, agriculture, forest, and wetland land-use for each decade from 1900-1970, for each decade from 2010 to 2040, and 2070 and 2100 (Figure 4.2). This iteration of the land transformation model assumes "business-as-usual" approaches to land development; current rates of land transformation will continue in the future. Measured land-use data were available for the pre-settlement era (approximately 1830) and 1978.

Since the land transformation model data and 1978 data were based on aerial photos but the 2001 data (upon which the models were created) were based on satellite images (MCGI 2007), I needed to transform 1978 land-use and land transformation model land-use into a data format compatible with 2001 coverage. To do this, I used simple linear regression equations for the years 2001 (satellite, dependent variable) and 1998 (aerial photos, independent variable) for each land-use category (urban, agriculture, forest, and wetland) and applied them to the 1978 land-use, the backcasted land-use, and the forecasted land-use.

Ninety and ten percent exceedence flows and stream power, which were predicted from regressions based on urbanization, agriculture, and surficial geology, were adjusted for each year of the backcasted and forecasted land-use (Brenden et al 2006). Also, total phosphorus, which depended heavily on agriculture (Chapter 3), was predicted for each year of the backcasted and forecasted land-use.

The variable DOWN_LENGTH, which measured the distance from stream reach to Great Lake and also identified whether a dam interrupted that connection, was updated for the historical model years. For 1830, all dams were removed from the Muskegon system. Other dams were replaced back into the DOWN_LENGTH variable as time progressed. Of these changes, of greatest note is the building of the Croton Dam in 1906, which separates about two-thirds of the Muskegon stream system from Lake Michigan.

I developed three potential scenarios for what could happen to habitat variables in the future. The baseline scenario assumes no global warming; future air and water temperatures remain the same from the 2001 levels although hydrology, land-use, and phosphorus change as described above. The “slow temperature-change” scenario uses these changes but also adds the assumption that air temperature warms 3 °C linearly from 2001 to 2100 (0.03 °C per year). The “fast temperature-change” scenario assumes that air temperature warms 5 °C linearly from 2001 to 2100 (0.05 °C per year). These values are used because studies of air temperature change predict an increase of 3-5 °C by 2100 (Thomson et al. 2005). Given the difficulty of determining how climate change will alter precipitation, I decided to concentrate on temperature change and not implement water quantity changes except for those changes caused by land-use alteration in the flow regression models.

I expected that the stream water temperature will not warm as quickly as air temperature for two reasons: First, water has a higher specific heat than the atmosphere so it takes more energy input per unit of mass to raise its temperature. Secondly, a portion of the water in a stream comes via groundwater routing. Because the temperature of groundwater is approximately equal to the mean annual air temperature, groundwater temperature will increase over time, but the increase rate will be less than the air temperature increase rate given that groundwater is beneath the surface and insulated from changes in the atmosphere. Stefan and Preud'homme (1993) found that in the surface run-off driven streams of northern and central portions of the Mississippi River basin, weekly water temperature increased by 0.86 times the weekly air temperature. However, they indicated that this value would be too high for groundwater systems. Glacial processes deposited large amount of sand and gravel where the Muskegon River system now flows, so groundwater is a major source of stream discharge for Muskegon streams. Therefore, I made the assumption that water temperature will increase by 0.8 times the rate of air temperature increase (Stefan and Preud'homme 1993; Eaton and Scheller 1996; Schindler 1997). This keeps the conversion rate similar to that which was reported in the literature but adjusts it slightly for changes in the water source.

Classification tree models and the altered land-use/cover, stream discharge and power, and phosphorus variables were used to make historical predictions (pre-settlement to 1978) for each game fish mentioned above. Classification tree models and the altered variables appropriate to each future scenario were used to make future predictions for each game fish. From these predictions we determined the percent chance of occurrence of each fish in the Muskegon River system and the five study units and examined how the percent chance of occurrence changed over time.

Not all fish species examined in this chapter are native to Michigan and therefore were not present in Michigan in the pre-settlement era. Brown trout and rainbow trout were introduced into Michigan in the late 19th century; therefore, predictions were not made for these species during the pre-settlement era. Chinook and Coho salmon were introduced to the Great Lakes basin in 1967; therefore application of these models to the Muskegon watershed began at 1970. It is not known whether brook trout were widespread throughout the Muskegon in the pre-settlement era due to competition with the Arctic graying, which occupied similar habitat (Scott and Crossman 1973). The model's pre-settlement predictions of brook trout may apply to this species instead. However, by 1900 Arctic grayling had disappeared throughout Michigan so predictions from 1900 and later apply only to brook trout.

## Results

### Brook Trout

The brook trout model predicted that the majority of the streams during the pre-settlement Muskegon River system were classified into terminal node 4 (frequency of occurrence, FO 0.80) and 2 (FO 0.84) (Appendix A). Therefore most of the river system had a high chance of brook trout (or Arctic grayling) presence during the pre-settlement era. However, as agriculture and phosphorus increased through the turn of the century, the model predicted that many of these streams lost some of their brook trout potential (terminal node 5, FO 0.44). From 1900 to 2001, there was approximately a 35% chance of finding brook trout in any random kilometer of stream in the Muskegon River system (Table 4.2A).

In the baseline future scenario, a reduction in phosphorus due to agriculture loss resulted in reclassifying streams from terminal node 5 (FO 0.15) to terminal nodes 1 (FO 0.51) and 4 (FO 0.80). Therefore, under future land-use change, the model predicted brook trout populations increasing slightly across the Muskegon system (Figure 4.3). However, under this scenario the model also predicted a decrease of percent chance of occurrence by 19% in Bigelow Creek (Table 4.2A) due to forest reduction and subsequent stream reclassification from terminal node 2 (FO 0.84) to 1 (FO 0.51).

A mean July water temperature value of over 19.4 °C resulted in classification of streams into terminal node 6 (FO 0.07). Therefore, in both of the temperature-change scenarios, the model predicted that brook trout prevalence was drastically cut through the 21st century (Figure 4.3). Eventually, whether the temperature change was slow or fast, the model predicted virtual eradication of brook trout . The exception to this was the Cedar River, which was cold enough to withstand the water temperature increase in the slow temperature-change scenario (at least through 2100) (Table 4.2A).

## Brown Trout

The models predicted that past and future land-use changes only cause minor fluctuations of brown trout population in the Muskegon watershed (Figure 4.3). However, in the future climate warming scenarios, the model predicted an eventual shift in stream classification to terminal node 4 (FO 0.16) and 5 (FO 0.36) because most streams increased in temperature above 20.2 °C (Appendix B). By 2100, brown trout in Bigelow Creek and the Cedar River were predicted to decline rapidly under the high temperature-change scenario as higher temperatures combined with deforestation resulted in streams classified into terminal node 4 (FO 0.16) (Table 4.2B). However, streams with greater than 30% forest land-cover in the watershed (terminal node 5, FO 0.36) have some potential to hold brown trout despite the high water temperatures. Therefore, brown trout across the Muskegon system were predicted decrease overall but maintain populations in the lower Muskegon, Clam, and Middle Branch Rivers (Table 4.2B).

### Rainbow Trout

The rainbow trout model had a high frequency of occurrence for streams with a July mean water temperature less than 19.7 °C and without a dam blocking passage to a Great Lake (terminal node 1, FO 0.69) (Appendix C). In 1900, the models predicted that 78% of the Muskegon River system was classified into this category. In 1906, the Croton Dam was built on the main branch of the Muskegon River, and subsequently the percent chance of species occurrence in the Muskegon River system was reduced by about 40% (Figure 4.3). The Middle Branch and Clam Rivers, which are above Croton dam, drop 50% and 30% respectively in their percent chance of presence during this time (Table 4.2C).

Predictions made under the baseline future scenario indicated that the percent chance of rainbow trout occurrence in lower Muskegon River decreased in half due to reductions in baseflow caused by increased urbanization and subsequent stream reclassification from terminal node 7 (FO 0.21) to 5 (FO 0.09) (Table 4.2C). The increase of temperature in the warming scenarios resulted in reductions similar to brook trout. Bigelow Creek maintained a 70% percent chance of presence per kilometer until 2040 in the fast temperature-change scenario and until 2070 in the slow temperature-change scenario, due to its cold water (Table 4.2C). Rainbow trout started to decrease in the Cedar by 2020, stabilized until 2070, and then declined rapidly. By 2100, under the fast temperature-change scenario, rainbow trout was virtually gone from the Muskegon River system (Table 4.2C, Figure 4.3). In the slow temperature-change scenario, the species was able to maintain a presence below the Croton dam in 2100, but was on a trajectory towards extirpation by 2130.

### Chinook and Coho Salmon

Dam location was very important for both of these species: the best Chinook streams (terminal node 2, FO 0.39) (Appendix D) and best Coho streams (terminal node 1, FO 0.33) (Appendix E) were only found below Croton Dam. Upstream from Croton Dam, the Chinook model predicted a low chance of presence in streams within a kilometer upstream of lakes greater than .04 $km^2$ (10 acres) (terminal 4, FO 0.14). Only four river reaches upstream from the Croton Dam have a chance of Coho presence, and they have a low probability of occurrence (terminal node 5, FO 0.15).

In the baseline future scenario, Chinook was predicted to disappear by 2100 in both Cedar Creek and Bigelow Creek due to the decrease of forest land-cover in their watersheds and subsequent switch from terminal node 2 (FO 0.39) to terminal node 1 (FO 0.0) (Table 4.2D). The lower Muskegon River maintained a good Chinook population throughout the predicted years. Temperature-change scenarios were not applied to this species because the classification tree model did not include a temperature variable.

Future land-use changes are not expected to affect the Coho. An increase of water temperature above 18.0 °C, however, reclassified streams below Croton Dam from terminal node 1 (0.33) to terminal node 2 (0.09) in the global warming scenarios. Bigelow and Cedar Creeks are predicted to lose nearly all potential for Coho by 2100 due to the future temperature increase (Table 4.2D).

### Smallmouth and Largemouth Bass

While variables affected by land-use change (total phosphorus and ten percent exceedence flow yield) were included in smallmouth bass and largemouth bass models, they did not change enough throughout the years of model application to change occurrence results from pre-settlement levels. Therefore, from 1830-2100 under land-use change only, the models did not predict that the distribution of these two species in the Muskegon River system would change (Figure 4.3). During this time period, the model predicted smallmouth bass to be found in warm, large rivers, such as the lower Muskegon River and in pieces of the Clam River (Appendix F). Largemouth bass was predicted to be in these streams as well as in cool and warm-water rivers (>18.9 °C) within 20 km of ponds and lakes (Appendix G).

Under the climate warming scenarios, smallmouth bass were able to move into smaller streams that used to be too cold to support them. By 2100, across the watershed, the percent chance of smallmouth presence was predicted to increase by 8% under the fast temperature-change scenario (Table 4.2E). The percent chance of smallmouth presence in the Clam River was expected to increase by 19% by 2100, while the percent chance of smallmouth presence in the colder streams of Bigelow Creek and Middle Branch River just started to increase upon reaching 2100 (Figure 4.3E). If the models were run through 2200, the predictions would show smallmouth bass widely prevalent throughout the watershed.

By 2100, the model predicted an increase across the watershed in the percent chance of largemouth bass occurrence by 12% for the fast change-temperature scenario and 10% for the slow-change scenario (Figure 4.3F). The model predicted an approximately 25% increase in the probability of largemouth presence for Bigelow Creek and Middle Branch, and a very small increase for Cedar Creek. The probability of largemouth presence in the Clam or the lower Muskegon Rivers did not increase with temperature change in the future scenarios, since in 2001 these streams already had a July mean water temperature greater than 18.9 °C, the temperature threshold identified by the largemouth bass model (Appendix G).

### Northern Pike

The northern pike model had a high frequency of occurrence of the species in streams with a July mean water temperature greater than 21.9 °C (terminal node 6, FO 0.74) (Appendix H). As the Muskegon is a largely cold and cool-water system, from the pre-settlement era through 2001 the only river reaches with this high frequency of occurrence were pieces of lower Muskegon River (Table 4.2H). In the baseline future scenario, it was predicted that northern pike distribution would not change (Figure 4.3).

Water temperature increases in the fast temperature-change future scenarios, though, caused the percent chance of northern pike to increase by 22% across the entire Muskegon system (Figure 4.3). Individual streams vary: the lower Muskegon River remained unchanged because its temperature was already above the 21.9 °C threshold given by the classification model (Appendix H). Bigelow Creek remained unchanged due to its very low water temperature, and Cedar Creek only increased by 15% percent chance of presence in the fast temperature-change scenario. The Clam River, however, had July mean water temperatures very close to 21.9 °C, and increases in water temperature caused an increase of nearly 50% in percent chance of northern pike for both climate-warming scenarios. The Middle Branch, which has a water temperature between the Clam and Cedar, increased by 30% in percent chance of northern pike occurrence by 2070 in the fast-change temperature scenario.

### Walleye

The walleye model has a frequency of occurrence of 0.57 in streams with a catchment area greater than 656 kilometers and with less than 8.5% of urbanization in the watershed (terminal node 4) (Appendix I). From the pre-settlement era until 2030, the main branch of the Muskegon River was the only stream that met this criterion. However, due to expected future urban expansion, urbanization land-cover in the watershed of the lower Muskegon River was predicted to be greater than 8.5% by 2040, and the rest of the main branch Muskegon River, above and below Croton Dam, was expected to be above 8.5% by 2070. This change reclassified these streams from terminal node 4 (FO 0.57) to terminal node 5 (FO 0.26); therefore, the model predicted the prevalence of walleye was cut in half by 2070 (Table 4.2I, Figure 4.3). Temperature-change scenarios were not applied to this species because the walleye classification tree model did not include a temperature variable.

## Discussion

### Model Application

Across the Muskegon River system, the classification tree models predicted substantial changes in the structure of the fish community by 2100. Under land-use change scenarios, models predicted the decline of walleye and Chinook salmon across the system. Under future climate change scenarios, models predicted decreases of Coho salmon, brook, brown, and rainbow trout, and increases of smallmouth bass, largemouth bass, and northern pike.

There was spatial variance on the overall effects of the system; some streams were predicted to change more, and others changed less. In the high temperature-change scenario, Bigelow Creek lost virtually all brook, brown, and rainbow trout, Chinook and Coho Salmon, but gained largemouth bass and did not warm up enough to gain northern pike. Cedar Creek was predicted to respond like Bigelow Creek, but due to higher initial temperatures was able to gain northern pike. The Middle Branch and Clam River lost brook trout but were expected to maintain small populations of brown trout due to high amounts of forest in their watersheds. Both of these rivers were expected to develop substantial populations of northern pike, smallmouth, and largemouth bass. The lower Muskegon River saw declines of walleye and Chinook due to increased urbanization throughout the watershed

Changes in temperature were predicted to cause much greater shifts in fish occurrence than changes in land-use. Typically, the water temperature variable was brought into the classification tree models earlier in model formulation than land-use, influencing a greater number of observations and indicating its greater relative importance. The dominant change in the Muskegon watershed was that the system was predicted to switch from a system dominated by cold-water fish to a system dominated by warm-water fish.

However, adjusting air and water temperature is a global issue and beyond the power of fisheries managers. For these models to be applied in the real world, such as being used to prevent the predictions of this study from coming true, managers need to concentrate on the habitat variables that can be altered. According to these models, increasing forest size, limiting urban areas, and decreasing agricultural land can increase the potential of fish to live in the Muskegon River system. In addition, the phosphorus, flow, and stream power variables in our models are predicted from regression equations based on land-use and can be controlled by urban and agriculture levels (Brenden et al. 2006).

Changes in the habitat variables highlighted by our models do not guarantee the species can live in the manipulated area, yet such changes at least allow for increased potential of species establishment. For example, rainbow trout is found in 26% of streams that have a water temperature less than 19.7 °C, are above a dam, have a ninety percent exceedence flow yield greater than 0.0043, and have greater than 19.5% agriculture in the watershed (Appendix C, terminal node 4). Reducing the agriculture in the watershed to a level less than 19.5% would place those streams in terminal node 3, for which the chance of rainbow trout presence is 49%. According to the model, if agriculture is reduced in the watersheds of these streams, the likelihood of rainbow trout presence will almost double.

Interpreting and applying the models in this manner works for other species as well. In streams greater than 20.2 °C , brown trout is only found in 6% of streams when there is less than 30% forest in the watershed. However, brown trout is found in 29% of these warm streams when the amount of forest in the watershed is greater than 30%. Increasing forest in the watershed to levels above 30% can quadruple the percent chance of finding brown trout in warm streams. Walleye is only found in 26% in large streams (catchment area > 650 square kilometers) with watershed urbanization greater than 8.5%, but across Michigan walleye is found in 57% of large streams with less than 8.5% urbanization. Keeping urbanization levels below 8.5% is essential for maintaining walleye. Chinook salmon, Coho salmon, and rainbow trout are far more likely to be found in streams directly connected to the Great Lakes than in streams separated from the Great Lakes by a dam. A continued emphasis on removing dams, especially those that are located near the Great Lakes, is important for managers who are interested in maintaining and expanding salmon populations.

### Classification tree justification

Classification trees are built through brute-force computer algorithms. For every variable, the computer divides the data into two groups and compares the frequency of the target classes in both groups. It does this for every possible split in the variable, splitting the data into two groups, one observation at a time. The final split that the computer chooses depends on the exact splitting rule the user picks, but in general the split chosen is the one where the two groups have the lowest amount of diversity possible for the predicted classes. Clearly the computer cannot possibly be using ecological mechanisms to determine the shape of the tree; the procedure is simply a pattern processor and has no ability to understand what is really happening in nature.

However, despite being a brainless process, this algorithm produces a tree that is both ecological meaningful and accurate. The models used in this study are able to accurately predict the presence or absence of a fish in a stream about 75% of the time (Chapter 3). For most of the nodes in the trees (though certainly not all), the variable chosen and the split made in that variable are consistent with our understanding of that species' physiological needs (e.g. temperature) or usual location in the landscape (e.g. distance from a lake).

Conceptually, the classification tree treats species as if they were constrained to live within certain variable ranges. Every split within the tree marks either a lower or upper bound of the range for a particular habitat variable. Usually, only one end of the range is recorded into the tree. For example, the Habitat Suitability Index for brook trout reports that the species is constrained to temperatures between 0 and 24 °C (Raleigh 1982). Our brook trout model creates a cut value at 19.4 °C, showing the upper endpoint of the temperature range (Appendix A). However, because our data for stream temperatures never goes below about 15°C, the model does not show the lower range boundary at all. For land-use data, a split creates a range of habitat from that split value to 100% or 0%. In fact, for any variable, the habitat range created from a single split goes from the split value to either the minimum or maximum value of that variable in the dataset. The exception to this would be on the occasion when the model includes two or more splits of a single variable, which would narrow the habitat range to a portion of the data. Since a classification tree model identifies of a series of habitat ranges, the pieces of the tree are conceptually very similar to a quantitative version of Hutchinson's (1957) n-dimensional niche- the habitat space in which a species is able to maintain a population. For example, a terminal node classified as "present", with three habitat variable splits above it, represents a potential 3-dimensional habitat space.

As the classification models used in this study give an estimate of the niche, we can conclude these models will accurately predict fish distributions whether they predict the occurrence of fish in the year 2001 or some year in the future, assuming that the niche of the fish does not change in the future and that real-world changes in the predictor variables will indeed match the variable changes given in the possible future scenarios.

## Abiotic Filters

The concept of the niche relates well to the abiotic filter framework (Tonn et al. 1990, Keddy 1992, Poff 1997), which explains that there are a series of filters, existing on different scales, which must be passed in order for a species to be present in any particular place. This framework gives levels of importance to the dimensions of the niche. The models used in this study quantitatively identify some of the coarse filters; for example, to pass the coarsest filter, brook trout requires water less than 19.4 °C. After fulfilling this habitat requirement, the fish moves down the classification tree to slightly finer filters; the brook trout is more likely to be found in low phosphorus streams with a high amount of forest in the watershed. Meeting a species' coarse scale "filter requirements” in an area does not guarantee that the species can live there, but failure to meet the requirements can give a convincing reason for why the species is *not* there.

The finest scale filters are those that these models do not specifically address, such as microhabitat, species interactions, and food web limitations. For example, Hanchin et al. (2007) reported a dramatic decline in walleye in the Muskegon River due to alewives feeding on walleye eggs and fry, but our models do not have the capability to address this issue. However, these biotic interactions do indeed affect the model because these variables have a part in controlling the training presence/absence data. Yet because they are not included as predictors, the model attempts to account for the variation left unexplained using the variables that are included. Therefore, the models will inherently contain error as they will be unable to fully explain the data since the variables we do include are not fully correlated with the important variables we do not have.

## Model Limitations

As with all models, there are limitations to the predictions that must be recognized. For example, in the two global warming future scenarios, I only altered mean annual air temperature and stream temperature. However, climate change is expected to affect many of the model input variables, such as total phosphorus and stream exceedence flows, which rely on water quantity and water quality (Regier and Meisner 1990). It is expected that changes in water quality and quantity will affect future fish distributions, but these changes are difficult to predict and apply in our models. Because changes in water quantity and water quality due to climate change are generally thought to be negative on fish, these predictions may be best-case scenarios (Schindler 2001).

All models in this study were built with present day temperature data. Typically in these models, warm-water fish have a lower bound on temperature but not an upper bound; for example, smallmouth bass is unlikely to be found in water less than 21.3 °C (Appendix F), but does not have an upper bound in the model. In 2001, the maximum July mean water temperature of Michigan streams is around 25 °C, which is below the maximum temperature a smallmouth bass can tolerate (approximately 32 °C, Edwards et al. 1983), and therefore due to how classification trees are built it was impossible to have an upper bound temperature. In the future predictions, smallmouth bass fish are predicted to live in streams no matter how hot they get; the models lose realism when applying them to water that due to global warming is outside the range of the temperature on which it was created. Having upper temperature bounds on warm-water species may be more realistic for future scenarios in which stream temperatures could become quite hot.

To determine the effect of temperature and warming effects across the Muskegon watershed, analyses were based on stream length. However, headwater streams one kilometer long and main stem river one kilometer long are obviously much different in overall stream size. Stream length was used instead of a more informative measurement such as stream area due to difficulties in obtaining stream width. Stream length is easily obtained with GIS. Potentially, due to this problem the predicted occurrence of larger stream fish such as walleye will be lower than reality and the predicted occurrence of headwater fish such as brook trout will be too large in comparison. Overall, this problem was minor as the overall trends for a particular species would not change despite the measurement used.

## Conclusion

The predictions given in this study indicate that the Muskegon River system will shift from cold-water fish communities to warm-water communities during the 21st century given temperature increases and business-as-usual land development. Future predictions such as this are useful because they provide both a warning and an incentive for action. The fish models indicate that landscape-scale habitat and disturbance can have both positive and negative effects on any particular species; a clear task of managers is to both restore and maintain stream and watershed habitat accordingly to maximize species potential and minimize species risk.

Table 4.1. List of habitat and land-use stressor variables used in the creation of the presence/absence models for Michigan stream fishes. The descriptive statistics summarize the entire Michigan stream population.

| Variable Code | Variable Description | Unit | Min | Max | Mean |
|---|---|---|---|---|---|
| *Temperature* | | | | | |
| WATER_TEMP | Water temperature, predicted July mean | Celsius | 12.3 | 26.2 | 19.5 |
| WT_MAAT | Mean annual air temperature | Celsius | 3.7 | 9.8 | 7.3 |
| *Position in Catchment* | | | | | |
| CATCHAREA | Area of the watershed | km$^2$ | 0.72 | 14103.5 | 721 |
| *Connectivity* | | | | | |
| UP_POND | Distance upstream to closest pond >=5 acres | meters | 0 | 57566.4 | 8948 |
| DOWN_POND | Distance downstream to closest pond >=10 acres or Great Lake | meters | 0 | 195470.1 | 29732.2 |
| LINKDCATCH | Distance from downstream reach with 10% >= catchment area than target reach | meters | 0 | 58851 | 2871 |
| DOWN_LENGTH | Distance to Great Lake from downstream end of reach | meters | 0 | 130093.1 | 31886.8 |
| *Geology/Hydrologic* | | | | | |
| WT_FINE | Fine-grain surficial geology - percentage of watershed | % | 0 | 1 | 0.11 |
| WT_COARSE | Coarse-grain surficial geology- percentage of watershed | % | 0 | 1 | 0.65 |
| TEN_YIELD | 10% exceedence flow yield | cms/km$^2$ | 0.0075 | 0.0416 | 0.0186 |
| NINETY_YIELD | 90% exceedence flow yield | cms/km$^2$ | 0.0001 | 0.0264 | 0.0039 |
| GRADIENT | Channel gradient | unitless | 0 | 0.0288 | 0.0026 |
| TEN_POWER | High flow-based specific power | cms/km$^2$ | 0 | 0.0073 | 0.0005 |
| NINETY_POWER | Summer flow-based specific power | cms/km$^2$ | 0 | 0.0021 | 0.0001 |
| *Land-use* | | | | | |
| WT_FOREST | Forest Land cover - percentage of watershed | % | 0.02 | 0.95 | 0.41 |
| WT_WETLAND | Wetland land cover- percentage of watershed | % | 0 | 0.56 | 0.15 |
| WT_AGR | Agricultural land-use- percentage of watershed | % | 0 | 0.95 | 0.28 |
| WT_URBAN | Urban land-use- percentage of watershed | % | 0 | 0.64 | 0.05 |
| *Water Quality* | | | | | |
| TOTAL_P__PPM | Total phosphorus, predicted | ppm | 0.01 | 0.25 | 0.05 |

Table 4.2. Percent chance of species occurrence for any stream kilometer in the Muskegon system, the lower Muskegon, and other study units. Species included are A) Brook trout, B) Brown trout, C) Rainbow trout, D) Chinook salmon, E) Coho Salmon, F) Smallmouth bass, G) Largemouth bass, H) Northern pike, and I) Walleye. With the exception of Chinook and walleye, three scenarios were run for each species: 1) baseline, 2) slow temperature change, and 3) fast temperature change. Chinook and walleye do not have temperature variables in the model and so were only predicted for the baseline scenario.

A. Brook Trout

| | Muskegon System | | | Lower Muskegon | | | Bigelow Creek | | | Cedar River | | | Clam River | | | Middle Branch River | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario: | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Year 1830 | 58 | | | 8 | | | 81 | | | 68 | | | 22 | | | 70 | | |
| 1900 | 34 | | | 8 | | | 81 | | | 48 | | | 22 | | | 60 | | |
| 1930 | 34 | | | 8 | | | 81 | | | 48 | | | 22 | | | 60 | | |
| 1960 | 36 | | | 8 | | | 81 | | | 48 | | | 22 | | | 60 | | |
| 1978 | 31 | | | 8 | | | 81 | | | 48 | | | 22 | | | 77 | | |
| 2001 | 40 | | | 8 | | | 81 | | | 50 | | | 16 | | | 41 | | |
| 2010 | 45 | 44 | 42 | 8 | 8 | 8 | 81 | 81 | 81 | 42 | 27 | 27 | 22 | 22 | 22 | 60 | 51 | 51 |
| 2040 | 47 | 40 | 30 | 8 | 8 | 8 | 81 | 81 | 81 | 45 | 37 | 45 | 18 | 14 | 18 | 60 | 35 | 25 |
| 2070 | 49 | 37 | 21 | 8 | 8 | 8 | 62 | 62 | 15 | 49 | 41 | 26 | 18 | 14 | 8 | 57 | 10 | 8 |
| 2100 | 50 | 27 | 9 | 8 | 8 | 8 | 62 | 29 | 8 | 49 | 41 | 8 | 16 | 8 | 8 | 54 | 8 | 8 |

B. Brown Trout

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1900 | 44 | | | 29 | | | 62 | | | 64 | | | 24 | | | 21 | | |
| 1930 | 44 | | | 29 | | | 62 | | | 64 | | | 31 | | | 21 | | |
| 1960 | 44 | | | 29 | | | 62 | | | 64 | | | 32 | | | 21 | | |
| 1978 | 46 | | | 29 | | | 62 | | | 64 | | | 32 | | | 21 | | |
| 2001 | 47 | | | 29 | | | 66 | | | 62 | | | 37 | | | 21 | | |
| 2010 | 46 | 45 | 45 | 29 | 29 | 29 | 68 | 68 | 68 | 64 | 64 | 64 | 37 | 38 | 33 | 21 | 21 | 21 |
| 2040 | 46 | 45 | 46 | 29 | 29 | 29 | 68 | 68 | 68 | 64 | 56 | 55 | 37 | 33 | 30 | 21 | 21 | 21 |
| 2070 | 48 | 45 | 41 | 29 | 29 | 29 | 68 | 68 | 56 | 66 | 53 | 53 | 39 | 29 | 29 | 21 | 29 | 29 |
| 2100 | 49 | 43 | 33 | 29 | 29 | 29 | 68 | 68 | 19 | 66 | 53 | 12 | 48 | 29 | 29 | 21 | 29 | 29 |

Table 4.2, continued.

C. Rainbow Trout

| | Muskegon System | | | Lower Muskegon | | | Bigelow Creek | | | Cedar River | | | Clam River | | | Middle Branch River | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario: | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Year 1900 | 56 | | | 21 | | | 69 | | | 67 | | | 25 | | | 69 | | |
| 1930 | 24 | | | 21 | | | 69 | | | 67 | | | 11 | | | 18 | | |
| 1960 | 24 | | | 21 | | | 69 | | | 68 | | | 11 | | | 18 | | |
| 1978 | 24 | | | 21 | | | 69 | | | 68 | | | 11 | | | 18 | | |
| 2001 | 26 | | | 50 | | | 69 | | | 68 | | | 20 | | | 18 | | |
| 2010 | 23 | 23 | 23 | 9 | 9 | 9 | 69 | 69 | 69 | 67 | 67 | 56 | 11 | 11 | 11 | 18 | 18 | 18 |
| 2040 | 23 | 25 | 23 | 9 | 9 | 9 | 69 | 69 | 69 | 67 | 65 | 56 | 11 | 11 | 11 | 18 | 14 | 18 |
| 2070 | 23 | 20 | 16 | 9 | 9 | 9 | 69 | 69 | 35 | 67 | 55 | 55 | 11 | 10 | 9 | 18 | 11 | 9 |
| 2100 | 23 | 17 | 11 | 9 | 9 | 9 | 69 | 39 | 9 | 67 | 55 | 9 | 11 | 9 | 9 | 18 | 9 | 9 |

D. Chinook Salmon

| | | Muskegon System | Lower Muskegon | Bigelow Creek | Cedar River | Clam River | Middle Branch River |
|---|---|---|---|---|---|---|---|
| Year | 1970 | 5 | 40 | 39 | 22 | 0 | 1 |
| | 1978 | 5 | 40 | 39 | 25 | 0 | 1 |
| | 2001 | 6 | 40 | 39 | 22 | 0 | 0 |
| | 2010 | 5 | 40 | 39 | 22 | 0 | 1 |
| | 2040 | 5 | 40 | 39 | 22 | 0 | 1 |
| | 2070 | 4 | 40 | 22 | 4 | 0 | 1 |
| | 2100 | 3 | 40 | 4 | 4 | 0 | 1 |

Table 4.2, continued.

E. Coho Salmon

| | | Muskegon System | | | Lower Muskegon | | | Bigelow Creek | | | Cedar River | | | Clam River | | | Middle Branch River | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario: | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Year | 1970 | 5 | | | 9 | | | 33 | | | 28 | | | 2 | | | 0 | | |
| | 1978 | 5 | | | 9 | | | 33 | | | 28 | | | 2 | | | 0 | | |
| | 2001 | 5 | | | 9 | | | 33 | | | 28 | | | 2 | | | 0 | | |
| | 2010 | 5 | 5 | 5 | 9 | 9 | 9 | 33 | 33 | 26 | 28 | 28 | 28 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 2040 | 5 | 4 | 5 | 9 | 9 | 9 | 33 | 19 | 15 | 28 | 28 | 28 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 2070 | 5 | 4 | 3 | 9 | 9 | 9 | 33 | 12 | 9 | 28 | 16 | 9 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 2100 | 5 | 3 | 3 | 9 | 9 | 9 | 33 | 9 | 9 | 28 | 9 | 9 | 2 | 2 | 2 | 2 | 2 | 2 |

F. Smallmouth Bass

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1830 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 1900 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 1930 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 1960 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 1978 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 2001 | 18 | | | 81 | | | 8 | | | 8 | | | 47 | | | 8 | | |
| 2010 | 18 | 18 | 19 | 81 | 81 | 81 | 8 | 8 | 8 | 8 | 8 | 8 | 47 | 47 | 58 | 8 | 8 | 8 |
| 2040 | 18 | 20 | 20 | 81 | 81 | 81 | 8 | 8 | 8 | 8 | 8 | 8 | 47 | 63 | 63 | 8 | 8 | 8 |
| 2070 | 18 | 20 | 23 | 81 | 81 | 81 | 8 | 8 | 8 | 8 | 8 | 9 | 47 | 63 | 64 | 8 | 8 | 11 |
| 2100 | 18 | 22 | 26 | 81 | 81 | 81 | 8 | 8 | 19 | 8 | 9 | 9 | 47 | 64 | 66 | 8 | 9 | 14 |

Table 4.2, continued.

### G. Largemouth Bass

| | Muskegon System | | | Lower Muskegon | | | Bigelow Creek | | | Cedar River | | | Clam River | | | Middle Branch River | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario: | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Year 1830 | 28 | | | 53 | | | 18 | | | 28 | | | 41 | | | 21 | | |
| 1900 | 29 | | | 55 | | | 18 | | | 28 | | | 42 | | | 24 | | |
| 1930 | 29 | | | 55 | | | 18 | | | 28 | | | 44 | | | 23 | | |
| 1960 | 29 | | | 55 | | | 18 | | | 28 | | | 44 | | | 23 | | |
| 1978 | 29 | | | 55 | | | 18 | | | 28 | | | 44 | | | 23 | | |
| 2001 | 28 | | | 53 | | | 18 | | | 28 | | | 41 | | | 23 | | |
| 2010 | 29 | 30 | 30 | 53 | 53 | 53 | 18 | 18 | 18 | 28 | 28 | 28 | 43 | 47 | 47 | 23 | 35 | 44 |
| 2040 | 29 | 33 | 35 | 53 | 53 | 53 | 18 | 18 | 18 | 28 | 28 | 28 | 42 | 46 | 46 | 23 | 49 | 57 |
| 2070 | 29 | 36 | 39 | 53 | 53 | 53 | 18 | 40 | 42 | 28 | 28 | 32 | 41 | 46 | 46 | 23 | 57 | 57 |
| 2100 | 28 | 38 | 40 | 53 | 53 | 53 | 18 | 42 | 43 | 28 | 30 | 32 | 41 | 46 | 46 | 23 | 57 | 57 |

### H. Northern Pike

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1830 | 21 | | | 74 | | | 13 | | | 14 | | | 19 | | | 26 | | |
| 1900 | 22 | | | 74 | | | 13 | | | 14 | | | 22 | | | 27 | | |
| 1930 | 22 | | | 74 | | | 13 | | | 14 | | | 22 | | | 27 | | |
| 1960 | 22 | | | 74 | | | 13 | | | 14 | | | 22 | | | 27 | | |
| 1978 | 24 | | | 74 | | | 13 | | | 14 | | | 23 | | | 27 | | |
| 2001 | 22 | | | 74 | | | 13 | | | 14 | | | 25 | | | 26 | | |
| 2010 | 22 | 22 | 22 | 74 | 74 | 74 | 13 | 13 | 13 | 14 | 14 | 14 | 19 | 19 | 19 | 26 | 26 | 26 |
| 2040 | 22 | 23 | 25 | 74 | 74 | 74 | 13 | 13 | 13 | 14 | 14 | 16 | 19 | 19 | 37 | 26 | 26 | 26 |
| 2070 | 22 | 24 | 30 | 74 | 74 | 74 | 13 | 13 | 13 | 14 | 16 | 27 | 19 | 37 | 63 | 26 | 26 | 31 |
| 2100 | 22 | 28 | 42 | 74 | 74 | 74 | 13 | 13 | 13 | 14 | 16 | 27 | 19 | 60 | 69 | 26 | 26 | 63 |

Table 4.2, continued.

I. Walleye

| | Muskegon System | Lower Muskegon | Bigelow Creek | Cedar River | Clam River | Middle Branch River |
|---|---|---|---|---|---|---|
| Year 1830 | 8 | 57 | 2 | 2 | 6 | 3 |
| 1900 | 8 | 57 | 2 | 1 | 11 | 5 |
| 1930 | 8 | 57 | 2 | 1 | 11 | 5 |
| 1960 | 8 | 57 | 2 | 1 | 7 | 5 |
| 1978 | 8 | 57 | 2 | 1 | 7 | 5 |
| 2001 | 8 | 57 | 2 | 2 | 6 | 6 |
| 2010 | 8 | 57 | 2 | 1 | 7 | 6 |
| 2040 | 7 | 35 | 2 | 1 | 5 | 6 |
| 2070 | 4 | 26 | 2 | 2 | 5 | 5 |
| 2100 | 4 | 26 | 2 | 2 | 5 | 5 |

Figure 4.1. The streams and rivers of the Muskegon watershed, with highlighted streams indicating the five stream study units in this analysis.

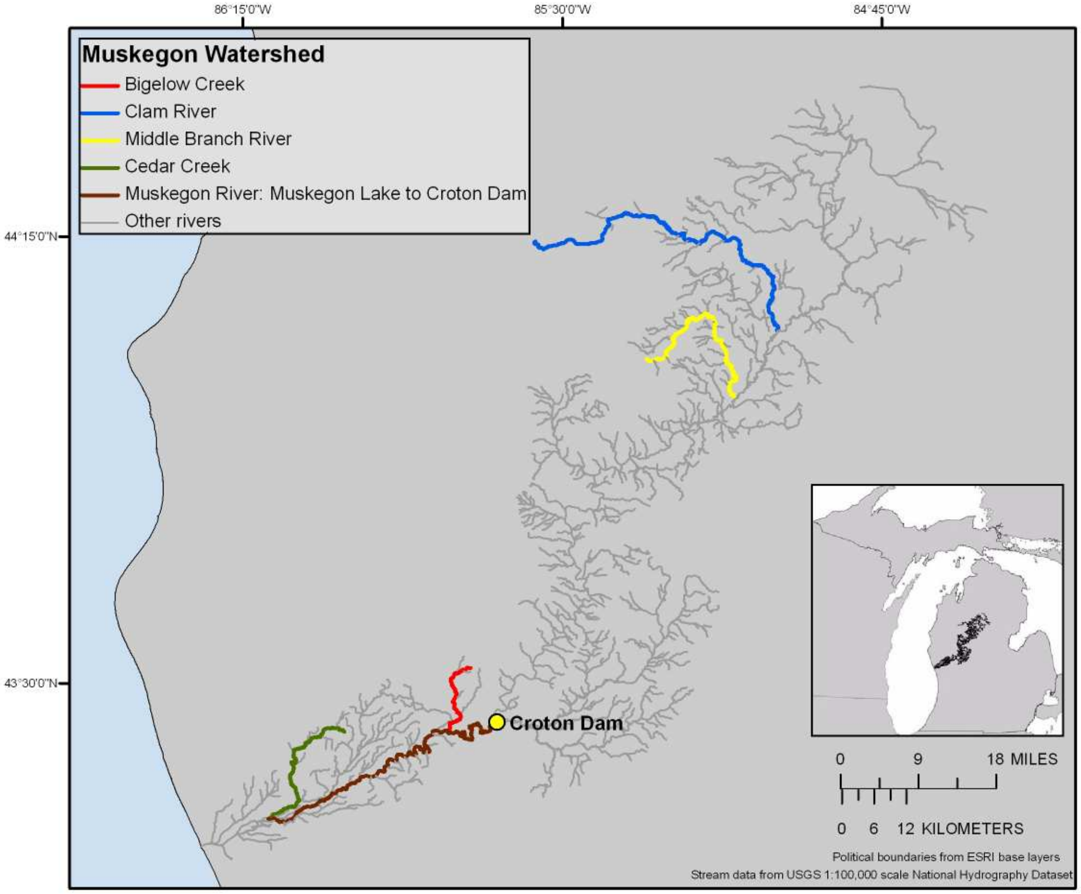


Figure 4.2. Measured land-use/cover in the Muskegon watershed, for both A) 1830 (pre-settlement) and B) 2001, and predicted land-use/cover for C) 2100.

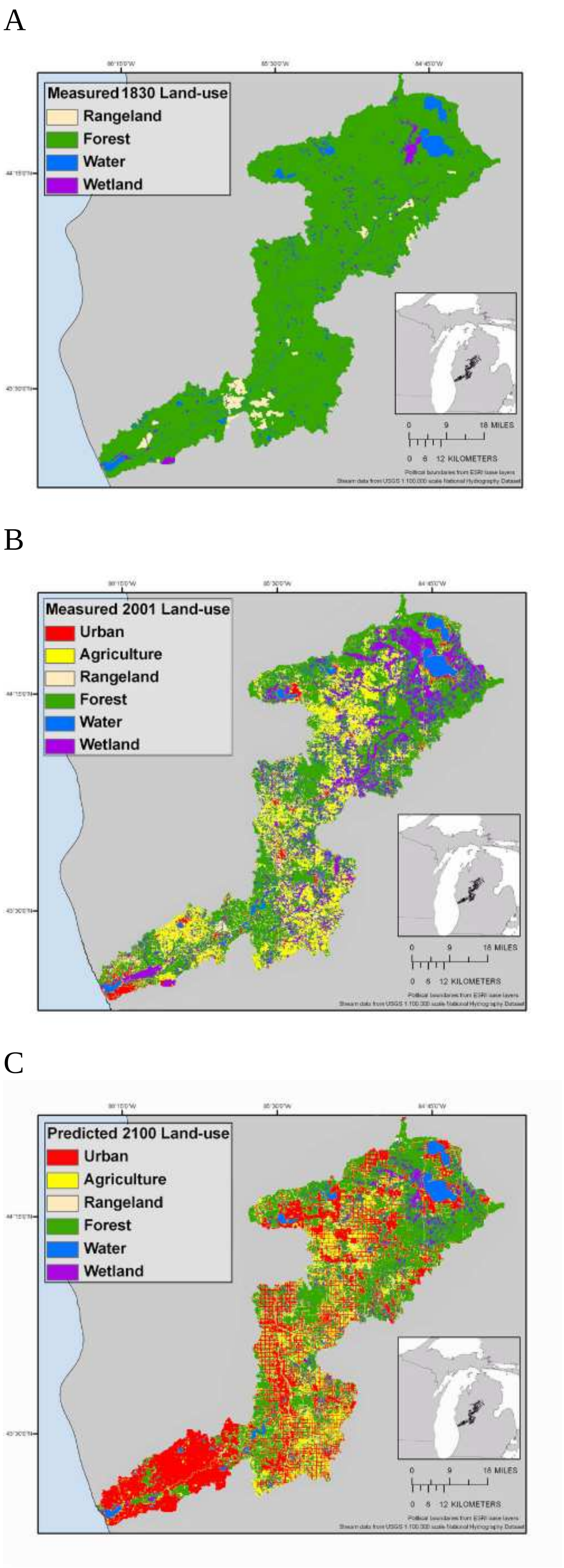
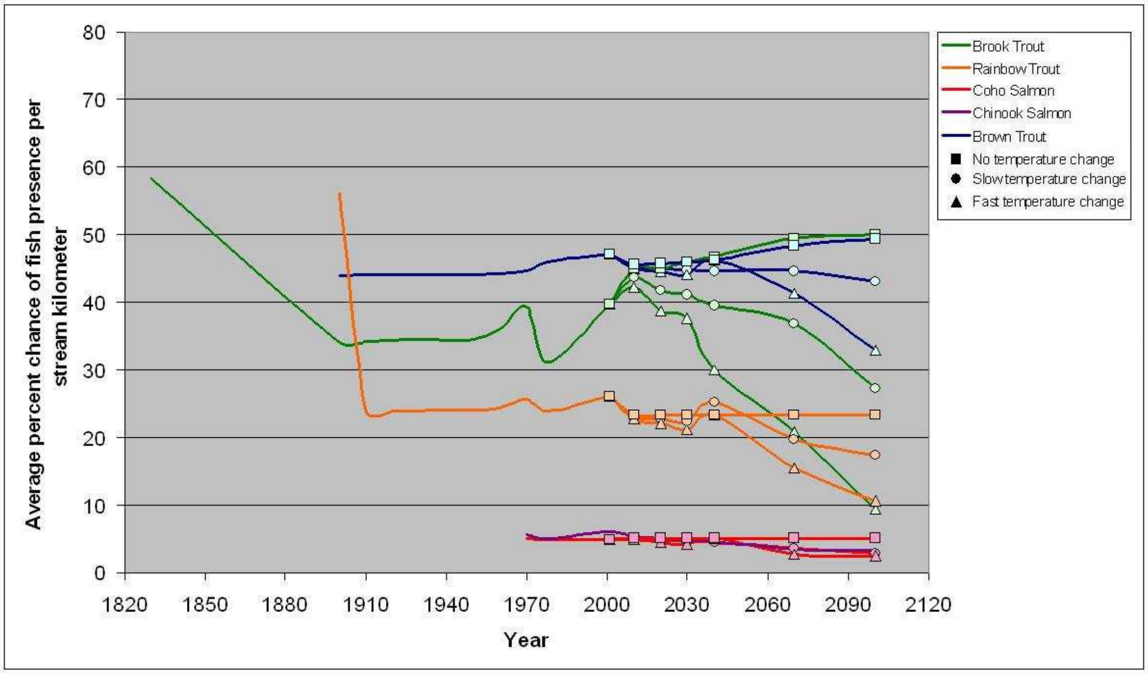
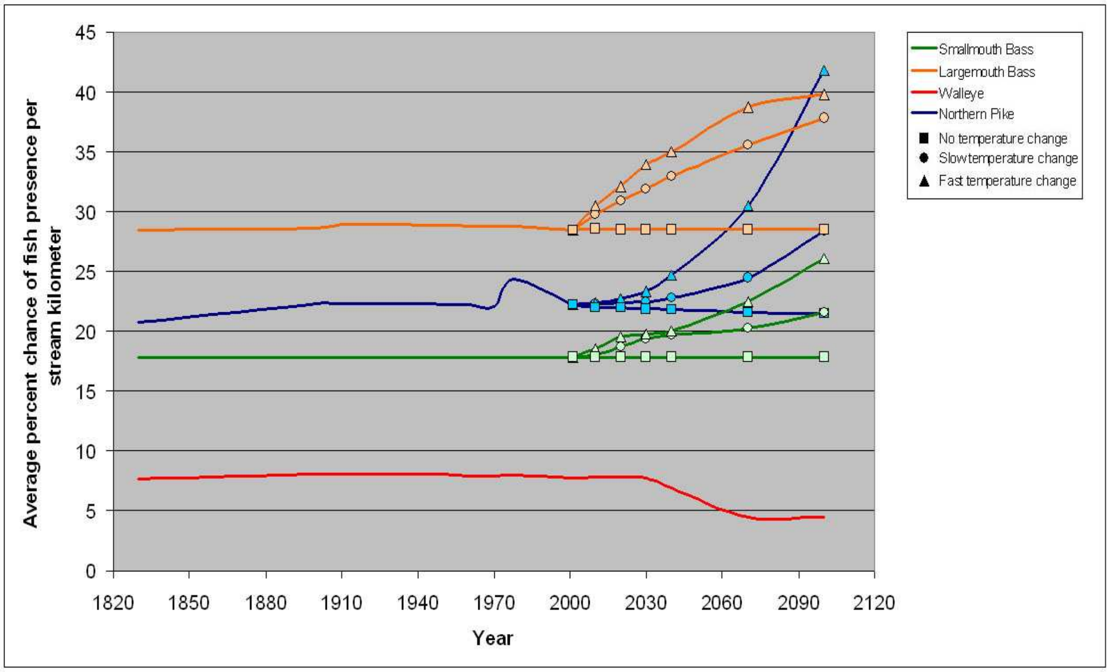


Figure 4.3. Predictions of the average chance of species presence, as weighted by stream length, for the entire Muskegon stream system. Represented here are both A) cold-water game fish, and B) warm-water game fish. Line markers indicate predictions of the species for the three future scenarios. Walleye and Chinook salmon models do not have a temperature variable and so do not have temperature-change future scenarios.

A



B



## References

Alig, R. J., J. D. Kline, and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. Landscape and Urban Planning 69:219-234.

Barker, L. S., G. K. Felton, and E. Russek-Cohen. 2006. Use of Maryland biological stream survey data to determine effects of agricultural riparian buffers on measures of biological stream health. Environmental Monitoring and Assessment 117:1-19.

Brenden, T. O., R. D. Clark, A. R. Cooper, P. W. Seelbach, L. Wang, S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49-74 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape Influences on Stream Habitat and Biological Assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Chu, C., N. E. Mandrak, and C. K. Minns. 2005. Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Candada. Diversity and Distributions 11:299-310.

Eaton, J. G., and R. M. Scheller. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. Limnology and Oceanography 41.

Edwards, E.A., G. Gebhart, and O.E. Maughan. 1983. Habitat suitability information: Smallmouth bass. US Department of Interior, Fish and Wildlife Service, FWS/OBS-82/10.36. 47 pp.

Flebbe, P. A., L. D. Roghair, and J. L. Bruggink. 2006. Spatial modeling to project southern Appalachian trout distribution in a warmer climate. Transactions of the American Fisheries Society 135:1371-1382.

Hanchin, P. A., R. P. O'Neal, R. D. Clark, and R. N. Lockwood. 2007. The walleye population and fishery of the Muskegon Lake System, Muskegon and Newayo counties, Michigan in 2002. Michigan Department of Natural Resources Fisheries Division, Special Report 40, Ann Arbor.

Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol. 1999. Effects of agriculture, urbanization, and climate on water quality in the northern Great Plains. Limnology and Oceanography 44:739-756.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22:415-427.

Jager, H. I., W. Van Winkle, and B. D. Holcomb. 1999. Would hydrologic climate changes in Sierra Nevada streams influence trout persistence? Transactions of the American Fisheries Society 128:222-240.

Keddy, P.A. 1992. A pragmatic approach to functional ecology. Functional Ecology 6:621-626.

King, J. R., B. J. Shuter, and A. P. Zimmerman. 1999. Empirical links between thermal habitat, fish growth, and climate change. Transactions of the American Fisheries Society 128:656-665.

Lyons, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. United States Department of Agriculture, Forest Service, General Technical Report NC-149.

Magnuson, J. J., J. D. Meisner, and D. K. Hill. 1990. Potential changes in the thermal habitat of Great Lakes fish after global climate warming. Transactions of the American Fisheries Society 119:254-264.

Mandrak, N. E. 1989. Potential invasion of the Great Lakes by fish species associated with climatic warming. Journal of Great Lakes Research 15:306-316.

McCauley, R. W., and D. M. Kilgour. 1990. Effect of air temperature on growth of largemouth bass in North America. Transactions of the American Fisheries Society 119:276-281.

MCGI (Michigan Center for Geographic Information). 2008. Land cover 2001 geographic theme: land cover/use. Available: www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=2&cat=Land+Cover+2001 (January 2008)

Meador, M. R., and R. M. Goldstein. 2003. Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition, and fish community structure. Environmental Management 31:504-517.

Meisner, J. D. 1990. Potential loss of thermal habitat for brook trout, due to climatic warming, in two southern Ontario streams. Transactions of the American Fisheries Society 119:282-291.

Miltner, R. J., D. White, and C. Yoder. 2004. The biotic integrity of streams in urban and suburbanizing landscapes. Landscape and Urban Planning 69:87-100.

NHD (National Hydrography Dataset), 2008. Available: http://nhd.usgs.gov/ (January 2008).

O'Neal, R. P. 1997. Muskegon River watershed assessment. Fisheries Division Special Report 19, Michigan Department of Natural Resources, Lansing.

Pijanowski, B. C., D. G. Brown, B. A. Shellito, and G. A. Manik. 2001. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems 26:553-575.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. The Annual Review of Ecology and Systematics 32:333-365.

Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. Journal of the North American Benthological Society 16:391-409.

Raleigh, R. F. 1982. Habitat suitability index models: Brook trout. FWS/OBS-82/10.24., U.S. Department of the Interior, Fish and Wildlife Service. 43 pp.

Regier, H. A., and J. D. Meisner. 1990. Anticipated effects of climate change on freshwater fishes and their habitat. Fisheries 15:10-15.

Riseng, C. M., M. J. Wiley, R. J. Stevenson, T. G. Zorn, and P. W. Seelbach. 2006. Comparison of coarse versus fine scale sampling on statistical modeling of landscape effects and assessment of fish assemblages of the Muskegon River, Michigan. Pages 555-575 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Roy, A. H., M. C. Freeman, B. J. Freeman, S. J. Wenger, J. L. Meyer, and W. E. Ensign. 2006. Importance of riparian forests in urban catchments contingent on sediment and hydrologic regimes. Environmental Management 37:523-539.

Schlinder, D. W. 1997. Widespread Effects of climatic warming on freshwater ecosystems in North America. Hydrological Processes 11:1043-1067.

Schindler, D. W. 2001. The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium. Canadian Journal of Fisheries and Aquatic Sciences 58:18-29.

Scott, W. B., and E. J. Crossman. 1973. Freshwater Fishes of Canada. Fisheries Research Board of Canada, Ottawa, Ontario.

Scott, J. B., C. R. Stewart, and Q. J. Stober. 1986. Effect of urban development on fish population dynamics in Kelsey Creek, Washington. Transactions of the American Fisheries Society 115:555-567.

Schlinder, D. W. 1997. Widespread effects of climatic warming on freshwater ecosystems in North America. Hydrological Processes 11:1043-1067.

Seelbach, P. W., and M. J. Wiley. 2005. An initial landscape-based system for ecological assessment of Lake Michigan tributaries. Pages 559-581 *in* T. Edsall and M. Munawar, editors. Aquatic Ecosystem Health and Management Society.

Snyder, C. D., J. A. Young, R. Villella, and D. P. Lemarie. 2003. Influences of upland and riparian land use patterns on stream biotic integrity. Landscape Ecology 18:647-664.

Stefan, H. G., and E. B. Preud'homme. 1993. Stream temperature estimation from air temperature. Water Resources Bulletin 29:27-45.

Stefan, H. G., X. Fang, and J. G. Eaton. 2001. Simulated fish habitat changes in North America lakes in response to projected climated warming. Transactions of the American Fisheries Society 130:459-477.

Tabit, C. R., and G. M. Johnson. 2002. Influence of urbanization on the distribution of fishes in a southeastern upper Piedmont drainage. Southeastern Naturalist 1:253-268.

Talmage, P. J., J. A. Perry, and R. M. Goldstein. 2002. Relation of instream habitat and physical conditions to fish communities of agricultural stream in the northern Midwest. North American Journal of Fisheries Management 22:825-833.

Thomson, A. M., N. J. Rosenberg, R. C. Izaurralde, and R. A. Brown. 2005. Climate change impacts for the coterminous USA: An integrated assessment. Part 2: models and validation. Climate Change 69:27-41.

Tonn, W.M., J.J. Magnuson, M. Rask, and J. Toivonen. 1990. Intercontinental comparison of small-lake fish assemblages: the balance between local and regional processes. American Naturalist 136: 345-375.

Trout Unlimited, 2008. Available: http://www.trailtotrout.com/blueribbon.html (January 2008).

Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 2001:255-266.

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

Wang, L., J. Lyons, P. W. Rasmussen, P. W. Seelbach, T. Simon, M. J. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. M. Stewart. 2003b. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, USA. Canadian Journal of Fisheries and Aquatic Sciences 60:491-505.

Weaver, L. A., and G. C. Garman. 1994. Urbanization of a watershed and historical changes in a stream fish assemblage. Transactions of the American Fisheries Society 123:162-172.

Zimmerman, J. K. H., B. Vondracek, and J. Westra. 2003. Agricultural land use effects on sediment loading and fish assemblages in two Minnesota (USA) watersheds. Environmental Management 32:93-105.

# Chapter 5

# Variation in the effect of urbanization on Michigan and Wisconsin stream fish: How can good fish communities exist in urban areas?

## Abstract

One of the primary goals in past investigations of urban stream analysis has been to understand the strength of the negative relationship between urbanization and biotic communities. However, little effort has been expended into understanding the variation that occurs around this effect; why streams with similar urban levels have fish communities of significantly different quality. In this study, I test the hypothesis that non-urban habitat features control the variance in the relationship between fish community quality and urbanization. To do this, Michigan and Wisconsin stream reaches were classified into groups based on fish community quality and amount of urbanization in their watershed and a series of univariate tests were performed to find how natural and anthropogenic features are related to fish biotic integrity. In addition, covariance structure analysis was used to provide multivariate insight into the complex relationships that control the quality of the stream fish community. Results indicated that urban streams with a higher percentage of natural land-cover in the watershed, more point source discharges, better water quality, and a close proximity to non-urbanized streams were more likely to hold higher quality fish communities.

## Introduction

Urban development damages the integrity of aquatic ecosystems by causing changes in their hydrological, chemical, and thermal properties and thereby reducing the diversity and abundance of resident organisms. In the United States, over 130,000 kilometers of streams and rivers have already been affected by urbanization, and land-use change projections predict that developed area will increase by 80% in the next 25 years (Paul and Meyer 2001; Alig et al. 2004; Walsh et al. 2005). In order to ensure long-term sustainability of stream ecosystems, it is critical for scientists and managers to understand how urbanization affects aquatic ecosystems and implement rational management programs as soon as possible.

In the past thirty years, there have been numerous investigations into how urbanization affects river ecosystems. As a result, the physical impacts of urbanization on streams are well understood (Klein 1979, Lenat and Crawford 1994; Arnold and Gibbons 1996; Booth and Jackson 1997; Paul and Meyer 2001; Konrad and Booth 2005; Walsh et al. 2005), and many studies have shown how these physical changes have consequences for a streams' biota. Several authors have reported a negative association between urbanization and fish abundance, richness, or number of intolerant fish species (Weaver and Garman 1994; Kemp and Spotila 1997; Tabit and Johnson 2002; Walters et al. 2003; Morgan and Cushman 2005). Similar results have been found for macroinvertebrates; watershed urbanization is often negatively correlated with Ephemeroptera-Plecoptera-Trichoptera (EPT) abundance and filterer, scraper, and EPT species richness (Wang and Kanehl 2003). Also, urbanization has been identified as a cause of homogenization; as disturbances create different selective pressure, generalist fish species tend to find the new conditions more suitable than fish with more specific habitat requirements, resulting in replacement of regionally distinct species with tolerant fish (Walters et al. 2003; McKinney 2006; Olden 2006; Scott 2006).

One of the primary goals in past investigations has been to document the strength of the negative impacts of urbanization on biotic communities. A typical method used in these studies was to select sampling locations to minimize variation in natural stream attributes (e.g. temperature, land-use, geology) and to maximize variation in urbanization (Wang et al. 2001; Wang et al. 2003a; Cuffney et al. 2005; Fitzpatrick et al. 2005; Meador et al. 2005; Tate et al. 2005). This enabled researchers to detect how biotic community quality changed as a function of urbanization while reducing complicating impacts of natural features on the analysis.

However, since studies have concentrated on identifying the strength of this negative "urbanization effect", little effort has been expended into understanding the variation that occurs around this effect. For example, while the relationship between urbanization and the integrity of the biotic community has been found to be strongly negative, some streams contain biotic communities of apparent high integrity while others contain degraded biotic communities despite having the same amount of urbanization within their watershed. I believe that understanding this noise (residual variation) is critical to intelligent ecosystem management. As stopping the spread of urbanization altogether is not a politically feasible or necessarily desirable goal, it is important to understand how to maximize the potential of fish communities for a given amount of urban disturbance. Understanding how good fish communities can sometimes exist even in highly urbanized streams can inform management and conservation agencies about how to adjust their practices in growing urban areas in ways that maintain good quality fish communities.

In this chapter, I tested the how the variance in the relationship between fish community quality and urbanization is controlled by non-urban habitat features. To do this, 1) I classified Michigan and Wisconsin streams into groups based on fish community quality and amount of urbanization in their watershed and performed a series of univariate tests designed to explore the relationship between natural and anthropogenic features, and fish biotic integrity and 2) used covariance structure analysis to provide insight into the complex relationships that ultimately control the quality of stream fish communities.

## Methods

### Data Description

Two regional conservation projects, the Great Lakes Aquatic GAP Project (GLGAP, GLSC 2006) and Classification and Impairment Assessment of Upper Midwestern Rivers (CIAUMR, UM 2006) provided habitat data for this study. Working in concert, these groups have established a high-resolution, GIS-linked database that contains characteristics of Michigan and Wisconsin rivers. The database was referenced to a group of ArcGIS line coverages (ESRI 2007), in which each river was divided into confluence-to-confluence reaches. Line coverages were based on the USGS National Hydrography Dataset (NHD, 2006) at the 1:100,000 scale, but were updated to provide more accurate representation of the rivers (Brenden et al. 2006). There are 31,817

Michigan stream reaches (86,983 kilometers of stream length) and 36,614 Wisconsin stream reaches (89,716 kilometers of stream length) included in the database.

For each stream reach, GLGAP and CIAUMR provided data on a variety of habitat and landscape variables (Table 5.1), including predicted July mean water temperature, predicted exceedence flows, percent of coarse surficial geology in the watershed (Chapter 3; Brenden et al. 2006), 1992 WI land-use/cover (WNDR 2007), and 2001 MI land-use/cover (Brenden et al. 2006). Land-use/cover was measured as a percentage of watershed area and riparian buffer area (30 meters to each side of the stream, for the reach of interest and all reaches upstream). Wang et al. (2007) and Jana Stewart (USGS Water Resources Division, personal communication) provided human disturbance variables representing population density, nutrient enrichment, agricultural pollution, and point source pollutants (Table 5.1). Hereafter, variables discussed in the text will be followed with the Table 5.1 variable code in parentheses.

I compiled fish community sample data from the Wisconsin Department of Natural Resources, the Michigan Department of Natural Resources, and the Michigan Rivers Inventory (Figure 5.1) (Chapter 3; Seelbach and Wiley 1997; John Lyons, Wisconsin Department of Natural Resources, personal communication). Fish data selected for this study had been collected from 1980 to 2004 and were entire community samples obtained through electroshocking or rotenone methods.

Each fish community sample was linked to the NHD and database and attributed with the habitat data discussed above. For stream reaches that had numerous sampling efforts over the years, I deleted the samples that had a lower total fish count so that a stream reach was represented by the one sampling effort that produced the most fish. This action made the assumption that the stream reach had the potential to hold the higher amount of fish, and lower fish counts were a result of disturbance or natural variation unrelated to the measured habitat factors.

I calculated a warm-water fish index of biotic integrity (IBI) for fish community samples using methods given by Lyons et al. (1992) for Wisconsin sites and Michigan Department of Environmental Quality's Procedure 51 (Grant 2002) for Michigan sites. These two procedures are comparable because they use the same fish metrics and scoring scale to calculate IBI. IBI metrics used were: number of native species, number of darter species, number of sucker species, number of intolerant species, percent of tolerant species, percent of omnivores, percent of carnivores, percent of insectivores, and percent of lithophilic spawners. Each IBI metric was scored from 0 to 10, with 0 indicating very poor fish condition and 10 indicating excellent fish condition. The metrics were added together to create a score that ranged from 0 to 90.

I also calculated a coldwater IBI for those sites with a predicted water temperature (WATER_TEMP) less than 22 °C (Lyons et al. 1996). I excluded those sites with a higher cold-water IBI than warm-water IBI from this study; these sites were considered to be cold-water streams that cannot be compared to warm-water streams due to large differences in fish communities (Lyons et al. 1992).

### Classifying Observations and Univariate Analysis

I compared total IBI score against the percent of urbanization in the stream's watershed (% URBAN) (Figure 5.2A). Similar comparisons in Wisconsin, Maryland, and Washington have indicated that streams with watershed imperviousness values less than 8-12% had fish communities ranging from a very bad quality to very good, but above this threshold degradation of the biotic community was “rapid and dramatic” (Scott et al. 1986; Wang et al. 2001; Barker et al. 2006). I found the threshold in this dataset (9% urbanization in watershed) by identifying the largest decrease of maximum total IBI score as I increased total urbanization in the watershed one percentage-point at a time. The 9% urban threshold was used to divide sampled fish communities into two groups: streams minimally affected by urbanization and streams substantially affected by urbanization.

I used a cluster analysis to identify the fish community samples in order to have groups of sites with different IBI scores but similar watershed urbanization (% URBAN). I used SPSS v.15 (SPSS 2007) to perform a k-means cluster with variables “total fish IBI score” and “% URBAN” (Figure 5.2B). Only fish community samples above the urban threshold were clustered. To increase statistical power, the elements of the clusters were manually adjusted after the clustering process in order to have groups with equal sample sizes. Throughout this study, I used cluster 1, 2, and 3 sites as examples of minimally urbanized (low) streams and compared clusters 4 and 5 as examples of highly urbanized (high) streams.

To determine how habitat differed between clusters 1, 2, and 3, I examined variables listed in Table 5.1 using a Kruskal-Wallis nonparametric test of means and a post-hoc multiple comparison Nemnyi test (Zar 1999). Similarly, I tested differences in habitat variables between cluster 4 and 5 using the Mann-Whitney U nonparametric test of means. Different tests are used because the Mann-Whitney U is more appropriate for comparisons of two groups than Kruskal-Wallis, but is not conducive for post-hoc comparisons with multiple groups. ANOVA methods could not be used on this data due to normality assumption violations; however, the Kruskal-Wallis and Mann-Whitney U are nonparametric equivalents of ANOVA. Similar analyses were also performed to understand how fish species differed between the clusters, though these results are not discussed in the text (Appendices J, K).

### Covariance Structure Analysis

While the univariate approaches above were useful for determining how habitat variation affect fish IBI given a certain range of urbanization, correlation between these habitat variables can obscure actual habitat-IBI relationship in a web of direct and indirect effects (Zorn and Wiley 2004). Therefore, I used covariance structure analysis (CSA) to take a multivariate approach in understanding how these variables relate to each other.

CSA is a powerful tool for ecological studies (Fjeld and Rognerud 1993; Wotton 1994a; Wotton 1994b; Sheldon and Meffe 1995; Issac and Hubert 2001; Riseng et al. 2004; Zorn and Wiley 2004; Infante et al. 2006; Riseng et al. 2006; Wehrly et al. 2006). In CSA, researchers use logic and prior knowledge to build a series of linear equations that represent their hypothesis about how pieces of their system are causally related. This system of equations can be represented graphically by a path diagram and is tested by determining if implied covariance relationships in the path diagram are consistent with the sample covariance matrix seen in the actual data (Bollen 1989, Wooton 1994a, Wooton 199b).

As a dependent variable in the CSA representing the impact of urbanization, I used a deviation calculated from the observed IBI scores and a regression equation based on the four highlighted points (hand-picked) shown in Figure 5.2C. The regression equation represented the maximum expected IBI score given any particular level of urbanization. For each community sample greater than or equal to 9% urban (the urban threshold), I calculated the difference between the IBI score of the sample and the potential IBI score predicted by the regression equation. This value (hereafter, referred to as “IBI deviation”) was the variation in the relationship between urbanization and IBI for each sample (e.g. sample IBI score= 50, maximum IBI score= 60, IBI deviation= 50 –60 = -10). Therefore, an IBI deviation close to zero represented a small difference between the potential and actual IBI score, and as IBI distance became more negative, there was a larger difference between the potential and actual IBI score.

A path diagram was built with IBI deviation as the response and the non-urban habitat variables in Table 5.1 as predictors. Initially, the exogenous (independent) variables were allowed to freely correlate because all of these variables were calculated using the same GIS techniques (Brendan et al. 2006; Wang et al. 2007). In addition, land-use data was allowed to freely correlate because each land-use variable is a piece of the entire watershed or riparian zone.

I used AMOS 7.0 (Arbuckle 2006) to test the system of equations as represented by the path diagram. This program calculated maximum likelihood estimates of covariances and regression weights that represent direct and indirect effects. AMOS estimated overall model fit with the $\chi^2$ (chi-squared) statistic, where a p-value of less than 0.05 indicated that the model did not fit the data. I iteratively altered the original path diagram by removing habitat variables, insignificant effects, and insignificant covariance until the $\chi^2$ p-value was greater than 0.05 and until the amount of variance explained in IBI deviation was as high as possible. Notably, I did not achieve this 0.05 benchmark until riparian buffer land-use variables were removed from the model.

I assessed final model fit with $\chi^2$, Goodness of Fit Index (GFI), Tucker-Lewis Index (TLI), and root mean square error of approximation (RMSEA). Squared multiple correlations were used to indicate the amount of variance explained in an endogenous (dependent) variable by its predictors. Significance of direct effects was assessed using the 95% biased corrected confidence interval and t-distribution based on the degrees of freedom of the model (Arbuckle 2006, Zorn and Wiley 2004).

## Results

### Classifying Observations

I determined the location of a threshold value in watershed urbanization (% URBAN) at which IBI scores consistently declined. From 0-6% watershed urbanization, the maximum IBI score of the fish community samples was 90, the highest score possible. At 7-8% watershed urbanization, the maximum score (85) began to decrease, and at 9% watershed urbanization, the maximum score was 75 (Figure 5.2A). This 10-point drop represented the largest decrease in IBI score as watershed urbanization was incrementally increased. Therefore, 9% watershed urbanization was used as the threshold value to divide non-urban sites from urbanized sites.

Through the urban threshold and cluster process, I divided Michigan and Wisconsin fish community samples into 3 main groups. Sites in the first group (N= 1829, Total IBI score range= 0-90) had urbanization values below the urban threshold; they were considered to be located at non-urbanized sites and were not used for the remainder of the study. The second group was composed of fish community samples in clusters 1-3, which range from 9 to 27% watershed urbanization and from 0 to 75 in IBI score. Samples in cluster 1 had the highest IBI (N=30, Total IBI score range = 45–75), samples in cluster 3 have the lowest (N=30, Total IBI score range = 0– 25), and samples in cluster 2 are in between (N=30, Total IBI score range = 27 –45) (Figure 5.2B). The third group, clusters 4 (N=19, Total IBI score range = 20-40) and 5 (N=19, Total IBI score range = 5-20), contains fish community samples greater than 27% watershed urbanization (Figure 5.2B). Cluster 6, which is composed of a single point, will be addressed in the discussion.

### Univariate Analysis

Numerous mean habitat differences were detected between clusters. Cluster 1 (low urban, high IBI sites) had significantly more forest land-cover (WT_FOREST, RT_FOREST) and less agricultural land-use (WT_AGR22, RT_AGR22) than cluster 2 (low urban, medium IBI sites) and 3 (low urban, low IBI sites) on both a watershed and riparian scale (Table 5.2, Appendix L). Also, cluster 1 had significantly more wetland land-cover (RT_WETLAND612) and open water land-cover (WT_WATER, RT_WATER). Cluster 1 had significantly higher values of urban due to parking lots and transportation (WT_URBAN14, RT_URBAN14), while cluster 2 had higher amount of urban due to residential areas (WT_URBAN12).

In the comparison of cluster 4 (high urban, medium IBI) and 5 (high urban, low IBI), I saw similar results: cluster 4 had significantly more forest (WT_FOREST, RT_FOREST), more wetlands (WT_WETLAND611), and less agriculture (WT_AGR22) (Table 5.3, Appendix M). Cluster 4 also had a higher base-flow (90_YIELD) and higher amount of coarse surficial geology (WT_COARSE) than did cluster 5.

I did not detect any significant differences in human disturbance variables between cluster 4 and 5. Cluster 1 had consistently lower nitrogen and phosphorus yields than cluster 2 and 3 (Table 5.2). Some of these yields were associated with agriculture (TNY_LIVE, TPY_LIVE), but cluster 1 also had significantly lower point source nitrogen yields (TNY_POINT). Interestingly, even though cluster 1 had a lower TNY_POINT, it had a significantly higher density of permitted point discharge locations (OUTFALL) than cluster 2 and 3. Significant differences in other variables also indicated that cluster 1 sites had an overall better water quality than cluster 2 and 3 sites (lower in EPATOXIC, MANURE, and INSECT) (Table 5.2).

In summary, the sites with higher IBI scores tended to have more natural land-cover and less agriculture in the watershed and upstream riparian zones. In addition, sites with higher IBI scores had less nutrient and pollutant inputs, although they had a greater density of permitted point discharge locations in the watershed.

## Covariance Structure Analysis

The system of structural equations as represented by the path diagram fit the data well according to the statistics used to test model fit (Figure 5.3)($\chi^2$= 6.3, d.f.= 9, p-value=.710 (want to fail), GFI = 0.99 , TLI = 1.0, RMSEA = 0.0). Because the data fit the model, the model can be used to show how the habitat affected the variation around the relationship between urbanization and IBI score. Variables significant at a p-value less than 0.10 are discussed here and unless otherwise noted the value in parentheses represents a variable’s standardized total effect.

Overall, the CSA explained 39% of the variation in IBI deviation (Figure 5.3). The strongest negative effect on IBI deviation was the amount of row crop agriculture in the watershed (Table 5.4) (WT_AGR22, -0.51). This variable alone is able to explain 20% of the variation in IBI deviation. Both total nitrogen yield from non-agriculture sources (TNY_NONAG, -0.19) and density of road crossings (RDCROSS, -0.17) are anthropogenic disturbances, so it was logical to see that these two variables were also negatively related to IBI deviation. However, because natural land-use is generally thought to be good for fish communities, it was unexpected that wetlands dominated by shrubs (WT_WETLAND610, -0.42) and open land (WT_OPEN, -0.28) also had a negative relationship with IBI deviation. In Michigan and Wisconsin, it is possible that land classified as open-land is actually composed of old agricultural fields, and much land classified as shrubby wetland may be in a constant state of disturbance. For example, riparian borders of disturbed streams and lakes, restored wetlands, and ditches could potentially be classified as shrubby wetland.

Forests (WT_FOREST43, 0.22) and forested wetlands (WT_WETLAND, 0.12) were related to higher (less negative) IBI deviations and therefore higher IBI scores. The density of permitted outfall sites in the watershed (OUTFALL) had conflicting effects on IBI deviation. OUTFALL had a negative indirect effect on IBI deviation through non-agricultural total nitrogen yield (TNY_NONAG, -0.04), but had a positive direct effect (0.14). Overall, OUTFALL had a positive total effect on IBI deviation (0.10).

## Discussion

The urban threshold has been an important concept in the management of urban fish communities. I found a threshold at a similar level to other studies (Paul and Meyer 2001; Wang et al. 2001, Riseng et al. 2006), appearing at about 9% urbanization for Michigan and Wisconsin streams, and used this threshold to divide our data into groups for comparison purposes. For fish communities above this threshold, I found that fish IBI score varies from zero to some maximum level that is apparently controlled tightly by the amount of watershed urbanization (% URBAN). Viewed on a scatterplot, the relationship between urban and fish IBI creates a wedge-shaped clump of data points where the diagonal edge of this wedge forms a ceiling indicating the maximum potential of the fish community (Figure 5.2A). This ceiling effect is very strong in our study and in others; in 10 studies that visualized data in this manner, only 4 sampled streams had a good enough biotic integrity (fish or macroinvertebrate) to place the point high above the ceiling (Weaver and Garman 1994; Wang et al. 2001; Wang and Kanehl 2003; Wang et al. 2003a; Walsh 2004; Carter and Fend 2005; Fitzpatrick et al. 2005; Kennan et al. 2005; Limburg et al. 2005; Walters et al. 2005). In this study, of the 1,857 sampled sites, only one was truly an outlier (cluster 6 in Figure 5.2B, discussed further below).

The potential of the biotic community may be limited by urbanization, but other factors controlled the variation beneath this ceiling. For a given range of urbanization within this wedge of data, some fish community samples had high measures of biotic integrity, others had low, and this variation could not be explained by overall urbanization measurements. Understanding this variation is important for management purposes as factors that control it may conceivably be within human control. A series of nonparametric univariate analyses and a covariance structure analysis were used to understand the residual variation in fish community quality. Based on these analyses and from work done in other studies, I can provide four possible hypotheses about how the variation in the urbanization effect is controlled.

How can good fish communities exist in urban areas?

1. In urban areas, high quality fish communities need good water quality.

Urbanization increases the concentration of nearly every chemical constituent in the water of urban streams (Paul and Meyer 2001). Phosphorus and nitrogen sources include wastewater, fertilizer, and leaking septic and sewage systems. Metals such as cadmium, lead, and mercury are routinely found in high concentrations in the sediment of urban streams (Paul and Meyer 2001). In our study, pollution was measured in a variety of ways; including upstream agricultural pollution (fertilizer, livestock waste, insecticide) and pollution from the urbanized area itself (point source nutrient discharge, discharges from EPA toxic release inventory sites).

As water and sediment pollution is spatially variable in urban streams, a logical hypothesis would be that good urban fish communities are found in those streams that have lower levels of contamination. Our results support this reasoning. Nitrogen and phosphorus yields from fertilizer, livestock waste, and point sources are significantly higher in the poor fish communities of cluster 3 than in the good fish communities of cluster 1. For example, means of phosphorus and nitrogen yields due to livestock waste were twice as high in cluster 3 as cluster 1, and means of phosphorus yields due to point source pollution were about three times high in cluster 3 than cluster 1. EPA toxic inventory sites had a mean density five times higher in cluster 3 than cluster 1. The CSA analysis indicated that the proportion of agriculture in the watershed and non-agriculture nitrogen yield help to explain why some urban fish communities had a large gap between their actual and potential IBI score.

Management actions that increase water quality by lowering nutrient and metal inputs will have positive effects on fish quality of urban streams. However, to see improvements in an urban stream, change needs to be made throughout the stream's entire watershed. The upstream and downstream areas are inextricably linked; my results seem to indicate that agriculture is strongly related to poor fish communities in urban streams.

2. In urban areas, point-source discharge locations can increase water flow and fish quality.

Many studies report that urbanization and impervious surface cause an increase of water runoff, lower groundwater recharge, and subsequently lower stream base-flow (Klein 1979, Paul and Meyer 2001, Wang et al 2001; Riseng et al. 2004). Increased peak flow is perhaps the most severe urban stream disturbance as high flows from storm events can wash away in-stream habitat and scour the streambed (Scott et al. 1986, Miltner et al. 2004). Roy et al. (2005) were able to link hydrologic disturbance to decreases in fish quality: they found that increases in the magnitude and frequency of storm events and in prolonged duration of low-flow conditions resulted in reduced number of sensitive fish and increases in tolerant fish. However, some studies have suggested that point source discharges in urban streams can offset or mitigate these negative hydrological effects by providing constant water input (Horowitz et al. 1999; Paul and Meyer 2001; Fitzpatrick et al. 2005). Consistent point source discharge would be mechanistically similar to groundwater inputs and would provide a stable environment for fish communities.

Our results support this hypothesis; in the univariate analysis, the mean density of permitted point discharge locations (OUTFALL) was significantly higher in fish community samples with higher total IBI score. In the CSA, a higher OUTFALL has a total positive effect on IBI deviation. Therefore, a higher density of outfall sites (and more stable water flow) is related to more healthy fish communities. The main concern regarding point source contributions is the amount of N and P that discharged water carries, because as the CSA indicates, OUTFALL has an weak indirect negative effect on IBI deviation through non-agricultural nitrogen yield (TNY_NONAG).

3**.** High quality urban fish communities need natural-land use throughout the watershed to mitigate urban changes in hydrology and water quality.

Our results indicate that natural land-use is critical to sustaining quality fish communities in an urbanized stream. The streams of cluster 1 (high IBI) had significantly higher forest land-cover than streams of cluster 3 (low IBI), and streams of cluster 4 (medium IBI) were significantly higher in forest and wetlands than streams of cluster 5 (low IBI). In the CSA, both forests and forested wetlands were important reducing the difference between observed IBI score and potential IBI score.

Agricultural impacts on fish IBI were quite clearly negative. The univariate analysis showed that poorer fish communities are associated with higher agriculture. The agricultural variable in the CSA was the most influential factor in decreasing IBI deviation. However, on a per-unit area basis, agriculture has less of an impact on fish than urban (Wang et al. 2000, Wang et al. 2003a); so any land that is not urbanized is better for fish than urbanized land.

From a conservation context, it would be very useful to know the most effective scale of land-use management. Historically, the standard stream management practice in urban areas was to maintain or build intact riparian forest zones around urbanized streams in order to mitigate urbanization effects (Steedman 1988, May et al 1997, Castelle 1994, Wang et al. 2001, Miltner 2004). About one-third of stream restoration projects in the United States are focused on riparian buffers (Bernhardt et al. 2005). Studies have hypothesized that riparian buffers moderate water run-off, absorb nutrients from run-off and through-flow, minimize erosive effects, and provide in-stream habitat such as woody debris and overhanging vegetation (Osborne and Kovacic 1993; Richards et al. 1996; Lammert and Allan 1999; Miltner 2004; Barker et al. 2006).

However, I expect that land-use throughout the entire stream catchment is more important to overall fish community quality in urban streams than land-use within the riparian buffer. Riparian buffers only have minimal mitigation effects on what is known to be two major sources of urban disturbance: water quality and water flow. For example, studies have found that riparian buffers provided minimal benefits for streams with highly altered sediment and hydrologic regimes (Fitzpatrick et al. 2005; Roy et al. 2006). High peak flows caused by run-off coming from impervious surfaces will not be moderated by riparian buffers, because in urban areas the water is discharged directly to the stream (Fitzpatrick et al. 2005). Forested land-cover and other permeable surfaces across the watershed are more important for decreasing overall surface run-off, and vegetation across the watershed improves water quality though nutrient uptake. In our CSA model, riparian variables were not significantly associated with IBI deviation. This indicates that riparian buffer land-use was either not associated with fish quality or that the variance in fish quality was already accounted for by the watershed land-use variables and riparian buffer land-use had nothing further to contribute.

It is my conclusion that while land-use in the riparian buffer may be important for local habitat structure, overall watershed land-use is a better determinant of the quality of urban fish communities. Wang et al. (2003b) also reached this conclusion in a study based on Minnesota, Michigan, and Wisconsin stream fish. To maintain high quality fish communities in urban environments, less emphasis should be placed on riparian buffer management and more on plans that build and maintain natural land-cover across the watershed. This may involve groups implementing fewer restoration work projects and taking more action in political and planning processes.

4. High quality urban fish communities can be supported by a close proximity to high quality, non-urbanized streams (anecdotal evidence).

Outliers can have interesting properties, and it is unfortunate the ceiling effect of urbanization on fish was so strong, as outliers were very rare. Only one fish community sample with high biotic integrity was located on a stream with high urbanization. This sample, located on Lincoln Creek in northern Milwaukee (cluster 6 in Figure 5.2B), had

82% watershed urbanization (% URBAN), the highest amount of urbanization in the study. However, with a total IBI score of 55, the fish community was extremely healthy for such a highly urbanized site. Fourteen species were caught here, including three species of redhorse, largemouth bass, smallmouth bass, rock bass, northern pike, hornyhead chub, and sand shiner.

It appears that an unusual set of circumstances allowed the stream to have such a high quality fish community. About 1.5 kilometers downstream from the sampled point on Lincoln Creek was another sampled river that was only 7% urban and had high amounts of agriculture, forests, and wetlands in its watershed. As every fish found in Lincoln Creek was also found in this downstream, less-disturbed river, it is entirely possible that the fish moved upstream to the Lincoln Creek site from the less-disturbed river. The effect of the downstream channel on upstream channels has been referred to as the downstream link; several studies have previously identified and used this effect to predict and analyze fish communities (Osborne and Wiley 1992; Grenouillet et al. 2004; Smith and Kraft 2005**).**

It was clear from aerial photographs (Google Earth 2007) that the sampled site on Lincoln Creek could have appealing local habitat due to 20-meter wide grassy riparian buffers and a small forested park where the sampling actually took place. Given my argument from the previous section, it is unlikely that these riparian buffers are sufficient to reduce negative hydrologic effects and allow the stream to maintain permanent resident populations. However, it is possible that Lincoln Creek has a transient good quality fish population due to the higher quality river downstream. I hypothesize that other small urbanized streams may also have the potential to hold good quality fish communities given appealing local habitat and close proximity to less disturbed “feeder-streams”. A wise management action would be to make sure that these high quality “feeder-streams” near urban areas are protected and undeveloped.

Numerous studies have found how and why urbanization has negative impacts on the physical structure of a stream and on its biotic integrity. To move toward mitigation, I suggest that future studies look at other variables that may affect the variation around the negative urban effect. As the CSA in this study was only able to explain 39% of the variation, it is certain there are other important factors that I have not considered here. Future investigations that measuring variables on local scales (e.g., in-stream habitat structure, sedimentation, water quality) as well as studies that examine interactions between fish and their food base may also prove useful in developing practical and feasible management tools that can help create or maintain good fish communities in urban streams.

Table 5.1. List of variables that were used in this study, their codes referred to in the text, their units, and the source of the data. All land-use variables listed below with the prefix "WT" have also been measured as a percent of the riparian buffer (entire upstream corridor, 30 meters to each side of stream, prefix "RT"). Citation numbers are as follows 1) Brenden et al. 2006, 2) MCGI 2007, 3)WDNR 2007, 4) Wang et al. 2007

| Variable Description | Variable Code | Unit | Citation |
|---|---|---|---|
| *Variables used for classification of fish community samples* | | | |
| Predicted mean July water temperature | WATER_TEMP | Celsius | 1 |
| Total urban land-use | % URBAN | % of watershed | 1,2,3 |
| *Variables used to find differences between clusters* | | | |
| Fine-grain surficial geology | WT_FINE | % of watershed | 1 |
| Coarse-grain surficial geology | WT_COARSE | % of watershed | 1 |
| 10% exceedence flow yield | TEN_YIELD | cms/km$^2$ | 1 |
| 90% exceedence flow yield | NINETY_YIELD | cms/km$^2$ | 1 |
| Total urban, riparian buffer | RT_URB | % of riparian buffer | 1,2,3 |
| Urban, commercial/industrial | WT_URBAN11 | % of watershed | 1,2,3 |
| Urban, residential | WT_URBAN12 | % of watershed | 1,2,3 |
| Urban, transportation and parking lots | WT_URBAN14 | % of watershed | 1,2,3 |
| Total agricultural land-use | WT_AGR | % of watershed | 1,2,3 |
| Agriculture, non-row crop | WT_AGR21 | % of watershed | 1,2,3 |
| Agriculture, row crop | WT_AGR22 | % of watershed | 1,2,3 |
| Open/non-forest | WT_OPEN | % of watershed | 1,2,3 |
| Forest land cover | WT_FOREST | % of watershed | 1,2,3 |
| Forest, deciduous, upland | WT_FOREST41 | % of watershed | 1,2,3 |
| Forest, coniferous, upland | WT_FOREST42 | % of watershed | 1,2,3 |
| Forest, mixed, upland | WT_FOREST43 | % of watershed | 1,2,3 |
| Open water | WT_WATER | % of watershed | 1,2,3 |
| Total wetland land cover | WT_WETLAND | % of watershed | 1,2,3 |
| Wetland, wooded, shrubland | WT_WETLAND610 | % of watershed | 1,2,3 |
| Wetland, wooded, lowland deciduous forest | WT_WETLAND611 | % of watershed | 1,2,3 |
| Wetland, wooded, lowland coniferous forest | WT_WETLAND612 | % of watershed | 1,2,3 |
| Wetland, wooded, mixed lowland forest | WT_WETLAND613 | % of watershed | 1,2,3 |
| Wetland, non-wooded | WT_WETLAND62 | % of watershed | 1,2,3 |
| Total nitrogen atmospheric yield | TNY_ATMOS | kg/km$^2$/yr | 4 |
| Total nitrogen fertilizer yield | TNY_FERT | kg/km$^2$/yr | 4 |
| Total nitrogen livestock waste yield | TNY_LIVE | kg/km$^2$/yr | 4 |
| Total nitrogen non-agriculture yield | TNY_NONAG | kg/km$^2$/yr | 4 |
| Total nitrogen point source yield | TNY_POINT | kg/km$^2$/yr | 4 |
| Total nitrogen yield | TNY_TOTAL | kg/km$^2$/yr | 4 |
| Total phosphorus fertilizer yield | TPY_FERT | kg/km$^2$/yr | 4 |

Table 5.1, continued.

| Variable Description | Variable Code | Unit | Citation |
|---|---|---|---|
| Total phosphorus livestock waste yield | TPY_LIVE | $kg/km^2/yr$ | 4 |
| Total phosphorus non-agriculture yield | TPY_NONAG | $kg/km^2/yr$ | 4 |
| Total phosphorus point source yield | TPY_POINT | $kg/km^2/yr$ | 4 |
| Total phosphorus yield | TPY_TOTAL | $kg/km^2/yr$ | 4 |
| Proportion of watershed treated with fertilizers | FERT | % of watershed | 4 |
| Proportion of watershed treated with herbicides | HERB | % of watershed | 4 |
| Proportion of watershed treated with insecticides | INSECT | % of watershed | 4 |
| Proportion of watershed treated with manure | MANURE | % of watershed | 4 |
| Density of permitted outfalls in watershed | OUTFALL | $\#/km^2$ | 4 |
| Density of active mines in watershed | MINES | $\#/km^2$ | 4 |
| Population density (2000 census) in watershed | POPDENS | $\#/km^2$ | 4 |
| Road crossing density in watershed | RDCROSS | $\#/km^2$ | 4 |
| Road density in watershed | RDDENS | $km/km^2$ | 4 |
| Density of EPA Toxic Release Inventory sites | EPATOXIC | $\#/km^2$ watershed | 4 |

Table 5.2. Results from the Kruskal-Wallis nonparametric test of means and the multiple comparison Nemnyi test of cluster 1, 2, and 3. Included in this table are the cluster means of the variables, $\chi^2$ of Kruskal-Wallis test, associated degrees of freedom, and associated significance. Multiple comparison results are interpreted as follows: q (1-3) is the Studentized range q statistic for the difference between cluster 1 and 3, and p (1-3) is the associated significance. Critical values for q are as follows: $q_{0.1, \infty, 3} = 2.902$, $q_{0.05, \infty, 3} = 3.313$, $q_{0.01, \infty, 3} = 4.200$ (Zar 1999). Metrics significant at $\alpha < .10$ are in bold. Only significant results are shown, all other results are recorded in Appendix L.

| Habitat Variable | Cluster 1 | Cluster 2 | Cluster 3 | $\chi^2$ | d.f. | p | q (1-3) | p (1-3) | q (1-2) | p (1-2) | q (2-3) | p (2-3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WT_URBAN12 | 3.93 | 5.77 | 4.76 | 5.33 | 2 | **0.07** | 2.28 | >0.1 | 3.12 | **<0.10** | 0.08 | >0.1 |
| WT_URBAN14 | 3.93 | 2.41 | 1.66 | 11.95 | 2 | **0.00** | 4.55 | **<0.01** | 2.98 | **<0.10** | 1.56 | >0.1 |
| WT_AGR22 | 17.20 | 20.04 | 25.75 | 6.91 | 2 | **0.03** | 3.66 | **<0.05** | 1.51 | >0.1 | 2.15 | >0.1 |
| WT_FOREST | 22.90 | 15.60 | 15.0 | 10.89 | 2 | **0.00** | 4.11 | **<0.05** | 3.86 | **<0.05** | 0.03 | >0.1 |
| WT_FOREST41 | 18.10 | 13.06 | 13.19 | 9.31 | 2 | **0.01** | 3.51 | **<0.05** | 3.85 | **<0.05** | 0.03 | >0.1 |
| WT_FOREST42 | 2.14 | 1.59 | 1.08 | 8.32 | 2 | **0.02** | 3.92 | **<0.05** | 2.24 | >0.1 | 1.68 | >0.1 |
| WT_FOREST43 | 2.66 | 1.89 | 1.33 | 11.73 | 2 | **0.00** | 4.75 | **<0.01** | 2.49 | >0.1 | 2.25 | >0.1 |
| WT_WATER | 2.55 | 1.03 | 1.25 | 10.25 | 2 | **0.01** | 3.30 | **<0.10** | 4.18 | **<0.05** | 0.88 | >0.1 |
| RT_URBAN14 | 2.85 | 1.94 | 1.01 | 13.79 | 2 | **0.00** | 4.96 | **<0.01** | 2.59 | >0.1 | 2.37 | >0.1 |
| RT_AGR22 | 12.66 | 155 | 19.33 | 5.22 | 2 | **0.07** | 3.16 | **<0.10** | 1.97 | >0.1 | 1.20 | >0.1 |
| RT_FOREST | 19.70 | 16.80 | 15.20 | 5.63 | 2 | **0.06** | 3.05 | **<0.10** | 2.64 | >0.1 | 0.411 | >0.1 |
| RT_FOREST42 | 1.97 | 1.43 | 1.17 | 6.00 | 2 | **0.05** | 3.29 | **<0.10** | 2.13 | >0.1 | 1.16 | >0.1 |
| RT_FOREST43 | 2.61 | 1.86 | 1.25 | 11.85 | 2 | **0.00** | 4.76 | **<0.01** | 2.77 | >0.1 | 1.99 | >0.1 |
| RT_WATER | 10.34 | 10.34 | 4.87 | 8.66 | 2 | **0.01** | 2.16 | >0.1 | 4.08 | **<0.05** | 1.91 | >0.1 |
| RT_WETLAND612 | 0.62 | 0.39 | 0.22 | 11.11 | 2 | **0.00** | 4.41 | **<0.01** | 2.72 | >0.1 | 1.69 | >0.1 |
| TNY_LIVE | 94.00 | 210.65 | 208.25 | 11.54 | 2 | **0.00** | 2.91 | **<0.10** | 3.22 | **<0.10** | 0.03 | >0.1 |
| TNY_NONAG | 86.44 | 78.01 | 97.70 | 6.91 | 2 | **0.03** | 2.70 | >0.1 | 0.79 | >0.1 | 3.49 | **<0.05** |
| TNY_POINT | 176.31 | 326.21 | 414.76 | 6.32 | 2 | **0.04** | 3.31 | **<0.10** | 6.64 | **<0.01** | 0.66 | >0.1 |
| TPY_FERT | 33.13 | 32.13 | 26.72 | 5.33 | 2 | **0.07** | 2.75 | >0.1 | 0.06 | >0.1 | 2.81 | >0.1 |
| TPY_LIVE | 13.36 | 27.86 | 22.52 | 9.42 | 2 | **0.01** | 3.40 | **<0.05** | 3.94 | **<0.05** | 0.55 | >0.1 |
| INSECT | 2.18 | 4.40 | 3.77 | 11.16 | 2 | **0.00** | 3.86 | **<0.05** | 4.20 | **<0.05** | 0.35 | >0.1 |
| MANURE | 1.87 | 4.38 | 4.68 | 12.90 | 2 | **0.00** | 4.72 | **<0.01** | 3.85 | **<0.05** | 0.87 | >0.1 |
| OUTFALL | 0.21 | 0.12 | 0.08 | 12.14 | 2 | **0.00** | 4.69 | **<0.01** | 3.40 | **<0.05** | 1.29 | >0.1 |
| EPATOXIC | 0.04 | 0.18 | 0.20 | 10.99 | 2 | **0.00** | 4.24 | **<0.01** | 0.57 | >0.1 | 3.66 | **<0.05** |

Table 5.3. Results from the Mann-Whitney U nonparametric test of means of clusters 4 and 5. Included in this table are the cluster means of the variables, U statistic, and associated significance. Only results significant at a = 0.10 are shown; all results are recorded in Appendix M.

| Habitat Variable | Cluster 4 | Cluster 5 | Mann-Whitney U statistic | p-value |
|---|---|---|---|---|
| WT_COARSE | 16.70 | 8.60 | 108.5 | 0.02 |
| 90_YIELD | 0.0016 | 0.0012 | 118 | 0.07 |
| WT_AGR22 | 2.46 | 5.99 | 122.5 | 0.08 |
| WT_FOREST | 21.11 | 14.89 | 120 | 0.08 |
| WT_FOREST43 | 2.56 | 1.63 | 119.5 | 0.07 |
| WT_WETLAND610 | 0.67 | 0.23 | 118.5 | 0.04 |
| WT_WETLAND611 | 2.45 | 1.55 | 124 | 0.10 |
| WT_WETLAND612 | 0.06 | 0.01 | 121 | 0.04 |
| WT_WETLAND613 | 0.01 | 0.00 | 130 | 0.07 |
| RT_FOREST | 25.37 | 18.32 | 121.5 | 0.09 |
| RT_FOREST43 | 2.78 | 1.74 | 119.5 | 0.07 |
| RT_WETLAND610 | 2.63 | 1.93 | 133.5 | 0.05 |
| RT_WETLAND612 | 0.14 | 0.02 | 111 | 0.01 |

Table 5.4. Standardized total effects of the stressor variables (left of table) on their dependent variables (top of table) as computed by the CSA. The standardized effect of WT_OPEN on IBI DEVIATION (-0.28) indicates that when WT_OPEN decreases by one standard deviation, IBI DEVIATION decreases by 0.28 standard deviations. Effect significant at <0.05 effects are in bold; other effects are significant at <0.10.

| | IBI DEVIATION | TNY_NONAG | NINETY_YIELD |
|---|---|---|---|
| WT_OPEN | **-0.28** | | |
| OUTFALL | 0.10 | **0.19** | |
| WT_WETLAND610 | **-0.42** | | |
| WT_FOREST43 | **0.22** | | |
| WT_AGR22 | **-0.48** | | **-0.17** |
| WT_WETLAND612 | 0.12 | | |
| RDCROSS | **-0.17** | | |
| TNY_NONAG | **-0.19** | | |
| NINETY_YIELD | **-0.18** | | |

Figure 5.1. Locations of the sampled fish sites available from the Wisconsin Department of Natural Resources, the Michigan Department of Natural Resources, and the Michigan Rivers Inventory. The open white circles represent cold-water sites or species-targeted samples that were not used in this study. The filled black circles were those sites used in this study; they are sites with warm-water fish communities that were sampled with either electroshocking or rotenone.

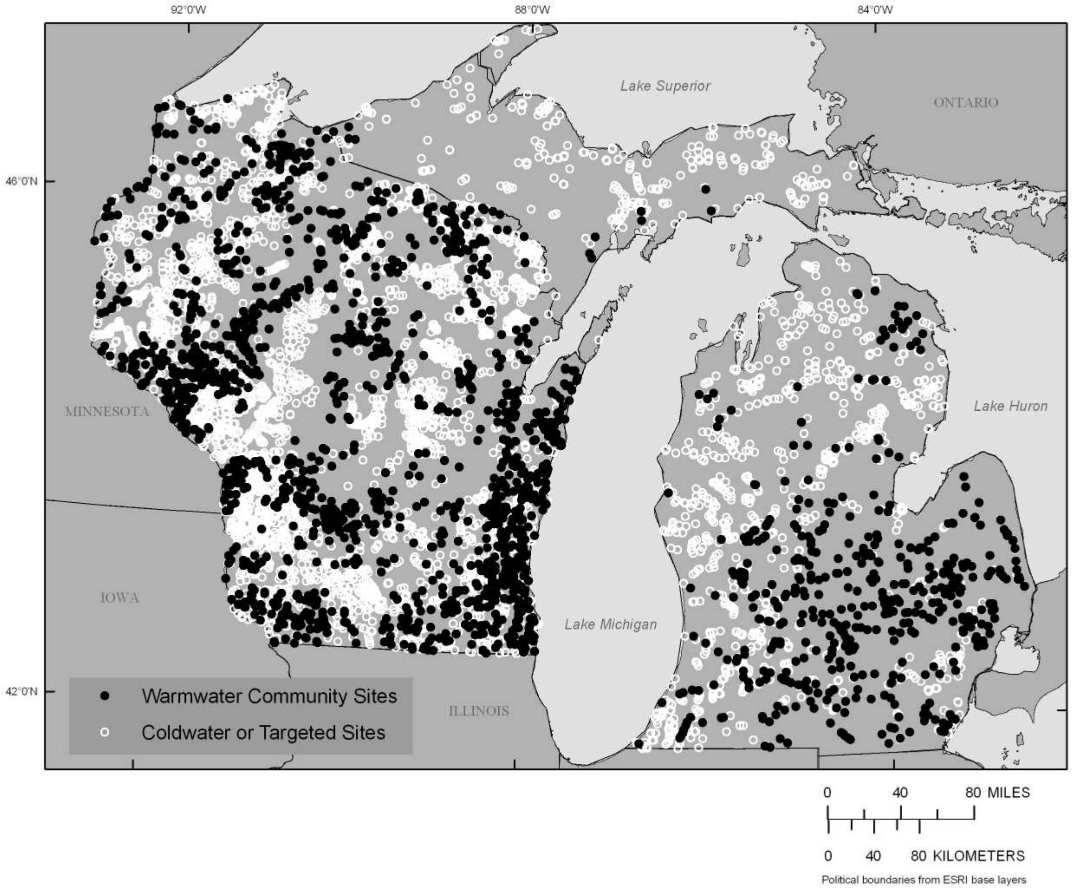


Figure 5.2. The relationship between Fish IBI Score and % URBAN for A) all fish community sample sites in study, B) sites equal to or above the urban threshold of 9% URBAN after being clustered by Fish IBI Score and % URBAN, and C) sites equal to or above the urban threshold of 9%, with a green regression line based off of the four red points. The blue line represents an example of “IBI deviation”, the difference in IBI score between the regression line and sample point.

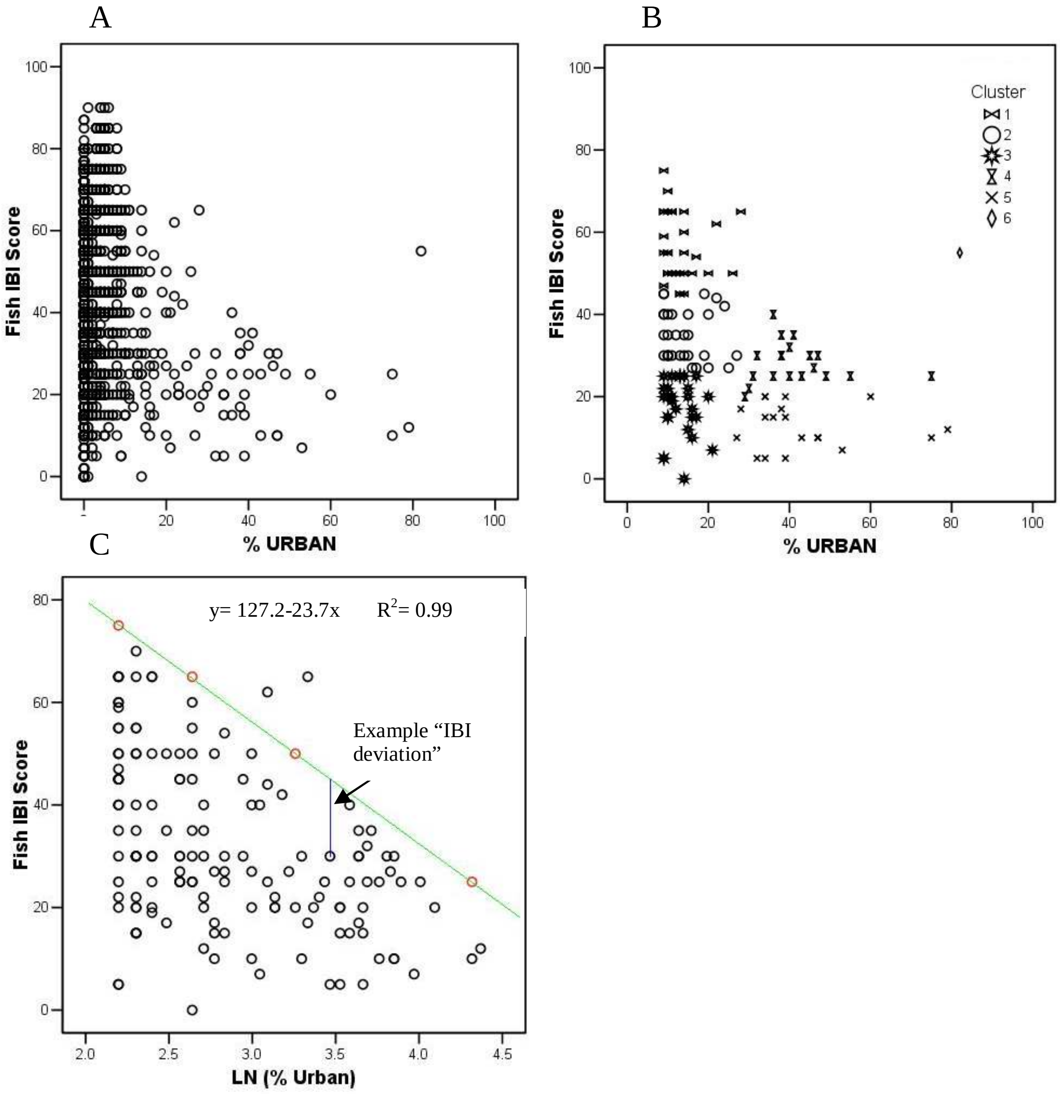
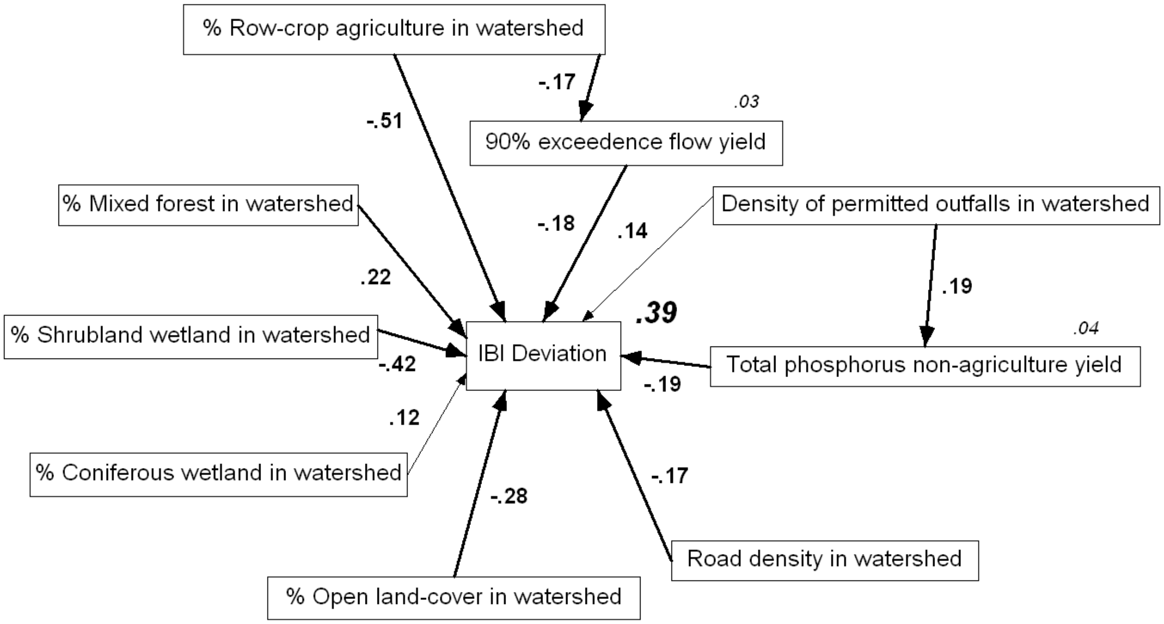


Figure 5.3. Simplified path diagram of the CSA of the fish community sample sites equal to or above the urban threshold (N= 128). Dark arrows indicate effects significant at p < 0.05, light arrows indicate significant effects at p <0.10, and the nearby numbers in bold are the corresponding standardized regression weight. Numbers in italics by the endogenous variables indicate the amount of variance explained by the predictor variables. Arrows representing covariance between variables were removed for simplicity.



# References

Alig, R. J., J. D. Kline, and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. Landscape and Urban Planning 69:219-234.

Arbuckle, J. L. 2006. Amos 7.0 User's Guide. Amos Development Corporation, Spring House, Pennsylvania.

Arnold, C., and J. C. Gibbons. 1996. Impervious surface coverage. Journal of the American Planning Association 62:243-258.

Barker, L. S., G. K. Felton, and E. Russek-Cohen. 2006. Use of Maryland biological stream survey data to determine effects of agricultural riparian buffers on measures of biological stream health. Environmental Monitoring and Assessment 117:1-19.

Bernhardt, E. S., M. A. Palmer, J. D. Allan, G. R. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, D. Galat, S. Gloss, P. Goodwin, D. Hart, B. Hassett, R. Jenkinson, S. Katz, G. M. Kondolf, P. S. Lake, R. Lave, J. L. Meyer, T. K. O'Donnell, L. Pagano, B. Powell, and E. Sudduth. 2005. Synthesizing U.S. river restoration efforts. Science 308:636-637.

Bollen, K.A. 1989. Structural equations with latent variables. Wiley, New York, New York.

Booth, D. B., and C. R. Jackson. 1997. Urbanization of aquatic systems: Degradation thresholds, stormwater detection, and the limits of migration. Journal of the American Water Resources Association 33:1077-1090.

Brendan, T. O., R. D. Clark, A. R. Cooper, P. W. Seelbach, L. Wang, S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49-74 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape Influences on Stream Habitat and Biological Assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Carter, J. L., and S. V. Fend. 2005. Setting limits: The development and use of factor-ceiling distributions for an urban assessment using macroinvertebrates. Pages 179-191 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Cuffney, T., H. Zappia, E. M. Giddings, and J. F. Coles. 2005. Effects of urbanization on benthic macroinvertebrate assemblages in contrasting environmental settings: Boston, Massachusetts, Birmingham, Alabama; and Salt Lake City, Utah. Pages 361-408 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors.

Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Betheda, Maryland.

ESRI. 2007. GIS and Mapping Software. Available: www.esri.com (April 2007).

Fjeld, E., and S. Rognerud. 1993. Use of path analysis to investigate mercury accumulation in brown trout (Salmo trutta) in Norway and the influence of environmental factors. Canadian Journal of Fisheries and Aquatic Sciences 50:1158-1167.

Fitzpatrick, F. A., M. W. Diebel, A. H. Mitchell, T. L. Arnold, M. A. Lutz, and K. D. Richards. 2005. Effects of urbanization on the geomorphology, habitat, hydrology, and fish index of biotic integrity of streams in the Chicago, Illinois, and Wisconsin. Pages 87-116 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

GLSC (Great Lake Science Center). 2007. The Great Lakes Aquatic GAP Project. Available: http://www.glsc.usgs.gov/GLGAP.html (December 2007).

Google Earth. 2007. Available: http://earth.google.com/ (December 2007)

Grant, J. 2002. Qualitative biological and habitat survey protocols for wadable streams and rivers. Michigan Department of Environmental Quality, Surface Water Quality Division, Great Lakes and Enviornmental Assessment Section Procedure #51, Lansing.

Grenouillet, G., D. Pont, and C. Herisse. 2004. Within-basin fish assemblage structure: the relative influence of habitat versus stream spatial position on local species richness. Canadian Journal of Fisheries and Aquatic Sciences 61: 93-102.

Horowitz, A. J., M. Meybeck, Z. Idlafkih, E. Biger. 1999. Variations in trace element geochemistry in the Seine River Basin based on floodplain deposits and bed sediments. Hydrological Processes 13:1329-1340.

Infante, D. M., M. J. Wiley, and P. W. Seelbach. 2006. Relationships between land use and stream ecosystems: A multistream assessment in southernwestern Michigan. Pages 339-358 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Isaak, D. J., and W. A. Hubert. 2001. Production of stream habitat gradients by montane watersheds: hypothesis tests based on spatially explicit path analyses. Canadian Journal of Fisheries and Aquatic Sciences 58:1089-1103.

Klein, R. D. 1979. Urbanization and stream quality impairment. Water Resources Bulletin 15:948-963.

Kemp, S. J., and J. R. Spotila. 1997. Effects of urbanization on brown trout Salmo trutta, other fish and macroinvertebrates in Valley Creek, Valley Forge, Pennsylvania. The American Midland Naturalist 138:55-68.

Kennen, J. G., M. Chang, and B. H. Tracy. 2005. Effects of landscape change on fish asseblage structure in a rapidly growing metropolitan area in North Carolina, USA. Pages 39-52 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Konrad, C. P., and D. B. Booth. 2005. Hydrologic changes in urban streams and their ecological significance. Pages 157-177 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Lammert, M., and J. D. Allan. 1999. Assessing biotic integrity of streams: Effects of scale in measuring the influence on land use/cover and habitat structure on fish and macroinvertebrates. Environmental Management **23**:257-270.

Lenat, D. R., and J. K. Crawford. 1994. Effects of land use of water quality and aquatic biota of three North Carolina Piedmont streams. Hydrobiologia 294:185-199.

Limburg, K. E., K. M. Stainbrook, J. D. Erickson, and J. M. Gowdy. 2005. Urbanization consequences: Case studies in the Hudson River watershed. Pages 23-37 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Lyons, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. United States Department of Agriculture, Forest Service, General Technical Report NC-149, St.Paul, Minnesota.

Lyons, J., L. Wang, and T. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. North American Journal of Fisheries Management 16: 241-255.

May, C. W., R. R. Horner, J. R. Karr, B. W. Mar, and E. B. Welch. 1997. Effects of urbanization on small streams in the Puget Sounds Lowland ecoregion. Watershed Protection Techniques 2:485-494.

MCGI (Michigan Center for Geographic Information). 2007. Land cover 2001 geographic theme: land cover/use. Available: www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=2&cat=Land+Cover+2001 (December 2007)

McKinney, M. L. 2006. Urbanization as a major cause of biotic homogenization. Biological Conservation 127:247-260.

Meador, M. R., J. F. Coles, and H. Zappia. 2005. Fish assemblage responses to urban intensity gradients in contrasting metropolitan areas: Birmingham, Alabama, and Boston, Massachusetts. Pages 409-423 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Morgan, R. P., and S. E. Cushman. 2005. Urbanization effects on stream fish assemblages in Maryland, USA. Journal of the North American Benthological Society 24:643-655.

Miltner, R. J., D. White, and C. Yoder. 2004. The biotic integrity of streams in urban and suburbanizing landscapes. Landscape and Urban Planning 69:87-100.

NHD (National Hydrography Dataset), 2007. Available: http://nhd.usgs.gov/ (December 2007).

NOAA (National Oceanic and Atmospheric Agency), 2007. Impervious Surface Analysis Tools. Available: http://www.csc.noaa.gov/crs/cwq/isat.html. (December 2007).

Olden, J. D. 2006. Biotic homogenization: a new research agenda for conservation biogeography. Journal of Biogeography 33:2027-2039.

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

Osborne, L. L., and D. A. Kovacic. 1993. Riparian vegetated buffer strips in water-quality restoration and stream management. Freshwater Biology **29**:243-258.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. The Annual Review of Ecology and Systematics 32:333-365.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influence on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Sciences **53**:295-311.

Riseng, C. M., M. J. Wiley, and R. J. Stevenson. 2004. Hydrological disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. Journal of the North American Benthological Society 23:309-326.

Riseng, C. M., M. J. Wiley, R. J. Stevenson, T. G. Zorn, and P. W. Seelbach. 2006. Comparison of coarse versus fine scale sampling on statistical modeling of landscape effects and assessment of fish assemblages of the Muskegon River, Michigan. Pages 555-575 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Roy, A. H., M. C. Freeman, B. J. Freeman, S. J. Wenger, W. E. Ensign, and J. L. Meyer. 2005. Investigating hydrologic alteration as a mechanism of fish assemblage shifts in urbanizing streams. The Journal of the North American Benthological Society 24:656-678.

Roy, A. H., M. C. Freeman, B. J. Freeman, S. J. Wenger, J. L. Meyer, and W. E. Ensign. 2006. Importance of riparian forests in urban catchments contingent on sediment and hydrologic regimes. Environmental Management 37:523-539.

Schueler, T. R. 1994. The importance of imperviousness. Watershed Protection Techniques 1:100-111.

Scott, J. B., C. R. Stewart, and Q. J. Stober. 1986. Effect of urban development on fish population dynamics in Kelsey Creek, Washington. Transactions of the American Fisheries Society 115:555-567.

Scott, M. C. 2006. Winners and losers among stream fishes in relation to land use legacies and urban development in the southeastern US. Biological Conservation 127:310-309.

Seelbach, P. W., and M. J. Wiley. 1997. Overview of the Michigan Rivers Inventory (MRI) project. Fisheries Technical Report 97-3, Michigan Department of Natural Resources, Ann Arbor.

Sheldon, A. L., and G. K. Meffe. 1995. Path analysis of collective properties and habitat relationships of fish assemblages in coastal plain streams. Canadian Journal of Fisheries and Aquatic Sciences 52:23-33.

Smith, T. A., and C. E. Kraft. 2005. Stream fish assemblages in relation to landscape position and local habitat variables. Transactions of the American Fisheries Society 134:430-440.

SPSS, Inc. 2007. SPSS for Windows v15.0. SPSS, Inc, Chicago, Illinois.

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 45:492-501.

Tabit, C. R., and G. M. Johnson. 2002. Influence of urbanization on the distribution of fishes in a southeastern upper Piedmont drainage. Southeastern Naturalist 1:253-268.

Tate, C. M., T. Cuffney, G. McMahon, E. M. Giddings, J. F. Coles, and H. Zappia. 2005. Use of an urban intensity index to assess urban effects of streams in three contrasting environmental settings. Pages 291-316 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Trautman, M. B. 1981. The fishes of Ohio. Ohio State University Press, Columbus.

UM (University of Michigan). 2007. Ecological classification of rivers for environmental assessment. Available: http://sitemaker.umich.edu/riverclassproject (December 2007).

Walsh, C. J. 2004. Protection of in-stream biota from urban impacts: minimise catchment imperviousness or improve drainage density? Marine and Freshwater Research 55:317-326.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. The Journal of the North American Benthological Society 24:706-723.

Walters, D. M., D. S. Leigh, and A. B. Bearden. 2003. Urbanization, sedimentation, and the homogenization of fish assemblages in the Etowah River Basin, USA. Hydrobiologia 494:5-10.

Walters, D. M., M. C. Freeman, D. S. Leigh, B. J. Freeman, and C. M. Pringle. 2005. Urbanization effects on fishes and habitat quality in a Southern Piedmont river basin. Pages 69-85 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization of stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 2001:255-266.

Wang, L., and P. Kanehl. 2003. Influences of watershed urbanization and instream habitat on macroinvertebrates in cold water streams. Journal of the American Water Resources Association 39:1181-1196.

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

Wang, L., J. Lyons, P. W. Rasmussen, P. W. Seelbach, T. Simon, M. J. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. M. Stewart. 2003b. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Canadian Journal of Fisheries and Aquatic Sciences 60:491-505.

Wang,L., T. Brenden, P. W. Seelbach, A. Cooper, D. Allan, R. Clark, Jr., and M. Wiley. 2007. Landscape based identification of human disturbance gradients and references for Michigan streams. Environmental Monitoring and Assessment. Online-First (Not yet assigned volume or page numbers).

WDNR (Wisconsin Department of Natural Resources). 2007. Landcover dataset (WISCLAND). Available: www.dnr.state.wi.us/maps/gis/datalandcover.html (December 2007).

Weaver, L. A., and G. C. Garman. 1994. Urbanization of a watershed and historical changes in a stream fish assemblage. Transactions of the American Fisheries Society 123:162-172.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2006. Influence of landscape features on summer water temperatures in Lower Michigan streams. Pages 113-127 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Wootton, J. T. 1994a. Predicting direct and indirect effects: An integrated approach using experiments and path analysis. Ecology 75:151-165.

Wootton, J. T. 1994b. The nature and consequences of indirect effects in ecologial communities. Annual Review in Ecology and Systematics 25:443-466.

Zar, J. H. 1999. Biostatistical Analysis, 4th edition. Prentice Hall, Upper Saddle River, New Jersey.

Zorn, T. G., and M. J. Wiley. 2004. Untangling relationships between river habitat and fish in Michigan's Lower Peninsula with covariance structure analysis. Michigan Department of Natural Resources, Fisheries Research Report 2073, Ann Arbor.

# Chapter 6

# Conclusions

Throughout this dissertation, I have shown how fish distribution models, with particular emphasis on those created through a classification tree methodology, can be used for two main purposes: formulating and testing hypotheses about the factors, disturbances, and processes that are important to organisms, and making predictions of species distributions and abundances for use in management contexts.

My first goal was to examine how different landscape factors and disturbance features relate to stream fish. In chapter 2, I found that a variety of techniques could be used to predict that brook trout have a preference for low water temperature, small streams, high amounts of forest, and high groundwater flow. The next step was to expand this modeling procedure to numerous fish species using a classification tree approach, an useful technique that does not rely on the underlying data distribution and produces models easy to interpret and apply to new scenarios. In chapter 3, I created distribution models for all common Michigan stream fish. The results indicated that when using variables measured on a landscape scale, it is possible to predict most stream species with a high level of accuracy; although certain groups of fish were more easily predicted than others. Water temperature, which has directly controls the level of dissolved oxygen, and catchment area, which is highly correlated with the size and flow of a river, were the two most influential variables that drove the species distribution. Models that predicted the fish most accurately were those controlled most directly by these two variables; fish found in big, warm rivers such as redhorse species, channel catfish, and common carp were predicted very well, as were fish found in small, cold streams such as brook trout, mottled sculpin, and slimy sculpin. The importance of these variables has also been supported by other studies (Hawkes 1975; Vannote et al. 1980; Fausch et al. 1998; Wiley et al. 1990; Lyons 1992; Zorn et al. 2002; Wehrly et al. 2003).

The classification tree model gives a quantitative prediction of the niche space of the species; the habitat ranges in which the fish is expected to reside. For most species the habitat space is first defined by water temperature and catchment area. The other variables included in the classification trees are examples of finer “filters”. Water temperature and catchment area control whether the stream system has the potential to hold the fish, while the other variables control the particular spatial location where the fish are found. For example, there is a low chance (10%) of largemouth bass being located in streams with a daily July mean water temperature less than 18.9 °C. Streams that are above this threshold however, have a moderate chance (62%) of holding largemouth if the stream is also 20 km upstream from a pond. In this example, water temperature is the first filter. Once a stream meets this requirement (in other words, passes through this filter), the distance to a pond becomes important in determining largemouth presence. In a cold stream (<18.9 °C), the distance to a pond does not matter, as there is only a low chance of largemouth occurrence despite this distance.

In chapter 4, I used the classification tree models to show how past and future land-use change and climate change are expected to shape the game fish communities in the Muskegon River stream system. For the future predictions, I developed three potential climate change and land-use change scenarios and applied them to the predictive fish models to create the potential distribution of fish from 2010-2100. Given a water temperature increase of approximately 4 °C, the models predicted virtual eradication of the brook trout, rainbow trout, and Chinook salmon in the Muskegon watershed, and a severe decline of brown trout. The distribution of warm-water fish is expected to spread, with the exception of walleye, which was predicted to decline in the Muskegon due to increases in urban development.

Chapter 5 examined how different landscape factors and in-stream processes affect urban stream fish communities. The increase of streams affected by urbanization is an issue of top concern for aquatic ecologists because a variety of urban impacts have negative consequences on stream communities. Analysis of the variation around the relationship between fish IBI and stream urbanization has shown that agricultural impacts are a major factor in whether urban streams are able to support good fish communities. Urban streams in watersheds that are primarily agriculture and urban are much more likely to have degraded fish communities than urban streams with sufficient forest in the watershed. Healthy fish communities in urban streams also need good water quality and stable flow. Discharges from point sources reduce water quality, but provide steady flow that is mechanistically similar to ground water.

Throughout this dissertation, the models that indicate how landscape factors and disturbances affect the stream fish community also can be used strengthen new concepts in stream fisheries management. For example, the analyses in this dissertation are based on the idea that fish species are influenced by processes that operate on larger spatial scales and slower temporal scale than those measured at the local-scale (Chapters 2-5; Richards et al. 1996; Leftwich et al. 1997; Rathert 1999; Allan 2004). Therefore, effective conservation management will need to be applied at the proper scale; evidence from this dissertation has indicated that managers need to plan on a watershed level, not on a riparian level. In chapter 3, land-use/cover on a watershed scale was used in the classification trees more often than land-use/cover on the riparian scale (16.7 % of presence/absence models vs. 13.4%). In chapter 5, riparian scale land-use/cover did not explain any variation in the difference between observed and potential IBI in urban streams that was not previously accounted for by the watershed land-use/cover. Managing on a watershed scale is not a new idea (Wang et al. 2001, Wang et al. 2003; Fitzpatrick et al. 2005) but the riparian management paradigm continues to be very popular (Bernhardt et al. 2005), probably due to the ease of working on a small-scale. While management at a local or riparian-scale certainly can produce favorable changes in fish communities, evidence from this dissertation suggests that operating on the watershed scale will be more effective.

Models built on a landscape-scale are decision-making tools able to be used in a variety of management and conservation applications. At their most basic use, these models predict the amount and location of the riverine habitat suitable for common fish species in Michigan. In situations where a manager has little information and needs a starting point or confirmation of an idea, these models and resulting maps provide baseline data. Inventory information is a vital component to fisheries management and species conservation, and the modeling described here is a good way to get this data on a large geographic scale. Managers can also use the models to aid their fish sampling and stream assessment work. The models can be used to identify potential high-quality "reference" streams and low-quality "impaired" sites. The models can also be used to identify streams that have a good restoration potential. For example, managers could predict if adding forest land-cover in the stream's watershed would have a positive effect on the fish community, or if the buffer would have little effect because the stream has low overall potential regardless of land-use management.

These models can be useful for the management of particular species. For some fish, a manager can rule out the presence of a fish based on a single factor. I found that trout species were unlikely to be found in streams with mean daily July water temperature over a particular value (brook trout, 19.4°C; brown trout, 20.2°C; rainbow trout, 19.6°C). This information combined with the ability to access water temperature on a GIS would be very useful to managers deciding whether to manage marginal streams for trout.

The models can be used identify streams that should be sampled for rare species or species of concern. Besides looking at streams where the fish has been found in the past, it is difficult to know where else the fish may reside. However, due to anthropogenic pollution and siltation impacts on streams over the past century, rare fish are not found where the models predict them to be located and therefore models of rare fish have high rates of false presence errors. While the predictions of rare fish were inaccurate compared to the test data in this study, these models still have practical management use because they predict the habitat space where the fish have the potential to reside.

To ensure long-term sustainability of aquatic resources, anticipating future changes in fish communities is an essential task. Knowledge of what may happen if we fail to act can provide both the motivation to act and indicate what steps may be necessary to prevent the predicted changes from occurring. In Chapter 4, I applied three “what-if” future scenarios to the classification tree models and saw that fish communities in the Muskegon River system, and by extrapolation, the fish communities throughout Michigan, will be much different in 2100 than today. While the problem of climate warming may be out of the hands of fisheries managers, it is not impossible to work for changes in land-use development in order to prevent some of the predicted future changes from coming true. The classification tree models give land-use thresholds that managers will not want to cross. For example, a big river (catchment areas greater than 657 $km^2$) with a watershed of less than 8.5% urban land-use has a much high probability to maintain walleye populations than a stream high in urban land (57% vs. 26%, Appendix I).

As mentioned above, the goals of this dissertation were to examine how models can be useful in formulating and testing hypotheses about the factors, disturbances, and processes that are important to organisms, as well as providing practical fisheries management tools. The models we have used do this are built on a landscape level, using correlation fish-habitat associations. Yet using these methods brings forth an inherent weakness. The models in chapters 2-4 have the ability to accurately predict fish distributions, but these models do not give any explanation of biological mechanisms. The analysis Chapter 5 represents a new and useful way of conceptualizing the issue of fish communities and urban streams, but the CSA analysis only explained 39% of the variation in the difference between observed and potential IBI score so there remains much to be explored. Throughout the entire dissertation, I do not consider the effect of competition, predator/prey relationships, or other species interactions, and am not able to include the effect of localized habitat features. The lack of clear biological relationships in the results is unfortunate. However, what this dissertation does provide is a large spatial scale, which relates better to temporally slower geological and hydrological mechanisms. Studies integrating local-scale and large-scale variables and using biological interactions as well as geological/hydrological processes will be difficult to carry out, but represent a possible avenue for improving this work in the future.

Temperature change, land-use change, and urbanization are realities that aquatic scientists are going to need to understand in order to develop the tools needed to conserve aquatic diversity. This dissertation shows that landscape-scale habitat variables partnered with GIS, classification trees, and covariance structure analysis can be used to sharpen our knowledge of how these disturbances affect stream fish and provide practical tools to aid in our management of aquatic systems.

## References

Allan, J. D., D. L. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology 37:149-161.

Bernhardt, E. S., M. A. Palmer, J. D. Allan, G. R. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, D. Galat, S. Gloss, P. Goodwin, D. Hart, B. Hassett, R. Jenkinson, S. Katz, G. M. Kondolf, P. S. Lake, R. Lave, J. L. Meyer, T. K. O'Donnell, L. Pagano, B. Powell, and E. Sudduth. 2005. Synthesizing U.S. river restoration efforts. Science 308:636-637.

Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream fish from habitat variables: 1950-1985. United States Department of Agriculture, Forest Service, Pacific Northwest Research Station, General Technical Report PNW-GTR-213, Portland, Oregon.

Fitzpatrick, F. A., M. W. Diebel, A. H. Mitchell, T. L. Arnold, M. A. Lutz, and K. D. Richards. 2005. Effects of urbanization on the geomorphology, habitat, hydrology, and fish index of biotic integrity of streams in the Chicago, Illinois, and Wisconsin. Pages 87-116 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Hawkes, H. A. 1975. River zonation and classification. Pages 312-374 *in* B. A. Whitton, editor. River Ecology. University of California Press, Berkeley.

Leftwich, K. N., P. L. Angermeier, and C. A. Dolloff. 1997. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Transactions of the American Fisheries Society 126:725-734.

Lyons, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. United States Department of Agriculture, Forest Service, General Technical Report NC-149, St.Paul, Minnesota.

Rathert, D., D. White, J. C. Sifneos, and R. M. Hughes. 1999. Environmental correlates of species richness for native freshwater fish in Oregon. Journal of Biogeography 26:257-273.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influence on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Sciences 53:295-311.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:131-137.

Wiley, M. J., L. L. Osborne, and R. W. Larimore. 1990. Longitudinal structure of an agricultural prairie river system and its relationship to current stream ecosystem theory. Canadian Journal of Fisheries and Aquatic Sciences 47:373-384.

Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 2001:255-266.

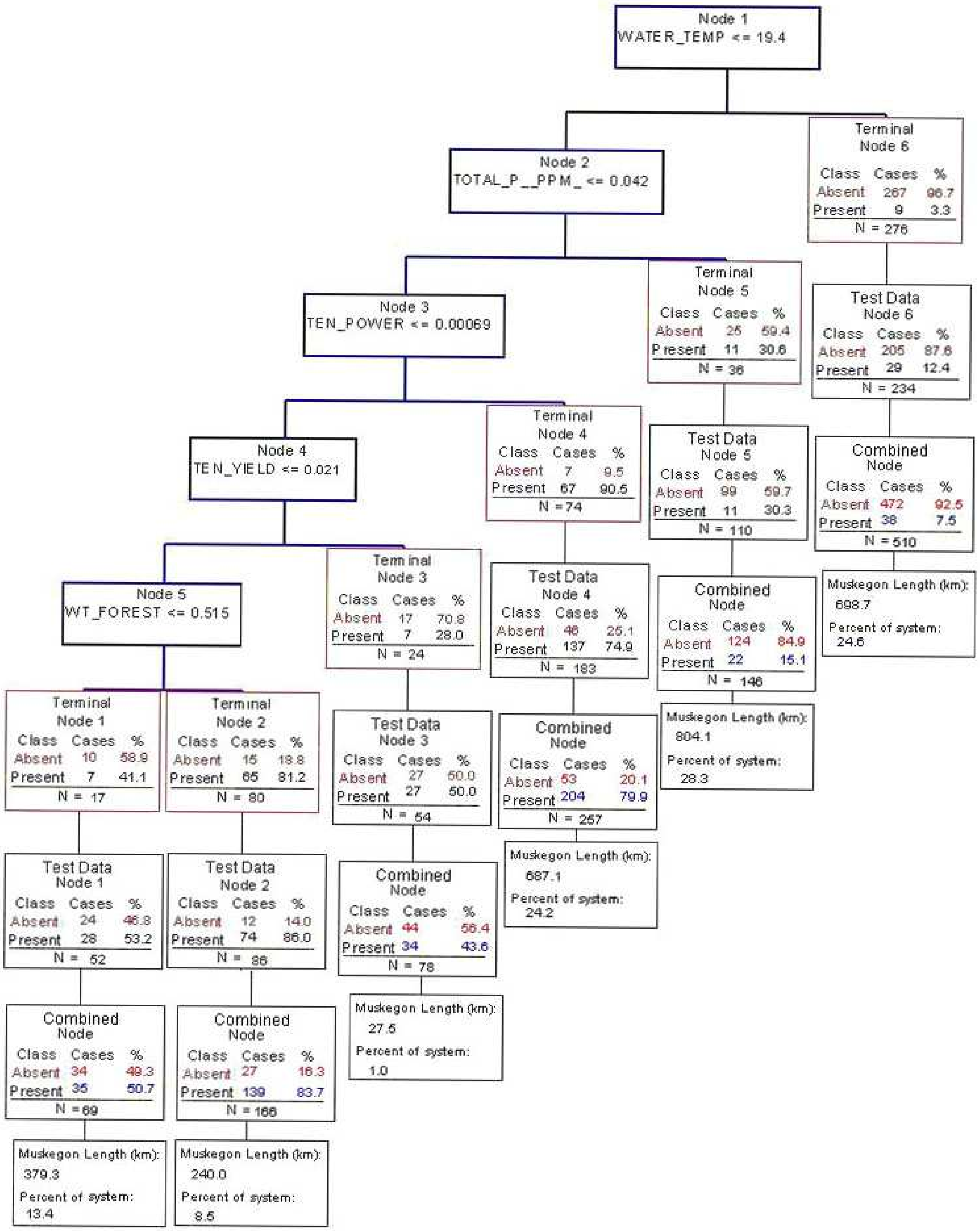
Wang, L., J. Lyons, and P. Kanehl. 2003. Impacts of urban land cover on trout streams in Wisconsin and Minnesota. Transactions of the American Fisheries Society 132: 825-839.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2003. Classifying regional variation in thermal regime based on stream fish community patterns. Transactions of the American Fisheries Society 132:18-37.

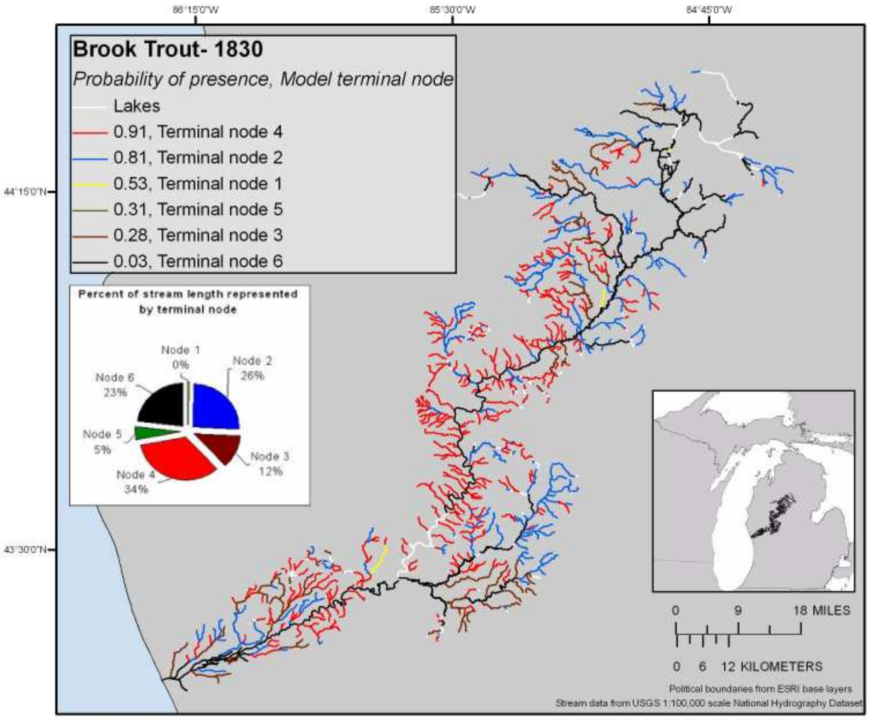
Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's lower peninsula. Transactions of the American Fisheries Society 131:70-85.

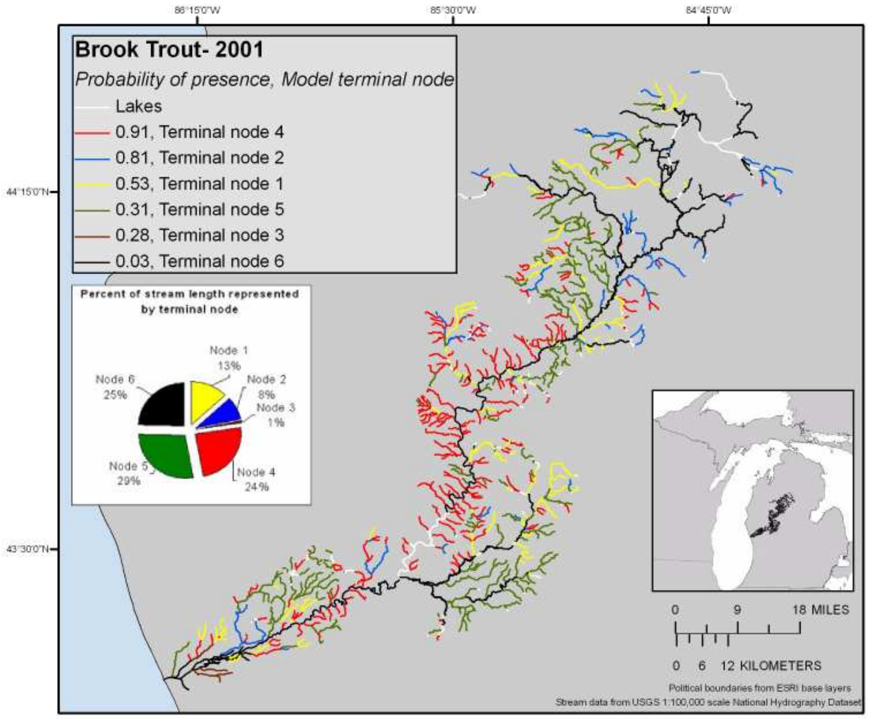
# Appendices

Appendix A. A) Classification tree model for brook trout, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the brook trout model as applied to the years 1830, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

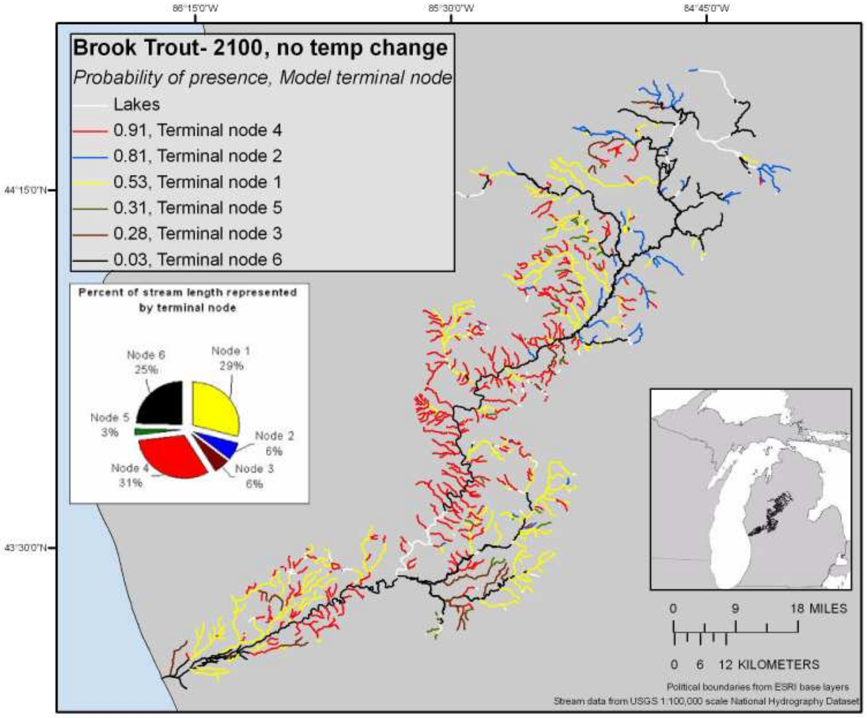


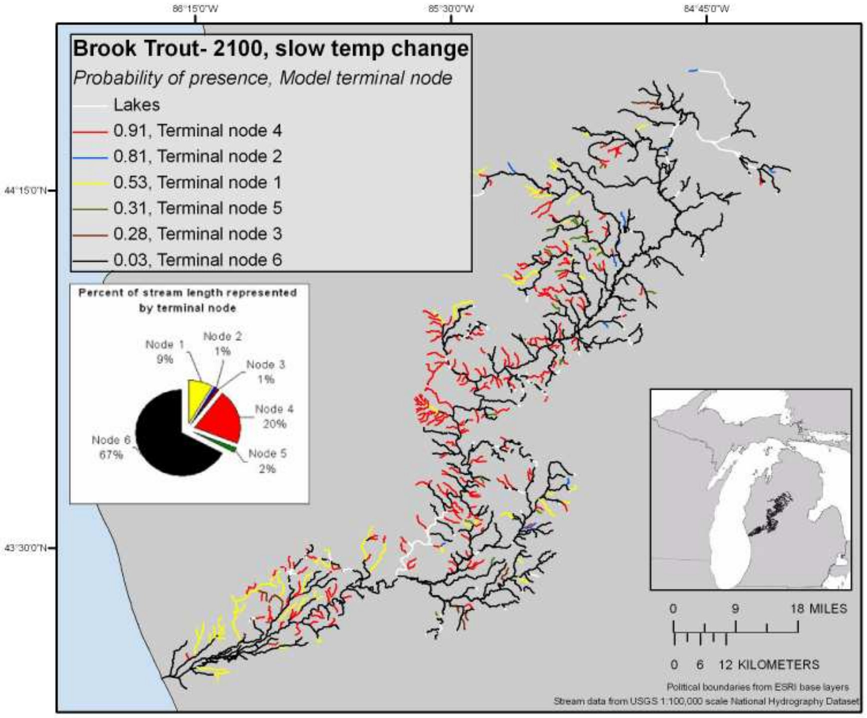
Appendix A, continued.



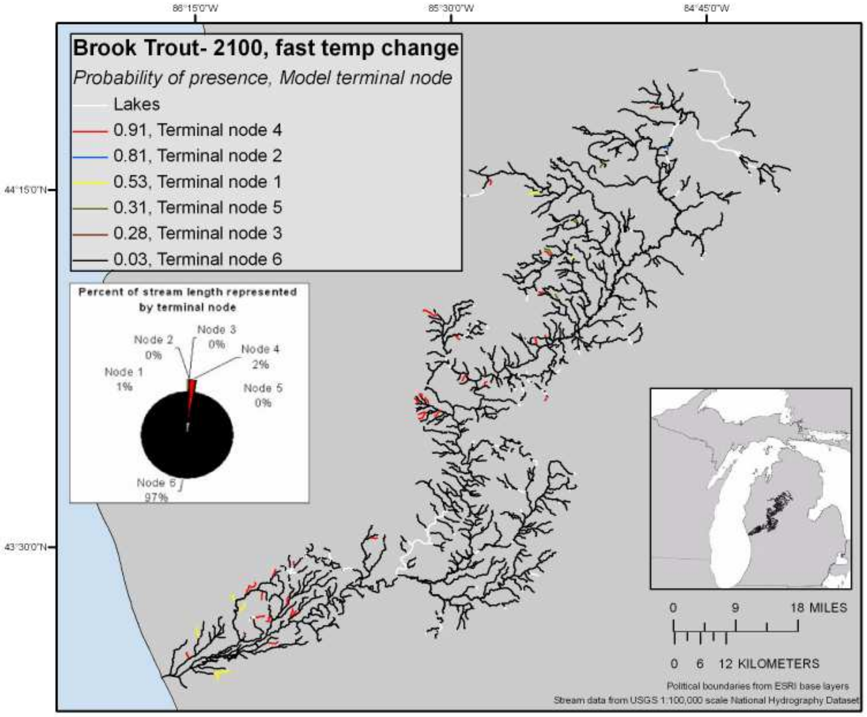


Appendix A, continued.



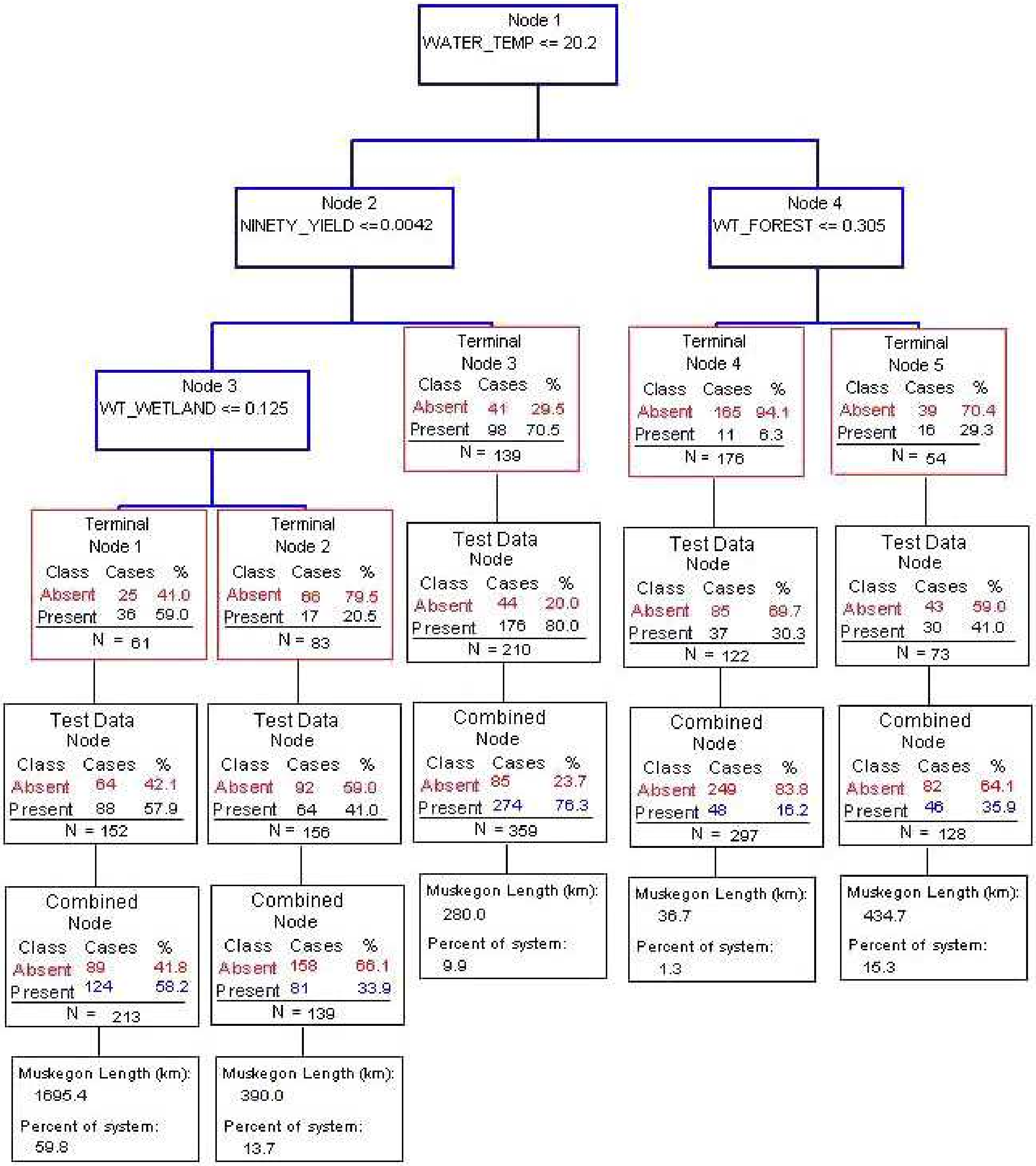


Appendix A, continued.

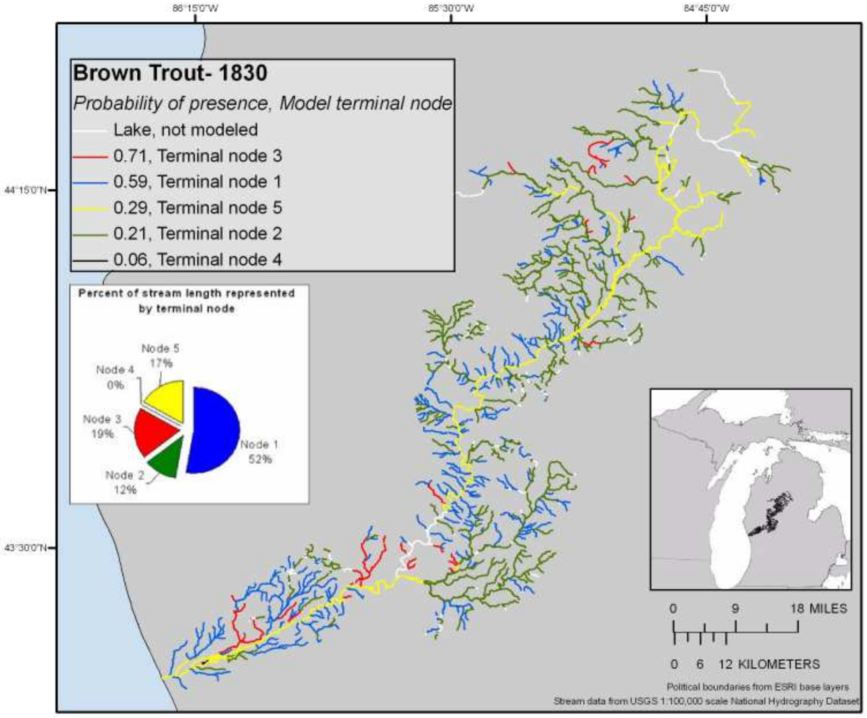


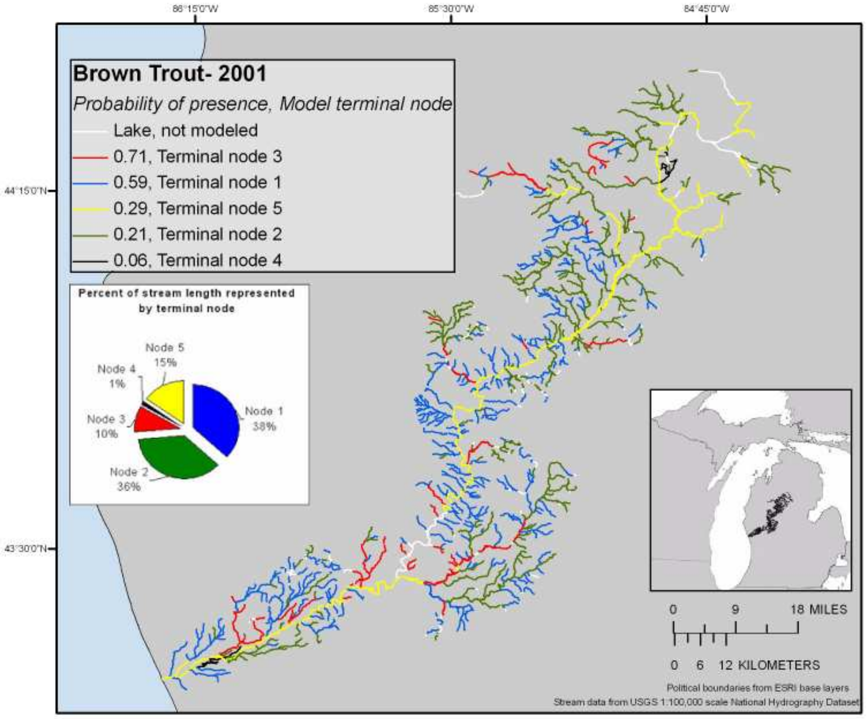
Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale National Hydrography Dataset

Appendix B. A) Classification tree model for brown trout, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the brown trout model as applied to the years 1830, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

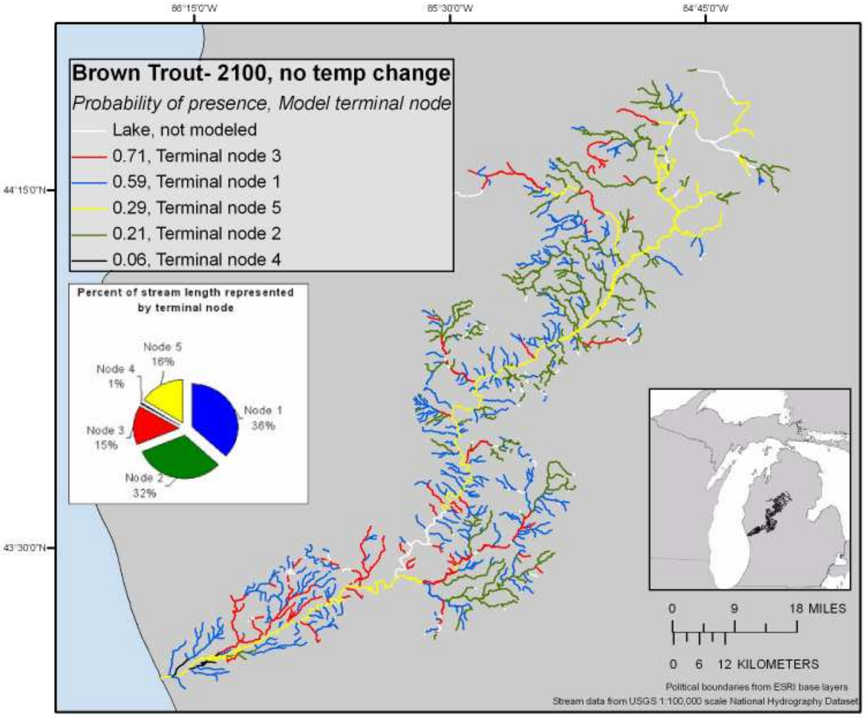


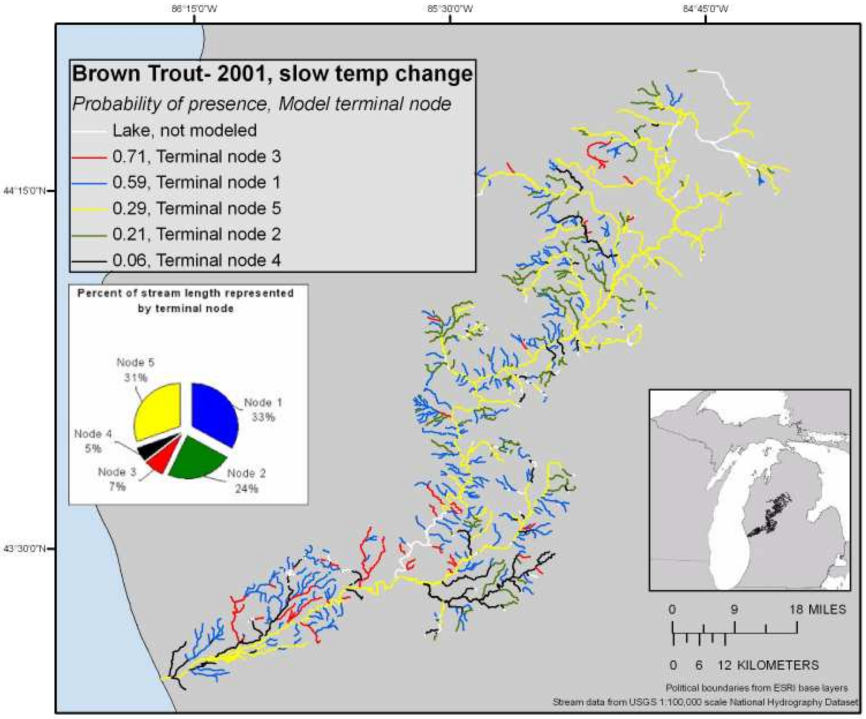
Appendix B, continued.



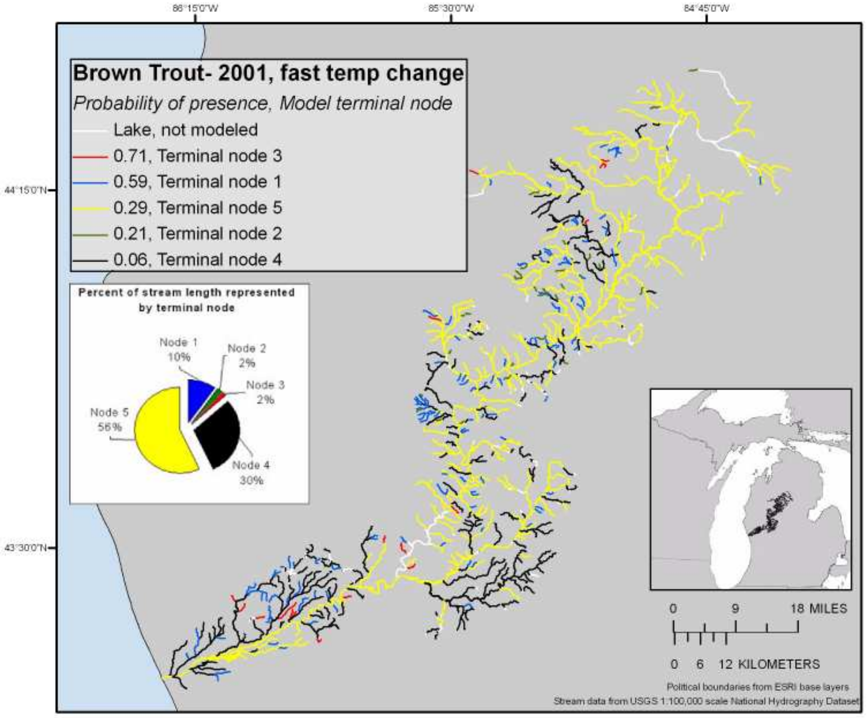


Appendix B, continued.

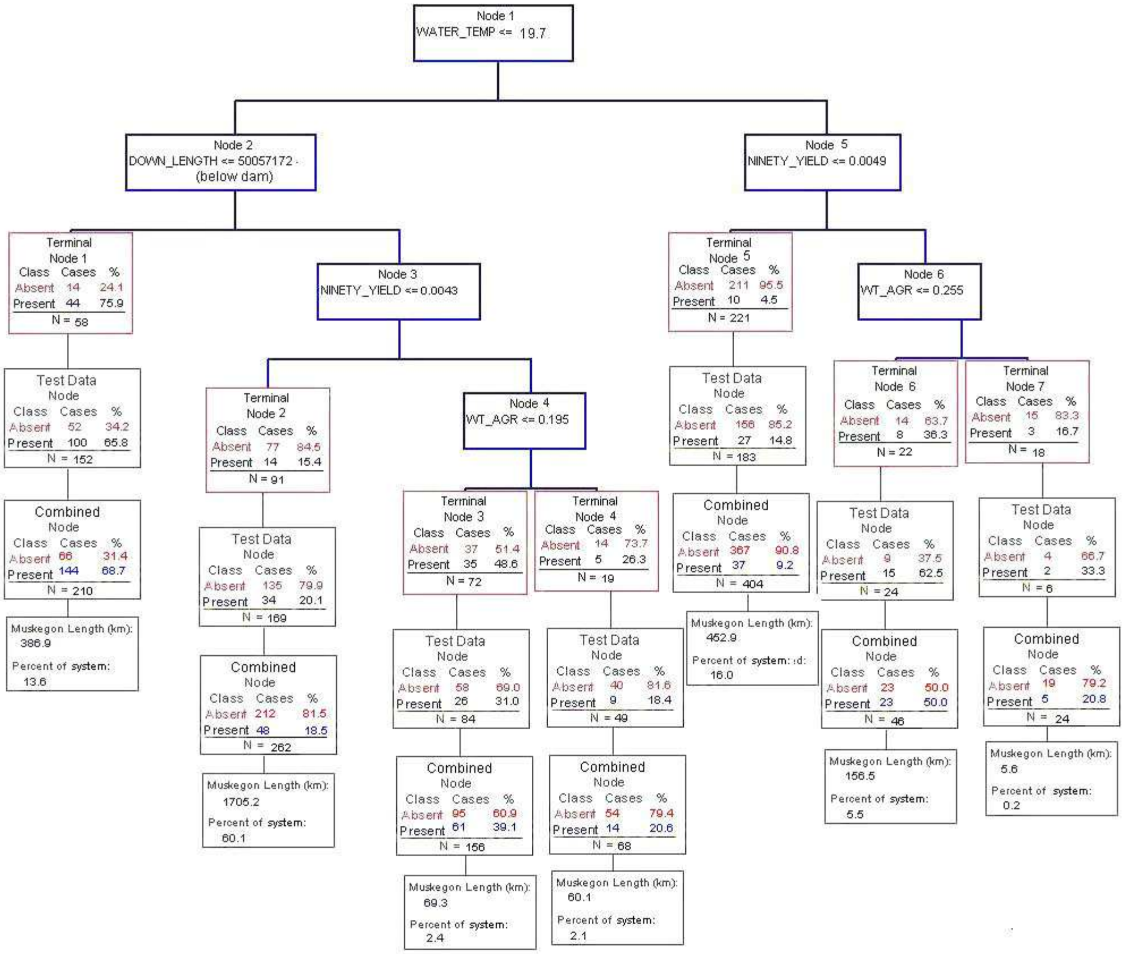




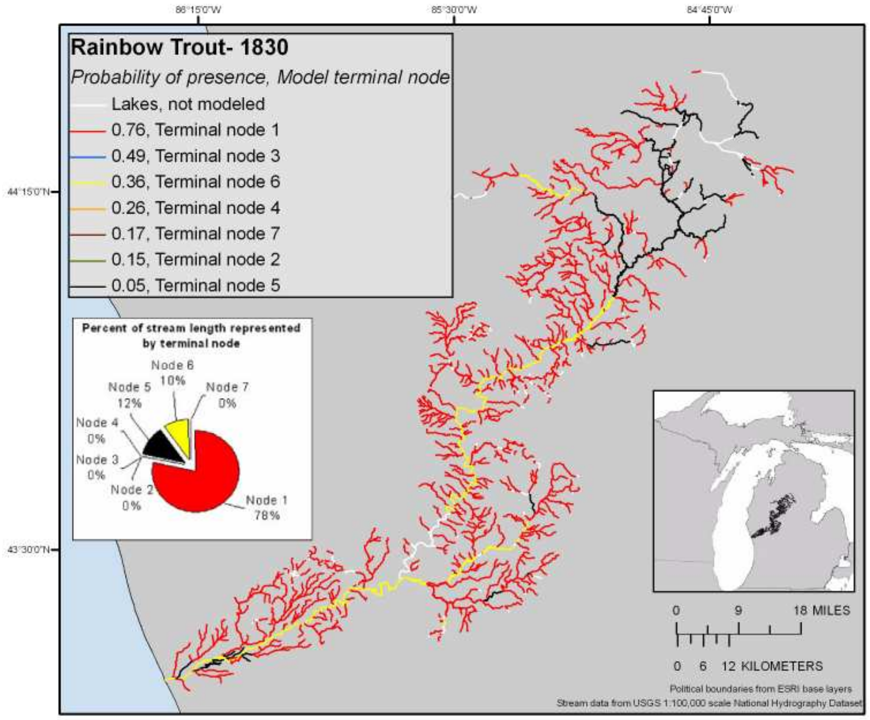
Appendix B, continued.

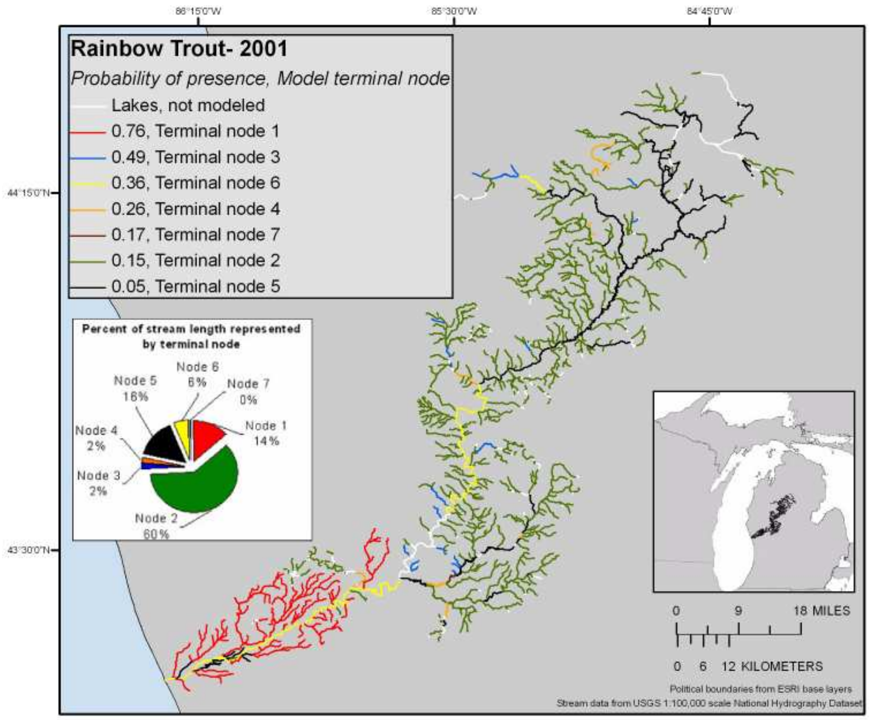


Appendix C. A) Classification tree model for rainbow trout, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the rainbow trout model as applied to the years 1830, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

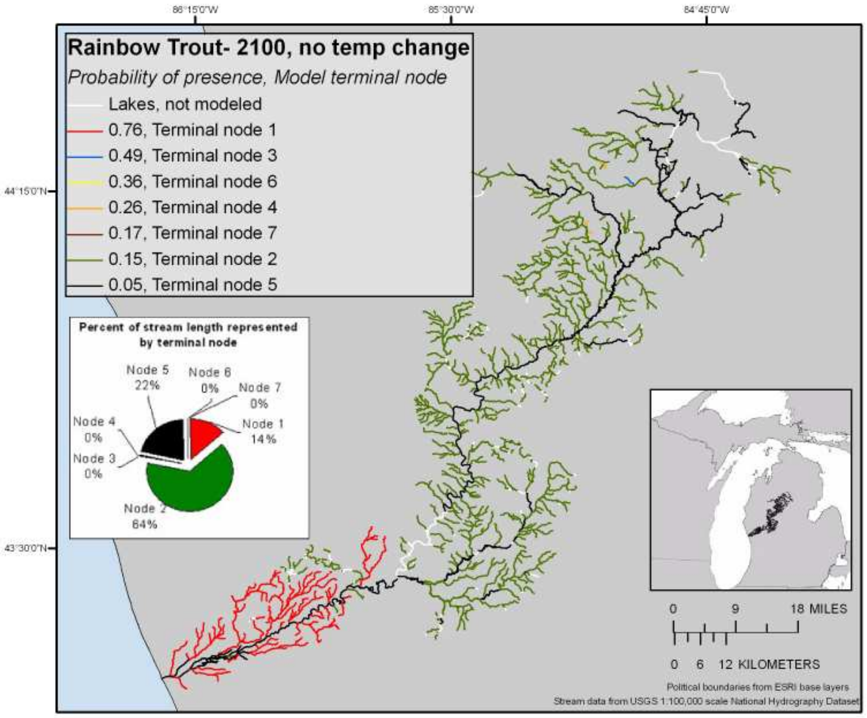


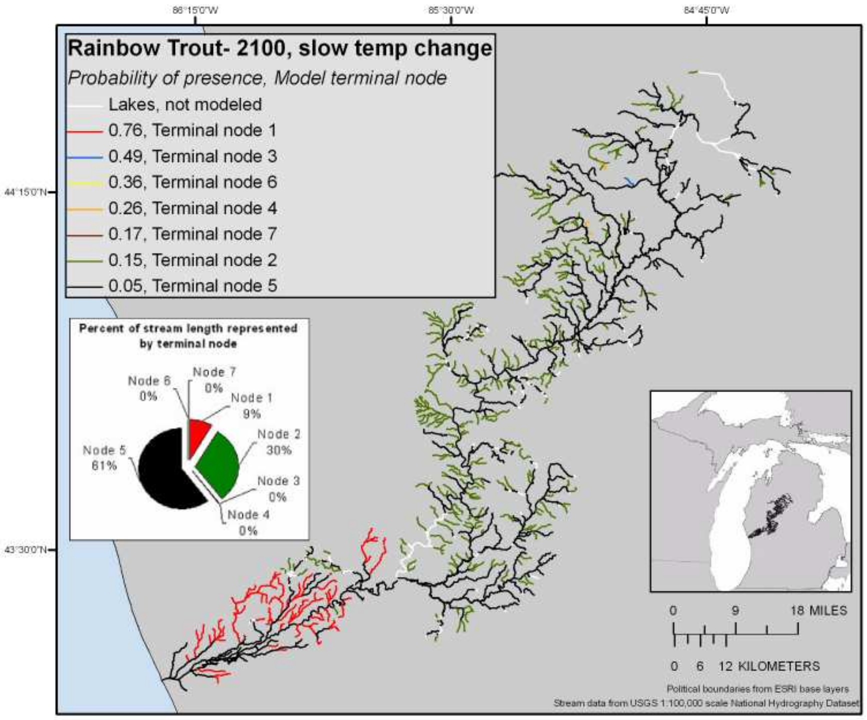
Appendix C, continued.



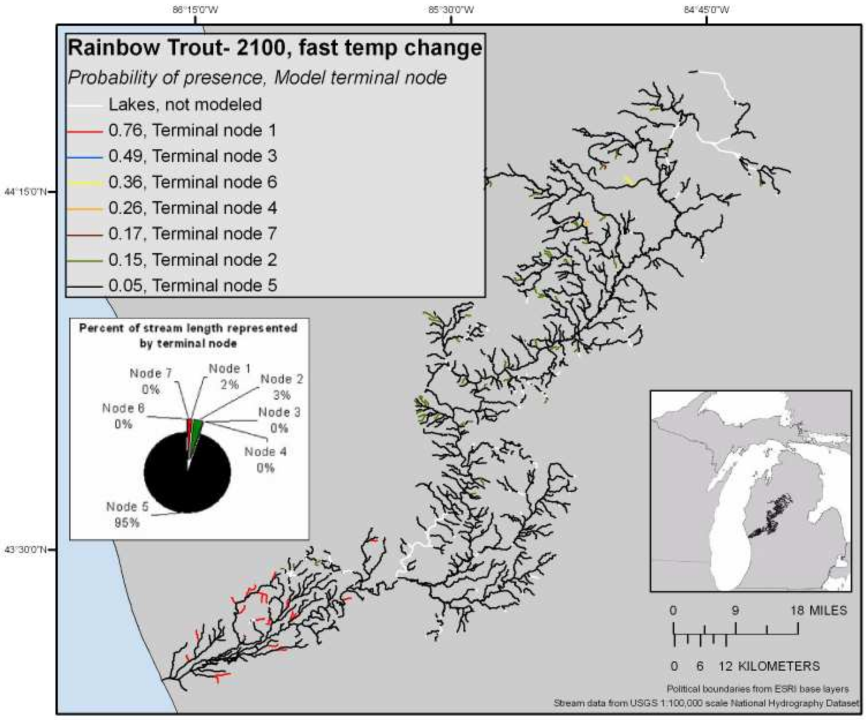


Appendix C, continued.

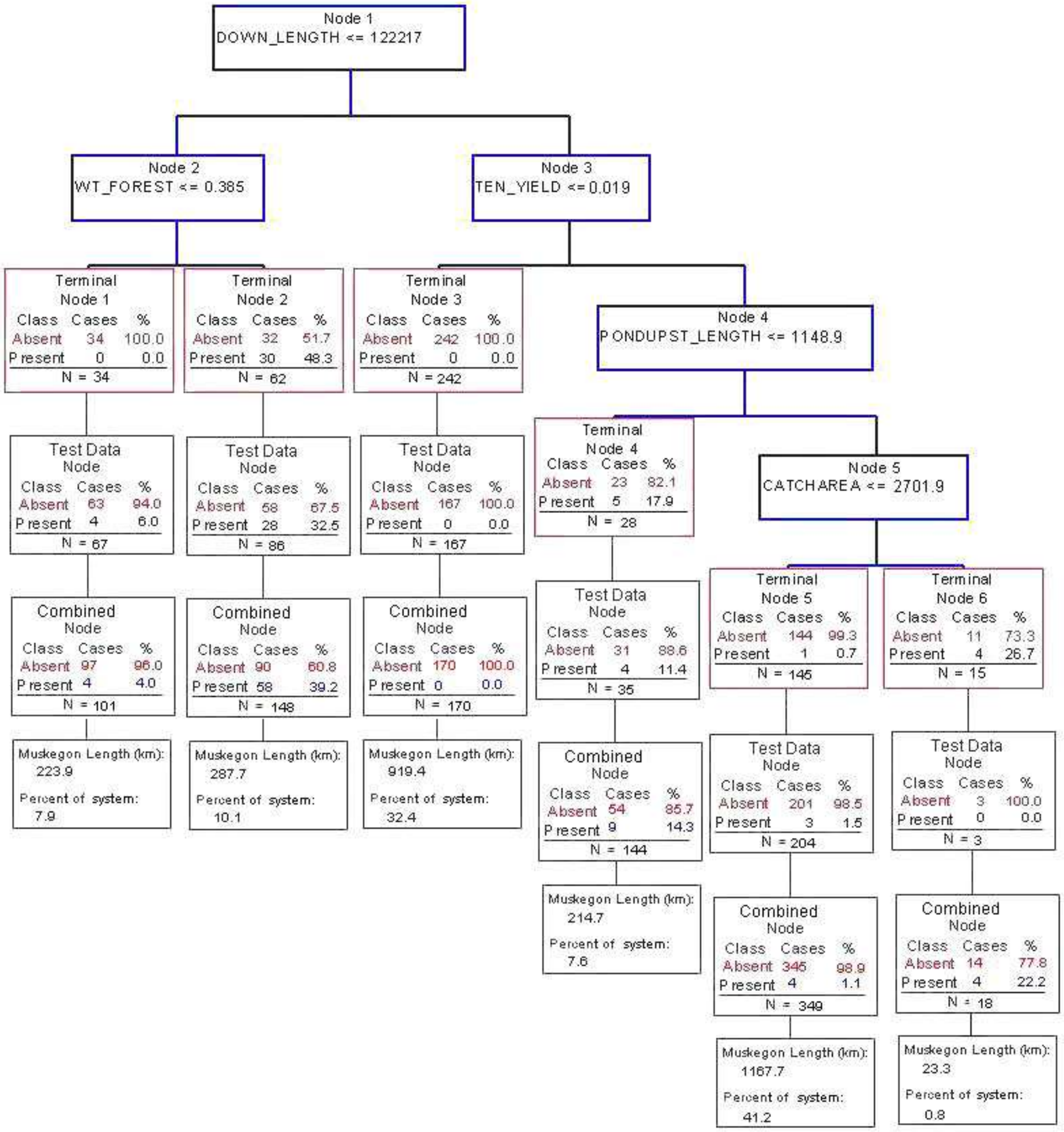




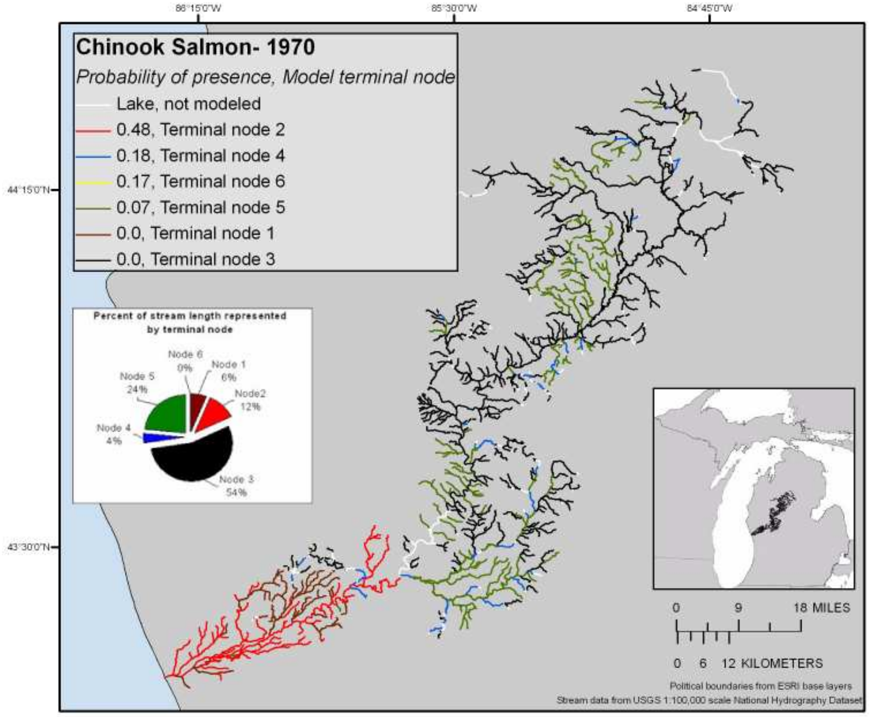
Appendix C, continued.



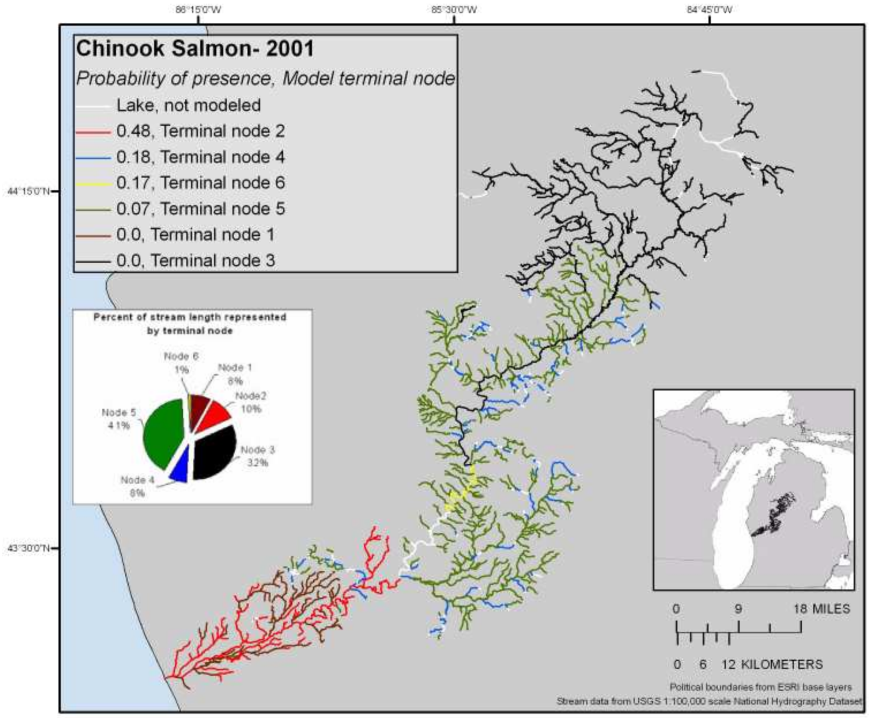
Appendix D. A) Classification tree model for Chinook salmon, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the Chinook salmon model as applied to the years 1970, 2001, and 2100. Since the Chinook salmon model does not include temperature, temperature change future models were not created for this species.



Appendix D, continued.

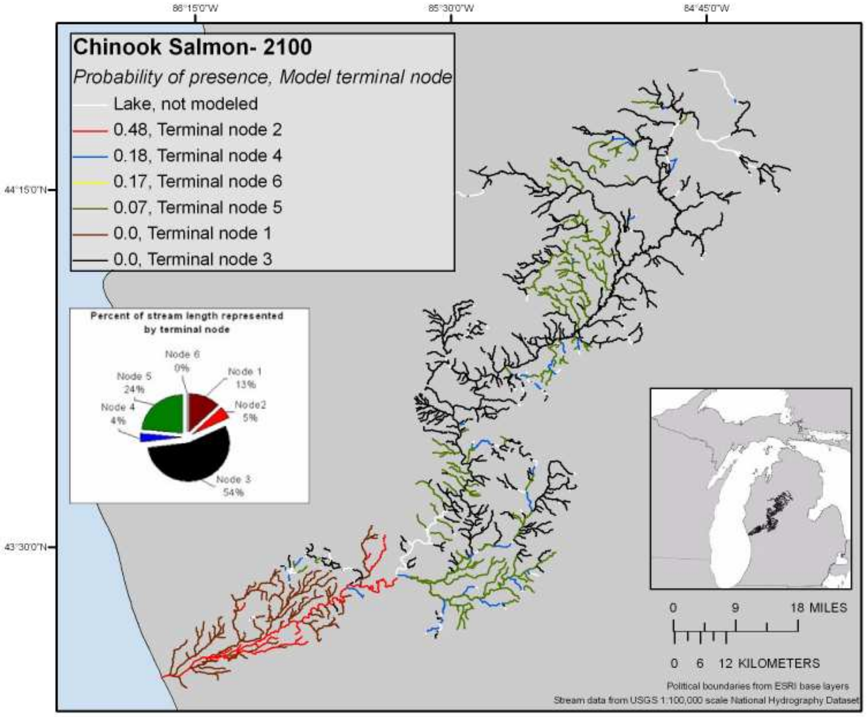


Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale National Hydrography Dataset



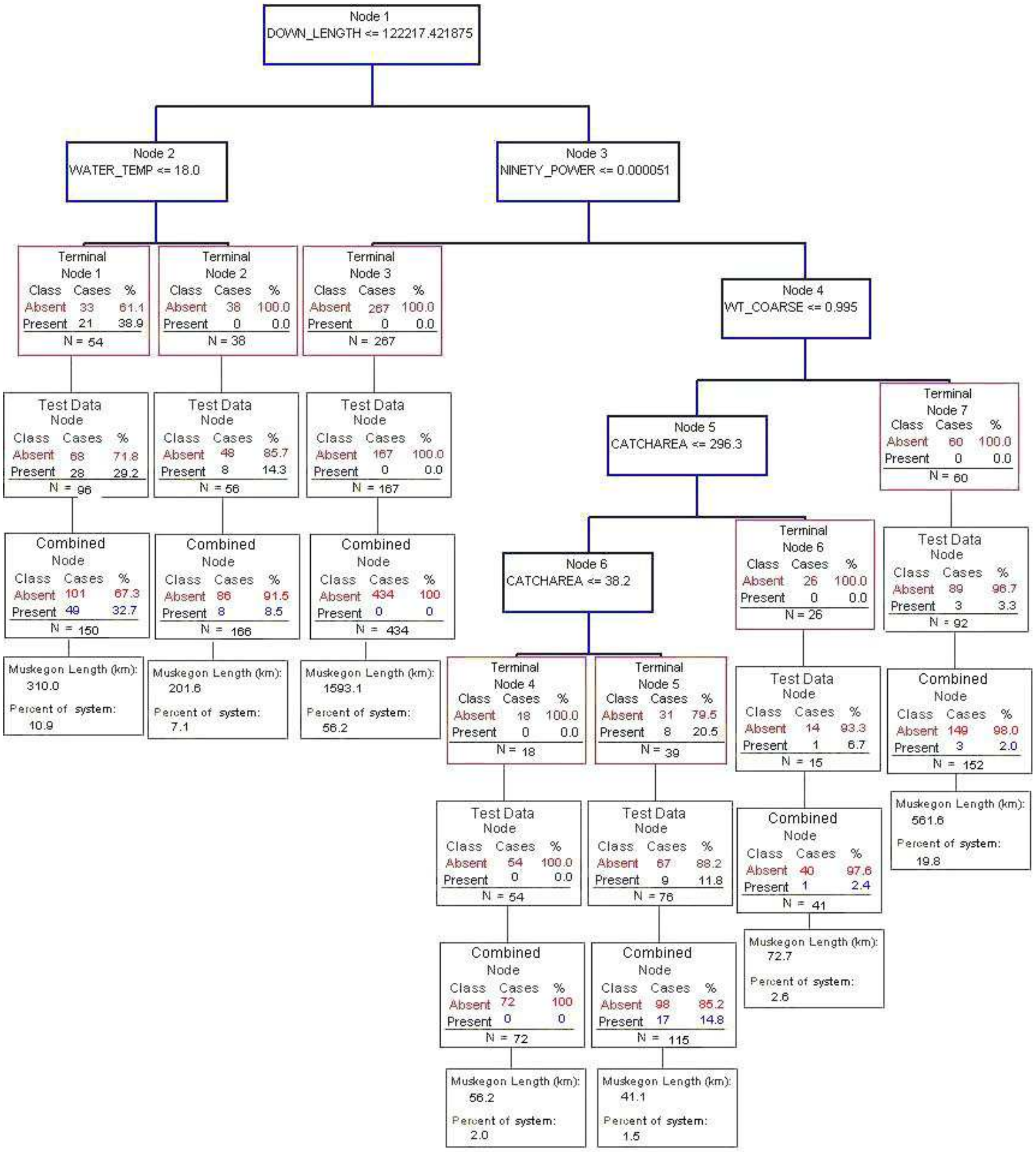
Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale National Hydrography Dataset

Appendix D, continued.

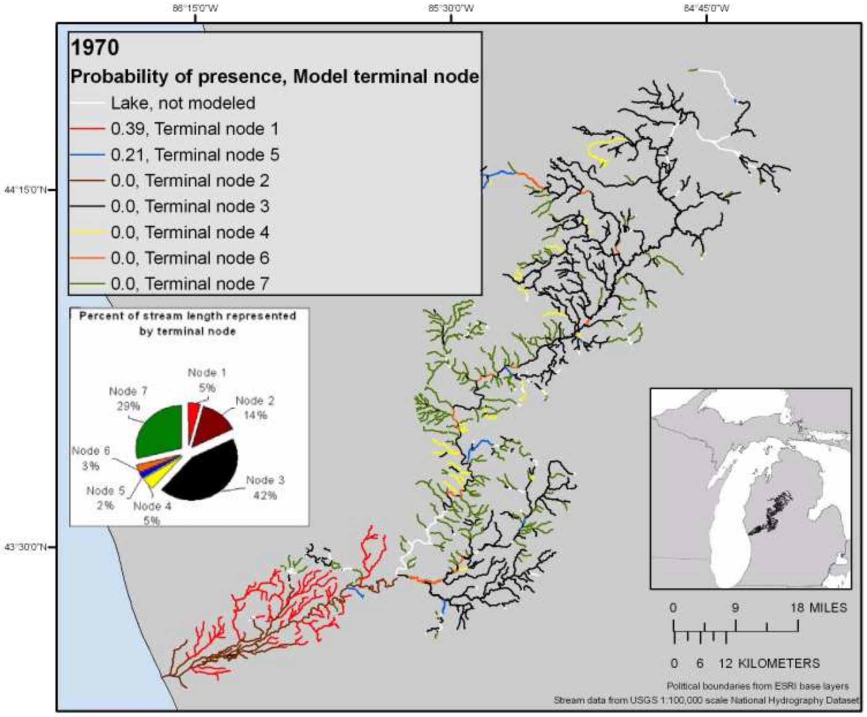


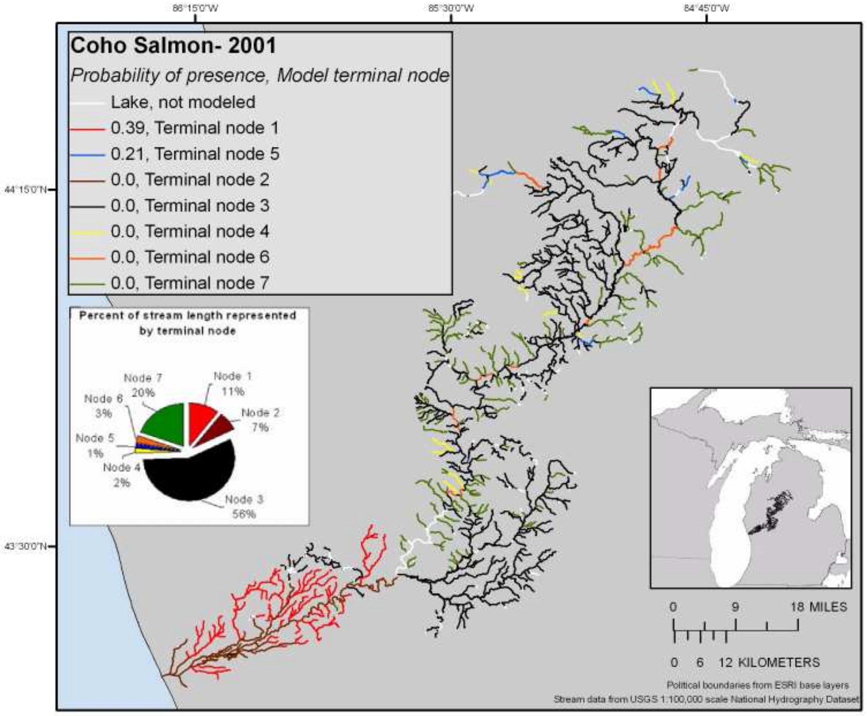
Political boundaries from ESRI base layers
Stream data from USGS 1:100,000 scale National Hydrography Dataset

Appendix E. A) Classification tree model for Coho salmon, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the Coho salmon model as applied to the years 1970, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

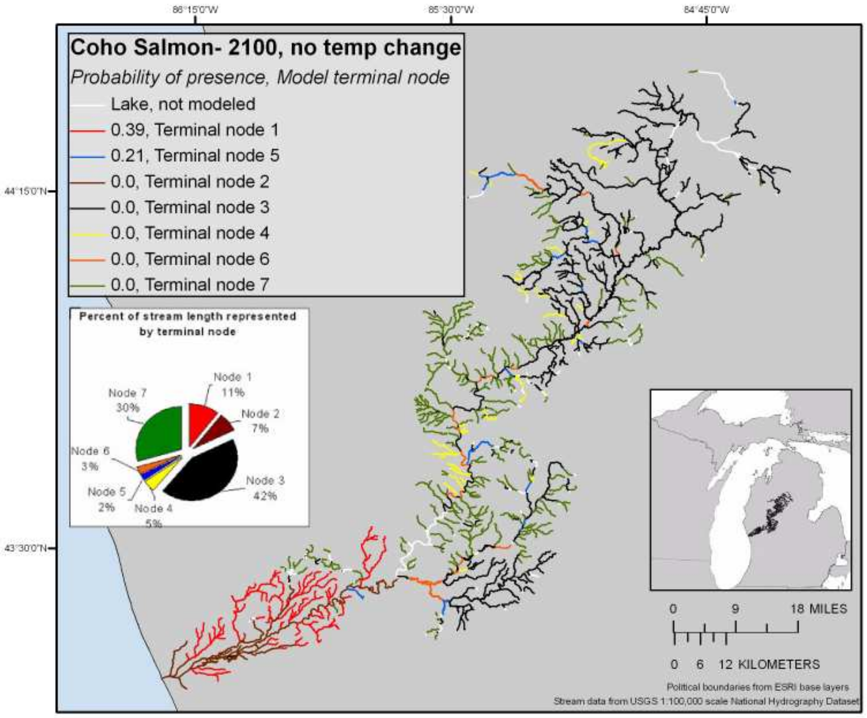


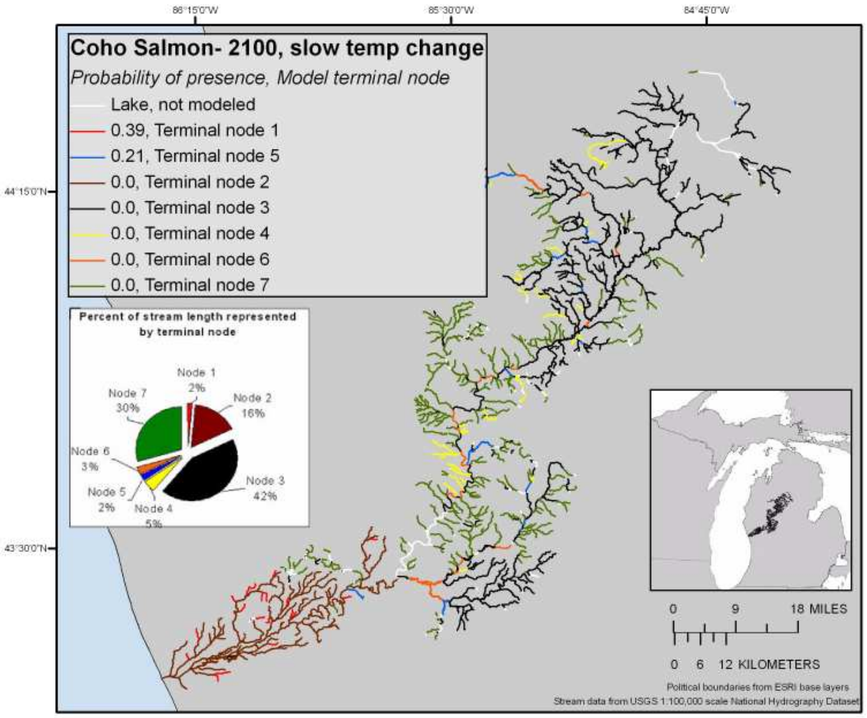
Appendix E, continued.



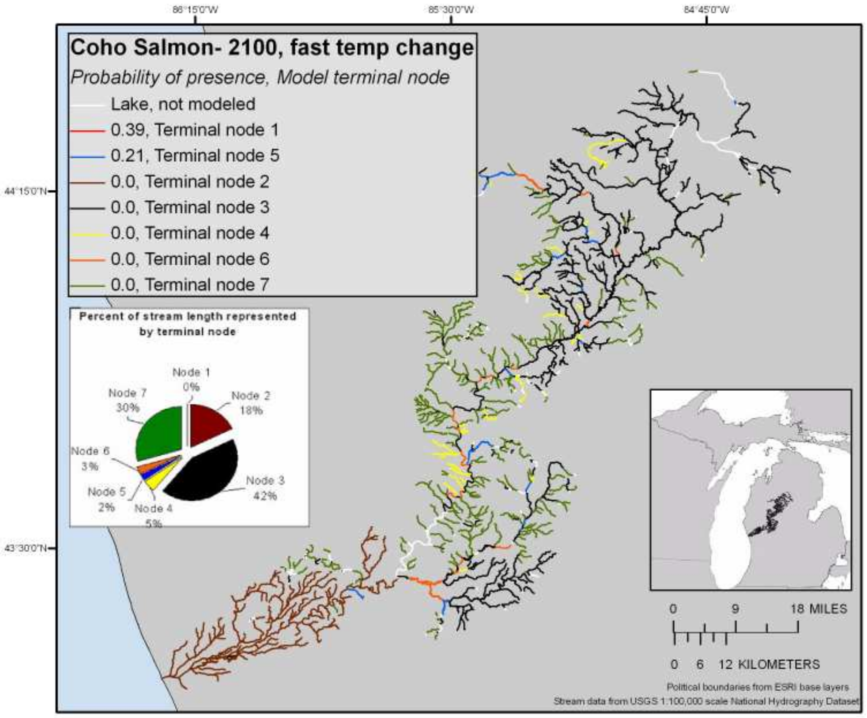


Appendix E, continued.

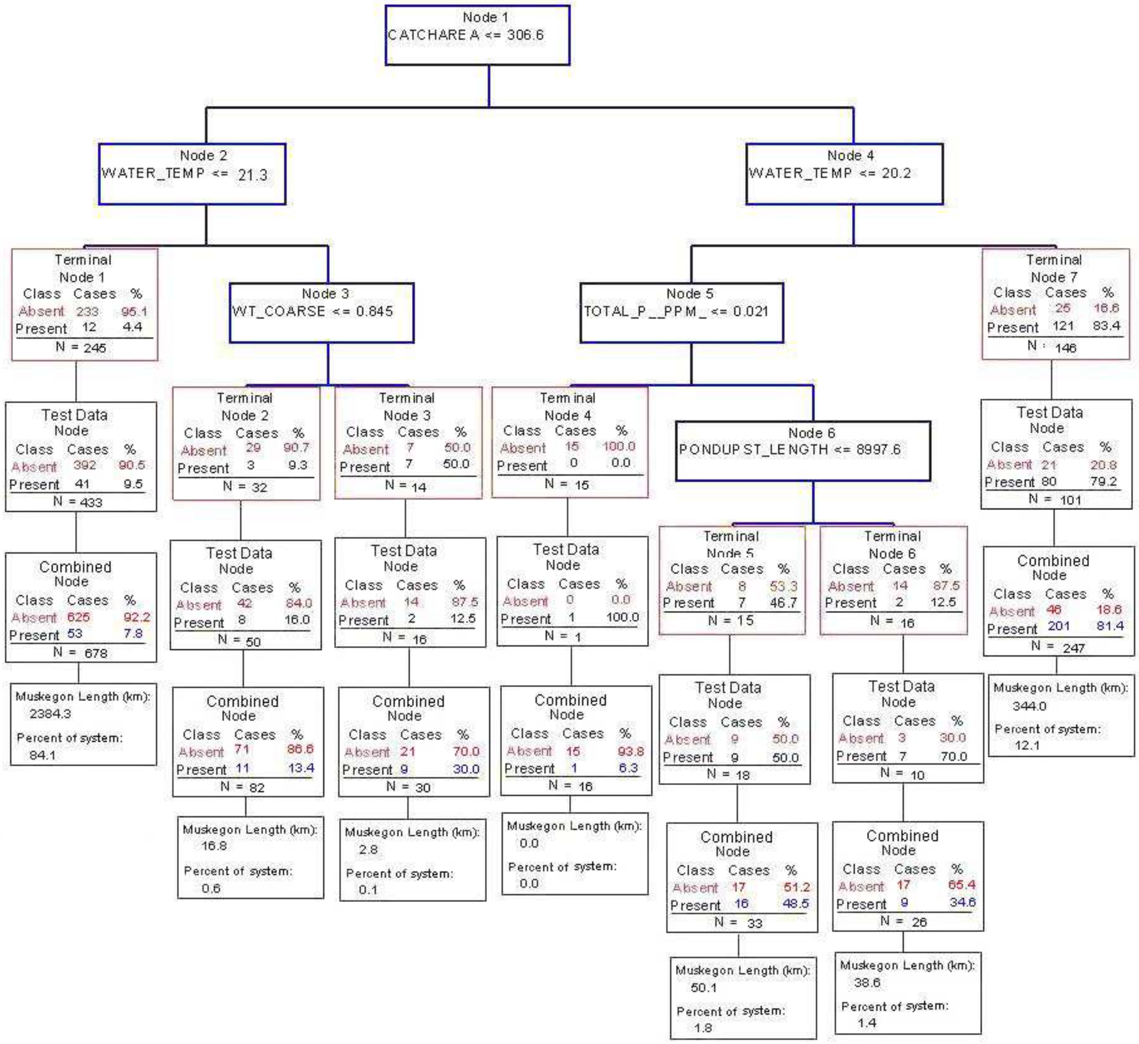




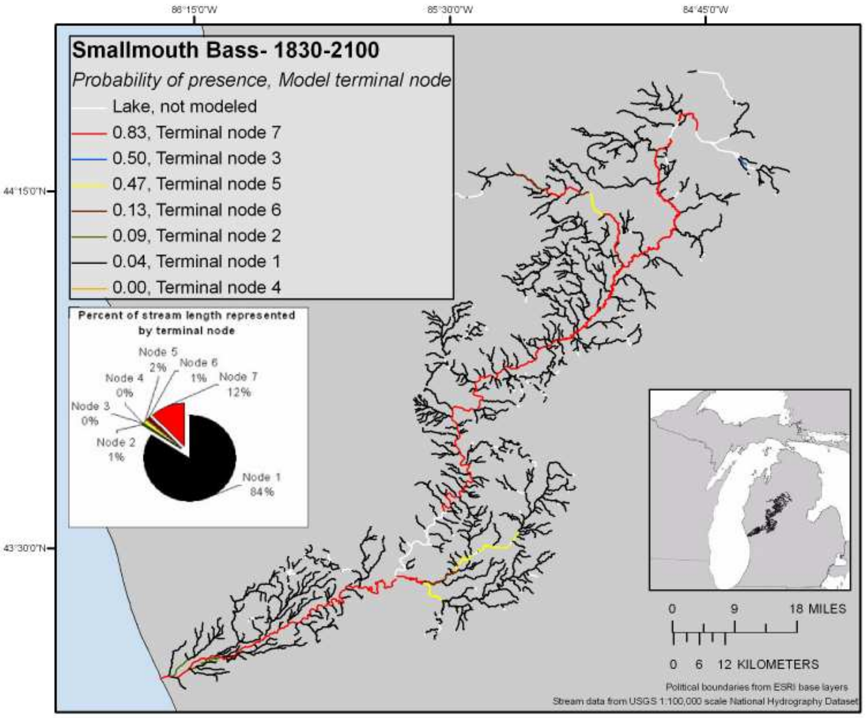
Appendix E, continued.

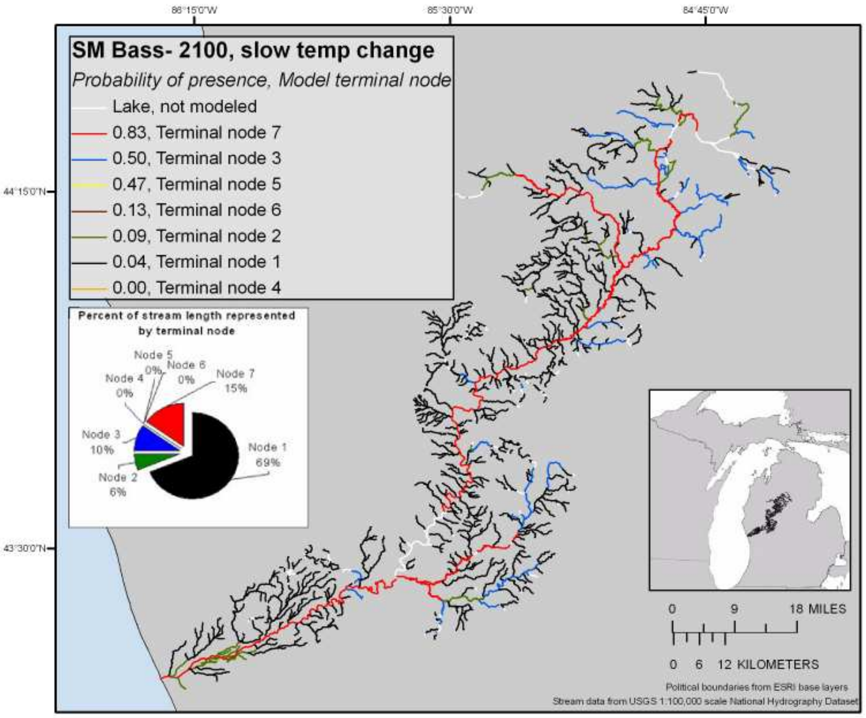


Appendix F. A) Classification tree model for smallmouth bass, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the smallmouth bass model as applied to the years 1830, 2001, and 2100. 1830, 2001, and 2100 (no temp change) have the same prediction.

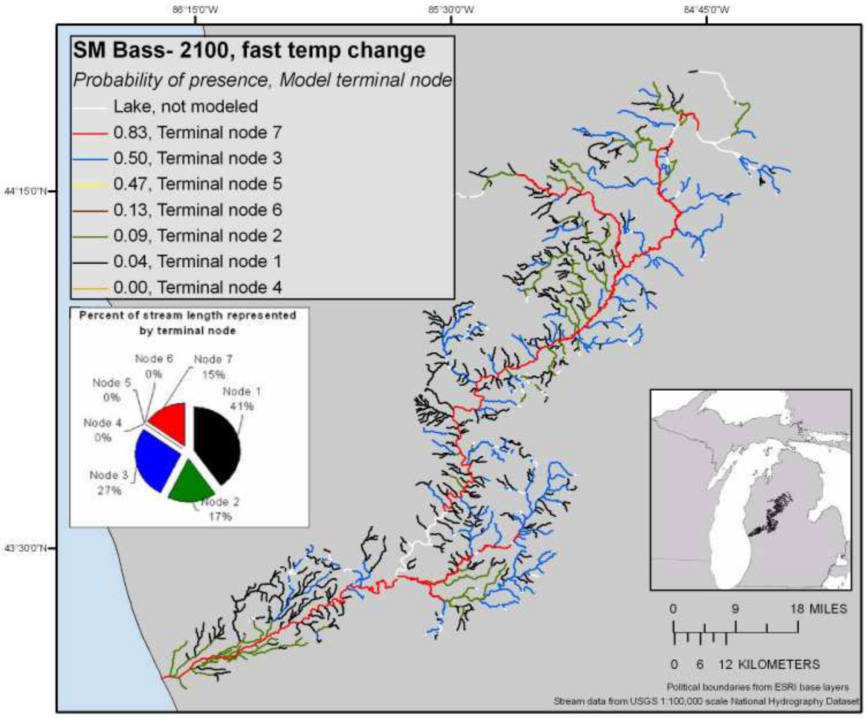


Appendix F, continued.

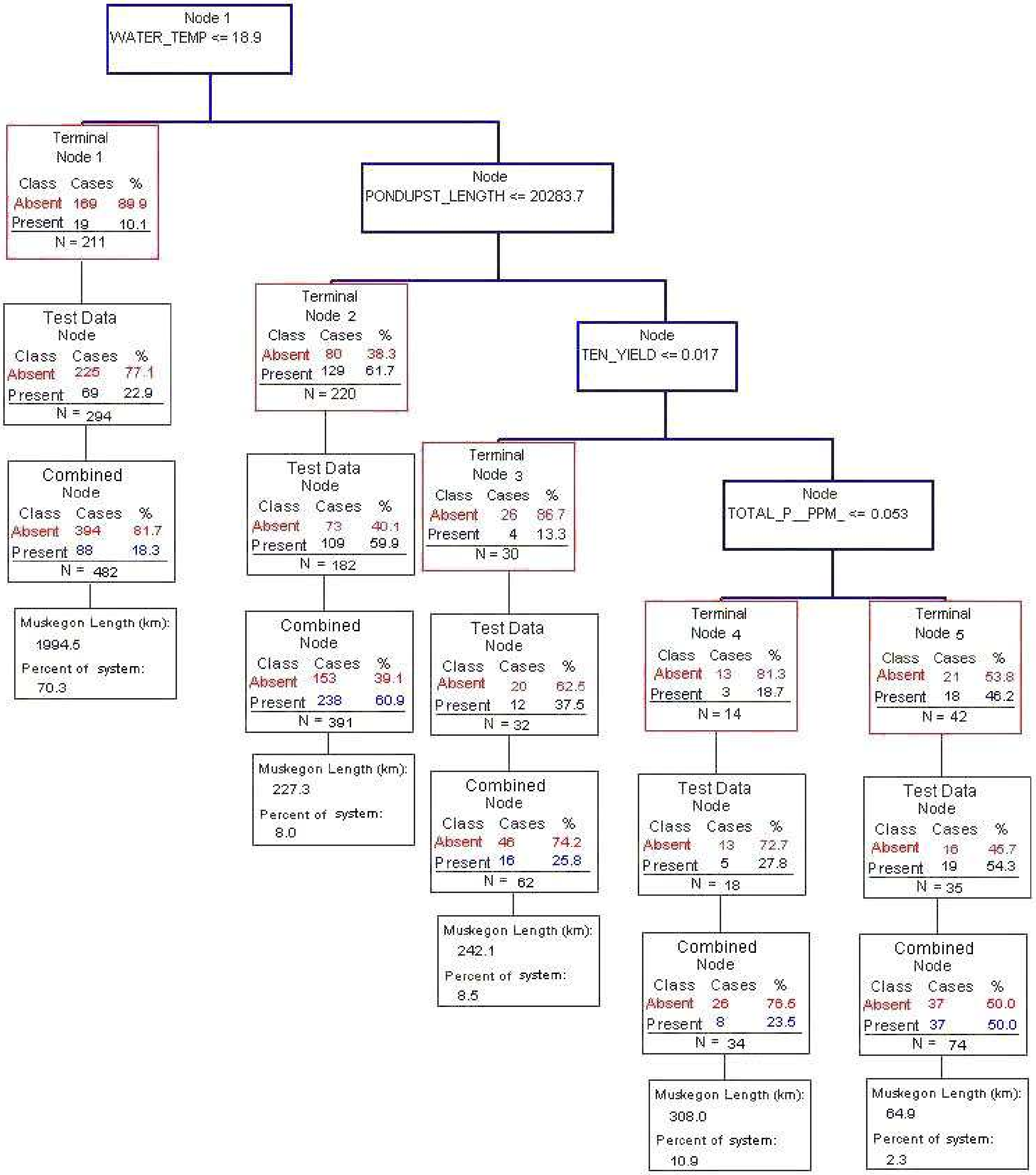




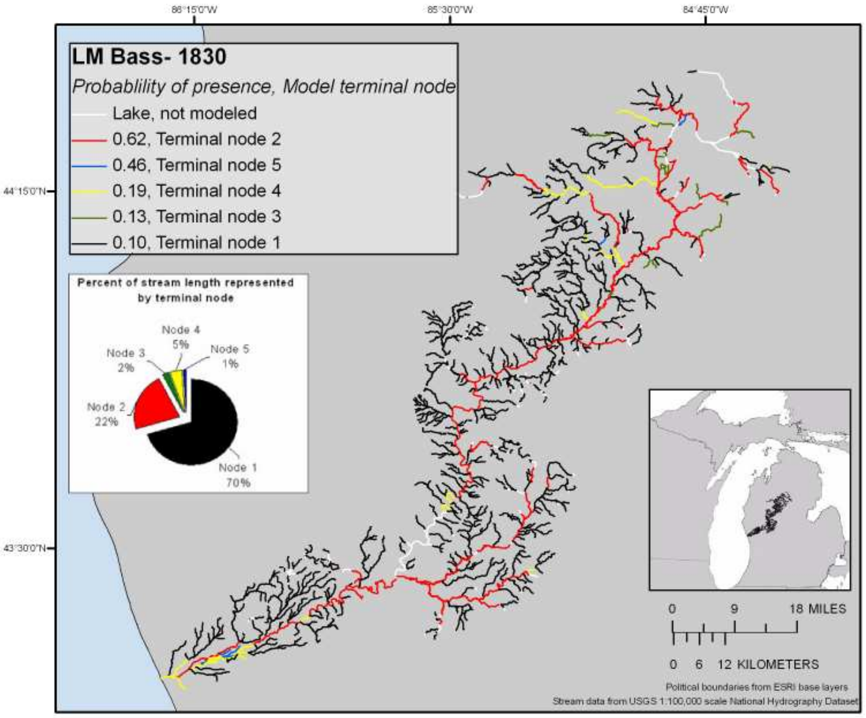
Appendix F, continued.

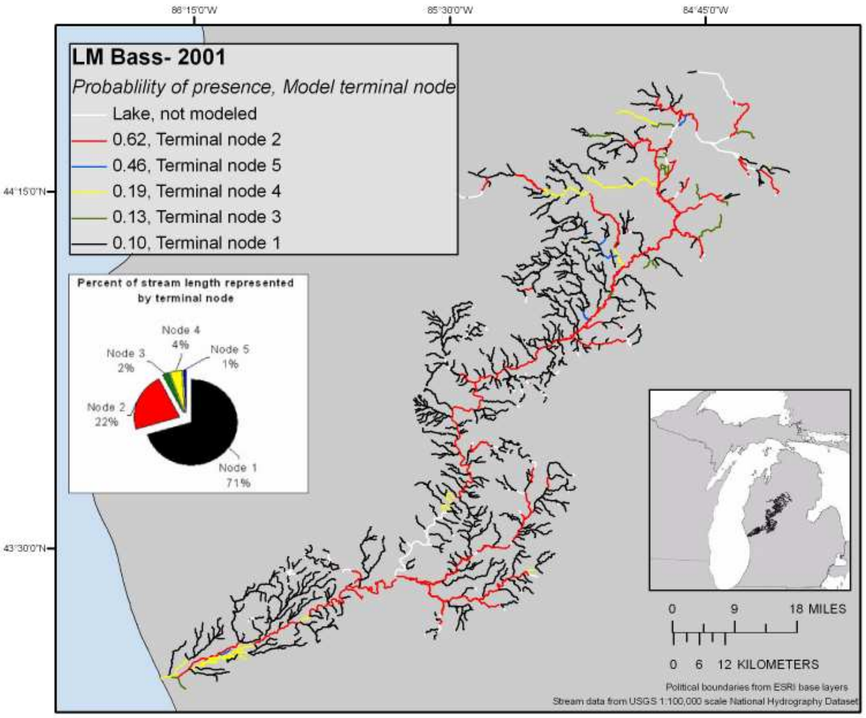


Appendix G. A) Classification tree model for largemouth bass, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the largemouth model as applied to the years 1830, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

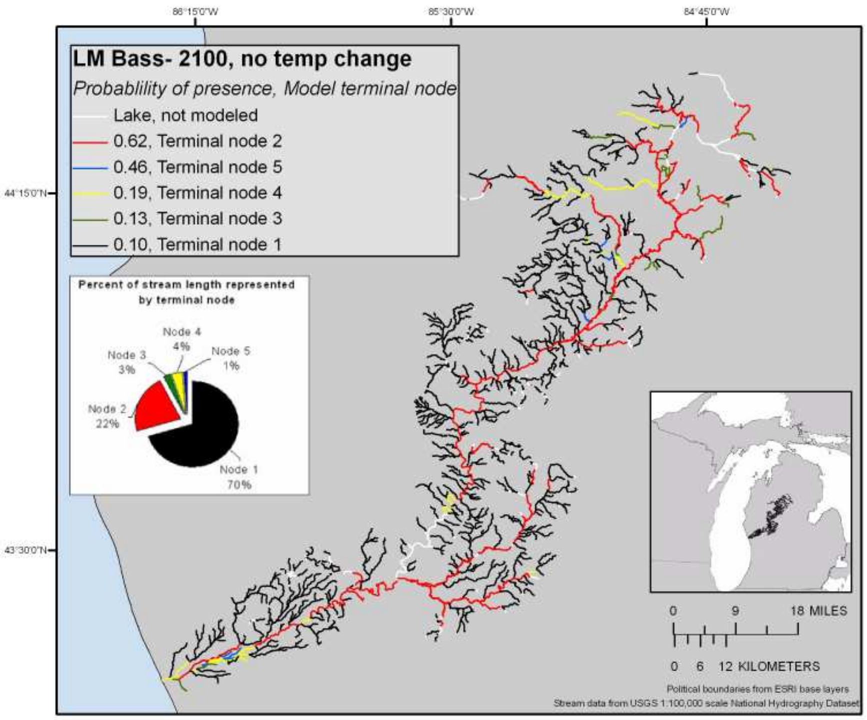


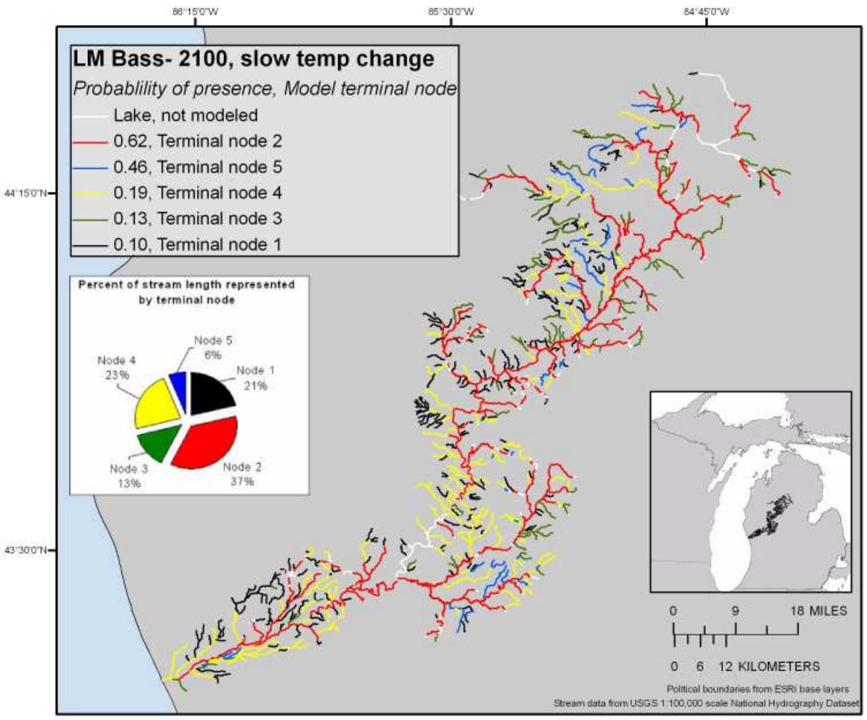
Appendix G, continued.



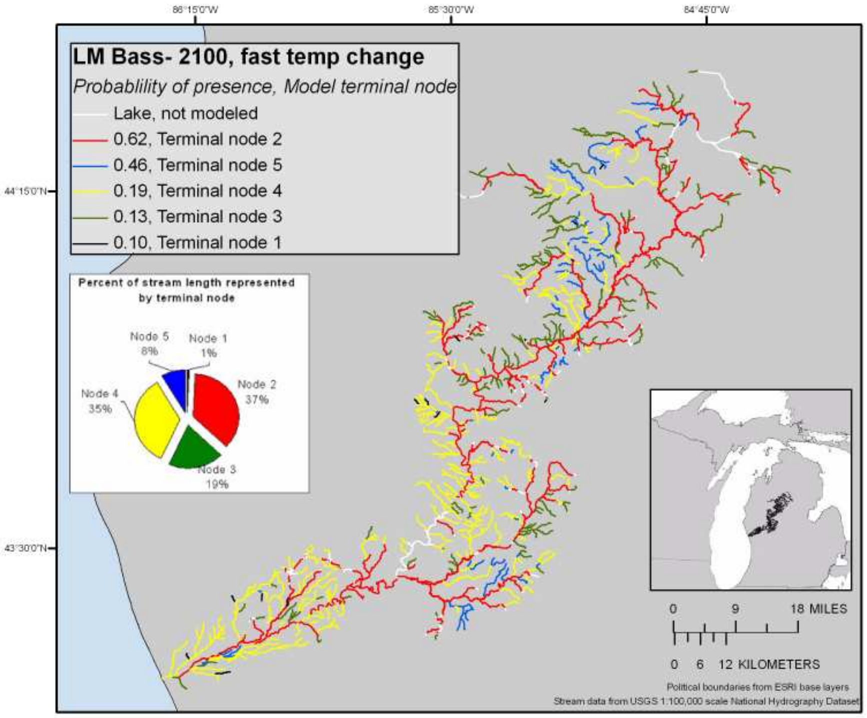


Appendix G, continued.

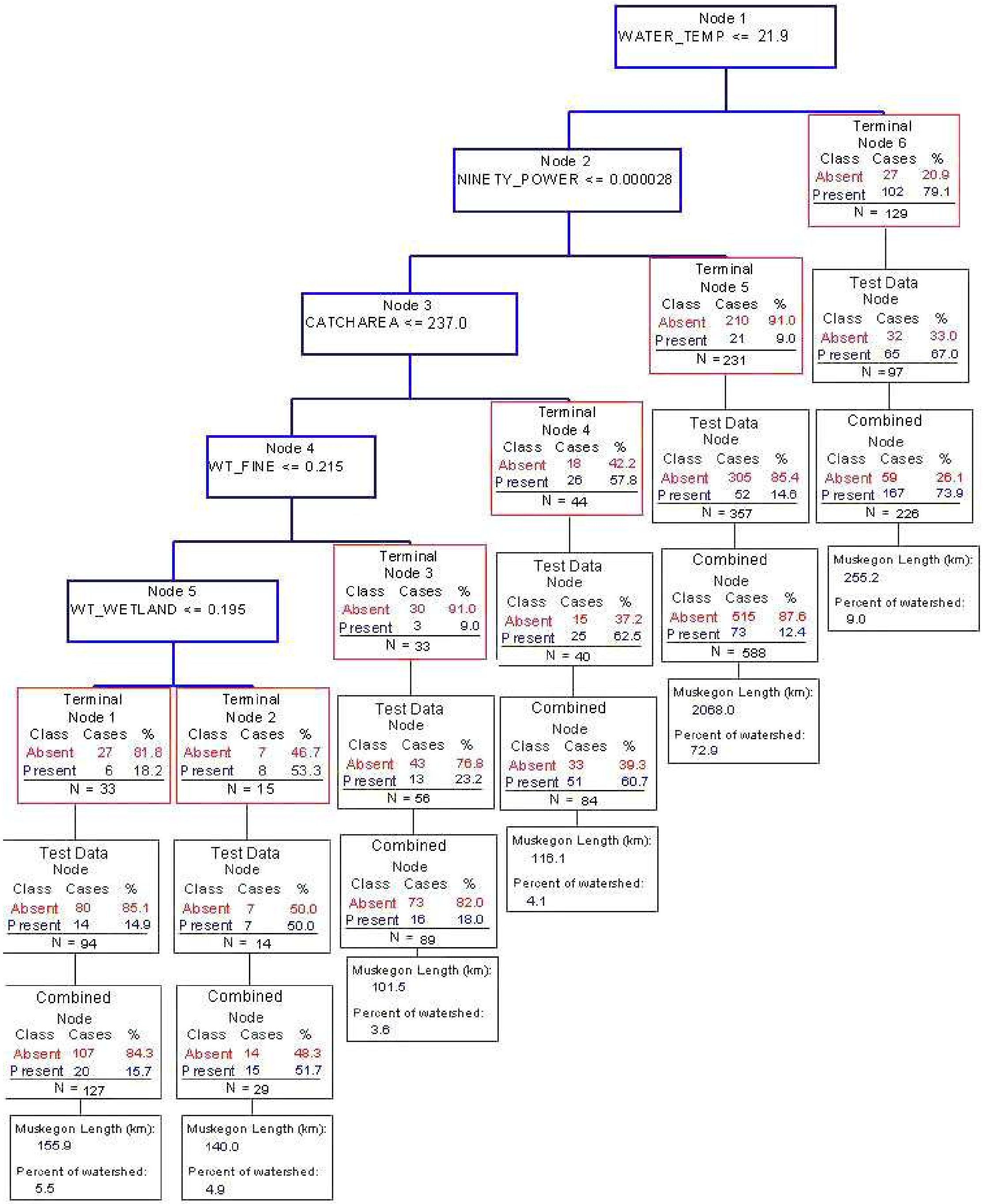




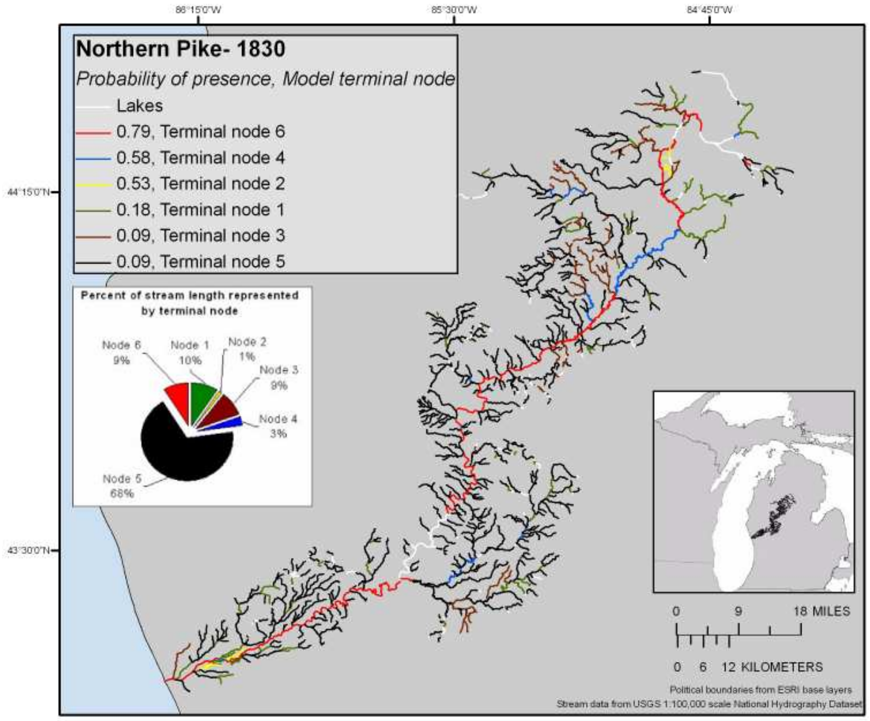
Appendix G, continued.

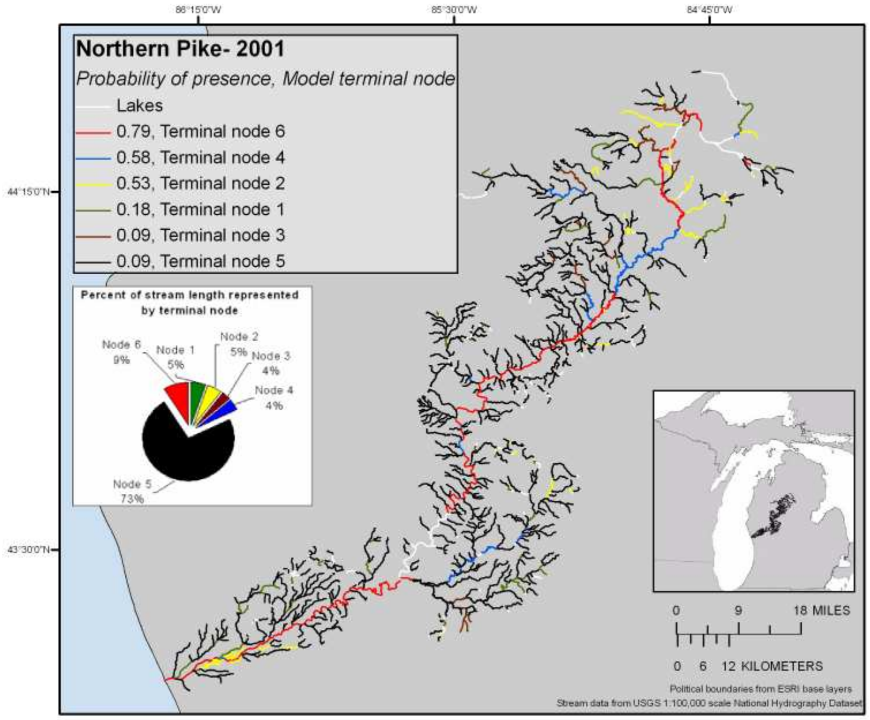


Appendix H. A) Classification tree model for northern pike, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the northern pike model as applied to the years 1830, 2001, and 2100. Three maps were made for 2100: 1) land-use change only, 2) slow temperature change, and 3) fast temperature change.

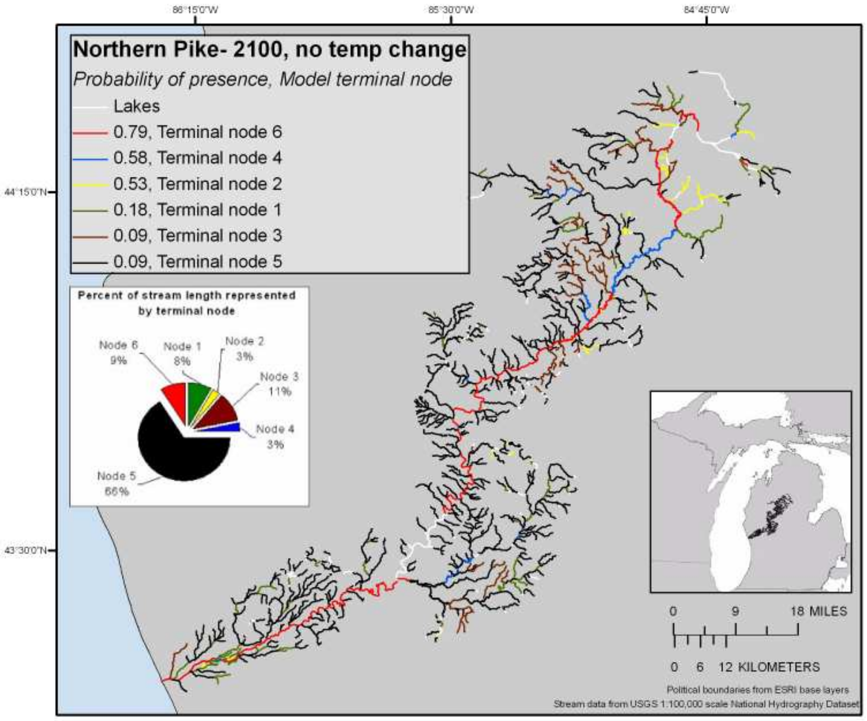


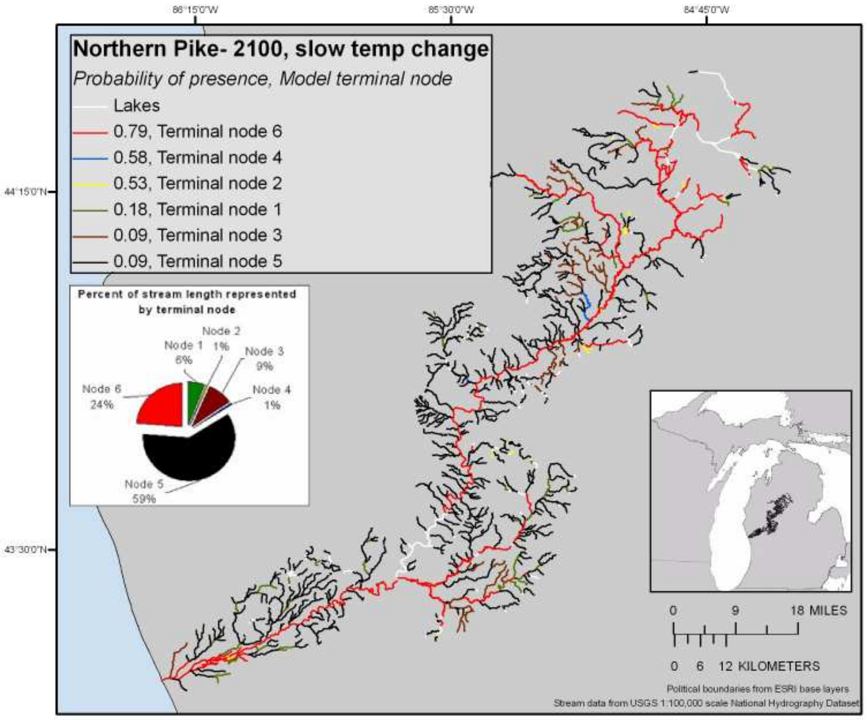
Appendix H, continued.



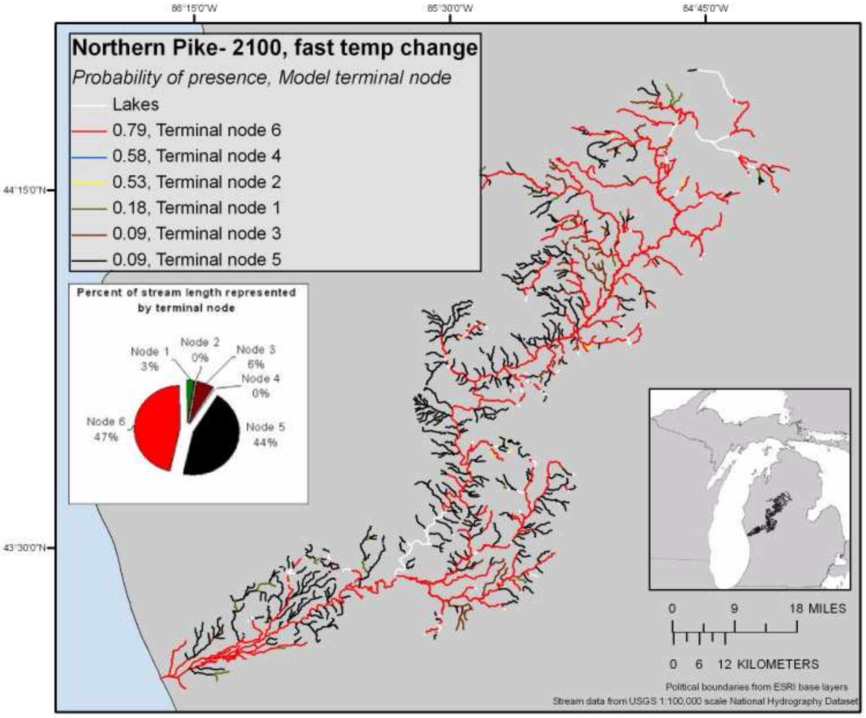


Appendix H, continued.

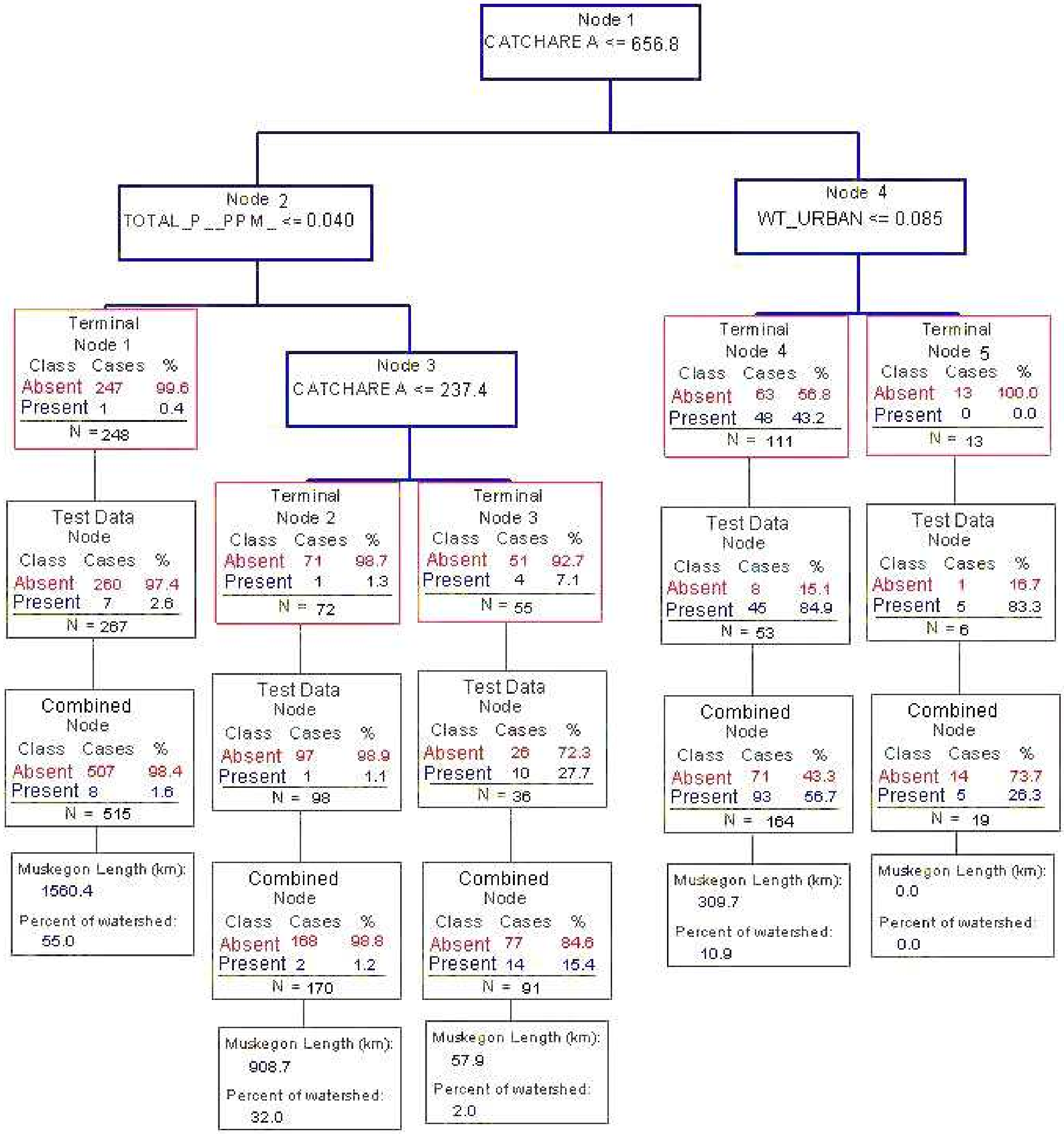




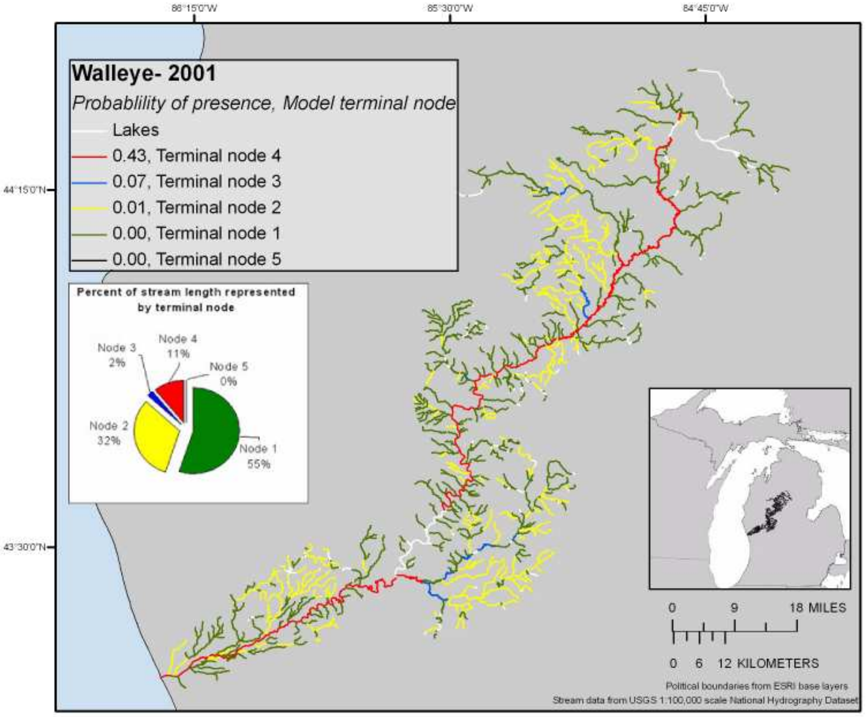
Appendix H, continued.

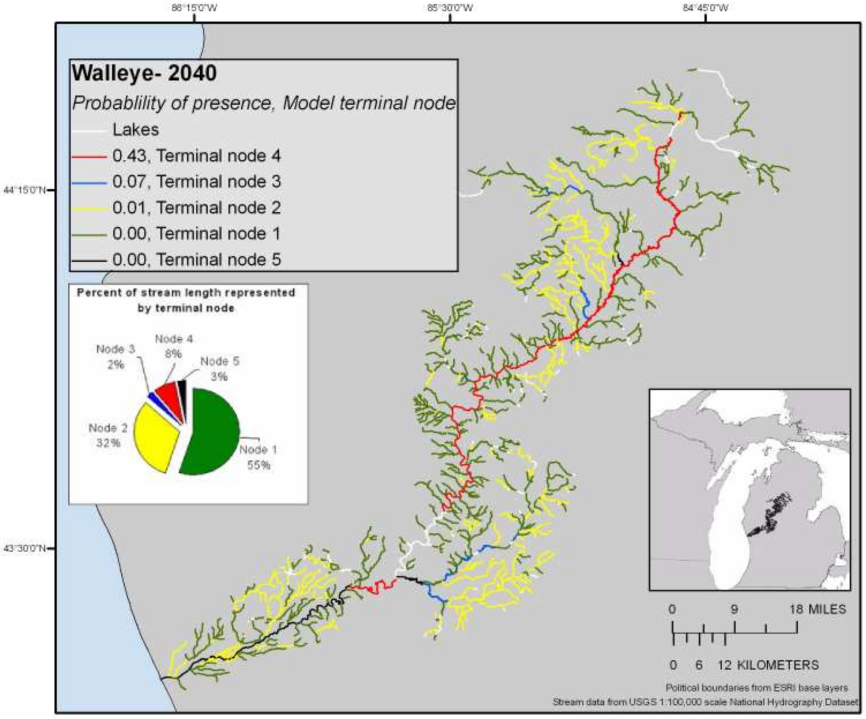


Appendix I. A) Classification tree model for walleye, developed by statewide fish samples and 2001 habitat data. Habitat variable code explanations are given in Chapter 2, Table 1. B) Predictive maps of the walleye model as applied to the years 1830, 2001, 2040, and 2100. Since the walleye model does not include temperature, temperature change future models were not created for this species.

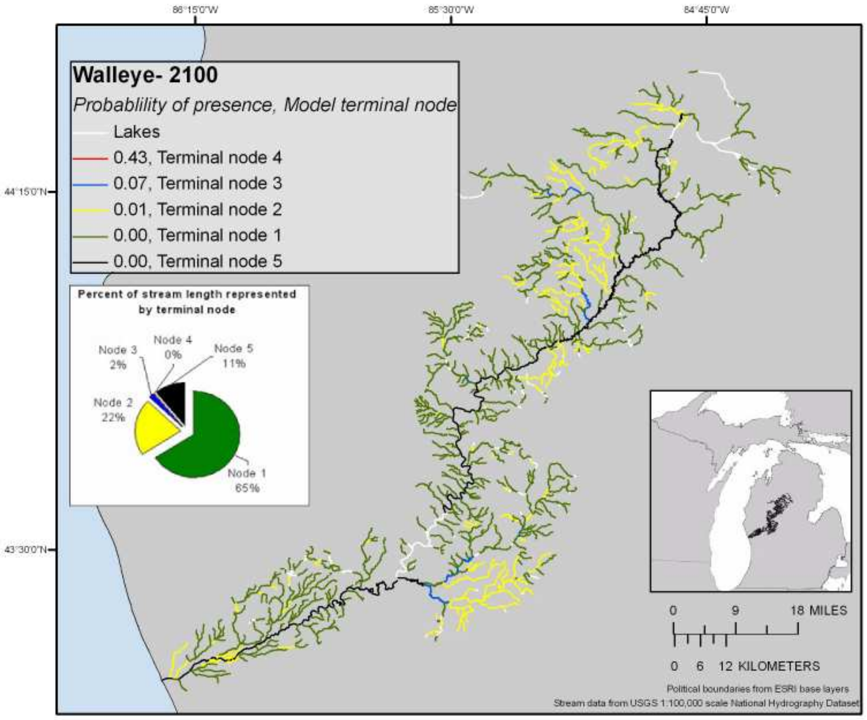


Appendix I, continued.





Appendix I, continued.



Appendix J. Contingency table counts and results for fish species of clusters 1,2, and 3. Tables A-D can be read as per this example from the first line of the table: Blackside Darter (from the darter species metric) is being compared between clusters 1 and 2+3. In cluster 1, this fish was absent from 21 sampling locations and present at 9 sampling locations and in cluster 2 and 3, there were 57 absent locations and 3 present locations. The Fisher's exact probability is 0.002, which is significant with the $\alpha$ level set at .03 (0.10/Number of species tested from the darter species metric). Tables E-H are based on fish species as a percentage of the total catch at a sampling site, but are interpreted in a similar manner to A-D.

A

| Darter Species | Clusters | Absent | Present | Absent | Present | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|
| **Blackside Darter** | 1 vs. 2+3 | 21 | 9 | 57 | 3 | | | 0.002 | 0.03 |
| **Johnny Darter** | 1 vs. 2+3 | 11 | 19 | 38 | 22 | 4.71 | 0.030 | | 0.03 |
| **Rainbow Darter** | 1 vs. 2+3 | 12 | 18 | 59 | 1 | 37.44 | 0.000 | | 0.03 |

B

| Sucker Species | Clusters | Absent | Present | Absent | Present | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|
| Northern Hog Sucker | 1 vs. 3 | 21 | 9 | 26 | 4 | 1.57 | 0.117 | | 0.05 |
| White Sucker | 1 vs. 3 | 6 | 24 | 5 | 25 | 0 | 1.000 | | 0.05 |

C

| Intolerant Species | Clusters | Absent | Present | Absent | Present | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|
| Mottled Sculpin | 1 vs. 3 | 26 | 4 | 29 | 1 | | | 0.353 | 0.02 |
| **Rainbow Darter** | 1 vs. 3 | 12 | 18 | 30 | 0 | 22.94 | 0.000 | | 0.02 |
| **Rock Bass** | 1 vs. 3 | 11 | 19 | 26 | 4 | 13.82 | 0.000 | | 0.02 |
| Smallmouth Bass | 1 vs. 3 | 22 | 8 | 25 | 5 | 0.39 | 0.531 | | 0.02 |
| (Northern Hog Sucker) | 1 vs. 3 | see Suckers | | | | | | | |

Appendix J, continued.

D

| Native Species | Clusters | Absent | Present | Absent | Present | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|
| Black Bullhead | 1 vs. 3 | 22 | 8 | 24 | 6 | 0.093 | 0.76 | | 0.005 |
| Black Crappie | 1 vs. 3 | 23 | 7 | 27 | 3 | 0.093 | 0.76 | | 0.005 |
| Blacknose Dace | 1 vs. 3 | 22 | 8 | 20 | 10 | 0.079 | 0.778 | | 0.005 |
| **Blackside Darter** | 1 vs. 3 | 21 | 9 | 30 | 0 | | | 0.002 | 0.005 |
| Bluegill | 1 vs. 3 | 15 | 15 | 17 | 13 | 0.067 | 0.796 | | 0.005 |
| Bluntnose Minnow | 1 vs. 3 | 18 | 12 | 25 | 5 | 2.955 | 0.086 | | 0.005 |
| Brook Stickleback | 1 vs. 3 | 21 | 9 | 22 | 8 | 0 | 1 | | 0.005 |
| Central Mudminnow | 1 vs. 3 | 12 | 18 | 18 | 12 | 1.667 | 0.197 | | 0.005 |
| Central Stoneroller | 1 vs. 3 | 23 | 7 | 27 | 3 | 1.08 | 0.299 | | 0.005 |
| Common Shiner | 1 vs. 3 | 14 | 16 | 20 | 10 | 1.697 | 0.193 | | 0.005 |
| Creek Chub | 1 vs. 3 | 6 | 24 | 13 | 17 | 2.773 | 0.096 | | 0.005 |
| Fathead Minnow | 1 vs. 3 | 26 | 4 | 17 | 13 | 5.253 | 0.022 | | 0.005 |
| Grass Pickerel | 1 vs. 3 | 21 | 9 | 28 | 2 | 4.007 | 0.045 | | 0.005 |
| Green Sunfish | 1 vs. 3 | 4 | 26 | 9 | 21 | 1.571 | 0.21 | | 0.005 |
| Horneyhead Chub | 1 vs. 3 | 20 | 10 | 28 | 2 | 5.104 | 0.024 | | 0.005 |
| **Johnny Darter** | 1 vs. 3 | 11 | 23 | 23 | 7 | 8.21 | 0.004 | | 0.005 |
| Largemouth Bass | 1 vs. 3 | 10 | 20 | 19 | 11 | 4.271 | 0.039 | | 0.005 |
| Northern Pike | 1 vs. 3 | 23 | 7 | 25 | 5 | 0.104 | 0.747 | | 0.005 |
| **Pumpkinseed** | 1 vs. 3 | 7 | 23 | 19 | 11 | 8.21 | 0.004 | | 0.005 |
| Yellow Bullhead | 1 vs. 3 | 19 | 11 | 25 | 5 | 2.1 | 0.144 | | 0.005 |
| (Rainbow Darter) | 1 vs. 3 | see Intolerant | | | | | | | |
| (Rock Bass) | 1 vs. 3 | see Intolerant | | | | | | | |
| (Mottled Sculpin) | 1 vs. 3 | see Intolerant | | | | | | | |
| (Smallmouth Bass) | 1 vs. 3 | see Intolerant | | | | | | | |
| (Northern Hog Sucker) | 1 vs. 3 | see Suckers | | | | | | | |

Appendix J, continued.

E

| % Carnivore | Clusters | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Black Crappie | 1 vs 2+3 | 1.2 | 23 | 4 | 3 | 54 | 2 | 4 |  |  | 0.175 | 0.014 |
| % Channel Catfish | 1 vs 2+3 | 0.71 | 26 | 2 | 2 | 54 | 3 | 3 |  |  | 1.000 | 0.014 |
| % Grass Pickerel | 1 vs 2+3 | 0.5 | 21 | 3 | 6 | 55 | 4 | 1 |  |  | 0.005 | 0.014 |
| **% Largemouth Bass** | 1 vs 2+3 | 1.57 | 10 | 8 | 12 | 38 | 13 | 9 | 8.51 | 0.014 |  | 0.014 |
| % Northern Pike | 1 vs 2+3 | 0.48 | 23 | 2 | 5 | 48 | 7 | 5 |  |  | 0.437 | 0.014 |
| **% Rockbass** | 1 vs 2+3 | 2.9 | 11 | 9 | 10 | 48 | 6 | 6 | 16.05 | 0.000 |  | 0.014 |
| % Smallmouth Bass | 1 vs 2+3 | 2.21 | 22 | 4 | 4 | 49 | 5 | 6 |  |  | 0.612 | 0.014 |

F

| % Tolerant | Clusters | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Blacknose Dace | 1 vs. 3 | 8.89 | 41 | 11 | 8 | 20 | 3 | 7 |  |  | 0.368 | 0.011 |
| % Bluntnose Minnow | 1 vs. 3 | 6.42 | 36 | 13 | 11 | 25 | 1 | 4 |  |  | 0.041 | 0.011 |
| % Central Mudminnow | 1 vs. 3 | 4.06 | 27 | 16 | 17 | 18 | 6 | 6 | 1.646 | 0.481 |  | 0.011 |
| **% Common Carp** | 1 vs. 3 | 1.9 | 37 | 16 | 7 | 16 | 2 | 12 | 12.31 | 0.002 |  | 0.011 |
| % Creek Chub | 1 vs. 3 | 12 | 12 | 27 | 21 | 13 | 5 | 12 | 8.08 | 0.018 |  | 0.011 |
| **% Fathead Minnow** | 1 vs. 3 | 3.44 | 51 | 6 | 3 | 17 | 5 | 8 |  |  | 0.005 | 0.011 |
| % Green Sunfish | 1 vs. 3 | 3.45 | 11 | 24 | 25 | 9 | 11 | 10 | 1.036 | 0.596 |  | 0.011 |
| % White Sucker | 1 vs. 3 | 7.28 | 11 | 26 | 23 | 5 | 11 | 14 |  |  | 0.767 | 0.011 |
| % Yellow Bullhead | 1 vs. 3 | 1.85 | 43 | 10 | 7 | 25 | 1 | 4 |  |  | 0.2 | 0.011 |

Appendix J, continued.

G

| % Insectivores | Clusters | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Black Bullhead | 1 vs. 3 | 1.43 | 22 | 4 | 4 | 24 | 3 | 3 | | | 0.829 | 0.007 |
| **% Blackside Darter** | 1 vs. 3 | 2.29 | 21 | 4 | 5 | 30 | 0 | 0 | | | 0.002 | 0.007 |
| % Bluegill | 1 vs. 3 | 4.19 | 15 | 6 | 9 | 17 | 8 | 5 | 1.444 | 0.486 | | 0.007 |
| % Brook Stickleback | 1 vs. 3 | 3.03 | 21 | 6 | 3 | 22 | 2 | 6 | | | 0.255 | 0.007 |
| % Central Mudminnow | 1 vs. 3 | 3.87 | 12 | 9 | 9 | 18 | 6 | 6 | 2.046 | 0.360 | | 0.007 |
| % Common Shiner | 1 vs. 3 | 3.94 | 14 | 5 | 11 | 20 | 8 | 2 | 7.666 | 0.022 | | 0.007 |
| % Green Sunfish | 1 vs. 3 | 3.03 | 4 | 13 | 13 | 9 | 10 | 11 | 1.875 | 0.392 | | 0.007 |
| % Horneyhead Chub | 1 vs. 3 | 3.36 | 20 | 4 | 6 | 28 | 2 | 0 | | | 0.018 | 0.007 |
| **% Johnny Darter** | 1 vs. 3 | 6.6 | 11 | 7 | 12 | 23 | 6 | 1 | 13.04 | 0.001 | | 0.007 |
| % Northern Hog Sucker | 1 vs. 3 | 3.6 | 21 | 4 | 5 | 26 | 2 | 2 | | | 0.343 | 0.007 |
| **% Pumpkinseed** | 1 vs. 3 | 0.08 | 7 | 9 | 14 | 19 | 8 | 3 | 12.00 | 0.002 | | 0.007 |
| **% Rainbow Darter** | 1 vs. 3 | 1.9 | 12 | 9 | 9 | 30 | 0 | 0 | | | 0.000 | 0.007 |
| % Sand Shiner | 1 vs. 3 | 10.5 | 26 | 2 | 2 | 27 | 1 | 1 | | | 1.000 | 0.007 |
| % Yellow Bullhead | 1 vs. 3 | 1.85 | 19 | 7 | 4 | 25 | 1 | 4 | | | 0.101 | 0.007 |
| % Yellow Perch | 1 vs. 3 | 0.74 | 25 | 1 | 4 | 28 | 2 | 0 | | | 0.112 | 0.007 |

H

| % Omnivores | Clusters | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **% Common Carp** | 1+2 vs. 3 | 1.9 | 37 | 16 | 7 | 16 | 2 | 12 | 12.31 | 0.002 | | 0.020 |
| **% Fathead Minnow** | 1+2 vs. 3 | 3.44 | 51 | 6 | 3 | 17 | 5 | 8 | | | 0.005 | 0.020 |
| **% Creek Chub** | 1+2 vs. 3 | 12 | 12 | 27 | 21 | 13 | 5 | 12 | 8.08 | 0.018 | | 0.020 |
| % Bluntnose Minnow | 1+2 vs. 3 | 6.42 | 36 | 13 | 11 | 25 | 1 | 4 | | | 0.041 | 0.020 |
| % White Sucker | 1+2 vs. 3 | 7.28 | 11 | 26 | 23 | 5 | 11 | 14 | | | 0.767 | 0.020 |

Appendix K. Contingency table counts and results for fish species of clusters 4 and 5. Table A can be read as per this example from the first line of the table: Black Bullhead (from the native species metric) is being compared between clusters 4 and 5. In cluster 4, this fish was absent from 13 sampling locations and present at 6 sampling locations and in cluster 5, there were 14 absent locations and 5 present locations. The $\chi^2$ value is 0.000 and associated probability is 1.000, which is insignificant with the $\alpha$ level set at .07 (0.1/Number of species tested from the "native species metric"). Tables B and C are based on fish species as a percentage of the total catch at a sampling site, but are interpreted in a similar manner to A.

A

| Native Species | Cluster | Absent | Present | Absent | Present | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|
| Black Bullhead | 4 vs. 5 | 13 | 6 | 14 | 5 | 0.00 | 1.000 | | 0.07 |
| Blacknose Dace | 4 vs. 5 | 10 | 9 | 15 | 4 | 1.87 | 0.171 | | 0.07 |
| Bluegill | 4 vs. 5 | 12 | 7 | 14 | 5 | 0.12 | 0.727 | | 0.07 |
| Bluntnose Minnow | 4 vs. 5 | 13 | 6 | 12 | 7 | 0.00 | 1.000 | | 0.07 |
| Brook Stickleback | 4 vs. 5 | 14 | 5 | 11 | 8 | 0.47 | 0.494 | | 0.07 |
| Central Mudminnow | 4 vs. 5 | 10 | 9 | 12 | 7 | 0.11 | 0.742 | | 0.07 |
| Central Stoneroller | 4 vs. 5 | 13 | 6 | 18 | 1 | 2.80 | 0.094 | | 0.07 |
| Common Shiner | 4 vs. 5 | 14 | 5 | 16 | 3 | 0.16 | 0.691 | | 0.07 |
| Creek Chub | 4 vs. 5 | 4 | 15 | 5 | 14 | | | 1.000 | 0.07 |
| Fathead Minnow | 4 vs. 5 | 12 | 7 | 5 | 14 | 3.83 | 0.050 | | 0.07 |
| Greensunfish | 4 vs. 5 | 4 | 15 | 5 | 14 | | | 1.000 | 0.07 |
| Johnny Darter | 4 vs. 5 | 11 | 8 | 13 | 6 | 0.11 | 0.737 | | 0.07 |
| Largemouth Bass | 4 vs. 5 | 11 | 8 | 17 | 2 | 3.39 | 0.065 | | 0.07 |
| Pumpkinseed | 4 vs. 5 | 11 | 8 | 14 | 5 | 0.47 | 0.494 | | 0.07 |
| White Sucker | 4 vs. 5 | 4 | 15 | 5 | 14 | | | 1.000 | 0.07 |

Appendix K, continued.

B

| % Omnivores | Cluster | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Common Carp | 4 vs. 5 | 0.9 | 12 | 3 | 4 | 16 | 2 | 1 | | | 0.349 | 0.33 |
| % Fathead Minnow | 4 vs. 5 | 4.76 | 12 | 5 | 2 | 5 | 5 | 9 | 7.34 | 0.026 | | 0.033 |
| % White Sucker | 4 vs. 5 | 13.4 | 4 | 10 | 5 | 5 | 4 | 10 | | | 0.125 | 0.33 |

C

| % Tolerant | Cluster | Median | 0% | <Med. | >Med. | 0% | <Med. | >Med. | $\chi^2$ | $\chi^2$ sig. | Fisher's sig. | Required sig. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Blacknose Dace | 4 vs. 5 | 20.79 | 10 | 5 | 4 | 15 | 1 | 3 | | | 0.197 | 0.02 |
| % Bluntnose Minnow | 4 vs. 5 | 4.98 | 13 | 3 | 3 | 12 | 3 | 4 | | | 1 | 0.02 |
| % Central Mudminnow | 4 vs. 5 | 17.1 | 10 | 6 | 3 | 12 | 2 | 5 | | | 0.402 | 0.02 |
| % Creek Chub | 4 vs. 5 | 21.2 | 4 | 8 | 7 | 5 | 6 | 8 | | | 0.843 | 0.02 |
| % Green Sunfish | 4 vs. 5 | 3.48 | 4 | 9 | 6 | 5 | 5 | 9 | | | 0.424 | 0.02 |
| % Fathead Minnow | 4 vs. 5 | see % Omnivores | | | | | | | | | | |
| % White Sucker | 4 vs. 5 | see % Omnivores | | | | | | | | | | |
| % Common Carp | 4 vs. 5 | see % Omnivores | | | | | | | | | | |

Appendix L. Results from the Kruskal-Wallis nonparametric test of means and the multiple comparison Nemnyi test of cluster 1,2, and 3 for habitat variables. Included in this table are the cluster means of the variables, $\chi^2$ of Kruskal-Wallis test, associated degrees of freedom, and associated significance. Multiple comparison results are interpreted as follows: q (1-3) is the Studentized range q statistic for the difference between cluster 1 and 3, and p (1-3) is the associated significance. Critical values for q are as follows: $q_{0.1, \infty, 3}$ = 2.902, $q_{0.05, \infty, 3}$ =3.313, $q_{0.01, \infty, 3}$ =4.200 (Zar 1999). Metrics significant at $\alpha$ <.10 are in bold.

| Habitat variables | Cluster 1 | Cluster 2 | Cluster 3 | $\chi^2$ | d.f. | p | q (1-3) | p (1-3) | q (1-2) | p (1-2) | q (2-3) | p (2-3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WT_FINE | 17.46 | 20.53 | 0.38 | 1.58 | 2 | 0.46 | | | | | | |
| WT_COARSE | 43.74 | 55.96 | 0.39 | 4.24 | 2 | 0.12 | | | | | | |
| 10_YIELD | 0.017 | 0.017 | 0.016 | 0.51 | 2 | 0.77 | | | | | | |
| 90_YIELD | 0.0017 | 0.0021 | 0.0018 | 2.86 | 2 | 0.24 | | | | | | |
| WT_URBAN11 | 3.90 | 5.69 | 5.66 | 4.33 | 2 | 0.12 | | | | | | |
| **WT_URBAN12** | 3.93 | 5.77 | 4.76 | 5.33 | 2 | **0.07** | 2.28 | **>0.1** | 3.12 | **<0.10** | 0.08 | >0.1 |
| **WT_URBAN14** | 3.93 | 2.41 | 1.66 | 11.95 | 2 | **0.00** | 4.55 | **<0.01** | 2.98 | **<0.10** | 1.56 | >0.1 |
| WT_AGR | 22.00 | 37.70 | 38.93 | 1.75 | 2 | 0.42 | | | | | | |
| WT_AGR21 | 15.84 | 17.63 | 13.19 | 3.74 | 2 | 0.15 | | | | | | |
| **WT_AGR22** | 17.20 | 20.04 | 25.75 | 6.91 | 2 | **0.03** | 3.66 | **<0.05** | 1.51 | >0.1 | 2.15 | >0.1 |
| WT_OPEN | 13.38 | 16.99 | 16.56 | 1.73 | 2 | 0.42 | | | | | | |
| **WT_FOREST** | 22.90 | 15.60 | 15.50 | 10.89 | 2 | **0.00** | 4.11 | **<0.05** | 3.86 | **<0.05** | 0.03 | >0.1 |
| **WT_FOREST41** | 18.10 | 13.06 | 13.19 | 9.31 | 2 | **0.01** | 3.51 | **<0.05** | 3.85 | **<0.05** | 0.03 | >0.1 |
| **WT_FOREST42** | 2.14 | 1.59 | 1.08 | 8.32 | 2 | **0.02** | 3.92 | **<0.05** | 2.24 | >0.1 | 1.68 | >0.1 |
| **WT_FOREST43** | 2.66 | 1.89 | 1.33 | 11.73 | 2 | **0.00** | 4.75 | **<0.01** | 2.49 | >0.1 | 2.25 | >0.1 |
| **WT_WATER** | 2.55 | 1.03 | 1.25 | 10.25 | 2 | **0.01** | 3.30 | **<0.10** | 4.18 | **<0.05** | 0.88 | >0.1 |
| WT_WETLAND | 28.57 | 9.80 | 9.87 | 3.00 | 2 | 0.22 | | | | | | |
| WT_WETLAND610 | 3.63 | 2.96 | 2.69 | 2.93 | 2 | 0.23 | | | | | | |
| WT_WETLAND611 | 5.10 | 3.55 | 4.03 | 3.93 | 2 | 0.14 | | | | | | |
| WT_WETLAND612 | 0.32 | 0.12 | 0.10 | 2.09 | 2 | 0.35 | | | | | | |
| WT_WETLAND613 | 0.02 | 0.01 | 0.01 | 1.19 | 2 | 0.55 | | | | | | |

Appendix L, continued.

| Habitat variables | Cluster 1 | Cluster 2 | Cluster 3 | $\chi^2$ | d.f. | p | q (1-3) | p (1-3) | q (1-2) | p (1-2) | q (2-3) | p (2-3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WT_WETLAND62 | 3.9 | 3.12 | 3.02 | 2.09 | 2 | 0.35 | | | | | | |
| RT_URBAN11 | 2.3 | 3.43 | 3.24 | 1.19 | 2 | 0.55 | | | | | | |
| RT_URBAN12 | 2.23 | 3.37 | 2.45 | 1.29 | 2 | 0.52 | | | | | | |
| **RT_URBAN14** | 2.85 | 1.94 | 1.01 | 13.79 | 2 | **0** | 4.96 | **<0.01** | 2.59 | >0.1 | 2.37 | >0.1 |
| RT_AGR | 22 | 27.57 | 27.33 | 2.39 | 2 | 0.3 | | | | | | |
| RT_AGR21 | 9.43 | 12.02 | 8.05 | 1.37 | 2 | 0.51 | | | | | | |
| **RT_AGR22** | 12.66 | 15.55 | 19.33 | 5.22 | 2 | **0.07** | 3.16 | **<0.10** | 1.97 | >0.1 | 1.2 | >0.1 |
| RT_OPEN | 9.26 | 12.97 | 13.67 | 4.06 | 2 | 0.13 | | | | | | |
| **RT_FOREST** | 19.7 | 16.8 | 15.2 | 5.63 | 2 | **0.06** | 3.05 | **<0.10** | 2.64 | >0.1 | 0.411 | >0.1 |
| RT_FOREST41 | 15.14 | 13.44 | 12.79 | 2.79 | 2 | 0.25 | | | | | | |
| **RT_FOREST42** | 1.97 | 1.43 | 1.17 | 6 | 2 | **0.05** | 3.29 | **<0.10** | 2.13 | >0.1 | 1.16 | >0.1 |
| **RT_FOREST43** | 2.61 | 1.86 | 1.25 | 11.85 | 2 | **0** | 4.76 | **<0.01** | 2.77 | >0.1 | 1.99 | >0.1 |
| **RT_WATER** | 10.34 | 10.34 | 4.87 | 8.66 | 2 | **0.01** | 2.16 | >0.1 | 4.08 | **<0.05** | 1.91 | >0.1 |
| RT_WETLAND | 28.57 | 24.93 | 25.5 | 0.48 | 2 | 0.79 | | | | | | |
| RT_WETLAND610 | 7.02 | 8.12 | 5.48 | 3.15 | 2 | 0.21 | | | | | | |
| RT_WETLAND611 | 11.42 | 8.13 | 10.8 | 3.05 | 2 | 0.22 | | | | | | |
| **RT_WETLAND612** | 0.62 | 0.39 | 0.22 | 11.11 | 2 | **0** | 4.41 | **<0.01** | 2.72 | >0.1 | 1.69 | >0.1 |
| RT_WETLAND613 | 0.04 | 0.01 | 0.01 | 4.42 | 2 | 0.11 | | | | | | |
| RT_WETLAND62 | 9.4 | 8.19 | 8.98 | 1.74 | 2 | 0.42 | | | | | | |
| TNY_ATMOS | 219.16 | 202.65 | 190.65 | 4.14 | 2 | 0.13 | | | | | | |
| TNY_FERT | 434.03 | 459.09 | 384.89 | 4.21 | 2 | 0.12 | | | | | | |
| **TNY_LIVE** | 94 | 210.65 | 208.25 | 11.54 | 2 | **0** | 2.91 | **<0.10** | 3.22 | **<0.10** | 0.03 | >0.1 |
| **TNY_NONAG** | 86.44 | 78.01 | 97.7 | 6.91 | 2 | **0.03** | 2.7 | >0.1 | 0.79 | >0.1 | 3.49 | **<0.05** |
| **TNY_POINT** | 176.31 | 326.21 | 414.76 | 6.32 | 2 | **0.04** | 3.31 | **<0.10** | 6.64 | **<0.01** | 0.66 | >0.1 |
| TNY_TOTAL | 1009.93 | 1276.59 | 1296.25 | 2.49 | 2 | 0.29 | | | | | | |
| **TPY_FERT** | 33.13 | 32.13 | 26.72 | 5.33 | 2 | **0.07** | 2.75 | >0.1 | 0.06 | >0.1 | 2.81 | >0.1 |
| **TPY_LIVE** | 13.36 | 27.86 | 22.52 | 9.42 | 2 | **0.01** | 3.4 | **<0.05** | 3.94 | **<0.05** | 0.55 | >0.1 |

Appendix L, continued.

| Habitat variables | Cluster 1 | Cluster 2 | Cluster 3 | $\chi^2$ | d.f. | p | q (1-3) | p (1-3) | q (1-2) | p (1-2) | q (2-3) | p (2-3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TPY_NONAG | 7.84 | 6.61 | 8.12 | 3.46 | 2 | 0.18 | | | | | | |
| TPY_POINT | 25.27 | 4.89 | 37.25 | 1.07 | 2 | 0.59 | | | | | | |
| TPY_TOTAL | 79.6 | 111.49 | 94.62 | 2.05 | 2 | 0.36 | | | | | | |
| FERT | 16.7 | 22.57 | 19.95 | 3.08 | 2 | 0.22 | | | | | | |
| HERB | 14.43 | 18.77 | 16.34 | 2.55 | 2 | 0.28 | | | | | | |
| **INSECT** | 2.18 | 4.4 | 3.77 | 11.16 | 2 | **0** | 3.86 | **<0.05** | 4.2 | **<0.05** | 0.35 | >0.1 |
| **MANURE** | 1.87 | 4.38 | 4.68 | 12.9 | 2 | **0** | 4.72 | **<0.01** | 3.85 | **<0.05** | 0.87 | >0.1 |
| **OUTFALL** | 0.21 | 0.12 | 0.08 | 12.14 | 2 | **0** | 4.69 | **<0.01** | 3.4 | **<0.05** | 1.29 | >0.1 |
| MINES | 0 | 0.01 | 0.01 | 4.27 | 2 | 0.12 | | | | | | |
| POPDENS | 197.39 | 270.33 | 252.57 | 4.63 | 2 | 0.1 | | | | | | |
| RDCROSS | 0.6 | 0.81 | 0.69 | 0.13 | 2 | 0.94 | | | | | | |
| RDDENS | 2.86 | 3.36 | 3.16 | 3.91 | 2 | 0.14 | | | | | | |
| **EPATOXIC** | 0.04 | 0.18 | 0.2 | 10.99 | 2 | **0** | 4.24 | **<0.01** | 0.57 | >0.1 | 3.66 | **<0.05** |

Appendix M. Results from the Mann-Whitney U nonparametric test of means of clusters 4 and 5, for habitat variables. Included in this table are the cluster means of the variables, U statistic, and associated significance. Variables significant at $\alpha < .10$ are in bold.

| Habitat variable | Cluster 4 | Cluster 5 | Mann-Whitney U statistic | p-value |
|---|---|---|---|---|
| WT_FINE | 51.30 | 70.57 | 127 | 0.116 |
| **WT_COARSE** | **16.70** | **8.60** | **108.5** | **0.023** |
| 10_YIELD | 0.019 | 0.020 | 147 | 0.328 |
| **90_YIELD** | **0.0016** | **0.0012** | **118** | **0.068** |
| WT_URBAN11 | 16.51 | 20.92 | 150 | 0.373 |
| WT_URBAN12 | 15.41 | 17.11 | 152 | 0.405 |
| WT_URBAN14 | 2.71 | 3.62 | 137 | 0.133 |
| WT_AGR | 4.63 | 9.79 | 138 | 0.193 |
| WT_AGR21 | 2.20 | 3.81 | 145.5 | 0.279 |
| **WT_AGR22** | **2.46** | **5.99** | **122.5** | **0.082** |
| WT_OPEN | 23.61 | 22.80 | 179 | 0.965 |
| **WT_FOREST** | **21.11** | **14.89** | **120** | **0.08** |
| WT_FOREST41 | 15.34 | 11.65 | 128 | 0.125 |
| WT_FOREST42 | 3.23 | 1.58 | 129 | 0.121 |
| **WT_FOREST43** | **2.56** | **1.63** | **119.5** | **0.072** |
| WT_WATER | 0.53 | 0.44 | 139 | 0.199 |
| WT_WETLAND | 5.32 | 3.37 | 127.5 | 0.117 |
| **WT_WETLAND610** | **0.67** | **0.23** | **118.5** | **0.044** |
| **WT_WETLAND611** | **2.45** | **1.55** | **124** | **0.096** |
| **WT_WETLAND612** | **0.06** | **0.01** | **121** | **0.04** |
| **WT_WETLAND613** | **0.01** | **0.00** | **130** | **0.066** |
| WT_WETLAND62 | 2.10 | 1.57 | 152 | 0.4 |
| RT_URBAN11 | 10.71 | 15.50 | 136.5 | 0.199 |
| RT_URBAN12 | 10.27 | 12.17 | 146.5 | 0.32 |
| RT_URBAN14 | 4.01 | 2.52 | 142 | 0.184 |
| RT_AGR | 3.68 | 7.21 | 145.5 | 0.262 |
| RT_AGR21 | 1.96 | 3.07 | 147 | 0.276 |
| RT_AGR22 | 1.74 | 4.17 | 151 | 0.351 |
| RT_OPEN | 20.90 | 23.80 | 145 | 0.3 |
| **RT_FOREST** | **25.37** | **18.32** | **121.5** | **0.085** |
| RT_FOREST41 | 18.92 | 14.41 | 131 | 0.148 |
| RT_FOREST42 | 3.71 | 2.12 | 133 | 0.152 |
| **RT_FOREST43** | **2.78** | **1.74** | **119.5** | **0.072** |
| RT_WATER | 1.79 | 0.80 | 129 | 0.122 |
| RT_WETLAND | 16.00 | 12.26 | 133.5 | 0.17 |
| **RT_WETLAND610** | **2.63** | **1.93** | **133.5** | **0.048** |