

INFLUENCE OF REACH AND WATERSHED CHARACTERISTICS ON FISH
DISTRIBUTIONS IN SMALL STREAMS OF EASTERN MONTANA

by

Melissa Renee Wuellner

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Fish and Wildlife Management

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2007

© COPYRIGHT

by

Melissa Renee Wuellner

2007

All Rights Reserved

APPROVAL

of a thesis submitted by

Melissa Renee Wuellner

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Christopher S. Guy

Approved for the Department of Ecology

David R. Roberts

Approved for the Division of Graduate Education

Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from the reproduction of this thesis in whole or in parts may be granted only by the copyright holder. I specifically grant Montana Fish, Wildlife and Parks the right to make and distribute copies as necessary to meet their needs.

Melissa Renee Wuellner

May 2007

This thesis is dedicated to my family for their support during the education process, and especially to my husband, Ryan, for his continued financial, emotional, and technological support.

ACKNOWLEDGMENTS

This work was made possible by grants from the U. S. Environmental Protection Agency and Montana Fish, Wildlife and Parks. I thank the numerous field technicians who assisted with data collection from 1999 through 2004 (even though I did not work with all of these individuals): Diane Alonzo, Christopher Bare, Mike Borgreen, Ann Capaul, Windy Davis, Steven Dawes, Lyndsay Hellekson, Tyson Holzheimer, Aaron Liberty, Derek Poinsette, Andrew Puls, Helen Schemm, Jack Tuomikoski, and Jennette Vander Jagt. Initial GIS and modeling consultation was provided by Keith Gido and Bob Oakes at Kansas State University and Brent Brock at the Wildlife Conservation Society in Bozeman. My deepest gratitude is extended to Jim Johnson of Confluence, Inc., in Bozeman for his patience and expertise in tutoring me in the finer points of GIS applications. Appreciation is given to my graduate committee: Dr. Alexander Zale for being one of the initial principal investigators for the project and providing insight along the way; Dr. David Roberts for assistance with statistical concepts and methodology; my two (though not officially) co-advisors Drs. Robert Bramblett and Christopher Guy for continued financial support of this study and providing direction for my project and graduate career. Finally, I extend my greatest appreciation to all of the eastern Montana landowners whose cooperation and assistance made this project successful.

TABLE OF CONTENTS

LIST OF TABLES.....	viii
LIST OF FIGURES	x
ABSTRACT.....	xiii
1) INTRODUCTION	1
2) STUDY AREA	7
3) METHODS	11
Site Selection	11
Fish Sampling	12
Environmental Characteristics	13
Reach Scale.....	13
Watershed Scale.....	18
Statistical Analysis.....	18
Reach-scale, Watershed-scale, and Combined Models	18
Habitat Categories.....	29
Scale Comparisons.....	29
4) RESULTS	32
Fish Collection	32
Models Results by Scale	32
Reach-scale Models	32
Watershed-scale Models	38
Combined Models.....	41
Scale Comparisons.....	46
5) DISCUSSION.....	47
Interpretations of Variables Used in Models	53
Scale Comparisons.....	60
CONCLUSION.....	62

TABLE OF CONTENTS - CONTINUED

MANAGEMENT IMPLICATIONS	65
LITERATURE CITED	66
APPENDIX A: GIS Methodology And Documentation.....	75

LIST OF TABLES

Table	Page
1. Physical attributes measured at each reach sampled according to EPA protocol (Kaufman and Robison 1998).....	15
2. Physicochemical habitat variables calculated at each sample site according to Kaufman et al. (1999), their acronyms, and units of measure.	16
3. Biotic variables measured at each site, their acronyms, and units of measurement. Scientific names of all species are found in Table 6.	17
4. List of watershed variables calculated for each site, their acronyms, and the data source.....	20
5. Categorization of variables by biological relevance and by scale. Variables were grouped a posteriori based on the general aspect of the environment being measured. Variables are listed in Tables 2, 3, and 4.....	30
6. Scientific name, common name, and acronym of native and nonnative species collected in prairie streams of Montana from 1999 through 2004.	33
7. Reach models for 19 species from 105 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. Numbers in parentheses following means of percent reduction in deviance, MER metrics, and percent improvement are the 95% confidence interval.....	36
8. Watershed models for 20 species from 106 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. All MER metrics are reported in percents. The number in parentheses following the means of percent reduction in deviance is the 95% confidence interval. Acronyms following FEco and Drain variables are: PORB = Powder River Basin; NWGP = Northwestern Glaciated Plains; LMR = Little Missouri River; YER = Yellowstone River.	39

LIST OF TABLES – CONTINUED

Table	Page
9. Combined models for 19 species from 93 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. Numbers in parentheses following the means of percent reduction in deviance, MER metrics, and percent improvement are the 95% confidence interval. Acronyms following FEco and Drain variables are: PORB = Powder River Basin; LMR = Little Missouri River; YER = Yellowstone River.	43
10. Residual mean deviance (D^2) by species and scale. Numbers in bold indicate the models with the D^2 values for that species. Numbers in parentheses indicate the 95% confidence interval of the D^2	46

LIST OF FIGURES

Figure	Page
1. Map of Great Plains within Montana. The Northwestern Glaciated Plains ecoregion (dark gray) within the state extends from the Canada border south to roughly the Missouri River. The Northwestern Great Plains (light gray) of the state extends roughly south from the Missouri River to the Wyoming border. Major rivers in this area include the Missouri and Yellowstone rivers.	8
2. Graphical representation of the sampling reach and the eleven transects within the reach. The upper and lower ends of the reach are counted in the eleven transects.	14
3. Map of the 120 sites used in the modeling database. All sites were sampled between July and September from 1999 to 2004.	22
4. Example of a tree classifier for a species modeled at the reach scale. The words TRUE and FALSE signify terminal nodes. Interpretation of this model indicates that this species is likely to be found where instream vegetative cover is less than 8.41%, the standard deviation of width is greater than 1.69, and the percent of riparian ground cover is greater than 62.44%.	24
5. Two examples of the graphical results of the cross-validation procedure from this study. Graph A met the deviance reduced criterion because there is an initial increase followed by a rapid decrease in residual deviance at a terminal node value greater than zero or one. The number of terminal nodes selected for the pruned tree in this example is three. Graph B did not meet the deviance reduced criterion because no such trend was detected.	27
6. Frequency distribution of species richness for prairie streams sampled in eastern Montana from 1999 through 2004.	34
7. Species occurrence for 38 species collected in eastern Montana. Solid circles represent nonnative species, and open circles represent native species. Species acronyms are found in Table 6.	35

LIST OF FIGURES – CONTINUED

Figure	Page
8. Average percent deviance reduced by reach-scale variables used in the six models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variable that appeared in more than one species model is listed in parentheses above the variable.	37
9. Weighted average percent deviance reduced in species models by reach-scale categories. Acronyms are as follows: water quality (WQ), channel morphology (Morph), instream cover (InCov), introduced non-piscivores (INPisc), velocity (Veloc), piscivores (Pisc), and species richness and diversity (RiDi). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of variable occurrences within that category.	38
10. Average percent deviance reduced by watershed-scale variables used in the four models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variable that appeared in more than one species model is listed in parentheses above the variable.	40
11. Weighted average percent deviance reduced in species models by watershed-scale categories. Acronyms are as follows: elevation (Elev), anthropogenic influences (Anthro), soil characteristics (Soil), position (Posit), and stream size (Size). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of variable occurrences within that category.	41
12. Average percent deviance reduced by reach- and watershed-scale variables in combined models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variables that appeared in more than one species model is listed in parentheses above the variable.	44

LIST OF FIGURES – CONTINUED

Figure	Page
13. Weighted average percent deviance reduced in species models by combined variable categories. Acronyms are as follows: introduced non-piscivores (INPisc), piscivores (Pisc), land cover within the watershed (Land), anthropogenic influences (Anthro), species richness and diversity (RiDi); soil characteristics within the watershed (Soil), instream cover (InCov), position (Posit), and elevation (Elev). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of occurrences of variables within that category.	45

ABSTRACT

The prairie biome is arguably the most endangered ecoregion in North America, and the chance of extinction for many fishes is greater than in other ecoregions. Fisheries scientists and managers must understand the ecology of fishes at multiple spatial scales for effective conservation. Statistical models designed to explain distributions of fishes using environmental characteristics measured at multiple scales have provided this critical information. The objectives of this study were to: i) identify the reach- (i.e., physicochemical and biotic) and watershed-scale characteristics that affect the distribution (i.e., presence or absence) of fishes in prairie streams of Montana; and ii) identify which scale best explains the distribution of fishes in Montana prairie streams: reach, watershed, or a combination of variables measured at either scale (i.e., combined models). Reach and watershed information from 120 sites sampled between 1999 and 2004 was used to model the presence or absence of 20 species using tree classifiers. Models were evaluated by strict criteria in this study as compared to similar studies of prairie fishes. In this study, variables used in models had to reduce a significant amount of the deviance in species distributions before they could be evaluated for their predictive ability; other studies have only examined whether reach- or watershed-scale variables could be used to predict the occurrence of species. Only six reach, four watershed, and seven combined models were found to be significant, suggesting that explaining and predicting the distribution of fishes in eastern Montana is difficult. Significant models did provide information that could be used to guide conservation efforts. Most of the reach- and watershed-scale variables used in these models described large-scale ecological gradients. Additionally, results indicated that significant models could predict individual species distributions with a high level of accuracy. Previous predictive modeling studies have indicated that GIS-derived watershed characteristics better explain fish distributions than reach-scale characteristics. However, this study did not find that watershed-scale variables were significantly better at explaining or predicting fish distributions in eastern Montana. Results from this study suggest further information is needed to understand how spatial scale affects the distribution of fishes in warmwater streams of Montana.

INTRODUCTION

A combination of reach- [physicochemical (e.g., substrate, water quality, discharge) and biotic (e.g., competition, predation)], and watershed-scale (e.g., elevation, latitude, dispersal barriers) factors influence the distribution of fishes (Johnson et al. 1977; Leonard and Orth 1988; Jackson and Harvey 1989; Rahel and Hubert 1991; Fausch et al. 1994; Rabeni and Sowa 1996; Angermeier and Winston 1998; Jackson et al. 2001). Understanding the relationship between species distributions and environmental characteristics at multiple scales provides insight into evolutionary history of fish species (Darlington 1957; Jackson and Harvey 1989) and is critical to the conservation of aquatic communities (Matthews 1998; Berra 2001; Gido et al. 2006). General patterns in fish distributions among watersheds result from both historical (e.g., post-glacial dispersion) and extrinsically regulated environmental factors (e.g., climate). However, physicochemical characteristics of individual stream reaches (Rabeni and Sowa 1996; Quist et al. 2005) and biotic interactions within the reach (Jackson et al. 2001; Quist et al. 2005) may further restrict the distributions of fishes.

The distributions of fish species in the Great Plains of North America have been influenced by these watershed- and reach-scale characteristics. At the watershed scale, the processes of flooding and glaciation have influenced the morphology and connectivity of streams. The northern Great Plains were inundated during three separate occasions in the Cretaceous period. Regressions of these epicontinental seas deposited extensive layers of carbonate rocks, sandstones, and shales (Alden 1932; Cross et al.

1986). During this period, the Missouri, Little Missouri, and Yellowstone rivers flowed northward into an Arctic drainage (Hudson Bay). No one definitively knows which fish species were present during this time, but several authors have speculated. Bailey and Allum (1962) hypothesized that the assemblages of streams within the Missouri River drainage during the Cretaceous period included pearl dace *Margariscus margarita*, finescale dace *Phoxinus neogaeus*, northern redbelly dace *P. eos*, lake chub *Couesius plumbeus*, longnose sucker *Catostomus catostomus*, and mountain sucker *C. platyrhynchus*. McPhail and Lindsay (1970) expanded this list to include flathead chub *Platygobio gracilis* and goldeye *Hiodon alosoides*. However, other authors believe an even greater number of species occupied small streams of the Missouri River drainage during the preglacial period (Metcalf 1966; Pflieger 1971).

The activity of glaciers also strongly influenced historical distributions of fishes in parts of the Great Plains. Glaciation of the northwest Great Plains (presently known as the Northwestern Glaciated Plains ecoregion) during the Pliocene era forced Arctic drainages south to their present locations and connected the Missouri River drainage to two unglaciated systems, the Laurentian (St. Lawrence River) and Mississippi Teays (Alden 1932; Cross et al. 1986). This connectivity allowed for fishes unique to the Missouri River, Laurentian, and Mississippi Teays systems to colonize new areas. However, subsequent disconnection of the Missouri River and Laurentian systems by the end of the glacial period once again segregated fish assemblages. This major separation as well as additional fracturing of drainage basins in both glaciated and unglaciated (presently known as the Northwestern Great Plains) regions of the northwestern Great

Plains increased species diversity through isolated speciation and species movement between adjacent drainages (Cross et al. 1986). As a result, the distributions of some species are presently restricted to major drainages. For example, the distribution of pearl dace *Margariscus margarita* is restricted to the Missouri drainage in the Great Plains of Montana (Holton and Johnson 2003)

Local adaptations (e.g., physiological and behavioral changes) of Great Plains fishes to the physicochemical and biotic conditions of prairie streams have also influenced the distributions of fishes in the region. The hydrologic and temperature regimes of streams in the region are highly variable. Turbidity can be high, and habitat diversity is low relative to other areas of North America (Cross et al. 1986; Matthews 1988; Fausch and Bestgen 1997). Native prairie fishes are adapted to these harsh conditions; these adaptations include migration to areas of permanent water, rapid reproduction that may occur several times a year, and high tolerances of poor water quality (Matthews 1988; Labbe and Fausch 2000; Dodds et al. 2004). The distribution of species within streams has also been influenced through inter- and intraspecific competition and predation (Matthews 1988; Fausch and Bestgen 1997). However, information regarding which physicochemical and biotic factors influence the distribution of Great Plains fishes is scarce (Matthews 1988; Zale et al. 1989; Fausch and Bestgen 1997).

Anthropogenic disturbances have further altered the distribution of fishes within Great Plains streams. Stream flow regimes have been altered and fish populations have become fragmented through the construction of small impoundments and subsequent

irrigation withdrawal (Matthews 1988; Dodds et al. 2004). Agricultural practices (mainly cattle grazing and row crop cultivation) have artificially increased nutrient (e.g., nitrogen and phosphorous) and sediment inputs into streams (Fausch and Bestgen 1997), which may reduce numbers of some intolerant fishes [e.g., Iowa darter *Etheostoma exile* and stonecat *Noturus flavus* (Bramblett et al. 2005)]. Additionally, introductions of several species within the Great Plains have created new species interactions that may have negative consequences for native fishes. Some nonnative species (e.g., common carp *Cyprinus carpio* and black bullhead *Ameiurus melas*) are highly tolerant of anthropogenic disturbances and may have a competitive advantage over some native species in these altered habitats (Matthews 1988; Dodds et al. 2004). Introductions of piscivorous species such as northern pike *Esox lucius*, largemouth bass *Micropterus salmoides*, and smallmouth bass *M. dolomieu* have displaced or consumed native fishes (Cross et al. 1986; Matthews 1988; Dodds et al. 2004).

As a consequence of these human disturbances, the prairie biome has become arguably the most endangered region of North America, and the associated aquatic organisms are at greater risk of extinction than in other regions of the continent (Samson and Knopf 1994; Ostile et al. 1997). Historically, prairie streams have been largely overlooked by management agencies across the Great Plains because they lack economic importance and angler interest (Matthews 1988; Samson and Knopf 1994). Evidence for this lack of interest has been found in Montana as well. A recent analysis of the Montana Fish, Wildlife and Parks (FWP) database of the Montana River Information System (MRIS) identified more than 28,000 km of unsampled streams throughout the state; the

majority of these streams were small prairie streams (FWP, unpublished data). However, Great Plains streams are ecologically important because they offer a unique opportunity to study fish species that are subjected to variable flow regimes and sometimes diel fluctuations in water quality and thus learn about disturbance ecology (Matthews 1988; Dodds et al. 2004). These streams are also home to a number of threatened or endangered fish species [e.g., Topeka shiner *Notropis topeka* (Haslouer et al. 2005)]. The importance of conservation of prairie streams is further evidenced in Montana as the biodiversity of prairie streams is higher than in the more studied mountainous streams west of the Continental Divide (Brown 1971). Recently, several state and federal natural resource management agencies in the Great Plains, including Montana, have made the conservation of small prairie fishes and their associated habitat a priority.

To effectively conserve and manage prairie fishes, fisheries scientists must understand their ecology (Schlosser 1991; Fausch et al. 2002). Understanding how ecosystem processes influence the distribution of fishes varies widely across spatial scales (Rabeni and Sowa 1996; Allan et al. 1997). The use of statistical models (e.g., regression, classification and ordination techniques, neural networks) to explain the distribution, abundance, or biomass of species using physicochemical or watershed variables has provided insight into how these characteristics affect fish assemblage structure (Tonn and Magnuson 1982; Tonn et al. 1983; Hawkes et al. 1986; Lanka et al. 1987; Leonard and Orth 1988; Matthews and Robison 1988; Rabeni and Sowa 1996; Maret et al. 1997). Similarly, biotic models have shown how competition and predation influence species distributions at the reach scale; most studies are related to the influence

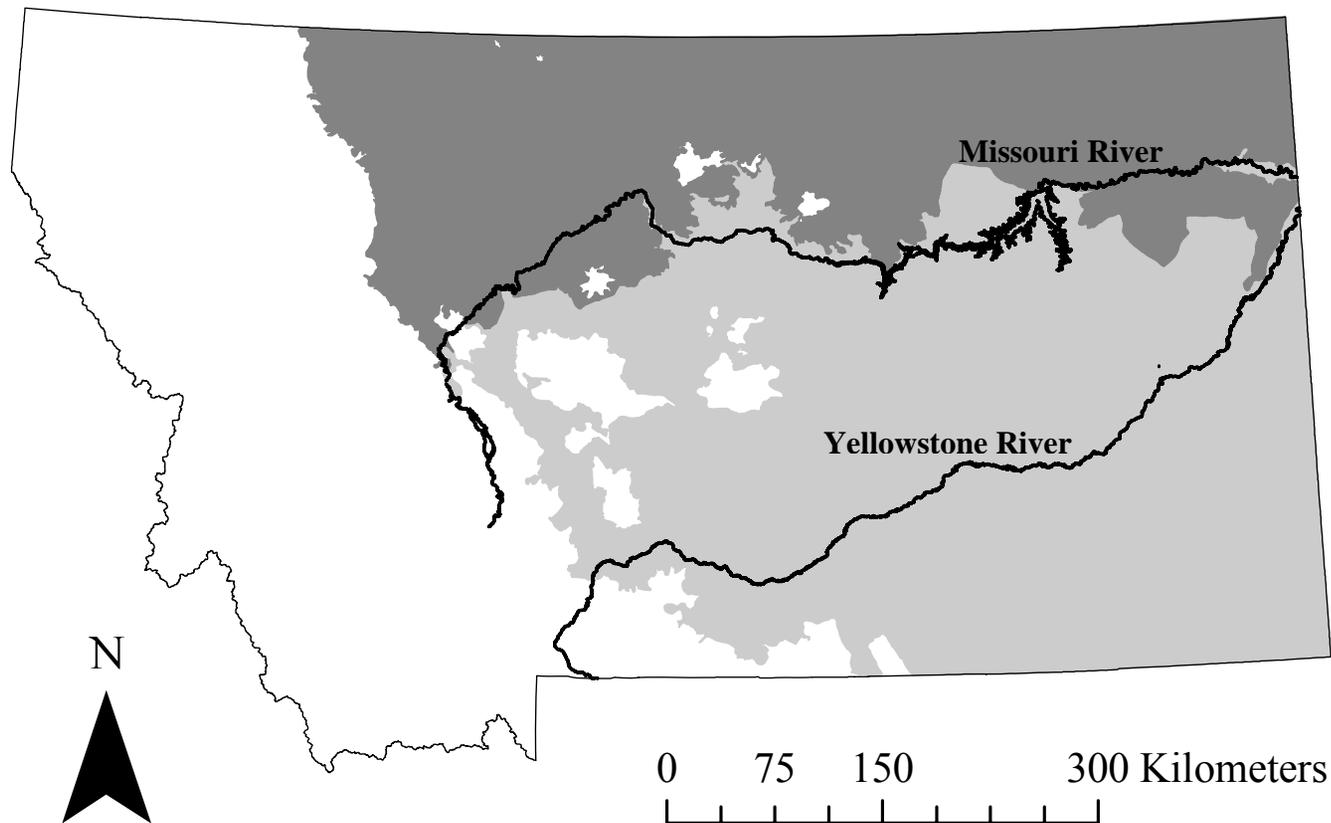
nonnative species may have on native species assemblages (Li and Moyle 1981; Grossman et al. 1998). Whereas biotic interactions and physicochemical characteristics of the stream reach have been identified as factors influencing fish assemblage structure, no study to date has compared the explanative or predictive power of species distribution models constructed using physicochemical and biotic variables measured at the reach scale concurrently to those constructed with watershed-scale variables. Additionally, no study has used all three factors simultaneously to understand how these factors operate in concert to influence the distribution (i.e., presence or absence) of fish species in streams. Therefore, the objectives of this study were to: i) identify the reach- and watershed-scale characteristics that affect the presence or absence of fishes in prairie streams of Montana; and ii) identify which scale best explains the presence or absence of fishes in prairie streams of Montana: reach, watershed, or a combination of variables measured at the reach and watershed scale.

STUDY AREA

The Great Plains of North America encompasses a triangular area of about 2.9 million km² (Omernik 1987). The western border of the Great Plains is the Rocky Mountain front, and the plains continue eastward to the middle of Iowa and Missouri. The north-south borders of the Great Plains of North America include the Canadian province of Saskatchewan and parts of Mexico.

This study examined small streams of the Great Plains in Montana. The Great Plains comprises about two-thirds of the area of the state and includes parts of the Missouri and Yellowstone river drainages (Figure 1). Two ecoregions are present within the Great Plains of Montana: the Northwestern Glaciated Plains and the Northwestern Great Plains (Level III; Omernik 1987). The Northwestern Glaciated Plains ecoregion in Montana is bordered on the north by Canada and on the south by the Missouri River. As previously mentioned, glaciation of this region occurred during the Pliocene era; thus, glacial till makes up the predominant soil type in this ecoregion. The soils are fertile and support native vegetation typical of both tallgrass (e.g., western wheatgrass *Pascopyrum smithii* and big bluestem *Andropogon girardi*) and shortgrass prairie (e.g., silver sagebrush *Artemisia cana* and prickly pear cactus *Opuntia* spp.). The soil is also productive enough for row crop cultivation; thus, land conversion for wheat and hay production is common (WSAL 2003).

Figure 1. Map of Great Plains within Montana. The Northwestern Glaciated Plains ecoregion (dark gray) within the state extends from the Canada border south to roughly the Missouri River. The Northwestern Great Plains (light gray) of the state extends roughly south from the Missouri River to the Wyoming border. Major rivers in this area include the Missouri and Yellowstone rivers.



The Northwestern Great Plains of Montana includes the area south of the Missouri River to the Wyoming border. Glacial activity did not extend this far south during the Pliocene; as a result, the soils are not as fertile as in the north. Livestock grazing is more prevalent in the region than in the Northwestern Glaciated Plains (Slagle 1984). Vegetation of this unglaciated area is typical of shortgrass prairie (WSAL 2003). In both ecoregions, woody vegetation is rare except in the cottonwood floodplains (Omernik 1987). Consequently, grasses are important sources of allochthonous inputs into prairie streams (Matthews 1988).

The Northwestern Great Plains and the Northwestern Glaciated Plains have similar climates. Both are semiarid and receive less than 400 mm of rain per year (Omernik 1987). Forty percent of the yearly precipitation occurs in the summer (i.e., May to August) and 10% occurs in the winter (Slagle 1984). The daily average high temperature is 18°C in the summer and -4°C in the winter (Slagle 1984).

Climate strongly influences the hydrology of Great Plains streams as it affects the frequency and duration of flooding and drying cycles (Matthews 1988). During spring, snowmelt from the montane and prairie regions flood some prairie streams, creating connectivity to the floodplain (Matthews 1988; Dodds et al. 2004). However, other prairie streams without montane headwaters or in dry years have little or no snowmelt stream flow. Many prairie streams are regularly desiccated during warmer, dryer months (July-September); this dehydration is likely related to transpiration and evaporation caused by summer ambient temperatures rather than a lack of precipitation (Matthews 1988). Drying periods commonly create intermittency in Great Plains streams (Matthews

1988; Fausch and Bestgen 1997; Dodds et al. 2004). However, thunderstorms during early to mid-summer may once again flood streams and reconnect intermittent stream reaches. By late summer, prairie soils are usually very dry, and only small amounts of the rain that does fall during this time may actually reach the stream bed. Intermittent conditions are restored as a result (Matthews 1988). The presence of groundwater inputs and the depth of scoured pools may be important factors in maintaining isolated pools during this drying period (Anderson 1973).

METHODS

Site Selection

The data used in this study came from two separate projects: the development of an Index of Biotic Integrity (IBI) for eastern Montana from 1999 through 2001 (Bramblett et al. 2005), and a biological survey of prairie streams from 2002 through 2004 (Bramblett et al. 2003). Thus, site selection varied between projects. For the IBI project, 66 streams were selected for evaluation. Fifty-four of these streams were selected based on a sampling frame consisting of second- through seventh-order streams (as classified by Strahler 1957) that were classified as perennial according to U.S. Geological Survey 1:100,000- scale topographic maps. Roughly equal proportions of stream orders were surveyed. An additional 12 streams were hand selected to represent reaches of high and low degrees of human influence based on information from unpublished biological and physicochemical data from the Montana Department of Environmental Quality (MDEQ) and other relevant publications (Bramblett et al. 2005).

One hundred and sixty-nine streams were selected for the biological survey study. To select random sites, the Northwestern Great Plains and Northwestern Glaciated Plains were first overlaid by fourth-code Hydrologic Cataloging Units (HUCs) in a geographic information system (GIS) to determine which HUCs were located within predominantly prairie areas. Any HUC that did not include at least 50% prairie habitat was eliminated. Within the remaining HUCs, second- through seventh-order (i.e., wadable) streams that were unsurveyed or that had not been surveyed in the past ten years were identified using

the MRIS database. The landscape surrounding each of these streams was then evaluated using GIS and local area maps, and all streams draining mountainous or forested regions were eliminated from site selection. The remaining streams were sorted in descending order by length. Longer streams were believed to be more likely to have water during the drier summer months than shorter streams. The number of streams selected for sampling in each HUC was proportional to the total area of the HUC relative to the total size of the prairie in eastern Montana. The sampling location on each stream was selected randomly by first generating a river kilometer followed by a random latitude and longitude within that river mile. Parcel ownership of the site was determined prior to sampling using the Montana Cadastral Mapping Program (<http://gis.mt.gov>), and permission to sample was requested from each landowner. If permission was denied, an alternate location on that stream was selected and the procedure for obtaining permission was repeated. If permission to sample at the alternative location was denied, then a sample site from the next longest stream in the HUC was selected.

Fish Sampling

Sampling for fish occurred from July to mid-September each year from 1999 through 2004. Length of stream sampled for fish differed depending on the project. During the IBI project (1999-2001), the length of stream sampled was equal to 40 times the mean channel width (range of stream lengths sampled: 150 – 500 m); this length is considered adequate enough to capture 90% of fish species present in all streams (Karr et al. 1986; McCormick and Hughes 1998). During the biological survey study, fish were

sampled by seining a 300-m reach, a length thought sufficient to capture 100% of fish species present in Wyoming prairie streams (Patton et al. 2000). Seine lengths of 4.6, 6.1, and 9.1 m were used, and the length of the seine used depended on the width of the stream. The mesh size of the seines was 6.4 mm, which can capture fish ≥ 30 mm in total length. Fish were identified to species and enumerated in the field. A maximum of ten individuals of each species were preserved in formalin as voucher specimens, and all voucher specimens were examined in the laboratory to verify field identifications.

Environmental Characteristics

Reach Scale

Variables measured at the reach scale included physicochemical and biotic characteristics. Physicochemical characteristics were defined as the chemical and physical attributes of each stream reach and the immediate 10-m riparian zone. Prior to fish sampling at each site, water quality was assessed by measuring dissolved oxygen (mg/L), conductivity ($\mu\text{S}/\text{cm}$), salinity (ppt), and temperature ($^{\circ}\text{C}$) using a YSI® 85 meter, pH using an Oakton® pH pen, and water turbidity [Nephelometric Turbidity Units (NTUs)] using a LaMotte® 2020 turbidity meter. Discharge was quantified using a Marsh-McBirney® flowmeter if the stream was flowing.

Physical attributes of the stream were assessed in each sampling reach as outlined by the U.S. Environmental Protection Agency's (EPA) protocol for characterization of the physical habitat of streams (Kaufman and Robison 1998) after fish sampling. Eleven transects crossed the wetted width of the stream and were evenly spaced within the reach

(Figure 2). Sixteen physical attributes were measured at each transect (Table 1).

Between each transect, thalweg depths, the presence of fine sediment (i.e., < 2 mm diameter), and an index of water velocity (as defined by Gorman and Karr 1978) were recorded at each of 10 evenly spaced intervals. All physical habitat measurements taken at each site were used to compute 30 metrics as defined by Kaufman et al. (1999); these were combined with the four water quality parameters to characterize the physicochemical environment for fishes at the reach scale (Table 2).

Figure 2. Graphical representation of the sampling reach and the eleven transects within the reach. The upper and lower ends of the reach are counted in the eleven transects.

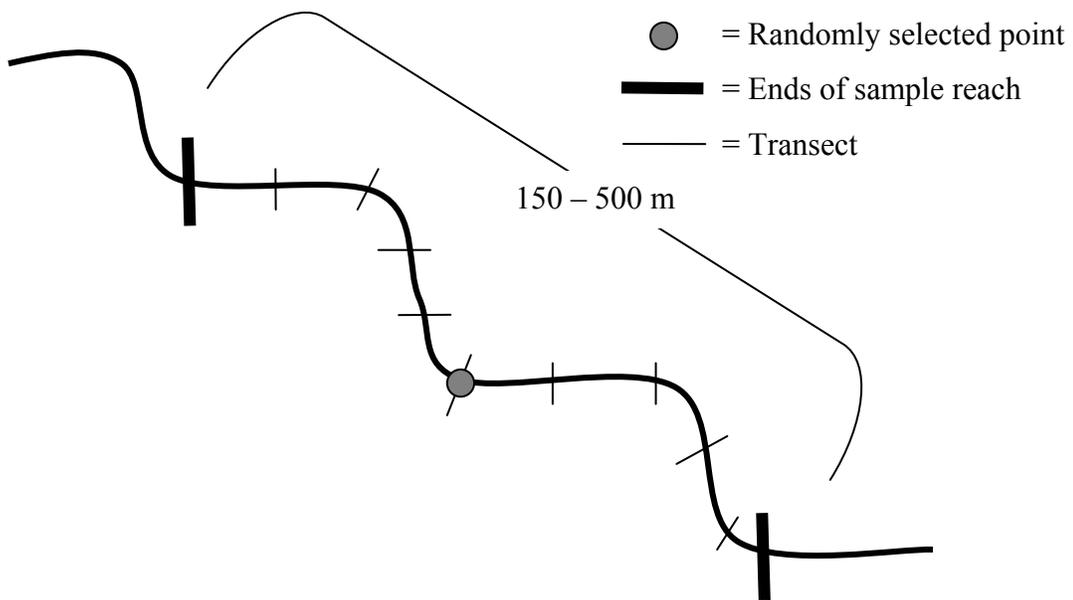


Table 1. Physical attributes measured at each reach sampled according to EPA protocol (Kaufman and Robison 1998).

Habitat measurement	Method or equipment	Units	Locations of measurement
Bank angle	Clinometer	Degrees	At right and left banks at each transect
Bankful height	PVC pipe marked in centimeter increments	Meters	At each transect
Bankful width	Tape measure	Meters	At each transect
Bearing	Compass	Degrees	At second through eleventh transects
Canopy cover	Densiometer	Percent	At right and left banks and in four directions in center of transect
Embeddedness of substrate	Visual estimation	Percent	Five evenly spaced intervals along each transect
Fish vegetation cover (algae, overhanging vegetation, etc.)	Visual estimation	Percent	At each transect
Human influence along stream banks (agriculture, roads, etc.)	Visual estimation	Classified according to distance from stream bank	At each transect
Incised bank height	PVC pipe marked in centimeter increments	Meters	At each transect
Large woody debris	Visual estimation	Counts	At each transect
Riparian vegetation cover (brush, grass, bare ground, etc.)	Visual estimation	Percent	Within a 100 m ² unit around the right and left banks of each transect
Slope	Clinometer	Percent	At second through eleventh transects
Substrate of stream bed (silt, gravel, etc.)	Visual estimation	Assigned size class code	Five evenly spaced intervals along each transect
Undercut banks	PVC pipe marked in centimeter increments	Meters	At right and left banks at each transect
Water depth	PVC pipe marked in centimeter increments	Centimeters	Five evenly spaced intervals along each transect and at ten evenly spaced intervals along the thalweg between each transect
Wetted width	Tape measure	Meters	At each transect

Table 2. Physicochemical habitat variables calculated at each sample site according to Kaufman et al. (1999), their acronyms, and units of measure.

Acronym	Variable	Unit
Agric	Agricultural disturbance within the riparian area	Proximity-weighted index
BankH	Mean bankful height	m
BankW	Mean bankful width	m
Big	Substrate larger than sand and gravel (diameter >2 mm)	Percent
Brush	Brush and small woody debris areal cover	Percent
Cond	Stream conductivity	$\mu\text{S}/\text{cm}$
Depth	Mean thalweg depth	cm
Dry	Dry channel	Percent
Embed	Mean embeddedness of channel substrate	Percent
Fast	Falls, cascades, rapids, or riffles	Percent
Fine	Substrate sand or fines (diameter <2 mm)	Percent
Ground	Riparian ground-layer with herbaceous and woody vegetation cover	Percent
Human	Anthropogenic disturbance within the riparian area (all types)	Proximity-weighted index
Incise	Mean incision height	m
LogSub	Log_{10} (estimated geometric mean substrate diameter)	Log_{10} mm
LWD	Large woody debris areal cover	Percent
MVeloc	Maximum velocity index	Index
OhVeg	Overhanging vegetation areal cover	Percent
Ph	pH	pH
Pool	Percent area of channel that is considered pool habitat	Percent
SDDepth	Standard deviation of thalweg depth	Standard deviation
SDVeloc	Standard deviation velocity	Standard deviation
SDWidDep	Standard deviation of ratio between wetted width and thalweg depth	Standard deviation
SDWidth	Standard deviation width	Standard deviation
Sinu	Channel sinuosity	Index
Slope	Water surface gradient over reach	Percent
Slow	Glides and all pools	Percent
Temp	Temperature	$^{\circ}\text{C}$
Turb	Turbidity	NTU
Under	Mean bank undercut distance	m
VegCover	Filamentous algae and aquatic macrophyte areal cover	Percent
Veloc	Mean velocity index	Index
WidDep	Mean ratio between wetted width and thalweg depth	m/m
Width	Mean wetted width	m

Biotic variables were defined as characteristics of the fish community sampled within the reach. Ten variables were selected based on hypotheses regarding how biotic interactions among species may influence the distribution of fishes in prairie streams of Montana (Table 3). If the presence of the species being modeled was included in one of the biotic variables, one of two options was selected: i) the variable was modified to exclude that species in the calculation; or ii) the variable was omitted from the model. For example, when modeling the distribution of northern pike, two variables were eliminated from the model (PresNopi and PerNopi; Table 3). In addition, two variables (IntPisc and PerInt) were calculated by omitting the abundance of northern pike. All variables were calculated using Structured Query Language (SQL) in Microsoft Access (2003).

Table 3. Biotic variables measured at each site, their acronyms, and units of measurement. Scientific names of all species are found in Table 6.

Acronym	Variable	Unit
IntPisc	Percent of introduced piscivores (i.e., green sunfish, northern pike, smallmouth bass, walleye, yellow perch)	Percent
NatPisc	Percent of native piscivores (i.e., channel catfish, sauger)	Percent
NatRich	Native species richness	Number of native species
PerBlbu	Percent of black bullhead	Percent
PerCoca	Percent of common carp	Percent
PerGrsu	Percent of green sunfish	Percent
PerInt	Percent of introduced non-piscivores (i.e., black bullhead, common carp, northern plains killifish, pumpkinseed, spottail shiner, white crappie)	Percent
PerNopi	Percent of northern pike	Percent
PresNopi	Presence of northern pike	0 or 1
SWDiv	Shannon-Weiner diversity index	Index

Watershed Scale

Watershed-scale characteristics were defined as the physical and biological attributes of the watershed upstream of the sampled reach. Twenty-six variables (Table 4) were quantified using GIS data layers in ArcGIS (Version 9; ESRI 2004) according to the methods in Appendix A. Several of the variables selected for the analysis [i.e., FEco, GwWNum, IrrDen, IrrTL, Lat, Long, MinPerm, PtElev, RelGrad, RdDen, RdLen, SoilErr, Strahl, UsDam, WsArea, and land cover variables (e.g., Bare, Crop, Shrub)] were selected based on other studies that evaluated the effects of watershed-scale characteristics on the distribution and abundance of fish species (Jackson and Harvey 1989; Angermeier and Winston 1998; Oakes et al. 2005; Smith and Kraft 2005). Other variables (i.e., AFlood, AveSalt, DsDam, MaxSalt, PCost, SoilPerm, WatYield, WsElev, WTDepA, WTDepL) were selected based on hypotheses regarding which variables may be important in explaining and predicting the distribution of fish in eastern Montana.

Statistical Analysis

Reach-scale, Watershed-scale, and Combined Models

All sites with no water at the time of sampling or where no fish were collected were omitted first from the modeling database. Additionally, any sites whose watersheds were partially within Canada were eliminated due to incomplete GIS coverage in Canada. One-hundred twenty sites were left in the modeling database (Figure 3). This database was further restricted by eliminating sites that lacked values for physicochemical, biotic, or watershed-scale characteristics. The number of sites where individual species were

present was calculated from this new data set. Any species that occurred at fewer than 10 sites was excluded from analyses because prediction models for extremely rare species are unreliable (Breiman et al. 1984). The presence or absence of the remaining species was related to variables measured at the reach and watershed scale using tree classifiers in R 2.0.1 [R Foundation for Statistical Computing (2004); TREE package, Ihaka and Gentleman (1996)]. This modeling method was selected for several reasons: i) tree classifiers have been shown to be especially useful in modeling ecological data that has several explanatory variables but only a single response (i.e., presence or absence; Breiman et al. 1984; Guisan and Zimmerman 2000); ii) tree classifiers have been shown to explain and predict species distributions as well or better than other statistical modeling techniques [e.g., regression, classification and ordinal techniques, neural networks (Olden and Jackson 2002; Oakes et al. 2005)]; iii) tree classifiers can use both continuous and categorical data to model species distributions (Breiman et al. 1984); iv) computation of the trees is relatively simple in comparison to more complex methods such as artificial neural networks (Keith Gido, Kansas State University, personal communication); v) the resulting tree for each species is simple to interpret for novice as well as expert users of tree classifiers (Vayssières et al. 2000); and vi) results from tree classifiers can be easily integrated into GIS to create potential distribution maps for species (Lees and Ritman 1991). Because tree classifiers have been rarely used in fisheries studies to date, it is important to provide background regarding how the models operate in explaining and predicting species distributions. Each species has an associated null deviance (D). This null deviance is a measure of the variability in the distribution of

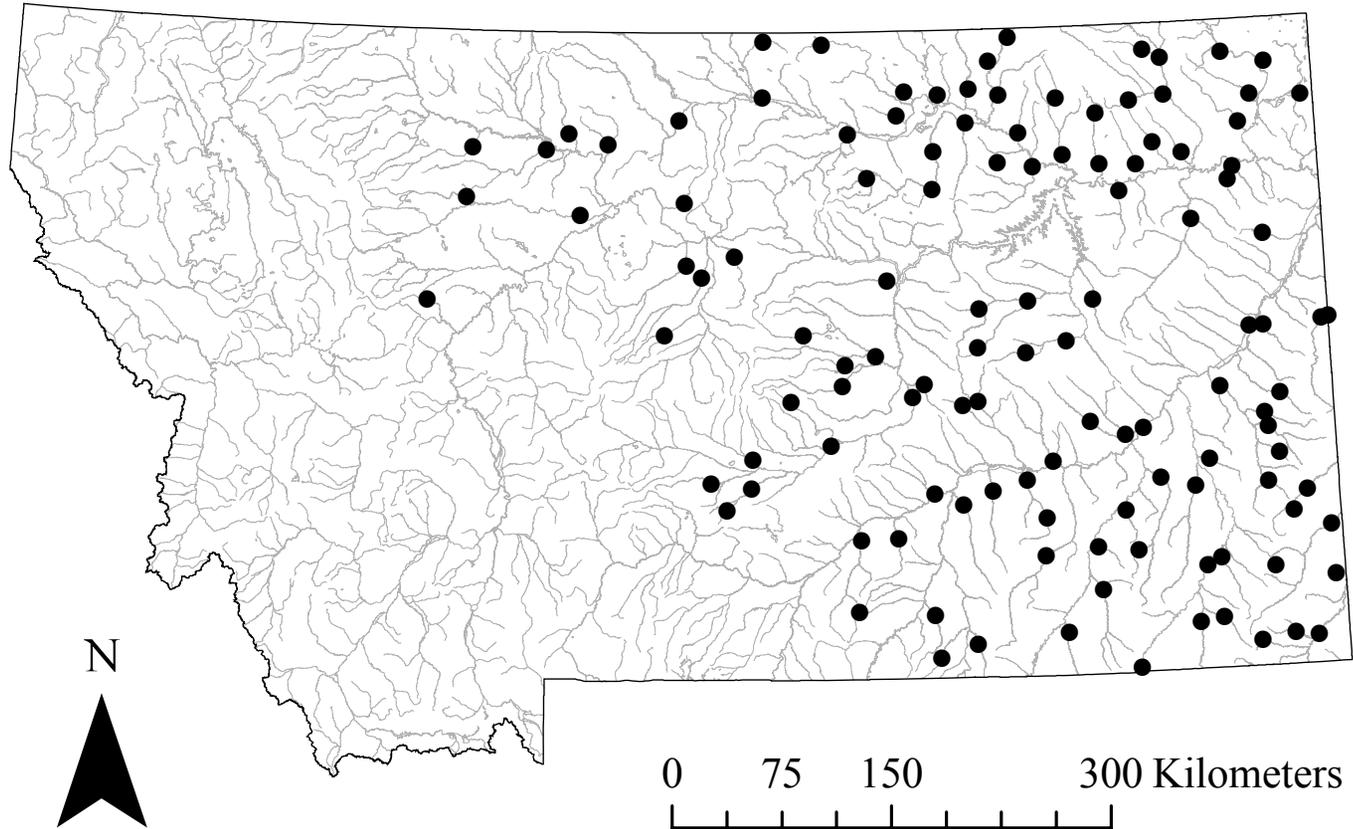
Table 4. List of watershed variables calculated for each site, their acronyms, and the data source.

Acronym	Variable	Data source
AnnFld	Annual flood frequency of soils in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
AveSalt	Average salinity of soil in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
Bare, Commer, Crop, Devel, Forest, Grass, Mines, Rock, Shrub, Snow, Trans, Water, Wetlan	Percentage of watershed covered by National Land Cover Database-defined type	Calculated by category of type (i.e., barren, commercial development, cropland, non-commercial development, forest, grassland, mines, rock, shrub, snow, transportation, water, or wetlands) using National Landcover Database (NLCD) (http://landcover.usgs.gov/landcoverdata.php)
Drain	Major drainage basin (i.e., Little Missouri, Missouri, or Yellowstone river)	Calculated in m ² using 5 th code HUC layers (http://nris.mt.gov/nsdi/nris/shape/hd109.zip)
DsDam	Presence of dams (i.e., yes or no) between the sample site downstream to the confluence of major river (i.e., Little Missouri, Missouri, Musselshell, Poplar, Powder, Teton, Tongue, and Yellowstone)	Calculated using data from Army Corps of Engineers National Inventory of Dams (http://crunch.tec.army.mil/nid/webpages/nid.cfm).
Eco	Level III ecoregion (Omernik 1987)	Provided by Jim Johnson, GIS Programmer/Analyst, Confluence, Inc., Bozeman, Montana
FEco	Level IV ecoregion (Omernik 1987)	Provided by Jim Johnson, GIS Programmer/Analyst, Confluence, Inc., Bozeman, Montana
GwWNum	Number of groundwater wells in watershed	Summarized using groundwater well location data (Montana http://nris.state.mt.us/nsdi/nris/shape/gwicwells.zip ; North Dakota: http://www.swc.state.nd.us ; South Dakota: Provided by Ron Duvall, Natural Resources Engineer, South Dakota Water Rights Program, Pierre, South Dakota; Wyoming: http://www.sdvc.uwyo.edu/metadata/wells.html)
IrrDen	Density of irrigation ditches within the watershed	Calculated from the National Hydrography Data set(NHD; (http://nris.mt.gov/nsdi/nhd/data/nhd_wb.zip) and watershed area
IrrTL	Total length of irrigation ditches in watershed	Calculated from the NHD (http://nris.mt.gov/nsdi/nhd/data/nhd_wb.zip)
Lat	Latitude of sample site	Provided by Garmin eTrex® GPS unit
Long	Longitude of sample site	Provided by Garmin eTrex® GPS unit
MaxSalt	Maximum soil salinity in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
MinPerm	Minimum soil permeability in watershed	Calculated from State Geographic Soil (STATSGO) data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
PCost	Cost of path for fish from site of collection to nearest major river	Calculated from Strahler stream order and network distance to nearest major river using the NHD (http://nris.mt.gov/nsdi/nhd/data/nhd_wb.zip)

Table 4. Continued.

Acronym	Variable	Data source
PtElev	Elevation of the sample site	Measured from 30-m digital elevation models (DEMs; Montana: http://nris.mt.gov/nsdi/nris/shape/el11.zip ; North Dakota: http://www.nd.gov/gis/mapsdata ; South Dakota: http://www.sdgs.usd.edu/digitalpubmaps/index.html ; Wyoming: http://www.wygisc.uwyo.edu/clearinghouse/land.html)
RelGrad	Relative gradient	A slope measurement calculated within 1 km above and below sample point along stream network using DEMs (Montana: http://nris.mt.gov/nsdi/nris/shape/el11.zip ; North Dakota: http://www.nd.gov/gis/mapsdata ; South Dakota: http://www.sdgs.usd.edu/digitalpubmaps/index.html ; Wyoming: http://www.wygisc.uwyo.edu/clearinghouse http://www.wygisc.uwyo.edu/clearinghouse/land.html)
RdDen	Road density by length within watershed	Calculated using TIGER road data (Montana; http://nris.mt.gov/nsdi/nris/shape/road2000.zip ; North Dakota: http://www.nd.gov/gis/mapsdata ; South Dakota: http://www.sdgs.usd.edu/digitalpubmaps/index.html ; Wyoming: http://www.wygisc.uwyo.edu/24k/road.html)
RdLen	Total length of roads in watershed	Calculated using TIGER road data (Montana; http://nris.mt.gov/nsdi/nris/shape/road2000.zip ; North Dakota: http://www.nd.gov/gis/mapsdata ; South Dakota: http://www.sdgs.usd.edu/digitalpubmaps/index.html ; Wyoming: http://www.wygisc.uwyo.edu/24k/road.html)
SoilEr	Average soil erodibility in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
SoilPerm	Average soil permeability in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
Strahl	Strahler stream order of the sample site	Calculated from NHD (http://nris.mt.gov/nsdi/nhd/data/nhd_wb.zip)
UsDam	Number of dams in the watershed upstream of sample site	Calculated using data from Army Corps of Engineers National Inventory of Dams (http://crunch.tec.army.mil/nid/webpages/nid.cfm)
WatYield	Maximum amount of water falling in the form of precipitation in the watershed	Volume of water calculated in m ³ from Oregon Climate Service PRISM precipitation (http://www.ocs.orst.edu/prism/products) for western U.S. and watershed area
WsArea	Watershed area	Calculated in km ² using 5 th (http://nris.mt.gov/nsdi/nris/shape/hd109.zip) and 6 th code(provided by Gerry Daumiller, Programmer/Analyst, Montana Natural Resource Information Center, Helena, Montana) HUC layers
WsElev	Average elevation of the watershed	Measured from DEMs (Montana: http://nris.state.mt.us/nsdi/nris/el10/dems.html ; North Dakota: http://www.state.nd.us/gis ; South Dakota: http://www.sdgs.usd.edu ; Wyoming: http://www.wygisc.uwyo.edu/clearinghouse)
WTDepA	Average water table depth in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov)
WTDepL	Minimum water table depth in watershed	Calculated from STATSGO data for Montana, North Dakota, South Dakota, and Wyoming (http://soildatamart.nrcs.usda.gov/)

Figure 3. Map of the 120 sites used in the modeling database. All sites were sampled between July and September from 1999 to 2004.



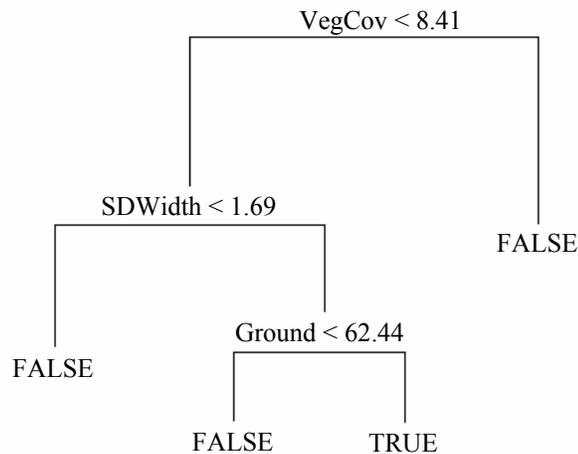
a species and is calculated based on the number of presences versus the number of absences of the species in the data set (Breiman et al. 1984):

$$D = -2N [p_i \ln(p_i) + (1 - p_i) \ln(1 - p_i)]$$

where D is the null deviance of the model, N is the sample size under consideration, and p_i is the proportion of the number of occurrences of the species to the total number of sites in the data set. Species that occur at exactly 50% of sites in the database have the highest null deviance, and null deviance decreases both for species present at <50% of sites and species present at >50% of sites. The TREE package (Ihaka and Gentleman 1996) in R (2004) uses a recursive partitioning algorithm to minimize the deviance in the individual species model and categorize sites where the species is present or absent into homogeneous categories based on values of the predictor variable. Thus, predictor variables and their associated values are selected by an automated procedure based on the amount of deviance reduced in individual species distributions. The result is a dichotomous key that is relatively easy to interpret (Figure 4). The minimum number of nodes in any tree is two.

Tree classifiers are “grown” until groups (i.e., sites where the species is present and sites where the species is absent) are homogeneous or until a minimum N (in this case, $N = 5$ sites) has been reached. As a result, the models have a tendency to overfit the data (Breiman et al. 1984) and must be “pruned” to explain the population rather than the sample. The process of pruning is done by first ordering the sequence of nested subtrees. This ordering sequence is based on a recursive removal of splits of the full tree and determining which splits are least important based on a cost-complexity measurement

Figure 4. Example of a tree classifier for a species modeled at the reach scale. The words TRUE and FALSE signify terminal nodes. Interpretation of this model indicates that this species is likely to be found where instream vegetative cover is less than 8.41%, the standard deviation of width is greater than 1.69, and the percent of riparian ground cover is greater than 62.44%.



that weighs the amount of deviance reduced in the subtree and the resulting number of terminal nodes. An automated tenfold cross-validation procedure (Ihaka and Gentleman 1996) was used to randomly select 90% of the data (i.e., the learning sample); these data were used to build a full tree and subsequently create the sequence of nested subtrees. The other 10% of the data (i.e., the test sample) was “dropped” through each of the trees in the sequence and the residual deviance between the learning and test samples was determined. The entire cross-validation procedure is done ten times leaving out a different 10% of the data each time. The graphical output produced by R (2004) compares the residual deviance, the number of terminal nodes, and the cost-complexity factor of a tree with that number of terminal nodes (Figure 4). The ideal result is somewhat of a U-shape (Figure 5A), whereby the residual deviance decreases as the number of terminal nodes increases then increases as the increasing number of terminal

nodes results in a higher cost-complexity. The lowest point in this U-shaped line was used to determine the number of terminal nodes in the pruned tree. The variables producing these nodes were interpreted to be important in explaining or predicting individual species distributions (Breiman et al. 1984). Multiple numbers of terminal nodes may produce the lowest residual deviance in the cross-validation procedure. In such cases, the largest number of terminal nodes was chosen in order to avoid missing any variables that may be important in explaining or predicting the distribution of the species.

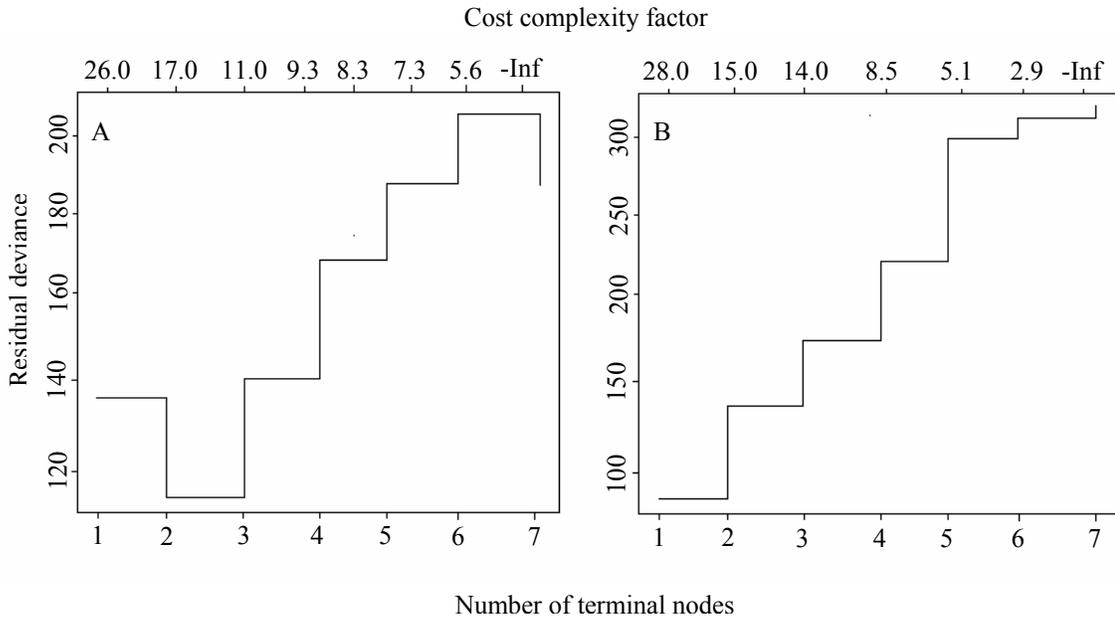
For each species, a hierarchy of models was developed: i) those that contained only reach-scale variables; ii) those that contained only watershed-scale variables; and iii) models that contained combinations of both reach- and watershed-scale variables (i.e., combined models). Models were first required to explain significant amounts of null deviance (i.e., models had to meet the deviance reduction criterion). Evaluating whether models met this criterion was determined from the cross-validation procedure explained above. The cross validations should result in a U-shaped residual deviance line in the resulting graph (Figure 5A). However, the initial decrease in residual deviance may not be seen, and the lowest residual deviance may be found where the number of terminal nodes equals zero or one (Figure 5B). The minimum number of terminal nodes in any functional tree is two. In cases where the suggested terminal node size is zero or one, a model that uses any variable and its associated value to explain the deviance in the distribution of a species may not perform better than the null model (i.e., a model that includes no splits and thus no terminal nodes). Therefore, models in which the U-shaped

trend in the residual deviance line was not detected (i.e., did not meet the deviance reduction criterion) were not considered to be statistically significant and were eliminated from further analysis.

Three metrics (i.e., sum of the percent deviance reduced; misclassification error rate (MER); MER reduced; improvement in MER) were calculated to evaluate significant models for their overall explanative and predictive ability. To summarize how well models explained species distributions, I calculated an overall percentage of the deviance reduced by all variables used in a given model. Calculating this metric was accomplished by first computing the percent deviance reduced by each variable within the model (i.e., the ratio between the amount of deviance reduced by a variable and the null deviance of that model multiplied by 100). Once this was completed, I summed the percent of deviance reduced for all variables used within the model. I assumed that larger values of summed deviance meant that the variables used in these models explained species distributions well.

Misclassification error rate and extensions of this metric (MER reduced and improvement in MER) were used to evaluate the predictive ability of models. Misclassification error rate was calculated as the ratio between the number of incorrect predictions of a model and the total number of sites in the data set. This metric has been used in other studies to measure the predictive ability of models (Olden and Jackson 2002; Oakes et al. 2005; Smith and Kraft 2005). Low MER values typically indicate that variables used within a given model may be used to predict species distributions.

Figure 5. Two examples of the graphical results of the cross-validation procedure from this study. Graph A met the deviance reduced criterion because there is an initial increase followed by a rapid decrease in residual deviance at a terminal node value greater than zero or one. The number of terminal nodes selected for the pruned tree in this example is three. Graph B did not meet the deviance reduced criterion because no such trend was detected.



Unlike MER, the MER reduced metric has never been used in a published study, but its use in evaluating models has been recommended (Olden and Jackson 2002). The amount of MER reduced determines how the inclusion of variables in a species model improved the predictive ability of the model compared to the null model (i.e., a model based on chance alone). To calculate MER reduced, two other metrics were calculated first: MER and the null error rate. The null error rate was calculated one of two ways depending on the occurrence of the species. For species present at <50% of the total number of sites, the null rate was calculated as the number of sites where a species was present divided by the total number of sites in the data set. For species present at >50% of the total number of sites, the null rate was calculated as the number of sites where a

species was absent divided by the total number of sites in the data set. The MER reduced metric was then calculated by subtracting the MER from the null error rate. Models with high MER reduction values likely demonstrate that variables contained within a given model may be used to predict species distributions.

Similar to MER reduced, metrics evaluating improvement in MER have not been used in fisheries literature to date. Percent improvement in MER was computed for individual species models by calculating the ratio between the MER reduced and the null error rate and multiplying by 100. Large improvements in MER likely indicate that variables within given model can be used to predict the distribution of that species.

Two metrics (i.e., the number of variable occurrences and the average percent deviance reduced) were used to evaluate individual variables contained within significant models. Average percent deviance reduced has often been used in ecological studies to determine which environmental characteristics have the greatest influence on species distributions (Breiman et al. 1984). Percent deviance reduced was calculated for each variable included in species models (see methods described above). An average of the percent deviance reduced by each variable was calculated for variables appearing in more than one model by summing the amount of deviance reduced by the variable each time it appeared within models of similar scale and dividing by the number of occurrences of that variable. I assumed that variables with larger average deviance reduction values were better at explaining prairie fish distributions than lesser variables.

The percent variable occurrence metric involved the number of times a particular variable appeared within each model type. This criterion has not been used in other

published studies. It was assumed that the more often a variable appeared within species models, the more important the variable was in explaining species distributions at a particular scale.

Habitat Categories

A large number of reach- and watershed-scale variables was used in this study to model the distribution of fishes in comparison to other studies. It is often difficult to find a balance between including enough variables to represent all species being modeled but not so many that biological relevance becomes unclear (Oakes et al. 2005). To determine the biological relevance of the variables used in species models, variables were grouped into distinct categories *a posteriori* based on which general aspect of the environment was being measured by each variable (Table 5). The average amount of deviance reduced by a category was calculated by multiplying the average deviance reduced by each variable within the category by the number of occurrences of that variable, summing the weighted averages, and dividing by the total number of occurrences of variables used in that category. The weighted averages of percent deviance reduced and the resulting 95% confidence interval were visually compared among all categories to determine which groupings of variables best explained species distributions.

Scale Comparisons

Average residual mean deviance was used to compare the explanative power of models at multiple scales. A residual mean deviance (D^2) was calculated for each species model. Calculation of this metric is similar to that of a generalized linear model (GLM;

Table 5. Categorization of variables by biological relevance and by scale. Variables were grouped *a posteriori* based on the general aspect of the environment being measured. Variables are listed in Tables 2, 3, and 4.

Scale	Category	Variables
Reach	Channel morphology	BankH, BankW, Depth, Dry, Incise, LogSub, Pool, SDDepth, SDWidDep, SDWidth, Sinu, Slope, WidDep, Width
	Instream cover	Brush, Ground, LWD, OhVeg, Under, VegCov
	Introduced non-piscivores	PerCoca, PerBlbu, PerInt
	Piscivores	IntPisc, NatPisc, PresNopi, PerGrsu, PerNopi
	Species richness and diversity	NatRich, SWDiv
	Substrate	Big, Embed, Fine
	Velocity	Fast, MVeloc, Salt, SDVeloc, Slow, Veloc
	Water quality	Agric, Cond, Human, Ph, Temp, Turb
Watershed	Anthropogenic influences	Commer, Crop, Devel, DsDam, Fallow, GwWNum, IrrDen, IrrTL, Mines, RdDen, RdLen, UsDam
	Elevation	PtElev, RelGrad, WsElev
	Land cover	Bare, Forest, Grass, Rock, Shrub, Snow, Trans, Water, Wetlan
	Position	Drain, Lat, Long
	Stream size	PCost, Strahl, WatYield, WsArea
	Soil characteristics	AnnFld, Ave Salt, FEco, MaxSalt, MinPerm, SoilEr, SoilPerm, WTDepA, WTDepL

Breiman et al. 1984) and is computed as the ratio between the amount of unexplained deviance within the model and the null deviance of that model. In GLM, D^2 is used to examine whether model is accurate in explaining or predicting the relationships between explanatory and response variables (McCullagh and Nelder 1989); tree classifiers calculate this metric for the same purpose. To compare among the model types, an average D^2 was calculated for each model type (i.e., reach-scale, watershed-scale, and combined models). The model type with the lowest average D^2 was determined to be the scale that best explained species distributions. No formal statistical tests were used to determine if there were significant differences between the D^2 values calculated for each model type since variables in the combined models were not independent of the variables used in reach- and watershed-scale models. Rather, visual comparison of the average D^2 values and their associated confidence interval was used to determine which scale best explained the distribution of fishes in eastern Montana.

RESULTS

Fish Collection

Thirty-nine fish species, including 27 native and 12 nonnative fishes, were sampled from the 120 sites (Table 6). The number of species collected per site varied between one and 17 (mean = 6; median = 5; mode = 5; Figure 6). Species occurrence (i.e., the number of presences for each species) varied from one to 90 (Figure 7). The two most common native species were fathead minnow ($N = 90$), and white sucker ($N = 79$). The most common nonnative species were common carp ($N = 48$) and black bullhead ($N = 42$). Mountain whitefish and sturgeon chub were the rarest native species ($N = 1$). Nonnative species collected only at a single site were rainbow trout, smallmouth bass, and white crappie (Figure 7).

Models Results by Scale

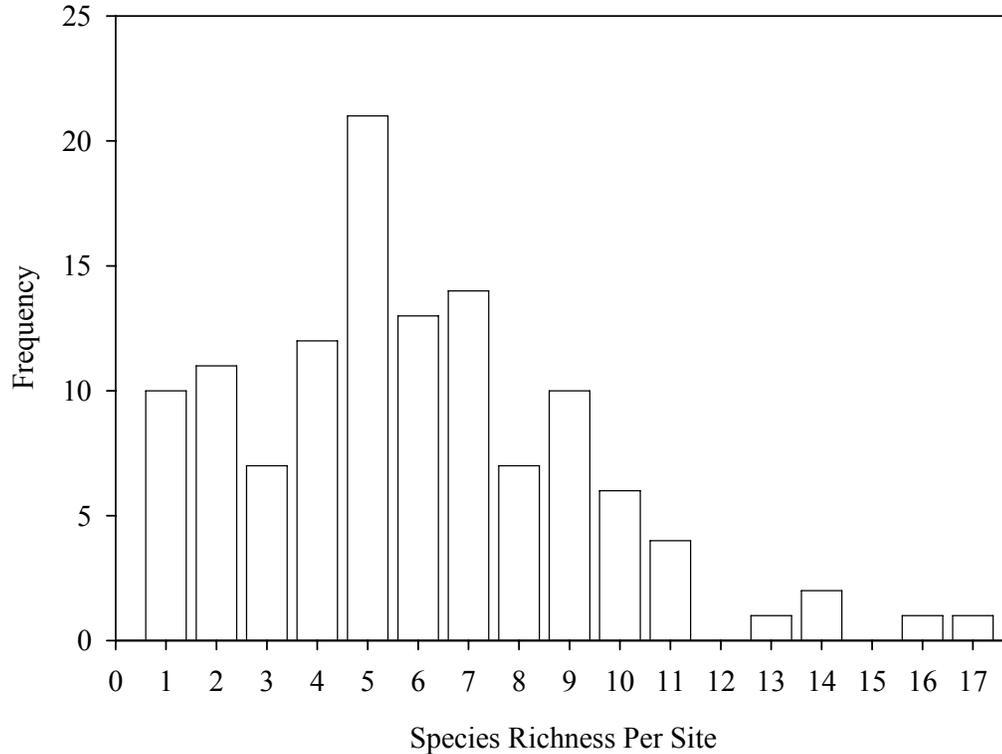
Reach-scale Models

Fifteen sites had missing physicochemical or biotic data at the reach scale. Most of the missing data were water quality measurements. In an attempt to increase sample size, water quality measurements were removed from the data set, and fish distributions were modeled using only physical and biotic measurements of the stream. However, this exercise made little to no difference in the results of the reach-scale models. Thus, results from reach-scale models reported below include both physical measurements and water quality variables.

Table 6. Scientific name, common name, and acronym of native and nonnative species collected in prairie streams of Montana from 1999 through 2004.

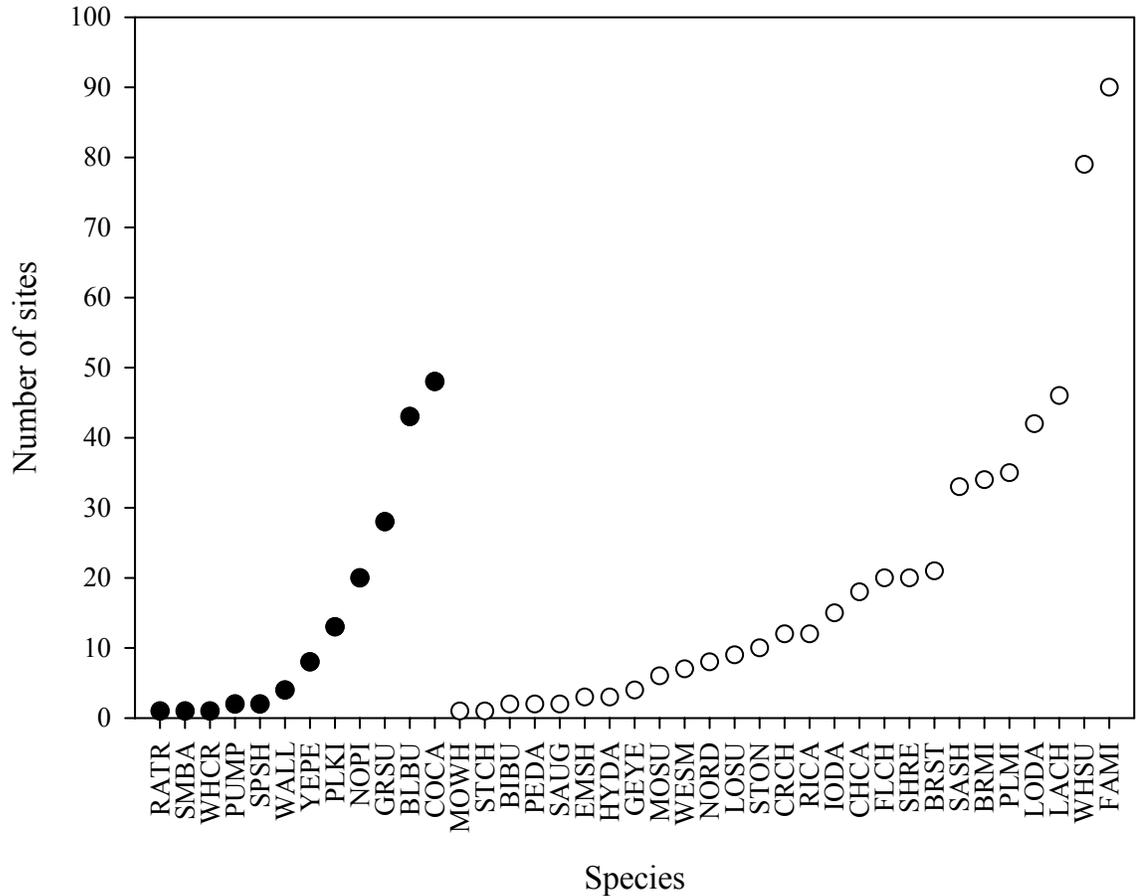
Status	Scientific Name	Common Name	Acronym	
Native	<i>Carpiodes carpio</i>	River carpsucker	RICA	
	<i>Catostomus catostomus</i>	Longnose sucker	LOSU	
	<i>Catostomus commersonii</i>	White sucker	WHSU	
	<i>Catostomus platyrhynchus</i>	Mountain sucker	MOSU	
	<i>Couesius plumbeus</i>	Lake chub	LACH	
	<i>Culea inconstans</i>	Brook stickleback	BRST	
	<i>Etheostoma exile</i>	Iowa darter	IODA	
	<i>Hiodon alosoides</i>	Goldeye	GOLD	
	<i>Hybognathus argyritis</i>	Western silvery minnow	WESM	
	<i>Hybognathus hankinsoni</i>	Brassy minnow	BRMI	
	<i>Hybognathus placitus</i>	Plains minnow	PLMI	
	<i>Ictalurus punctatus</i>	Channel catfish	CHCA	
	<i>Ictiobus cyprinellus</i>	Bigmouth buffalo	BIBU	
	<i>Macrohybopsis gelida</i>	Sturgeon chub	STCH	
	<i>Margariscus margarita</i>	Pearl dace	PEDA	
	<i>Moxostoma macrolepidotum</i>	Shorthead redhorse	SHRE	
	<i>Notropis atherinoides</i>	Emerald shiner	EMSH	
	<i>Notropis stramineus</i>	Sand shiner	SASH	
	<i>Noturus flavus</i>	Stonecat	STON	
	<i>Phoxinus eos</i>	Northern redbelly dace	NORD	
	<i>Phoxinus neogaeus</i> x <i>P. eos</i>	Northern redbelly x finescale dace hybrid	HYDA	
	<i>Pimephales promelas</i>	Fathead minnow	FAMI	
	<i>Platygobio gracilis</i>	Flathead chub	FLCH	
	<i>Prosopium williamsoni</i>	Mountain whitefish	MOWH	
	<i>Rhinichthys cataractae</i>	Longnose dace	LODA	
	<i>Sander canadense</i>	Sauger	SAUG	
	<i>Semotilus atromaculatus</i>	Creek chub	CRCH	
	Nonnative	<i>Amiurus melas</i>	Black bullhead	BLBU
		<i>Cyprinus carpio</i>	Common carp	COCA
		<i>Esox lucius</i>	Northern pike	NOPI
<i>Fundulus kansae</i>		Northern plains killifish	PLKI	
<i>Lepomis cyanellus</i>		Green sunfish	GRSU	
<i>Lepomis gibbosus</i>		Pumpkinseed	PUMP	
<i>Micropterus dolomieu</i>		Smallmouth bass	SMBA	
<i>Notropis hudsonius</i>		Spottail shiner	SPSH	
<i>Oncorhynchus mykiss</i>		Rainbow trout	RATR	
<i>Perca flavescens</i>		Yellow perch	YEPE	
<i>Pomoxis annularis</i>		White crappie	WHCR	
<i>Sander vitreum</i>		Walleye	WALL	

Figure 6. Frequency distribution of species richness for prairie streams sampled in eastern Montana from 1999 through 2004.



Characteristics measured at the reach scale from 105 sites were used in the construction of reach-scale models. Models were developed for 19 species. Six of the 19 (32%) reach models were successful at meeting the deviance reduction criterion (Table 7). Two to three variables were used in each of the six models to reduce null deviance (Table 7). These variables combined to reduce null deviance of individual species distributions between 25% and 59% (mean and 95% confidence interval = $41 \pm 12\%$; Table 7). Misclassification error rate (MER) for successful models varied between 6% and 24% (mean = $13 \pm 7\%$; Table 7). Reductions in MER varied from 2% to 22% (mean = $10 \pm 6\%$), which translated to improvements between 20% (northern plains killifish) and 68% (shorthead redhorse) from the null error rate (mean = $40 \pm 15\%$; Table 7).

Figure 7. Species occurrence for 38 species collected in eastern Montana. Solid circles represent nonnative species, and open circles represent native species. Species acronyms are found in Table 6.

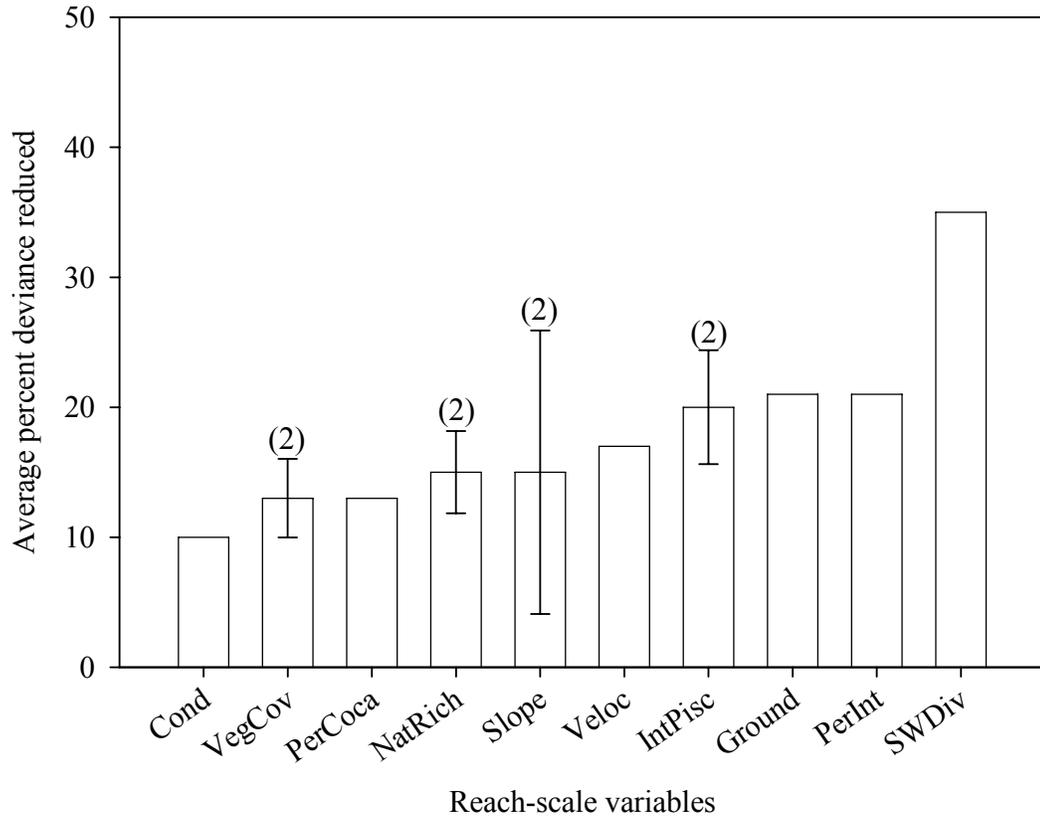


Null deviance of the six models varied from 70.44 (northern plains killifish) to 141.30 (lake chub; Table 7). Individual variables reduced, on average, 18% (± 4) of the null deviance in each significant model (Table 7). The Shannon-Weiner diversity index (SWDiv) explained the largest amount of deviance of all variables used in the successful models (Figure 8). Conductivity (Cond) reduced the least amount of deviance among all reach-scale variables used in successful models. Four variables (VegCov, NatRich, Slope, and IntPisc) occurred in more than one model (Figure 8).

Table 7. Reach models for 19 species from 105 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. Numbers in parentheses following means of percent reduction in deviance, MER metrics, and percent improvement are the 95% confidence interval.

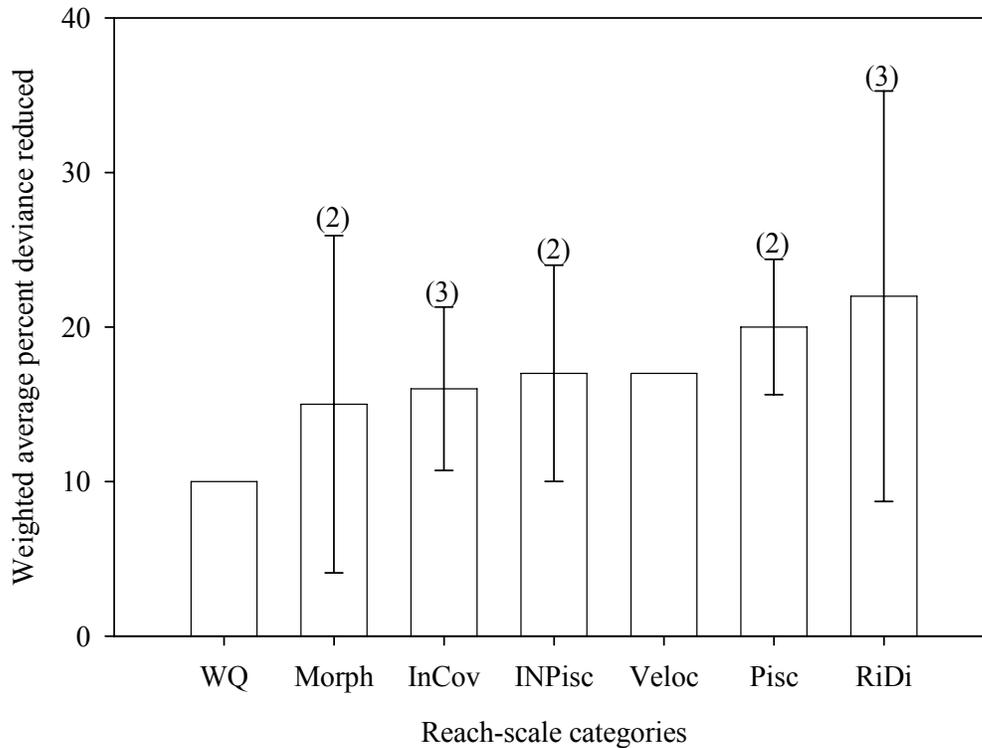
Species	N	Variable	Null deviance	Met deviance reduction criterion?	% Reduction in deviance	Total % reduction in deviance	MER (%)	MER reduced	% Improvement
CRCH	11	NatRich > 10	70.44	No					
		BankW > 6.92 m							
		PerInt > 0.52%							
PLKI	11	Ground < 64.09%	70.44	Yes	21	59	8	2	20
		Slope < 0.28%			21				
		NatRich > 10			17				
RICA	12	VegCov < 1.59%	74.63	Yes	12	25	8	3	27
		PerCoca > 3.08%			13				
IODA	15	BankH < 0.63 m	86.12	No					
CHCA	17	SWDiv > 1.74	92.99	Yes	35	50	10	6	38
		VegCov < 1.59%			15				
FLCH	18	NatRich > 8	96.21	No					
NOPI	18	BankH < 0.55 m	96.21	No					
BRST	20	SDWD > 25.73	102.30	No					
SHRE	20	NatRich ≥ 8	102.30	Yes	14	48	6	13	68
		Veloc > 0.88			17				
		IntPisc > 6.55%			17				
GRSU	27	PerInt > 3.84%	119.70	No					
SASH	31	SWDiv > 1.74	127.40	No					
PLMI	32	VegCov < 0.16%	129.10	No					
BRMI	34	Ground > 62.56%	132.20	No					
		Brush < 1.24%							
LODA	39	PerInt < 8.69%	138.50	Yes	21	31	24	13	35
		Slope > 0.32%			10				
BLBU	40	ProInt > 4.01%	139.60	No					
		SDDep > 24.43							
		OhVeg > 3.19%							
LACH	42	IntPisc < 0.63%	141.30	Yes	22	32	20	22	52
		Cond < 2060 μS			10				
COCA	43	IntPisc > 0.02%	142.10	No					
WHSU	76	Turb < 99.65 NTU	123.80	No					
FAMI	84	Pool > 9.09%	105.10	No					
		VegCov > 0.27%							
Means					18 (±4)	41 (±12)	13 (±7)	10 (±6)	40 (±15)

Figure 8. Average percent deviance reduced by reach-scale variables used in the six models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variable that appeared in more than one species model is listed in parentheses above the variable.



The 10 variables used in reach models represented seven categories (Figure 9). Variables within categories reduced between 10 and 22% of the null deviance in species models (Figure 9). Variables that measured water quality reduced the smallest amount of deviance in reach-scale models, and variables that measured species richness and diversity reduced the most (Figure 9). Two categories (water quality and velocity) were represented by a single variable; variables that measured instream cover and species richness and diversity occurred most often ($N = 3$) in reach-scale models.

Figure 9. Weighted average percent deviance reduced in species models by reach-scale categories. Acronyms are as follows: water quality (WQ), channel morphology (Morph), instream cover (InCov), introduced non-piscivores (INPisc), velocity (Veloc), piscivores (Pisc), and species richness and diversity (RiDi). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of variable occurrences within that category.



Watershed-scale Models

Fourteen sites had incomplete watershed-scale information. Therefore, the distribution of 20 species was modeled using watershed-scale data from 106 sites. Four models (20%) met the deviance reduction criterion (Table 8). Two to four variables were used in each of the four significant models to reduce null deviance (Table 8). These variables combined to reduce null deviance of individual species distributions between 38% and 60% ($50 \pm 9\%$; Table 8). Misclassification error rates of successful models

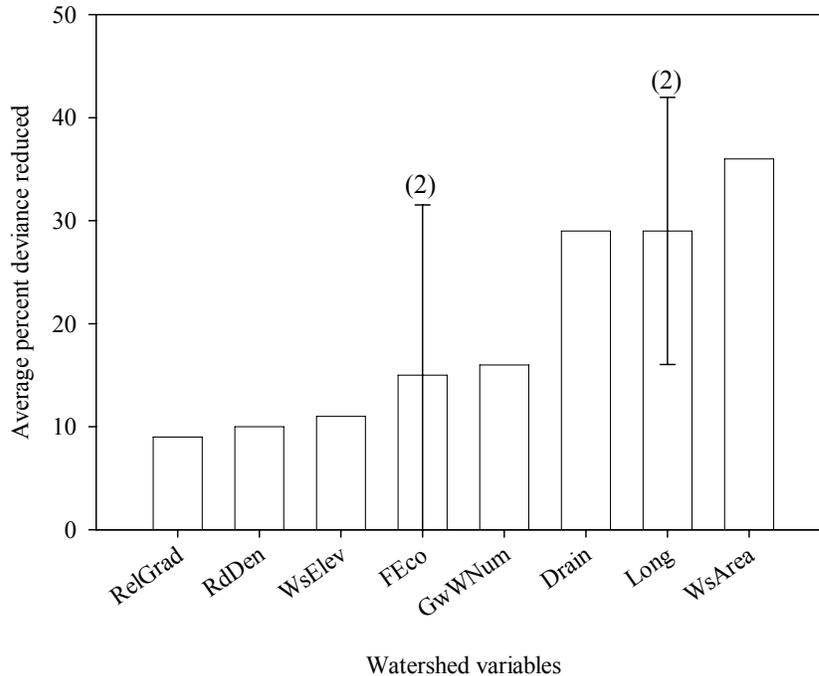
Table 8. Watershed models for 20 species from 106 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. All MER metrics are reported in percents. The number in parentheses following the means of percent reduction in deviance is the 95% confidence interval. Acronyms following FEco and Drain variables are: PORB = Powder River Basin; NWGP = Northwestern Glaciated Plains; LMR = Little Missouri River; YER = Yellowstone River.

Species	N	Variable	Null deviance	Met deviance reduction criterion?	% Reduction in deviance	Total % reduction in deviance	MER (%)	MER reduced	% Improvement
IODA	10	Fallow > 21.39%	66.42	No					
STON	10	GwWNum > 38	66.42	No					
CRCH	12	Long > -105.87	74.87	Yes	36	60	9	2	18
		FEco = PORB			24				
RICA	12	UsDam > 19	74.87	No					
PLKI	13	FEco = PORB	78.90	No					
NOPI	14	Fallow > 21.69%	82.75	No					
BRST	16	Fallow > 14.23%	89.96	No					
SHRE	17	WsArea > 3,131.49 km ²	93.34	Yes	36	52	8	8	50
		GwWNum > 258			16				
CHCA	18	UsDam ≥ 8	96.58	No					
FLCH	20	PCost < 193.12	102.70	No					
GRSU	28	Drain = LMR, YER	122.40	Yes	29	38	18	8	31
		RelGrad < 0.45%			9				
BRMI	29	WsArea < 288.64 km ²	124.40	No					
PLMI	32	Grass > 60.18%	129.80	No					
		AnnFld < 0.01							
SASH	39	Lat < 47.99	131.50	No					
		PCost < 155.65							
LODA	36	WsElev < 1,241.72 m	135.80	Yes	11	50	8	26	76
		RdDen < 0.09%			10				
		FEco = NWGP, PORB			7				
		-108.60 ≤ Long ≤ -106.11			22				
BLBU	38	Long > -107.60	138.30	No					
LACH	40	Long < -105.67	140.50	No					
COCA	44	UsDam > 35	143.90	No					
WHSU	69	WsElev > 1,262.86 m	159.70	No					
		Fallow > 19.06%							
FAMI	78	Water > 0.01%	122.40	No					
Means					20 (±7)	50 (±9)	11 (±5)	11 (±10)	44 (±25)

varied between 8% and 18% (mean = $11 \pm 5\%$; Table 8), and MER reduced varied from 2% to 26% (mean = $11 \pm 10\%$). Improvements in MER varied between 18 and 76% (mean = $44 \pm 25\%$) over the null error rate.

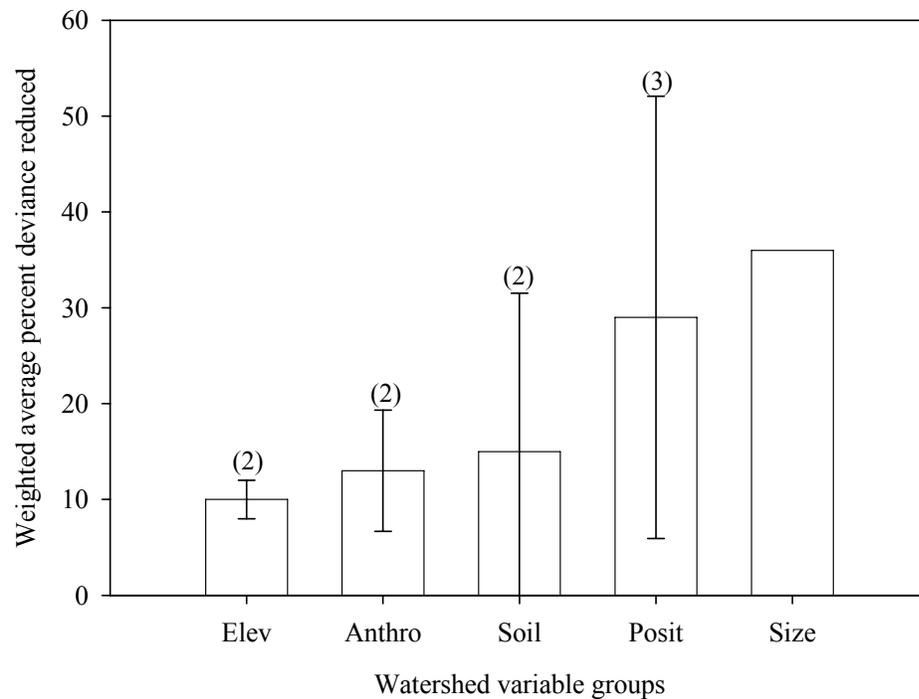
Null deviance of the four models varied from 74.87 (creek chub) to 135.80 (longnose dace; Table 8). Individual variables explained, on average, 20% (± 7) of the null deviance in each significant model (Table 8). Watershed area (WsArea) explained the largest amount of deviance of all variables used in the successful models, and relative gradient (RelGrad) reduced the least amount of deviance among all watershed-scale variables used in successful models (Figure 10). Two variables (FEco and Long) occurred in more than one model (Figure 10).

Figure 10. Average percent deviance reduced by watershed-scale variables used in the four models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variable that appeared in more than one species model is listed in parentheses above the variable.



The eight variables used in watershed models represented five categories (Figure 11). The variables within categories reduced between 10 and 36% of the null deviance in species models (Figure 11). Variables that measured elevation reduced the smallest amount of deviance in watershed-scale models, and variables that measured the size of the stream reduced the most (Figure 11). The category of soil characteristics was the only category represented in species models by a single variable; variables that measured position occurred most often ($N = 3$) in significant watershed-scale models.

Figure 11. Weighted average percent deviance reduced in species models by watershed-scale categories. Acronyms are as follows: elevation (Elev), anthropogenic influences (Anthro), soil characteristics (Soil), position (Posit), and stream size (Size). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of variable occurrences within that category.



Combined Models

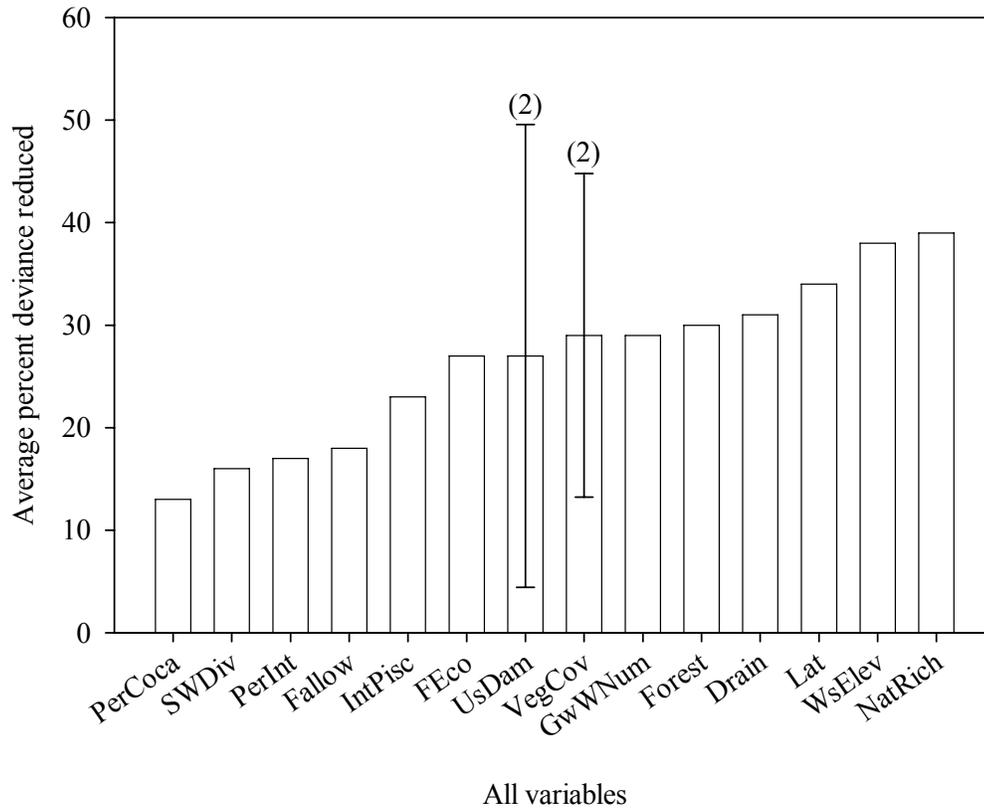
Ninety-three sites were used for construction of the combined models, and models were developed for 19 species. Seven models were successful at meeting the deviance reduction criterion. Two to three variables were used in these models (Table 9). When combined, these variables reduced null deviance of individual species distributions between 44% and 71% (mean = $61 \pm 8\%$; Table 9). Misclassification error rate (MER) for combined models varied between 3% and 13% (mean = $9 \pm 2\%$; Table 9), and MER reduced varied between 3% and 23% (mean = $11 \pm 5\%$). Reductions in MER translated to improvements between 25% and 81% (mean = $53 \pm 14\%$; Table 9) from null models.

Null deviance of the seven models varied from 67.35 (creek chub) to 120.10 (longnose dace; Table 9). Of 80 possible variables (44 reach- and 36 watershed-scale variables), 14 variables (18%) were used by tree classifiers to explain the deviance in the distributions of individual species (Figure 12). Fifty-seven percent of the variables used were measured at the watershed scale, and 43% were reach-scale variables. Individual variables explained, on average, 27% (± 4) of the null deviance in each significant model (Table 9). Native species richness within the reach (NatRich) reduced the largest amount of deviance and the percent of common carp within the reach reduced the least amount of deviance of all variables used in successful models (Figure 12.) Two variables (UsDam and VegCov) appeared in more than one model.

Table 9. Combined models for 19 species from 93 sites. The variables are those used in the pruned tree classifiers to explain the deviance in individual species distributions. Numbers in parentheses following the means of percent reduction in deviance, MER metrics, and percent improvement are the 95% confidence interval. Acronyms following FEco and Drain variables are: PORB = Powder River Basin; LMR = Little Missouri River; YER = Yellowstone River.

Species	N	Variable	Null deviance	Met deviance reduction criterion?	% Reduction in deviance	Total % reduction in deviance	MER (%)	MER reduced	% Improvement
IODA	10	Fallow > 21.39%	63.26	No					
CRCH	11	Long > -105.89	67.35	Yes	34	61	9	3	25
		FEco = PORB			27				
PLKI	11	FEco = PORB	67.35	No					
RICA	12	VegCov < 1.59%	71.25	Yes	37	52	8	5	38
		UsDam ≥ 47			15				
NOPI	13	PtElev < 760 m	74.95	No					
BRST	15	Forest < 0.19%	81.82	Yes	30	71	3	13	81
		Fallow > 18.78%			18				
		IntPisc < 0.19%			23				
CHCA	17	UsDam ≥ 8	88.06	Yes	38	59	8	10	56
		VegCov < 4.78%			21				
SHRE	17	NatRich ≥ 8	88.06	Yes	39	68	9	9	50
		GwWNum > 258			29				
FLCH	18	VegCov < 10.23%	90.95	No					
GRSU	27	Drain = LMR, YER	111.40	Yes	31	44	13	16	55
		PerCoca > 0.04%			13				
BRMI	29	WsArea < 68,8.37 km ²	114.70	No					
		Slope > 0.18%							
PLMI	29	VegCov < 16.03%	114.70	No					
SASH	31	Lat < 47.99	117.60	No					
LODA	33	WsElev > 1,241.72 m	120.10	Yes	38	71	12	23	68
		SWDiv ≥ 1.00			16				
		PerInt < 8.40%			17				
BLBU	35	Long > -107.60	122.20	No					
LACH	36	Long < -105.67	123.20	No					
COCA	37	IntPisc > 0.02%	125.40	No					
WHSU	66	Turb < 99.65 NTU	109.60	No					
FAMI	72	Pool > 9.09%	96.34	No					
Means					27 (±4)	61 (±8)	9 (±2)	11 (±5)	53 (±14)

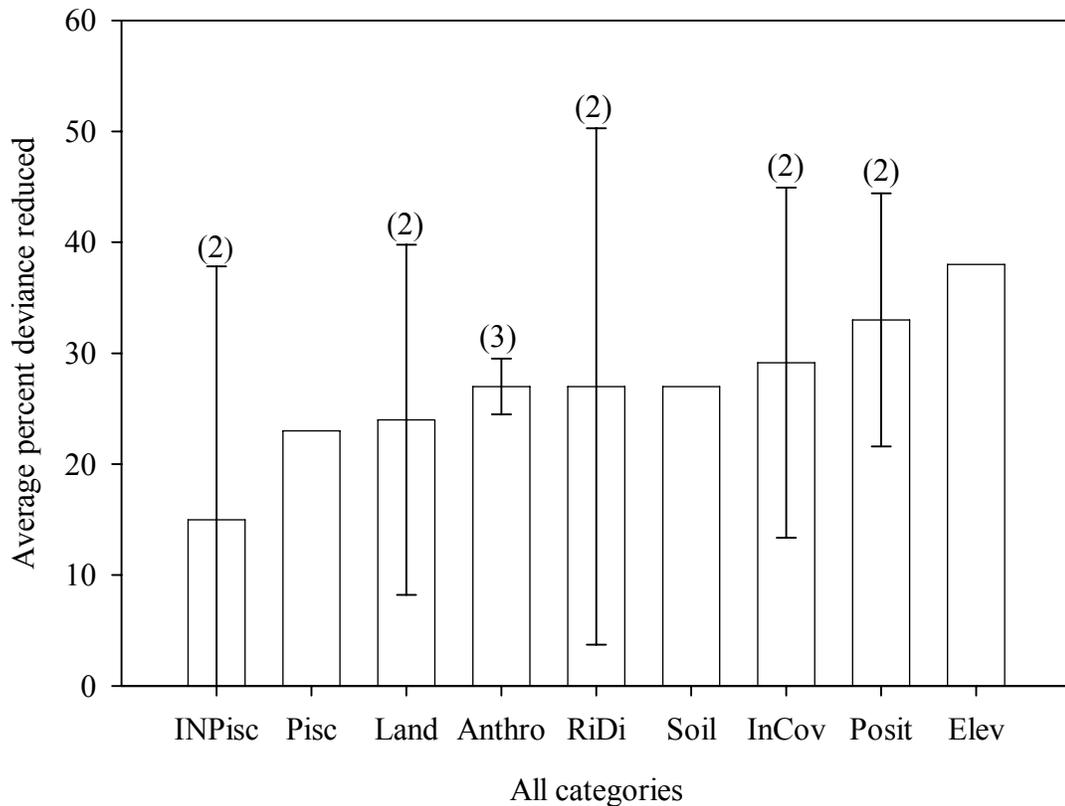
Figure 12. Average percent deviance reduced by reach- and watershed-scale variables in combined models that met the deviance reduction criterion. Error bars represent the 95% confidence interval. The number of occurrences for the variables that appeared in more than one species model is listed in parentheses above the variable.



The 14 variables used in combined models represented nine categories (Figure 13). Fifty-six percent of these categories were of the watershed scale, and 44% were reach-scale categories. Variables within the categories reduced between 15 and 38% of the null deviance in species models (Figure 13). Variables that measured elevation reduced the most deviance, and variables that measured the percent of introduced non-piscivorous fish in the reach reduced the least (Figure 13). Three of the nine categories

(Pisc, Soil, and Elev) were represented by a single variable; the anthropogenic influence category was represented the most in species models ($N = 3$; Figure 13).

Figure 13. Weighted average percent deviance reduced in species models by combined variable categories. Acronyms are as follows: introduced non-piscivores (INPisc), piscivores (Pisc), land cover within the watershed (Land), anthropogenic influences (Anthro), species richness and diversity (RiDi); soil characteristics within the watershed (Soil), instream cover (InCov), position (Posit), and elevation (Elev). Error bars represent the 95% confidence interval. Numbers in parentheses indicate the number of occurrences of variables within that category.



Scale Comparisons

Nine species (45%) were modeled at all scales. Residual mean deviances for individual models that met the deviance reduced criterion varied from 0.27 (brook stickleback combined model) to 0.95 (lake chub reach-scale model; Table 10). Three species only had successful models for a single scale; all other species had successful models for at least two scales (Table 10).

Average D^2 values by scale were smallest for combined models, followed by watershed- and reach-scale models (Table 10). Confidence intervals of these means overlapped. Four of six species (66%) who had successful models at multiple scales were best modeled at using a combination of reach- and watershed-scale variables. However, the distribution of longnose dace was best modeled using only watershed-scale variables, and river carpsucker distribution was slightly better modeled using only reach-scale variables.

Table 10. Residual mean deviance (D^2) by species and scale. Numbers in bold indicate the models with the lowest D^2 for that species. Numbers in parentheses indicate the 95% confidence interval of the average D^2 .

Species	Scale		
	Reach	Watershed	Combined
BRST			0.27
CHCA	0.55		0.40
CRCH		0.29	0.28
GRSU		0.73	0.70
LACH	0.95		
LODA	0.84	0.37	0.40
PLKI	0.44		
RICA	0.37		0.38
SHRE	0.32	0.44	0.31
Means	0.58 (± 0.21)	0.51 (± 0.19)	0.41 (± 0.11)

DISCUSSION

Research in the development of explanatory and predictive models has intensified as state and federal agencies have increased focus on the conservation of fishes. Efforts to model fish distributions have used sophisticated modeling techniques to minimize misclassification error rate (Olden and Jackson 2002; Wall et al. 2004; Oakes et al. 2005). However, I am aware of no published study in fisheries literature that has required that models must reduce a significant amount of deviance in species distributions prior to evaluation for their predictive ability. Nonetheless, I consider it important that variables reduce the null deviance before they are considered as potentially ecologically meaningful factors. For Montana prairie fishes, variables measured at the reach and watershed scale did not reduce the null deviance by significant amounts in most models (i.e., did not meet the deviance reduction criterion). This lack of a significant reduction in null deviance may be due to: i) unmeasured explanatory (reach- and watershed-scale characteristics) variables in the data set; ii) insufficient sample size; iii) the lack of habitat saturation by prairie fish species; iv) the generalist nature of most prairie fishes; and v) variability in both the explanatory and the response (fish presence or absence) variables.

One possible but unlikely reason for the lack of models that met the deviance reduction criterion in this study may be due to unmeasured explanatory variables. Most of the variables chosen for this study were selected based on their success in previous published studies that modeled the distribution of fishes at the reach or watershed scale. For example, channel morphology, instream habitat structure, and substrate explained

40% of the variation in fish assemblage structure in Midwestern prairie streams (Marsh-Matthews and Matthews 2000). Fourteen variables related to channel morphology, six variables related to instream habitat structure, and three variables related to substrate were chosen for my modeling efforts in eastern Montana. Similarly, several of the GIS-derived watershed-scale variables used in this study followed results of other published studies. Oakes et al. (2005) used GIS measurements in several types of predictive models (including tree classifiers) to predict fish distributions within the Big Blue River basin of Kansas and Nebraska. Variables related to stream size or the position of the sample site within the watershed often ranked high among models. This effort to model the distribution of prairie fishes of Montana used similar variables that measured stream size (i.e., watershed area and Strahler stream order) or position (i.e., latitude, longitude, and major drainage basin). In addition to using variables from published studies, this study also included several other variables I believed may affect prairie fish distributions. These variables were chosen based on examination of the literature, consultation with authors of other studies, and field experience. For example, the path-cost analysis variable has not been previously published in fisheries literature. However, this analysis reflects some of the basic principles of island biogeography (MacArthur and Wilson 1963), and these principles have been used to explain native fish zoogeography in the Missouri River basin within the Great Plains (Hoagstrom and Berry 2006). While large numbers of previously-tested and untested variables were included in efforts to model Montana prairie fishes, it is possible that one or more variables that strongly influence the distribution of prairie fishes are missing from the data set. However, the justification for

variable selection in this study makes the case for missing explanatory variables in the data set weak.

Insufficient sample size may be a more plausible explanation as to why the deviance in most species models was not significantly reduced in this study. Although the sample sizes in this study ($N_{\text{reach}}=105$; $N_{\text{watershed}} = 106$; $N_{\text{combined}} = 93$) were similar to or larger than other predictive fisheries modeling studies [Marsh-Matthews and Matthews 2000 ($N = 65$); Rashleigh 2004 ($N = 16$); Oakes et al. 2005 ($N = 120$)], they may be considered small by some standards. In plant ecology, where predictive models (such as tree classifiers) are more commonly used, sample sizes are often much larger [Franklin 1998 ($N = 906$); Vayssières 2000 ($N = 4,101$)]. Sample size may influence the construction of tree classifiers as larger sample sizes support more splits in the tree. As trees increase in size (i.e., as the number of splits increases), the amount of deviance explained by the tree also increases (Breiman et al. 1984). Therefore, the amount of null deviance reduced in species models is related to the sample size of the data set.

It may be possible that sample size was low in the predictive fisheries modeling studies listed above. To date, there are no published methods that suggest the minimum sample size needed for modeling with tree classifiers. However, substituting the null deviance in the distribution of a species for variance in the calculation of a required sample size from a single sample may provide an estimate of the minimum sample size needed in this study. The resulting formula would be as follows:

$$N = \frac{Dt^2}{(ap_i)^2}$$

where N is the required sample size, D is the deviance of the species, t is the tabular t -value at $n-1$ degrees of freedom for a desired significance level, a is the desired accuracy of the model (i.e., the level of accuracy desired in describing mean occurrence of the species), and p_i is the proportion of presences for a species. The sample size required would be largest where null deviance is greatest (i.e., species that occur near 50% of the total number of sites). For example, information in the common carp (present at 41% of sites) reach-scale model suggests that over 2,800 samples are required to model the species with 20% accuracy at the 95% confidence level $[(142.10)(1.28^2)/(0.20*0.41) = 2,840]$. This required sample size is approximately 27 times larger than the 105 sites used in modeling fish distributions at the reach scale in this study.

Models may not significantly explain prairie fish distributions due to the lack of “habitat saturation” by these species. All modeling techniques assume that a species occupies all suitable habitats (Breiman et al. 1984; Jongman et al. 1995), but this may not be true in prairie streams. For example, post-modeling comparisons between the presence and absence of brook stickleback and stream width indicated that this species was always sampled in streams whose mean wetted width ≤ 12 m. However, examination of all streams whose mean wetted widths ≤ 12 m demonstrated that brook stickleback were actually absent at more of these sites than they were present. Similarly, river carpsucker were likely to be found in reaches with a low path-cost score (i.e., are within or close to major streams), but there were many reaches with small costs which did not contain river carpsucker. This lack of habitat saturation can create unexplained

variability in fish distributions and make it difficult for any variable to reduce significant amounts of deviance (Jongman et al 1995).

The autecology of species may also contribute to the difficulty in reducing the deviance of fish distributions by significant amounts. Most prairie fishes in Montana are considered ecological generalists (Bramblett et al. 2005). Plains minnows can tolerate water temperatures as warm as $40.5 \pm 0.4^{\circ}\text{C}$ and dissolved oxygen levels as low as 2.08 ± 0.14 mg/L (Ostrand and Wilde 2001). Optimum survival and growth of creek chub likely occurs in waters whose pH values vary between 6.0 and 9.0, but populations of this species have been observed in more acidic and basic waters (McMahon 1982). Matthews (1987) noted that several species of the genus *Notropis* (including sand shiners) have persisted in many harsh prairie streams of the Canadian River drainage of Oklahoma at least partly due to their wide tolerance of physicochemical conditions. Because of their tolerant, generalist nature, prairie fishes do not occur exclusively in streams with a narrow set of environmental conditions. As a result, tree classifiers may not be able to produce significant models based on a one-time sample of the environmental that does not adequately describe the range of suitable conditions for the presence of a particular species because a single habitat sample would leave unexplained deviance in most species distribution models (Hubbell 2001).

Finally, variability in both the explanatory and response variables may be one reason why most models failed to explain a significant amount of deviance. Sources of variability include: i) variability in physicochemical characteristics at a site; and ii) the variability in the presence or absence of fish species at a site. The physicochemical

characteristics of a stream at a given site may vary between or within seasons due to unpredictable cycles of flooding and drying (Ostrand and Wilde 2002; Dodds et al. 2004). For example, in many prairie streams, an intermediate (e.g., a one-year return time) or more severe level flood during a summer thunderstorm may alter habitat by eliminating large woody debris, scouring new pools, and flooding riparian vegetation (Dodds et al. 2004). Conversely, drying during the summer months often eliminates riffle habitat and creates intermittent pools along the stream reach (Matthews 1988; Ostrand and Wilde 2004). Water quality parameters are often correlated with flow and may change as streams are desiccated during the summer months (Zale et al. 1989; Ostrand and Wilde 2004). Water quality of three creeks in Oklahoma varied widely with flow; dissolved oxygen and conductivity were high during low flow periods (Benham-Blair and Associates, Inc. 1976). Ostrand and Wilde (2004) found that turbidity decreased and conductivity increased as intermittent pools of the Brazos River drainage of Texas dried during the summer months (June-August). The variability of reach-characteristics affects the distribution of prairie fishes (Ostrand and Wilde 2001; Ostrand and Wilde 2004). However, a single sample of the physicochemical features of a stream does not adequately describe this variation and thus may not describe the distribution of prairie fishes in statistical models.

Similar to the explanatory variables of this study, the response variable (the presence or absence of fishes) may also vary temporally. This variability may be the result of seasonal migrations and stochastic events (i.e., flooding and drying) that may force fish movement. The effects of seasonal movements on prairie fish distributions in

eastern Montana have not been studied to date. However, flooding and drying have been recognized as major events influencing the distribution of prairie fishes by inducing local extirpations (Fausch and Bramblett 1991; Bramblett and Zale 2000; Ostrand and Wilde 2002). The rates of extirpation and eventual recolonization of fishes within a stream reach vary by species (Grossman et al. 1982; Fausch and Bramblett 1991; Fausch and Bestgen 1997; Matthews and Marsh-Matthews 2000). A stream reach may have suitable reach- and watershed-scale characteristics for a particular species, but the species may not be present within that reach due to flooding and drying processes that occurred prior to sampling. Again, predictive models that relate the occurrence of a species to habitat assume that the species of interest occupies all suitable habitats within its range (Breiman et al. 1984; Jongman et al. 1995). When all suitable habitats are not occupied, unexplained variability is introduced into the model. In prairie streams, stochastic flow events, reduced connectivity during intermittency, and variable recruitment may reduce the probability that this saturation will occur. However, more information on the frequency and duration of flooding and drying cycles on the temporal variability of fish distributions in eastern Montana and the effects of this variability on the explanative and predictive abilities of species models is needed.

Interpretations of Variables Used in Models

Although most models were not significant at any scale, the variables used in significant models did provide some insight as to what factors affect prairie fish distributions in Montana. Many variables used in significant and non-significant models

appear to indicate large-scale ecological gradients (e.g., stream size, elevation, watershed position), but were not always direct measurements of those gradients. At the reach scale, variables measuring species richness and diversity explained the largest amount of deviance (22%) and were present in the largest number of significant reach-scale models ($N=3$). When these variables were used, they were always positively related to the presence of that species; this trend could indicate the size of a stream where a species may be found. For example, shorthead redhorse were more likely to be present in stream reaches with at least eight native species, according to reach-scale models. This species is likely to be found in larger streams or moderate to large rivers throughout its range (Brown 1971; Pflieger 1997). Positive relationships between stream size and richness and diversity of the fish community have been demonstrated in other studies (Matthews 1987; Schlosser 1987). Further, the shorthead redhorse watershed-scale model in this study showed that this species was more likely to be found in streams that have a large watershed area ($WsArea > 3,131 \text{ km}^2$). Therefore, I believe that the use of species richness and diversity variables in this and other reach-scale models provides some insight on stream size gradients for a few species in eastern Montana.

It is interesting to note that other more direct reach-scale measurements of stream size (e.g., mean wetted width or mean stream depth) did not appear in significant models. Larger stream widths and depths have been previously linked to increased stream size (Strahler 1957; Vannote et al. 1980). However, this relationship may not always be true for prairie streams. Streams of the Great Plains frequently become desiccated during certain months or years due to geologic and climate factors (Matthews 1988; Fausch and

Bestgen 1997). Spatial patterns of dry reaches and pools may resemble “beads on a string” (Robert G. Bramblett, Montana State University, personal communication). Consequently, depths and widths of the stream within a given reach may vary widely. In this study, mean stream widths varied between 0.37 and 19.57 m and mean depths varied from 2.15 to 142.55 cm. Because of the large variability in these measurements, the tree classifiers could not use these variables to reduce deviance at the reach scale.

Many watershed-scale variables used in significant models directly or indirectly measure large-scale ecological gradients. Watershed variables that directly measured stream size (namely, watershed area), elevation, and position (as measured by latitude, longitude, and drainage) occurred in all four significant watershed-scale models and 43% of significant combined models. Some of these measurements were used in nonsignificant watershed-scale and combined models as well. However, other variables may serve as a proxy to these measurements. Ecoregion was used to explain the distribution of creek chubs in eastern Montana. Before analysis, I believed ecoregion would indicate soil characteristics and underlying geology of a watershed. Soil and geologic factors have been used in previous studies to predict fish-assemblage structure in other Great Plains streams (Oakes et al. 2005; Gido et al. 2006). However, additional examination of the relationship between creek chub occurrence and ecoregion demonstrates that this variable likely explains the zoogeographic range of creek chub. I verified the use of this variable in the creek chub watershed-scale model by overlaying ecoregion with records of creek chub presences; this exercise demonstrated that this species is largely confined to the Powder River Basin ecoregion. To date, no study has

examined the relationship between creek chub distributions and the inherent characteristics (e.g., lithology, geomorphology, and climate) of any ecoregion. However, other studies have evaluated the correspondence between ecoregion and spatial patterns of fish distributions (Whittier et al. 1988; Lyons 1989; Modde et al. 1991). The relationship between ecoregion and creek chub occurrence in this study is one example of how this variable may describe spatial gradients in fish distributions.

Variables selected to describe anthropogenic disturbance within the watershed may also be indicative of geographic gradients, particularly stream size. When such variables (namely, GwWNum and UsDam) were used in watershed-scale and combined models, they were always positively related to the presence of that species; thus, no detrimental effects of human disturbance were detected in this study. *A posteriori* regression analysis demonstrated a positive relationship between watershed area (a measure of stream size) and the two primary anthropogenic variables used in significant models (GwWNum: $r^2 = 0.79$, $p < 0.0001$; UsDam: $r^2 = 0.48$, $p < 0.0001$). Further, the models that included these variables were for species likely to inhabit larger streams or rivers [i.e., channel catfish, river carpsucker, and shorthead redhorse (Brown 1971; Pflieger 1997)]. I believe these variables were chosen over other direct measures of stream size due to the relatively smaller amounts of variability in anthropogenic disturbance variables (e.g., UsDam minimum and maximum = 0 and 122) compared to other measurements of stream size (e.g., WsArea minimum and maximum = 38 and 13,930 km²).

One direct measure of stream size (Strahler) measured at the watershed scale within a watershed was conspicuously absent in this study as compared to those studies described above. Two plausible explanations for this anomaly exist. First, the GIS data layer used to calculate stream order may not have been as sensitive as stream layers used in other studies. Other authors (e.g., Oakes et al. 2005; Gido et al. 2006) have used digital elevation models (DEMs) to derive stream networks; I used a previously-derived stream network from the National Hydrologic Database (NHD). The NHD layer may not have included some smaller headwater streams and, thus, negatively affected stream orders calculated for my sampled sites. Secondly, the method used to calculate stream order may not have been sensitive enough to detect changes in stream size. Strahler (1957) stream order increases when two streams of the same order converge. However, I noticed several occurrences within my stream layer where smaller stream orders would converge with larger ones. While the inflow of a larger stream would not increase stream order, it likely has some positive effect on stream size (e.g., wetted width or depth) below the confluence. The choice of another measure of stream order may be appropriate in future studies. Regardless of which reason is correct, coarseness of data appears to create difficulty in detected changes in stream size, but the effects of stream size on fish distributions is still detected through the use of other variables.

The use of stream size, position, and elevation in explaining or predicting fish distributions at large scales is common (Roth et al. 1996; Poff 1997; Wang et al. 2001; Joy and Death 2004), and their utility has been demonstrated for Great Plains streams (Rahel and Hubert 1991; Oakes et al. 2005; Quist et al. 2005; Gido et al. 2006).

Variables measuring stream size and position were typically ranked highest among 38 species-habitat models developed for the Big Blue River basin (Oakes et al. 2005). Similarly, Quist et al. (2005) determined that inclusions of stream size and elevation thresholds in a hierarchical faunal filter greatly reduced the number of potential stream reaches that some prairie fish could occupy as compared to measurements of reach-scale habitat in the North Platte River drainage of Wyoming. Results from this study at both the reach and watershed scale support the use of size, position, and elevation in explaining broad patterns of fish distributions.

In addition to large geographic ecological gradients, other variables do provide information on the ecology of fishes, particularly when examining models at the reach scale. Some physicochemical variables do appear to affect fish distributions. Results from the c river carpsucker and channel catfish reach-scale models highlight the importance of instream vegetation to this species; both species were more likely to be found in reaches where there was little to no submerged vegetation and algal cover ($\text{VegCov} < 1.59\%$). This trend is likely due to the fact that both species tend to be found in turbid streams and rivers (Pflieger 1997), and high turbidity may limit or exclude algae and submergent vegetation in prairie streams (Matthews 1988; Zale et al. 1989; Fausch and Bestgen 1997). The presence or absence of aquatic vegetation may play a role in structuring prairie fish communities (Fausch and Bestgen 1997), but further research into the relationship between instream cover and assemblage composition is needed (Matthews 1988; Fausch and Bestgen 1997).

Longnose dace distributions were also modeled with physicochemical factors, namely slope (Slope > 0.32%). In this case, slope is likely to reflect water velocity. Edwards et al. (1983) have demonstrated that longnose dace select fast water areas such as riffles and runs. The ability to rapidly regulate air in the swim bladder, and thus buoyancy, allows this species to occupy moderate- to swift-flowing water (Sigler and Miller 1963). Results from the above three models described above demonstrate how characteristics of the physical habitat at the reach scale may affect prairie fish distributions.

The importance of biotic interactions in determining fish distributions was also detected in significant models. The longnose dace reach-scale model indicated negative effects of introduced species, whether piscivorous or not (IntPisc < 0.63%; PerInt < 3.84%). Predation has been shown to influence the composition of stream fish assemblages (Lohr and Fausch 1996; Jackson et al. 2001), and the introduction of nonnative piscivores across the Great Plains is believed to be detrimental to many native fishes (Cross et al. 1986; Matthews 1988; Dodds et al. 2004). However, introductions of nonnative, nonpiscivorous fishes from other Great Plains watersheds could have a greater effect than piscivorous nonnative fishes on native fish assemblages (Fausch and Bestgen 1997). Other efforts to model fish distributions within the Great Plains have not considered biotic effects (Marsh-Matthews and Matthews 2000; Oakes et al. 2005; Gido et al. 2006). The use of physicochemical and biotic variables in significant models indicate that reach-scale measurements provide insight into species-habitat relationships in addition to variables measured at the watershed scale. Further research should

examine the relationships between reach-scale characteristics and the occurrence of prairie fishes to provide a mechanistic understanding of how these factors may limit distributions.

Scale Comparisons

The small number of significant models at all scales created difficulty when generalizing which scale best explained the distribution of Montana prairie fishes. As expected, combined models had the lowest average D^2 and thus were considered to be the best at explaining prairie fish distributions. The use of both reach- and watershed-scale variables to reduce the amount of null deviance in models does make sense. Watershed-scale variables may either be directly responsible for shaping the structure of fish assemblages (Jackson and Harvey 1989; Mandrak 1995; Gido et al. 2006) or may be closely related to physicochemical habitat at the reach scale (Ritter et al. 1995; Zorn et al. 2002). However, watershed-scale variables are an average of the variability in reach-scale factors and, as a result, may not explain all of the deviance in stream fish assemblages (Poff 1997; Gido et al. 2006). The inclusion of more reach-scale factors can help reduce the remaining deviance in stream fish distributions. Other studies have also found that combined models often have the greatest power in predicting fish distributions (Lanka et al. 1987; Poff 1997; Gido et al. 2006)

Based on previous research (Argent et al. 2003; Wall et al. 2004; Gido et al. 2006), I expected to find that watershed-scale variables would better explain prairie fish distribution than those measured at the reach scale. Residual mean deviance was, on

average, lower for watershed models than for reach models but the difference was not significant. Similar patterns were detected when comparing other model metrics such as the mean total percent reduced by variables in significant models, MER, MER reduced, and percent improvement of MER. However, fewer watershed models were significant compared to reach models (four versus six). Three species models (lake chub, northern plains killifish, and river carpsucker) were significant at the reach scale but not the watershed scale, and two species models (creek chub and green sunfish) were significant at the watershed scale but not the reach scale. These muddled results given above were somewhat surprising given the attention recently placed on modeling fishes using watershed variables provided by GIS (Roth et al. 1996; Poff 1997; Argent et al. 2003; Oakes et al. 2005; Gido et al. 2006). The use of GIS-derived watershed-scale variables has been shown to be superior to those measured at the reach scale in predicting the distribution of fishes in other studies (Jackson and Harvey 1989; Roth et al. 1996; Poff 1997; Joy and Death 2004; Gido et al. 2006). Further, GIS variables are less costly to measure and analyze than biotic or physicochemical variables measured at the reach scale, and thus their use in conservation efforts has been promoted (Wall et al. 2004; Gido et al. 2006). However, results from this study do not indicate that GIS-derived watershed variables alone are sufficient for explaining fish distributions in eastern Montana streams.

CONCLUSION

This study was the first fisheries study to my knowledge that required species-habitat models to reduce a significant amount of deviance in fish distributions before they were evaluated for their predictive ability. Strict evaluation of models showed that most models, regardless of scale, were not significant. The most likely explanation for this lack of significance was the lack of habitat saturation by prairie fishes due to temporal variability in fish distributions. Although not explored in this study, distributions of fishes in eastern Montana are variable and may be regulated by seasonal movements and the frequency and duration of stochastic flow cycles. Consequently, absence of fishes from a reach may be difficult to explain. Could a species be absent from a stream reach because it lacks suitable habitat characteristics? Is it possible that a species was extirpated from the reach due to some flow event? Was the species actually present but was not captured by the sampling gear? The reach may not be chosen as being a high conservation priority area if it is determined that a stream reach lacks suitable habitat characteristics for a given species based on a single sample. However, the species may be found in abundance within the reach at other times of the year or during another year depending on temporal environmental conditions. This suggests that reaches should be sampled several times among seasons and between years to delineate the relationship between habitat at multiple scales and fish distributions for effective management and conservation of prairie fishes (Angermeier et al. 2002; Dodds et al. 2004).

Despite that most models were not significant, some information that could be used to guide conservation efforts was demonstrated. Watershed-scale variables in significant and non-significant models mostly described large-scale ecological gradients, and some reach-scale models provide information on species-habitat relationships and the negative effects that introduced fishes may have on native fish distributions. Comparisons of D^2 values and MER metrics between reach- and watershed-scale models did not demonstrate that GIS-derived watershed-scale variables explain or predict prairie fish distributions with significantly greater accuracy than reach-scale characteristics measured in the field. Thus, results from this study suggest that conservation efforts for some species in eastern Montana streams should consider factors at the reach scale as well as at the larger scales and that GIS alone may not be an appropriate tool for use in the conservation of prairie fishes.

Most modeling efforts to date have primarily focused on reducing misclassification error rate in predicting the occurrence of fishes. However, a low error rate in predicting occurrences of species (in this case, MER) does not provide insight into the ecology of the species and may be inappropriately used to evaluate models unless the objective of the study is to either minimize the proportion of stream reaches that are misclassified or to compare the predictive ability of tree classifiers to other statistical methods (Vayssières 2000). The strict criteria of this study allowed for both ecological interpretation of model using statistical methods while still allowing for analysis of their predictive ability. The overall rate of misclassification in this study ($11 \pm 3\%$; 95% confidence interval) is lower compared to other studies [Filipe et al. 2002 (23.9%);

Welker and Scarnecchia 2004 (20%); Oakes et al. 2005 (21.4%)]. Perhaps more stringent evaluation of models may be related to a reduction in MER because models that significantly explain fish distribution have lower prediction error rates.

The overall MER, MER reduced, and improvement in MER metrics indicate that the models that met the deviance reduction criterion in this study perform better than chance and should be further evaluated for their ability to predict the distributions of those species for conservation efforts. However, the large number of models that were not evaluated for their predictive ability and the muddled interpretations of variables contained within models that were significant indicate that further information on the effects of reach- and watershed-scale characteristics on the distributions of these species is needed.

MANAGEMENT IMPLICATIONS

- 1) Test significant models from this study for their ability to predict fish distributions using an independent data set.
- 2) Determine variability in physicochemical characteristics of individual stream reaches by season and reaches within a single stream. Relate this variability to changes in fish assemblage structure.
- 3) Continue to sample habitat and fish community at the reach scale and use this information to further determine relationships between physicochemical characteristics and fish distributions.
- 4) Use experimental and field-based research to determine how and under what conditions nonnative fishes may exclude native fishes from a stream reach.
- 5) Determine seasonal migrations (e.g., direction, frequency) of prairie fishes in Montana.
- 6) Examine effects of flooding and drying cycles (e.g. frequency, duration) on prairie fish assemblages. Determine rates and spatial patterns of local extirpations and recolonization for prairie fishes.
- 7) Examine the effects of reach-scale habitat improvements on the structure of prairie fish assemblages.

LITERATURE CITED

- Alden, W. C. 1932. Physiography and glacial ecology of eastern Montana and adjacent areas. United States Department of Interior Professional Paper 174, Washington, D.C.
- 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.
- Anderson, R. J. 1973. Water and eastern Montana coal development. Montana Environmental Quality Council, Helena.
- Angermeier, P. L., and M. R. Winston. 1998. Local vs. watershed influences on local diversity in stream fish communities of Virginia. *Ecology* 79:911-927.
- Angermeier, P. L., K. L. Krueger, and C. A. Dolloff. 2002. Discontinuity in stream-fish distributions: implications for assessing and predicting species occurrence. Pages 519-527 *in* J. M. Scott, P. J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael, and B. Wall, editors. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Covelo, California.
- Argent, D. G., J. A. Bishop, J. R. Stauffer, Jr., R. F. Carline, and W. L. Myers. 2003. Predicting freshwater fish distributions using landscape-level variables. *Fisheries Research* 60:17-32.
- Bailey, R. M., and M. O. Allum. 1962. *Fishes of South Dakota*, University of Michigan, Ann Arbor.
- Benham-Blair and Affiliates, Inc. 1976. Environmental assessment report, Sooner Generating Station Units Number 1 and Number 2, Oklahoma Gas and Electric Company. Attachment 2.5.1-1, aquatic field studies, 1975-1976. Benham-Blair and Affiliates, Inc., Oklahoma City, Oklahoma.
- Berra, T. M. 2001. *Freshwater fish distributions*. Academic Press, San Diego.
- Bramblett, R. G., and A.V. Zale. 2000. The ichthyofauna of small streams on the Charles M. Russell National Wildlife Refuge, Montana. *Intermountain Journal of Sciences* 6:57-67.
- Bramblett, R. G., M. R. Jones-Wuellner, C. S. Guy, and A. V. Zale. 2003. Montana prairie riparian native species study. Montana Cooperative Fishery Research Unit, Bozeman.

- Bramblett, R. G., T. R. Johnson, A. V. Zale, D. G. Heggem. 2005. Development and evaluation of a fish assemblage index of biotic integrity for Northwestern Great Plains stream. *Transactions of the American Fisheries Society* 134:624-640.
- Breiman, L., J. H. Friedman, A. Olshen, and C. G. Stone. 1984. *Classification and regression trees*. Wadsworth International Group, Belmont, California.
- Brown, C. J. D. 1971. *Fishes of Montana*. Big Sky Books, Montana State University, Bozeman.
- Cross, F. B., R. L. Mayden, and J. D. Stewart. 1986. Fishes in the western Mississippi basin (Missouri, Arkansas, and Red rivers). Pages 363-412 *in* C. H. Hocutt and E. O. Wiley, editors. *Ecology and conservation of Great Plains vertebrates*. J. Wiley and Sons, New York City.
- Darlington, P. J. 1957. *Zoogeography: the geographical distribution of animals*. John Wiley and Sons, New York City.
- Dodds, W. K., K. Gido, M. R. Whiles, K. M. Fritz, and W. J. Matthews. 2004. Life on the edge: the ecology of Great Plains prairie streams. *BioScience* 54:205-216.
- Edwards, E. A., H. Li, C. B. Schreck, J. W. Terrell, and R. F. Raleigh. 1983. Habitat suitability index models: longnose dace. FWS/OBS-82/10.33. U.S. Fish and Wildlife Service, Washington, D.C.
- Environmental Systems Research Institute, Inc. (ESRI). 2004. *ArcGIS Desktop*, version 9.0. Redlands, California.
- Fausch, K. D., and R. G. Bramblett. 1991. Disturbance and fish communities in intermittent tributaries of a western Great Plains river. *Copeia* 1991:659-674.
- Fausch, K. D., S. Nakano, and K. Ishigaki. 1994. Distribution of two congeneric charrs in streams of Hokkaido Island, Japan: considering multiple factors across scales. *Oecologia* 100:1-12.
- Fausch, K. D., and K. R. Bestgen. 1997. Ecology of fishes indigenous to the central and southwestern Great Plains. Pages 131-166 *in* F. L. Knopf and F. B. Samson, editors. *Ecology and conservation of Great Plains vertebrates*. Springer-Verlag, New York City.
- 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 fishes. *BioScience* 52:483-498.

- Filipe, A. F., I. G. Cowx, and M. J. Collares-Pereira. 2002. Spatial modeling of freshwater fish in semi-arid river systems: a tool for conservation. *River Research and Applications* 18:123-136.
- 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.
- Gido, K. B., J. A. Falke, R. M. Oakes, and K.J. Hase. 2006. Fish-habitat relations across spatial scales in prairie streams. *American Fisheries Society Symposium* 48:265-285.
- Gorman, O. T., and J. R. Karr. 1978. Habitat structure and stream fish communities. *Ecology* 59:507-515.
- Grossman, G. D., R. E. Fatajczak, Jr., M. Crawford, and M. C. Freeman. 1998. Assemblage organization in stream fishes: effects of environmental variation and interspecific interactions. *Ecological Monographs* 68:395-420.
- Grossman, G. D., P. B. Moyle, and J. O. Whittaker, Jr. 1982. Stochasticity in structural and functional characteristics of an Indiana stream fish assemblage: a test of community theory. *The American Naturalist* 120: 423-454.
- Guisan, A., and N. E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Haslouer, S. G., M. E. Eberle, D. E. Edds, K. B. Gido, C. S. Mammoliti, J. R. Triplet, J. T. Collins, D. A. Distler, D. G. Huggins, and W. J. Stark. 2005. Current status of native fish species in Kansas. *Transactions of the Kansas Academy of Science* 108: 32-46.
- Hawkes, C. L., D. L. Miller, and W. G. Layher. 1986. Fish ecoregions of Kansas: stream fish assemblage patterns and associated environmental correlates. *Environmental Biology of Fishes* 17:267-279.
- Hoagstrom, C. W. and C. R. Berry, Jr. 2006. Island biogeography of native fish faunas among Great Plains drainage basins: basin scale features influence composition. *American Fisheries Society Symposium* 48:221-264.
- Holton, G. D., and H. E. Johnson. 2003. A field guide to Montana fishes, 3rd edition. *Montana Fish, Wildlife and Parks*, Helena.
- Hubbell, S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey.

- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299-314.
- Jackson, D. A., and H. H. Harvey. 1989. Biogeographic associations in fish assemblages: local vs. watershed processes. *Ecology* 70:1472-1484.
- Jackson, D. A., P. R. Peres-Neto, and J. D. Olden. 2001. What controls who is where in freshwater communities—roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences* 58:157-170.
- Johnson, M. G., J. H. Leach, C. K. Minns, and C. H. Oliver. 1977. Limnological characteristics of Ontario lakes in relation to associations of walleye (*Stizostedion vitreum vitreum*), northern pike (*Esox lucius*), lake trout (*Salvelinus namaycush*), and smallmouth bass (*Micropterus dolomieu*). *Journal of the Fisheries Research Board of Canada* 34:1592-1601.
- Jongman, R. H. G., C. J. F. Ter Braak, O. F. R. Van Tongeren. 1995. *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, United Kingdom.
- Joy, M. K., and R. G. Death. 2004. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages: an integrated GIS and neural network approach. *Freshwater Biology* 49:1036–1052.
- Karr, J. R., K. D., Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. *Illinois Natural History Survey Special. Publication, Number 5*.
- Kaufman, P. R., and E. G. Robison. 1998. Physical habitat characterization. Pages 77-118 in Lazorchak, J. M., D. J. Klemm, and D. V. Peck, editors. *Environmental monitoring and assessment program-surface waters: field operations and methods for measuring the ecological condition of wadable streams*. U. S. Environmental Protection Agency, 620/R-94/004F, Washington, D.C.
- Kaufman, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Quantifying physical habitat in wadeable streams. U. S. Environmental Protection Agency, 620/R-99/003, Washington, D.C.
- Labbe, T. R., and K. D. Fausch. 2000. Dynamics of intermittent stream habitat regulate persistence of a threatened fish at multiple scales. *Ecological Applications* 10:1774-1791.

- Lanka, R. P., W. A. Hubert, and T. A. Wesche. 1987. Relations of geomorphology to instream habitat and trout standing stock in small Wyoming streams. *Transactions of the American Fisheries Society* 116:109-117.
- Lees, B. G. and Ritman, K. 1991. Decision-tree and rule induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management* 15:823-831.
- Leonard, P. M., and D. J. Orth. 1988. Use of habitat guilds of fishes to determine instream flow requirements. *North American Journal of Fisheries Management* 8:399-409.
- Li, H. W., and P. B. Moyle. 1981. Ecological analysis of species introductions into aquatic systems. *Transactions of the American Fisheries Society* 110:772-783.
- Lohr, S. C., and K. D. Fausch. 1996. Effects of green sunfish (*Lepomis cyanellus*) predation on survival and habitat use of plains killifish (*Fundulus zebrinus*). *Southwestern Naturalist* 41:155-160.
- Lyons, J. 1989. Correspondence between the distribution of fish assemblages in Wisconsin streams and Omernik's ecoregions. *The American Midland Naturalist* 122:163-182.
- MacArthur, R. H., and Wilson, E. O. 1967. *The Theory of Island Biogeography*. Princeton University Press.
- Mandrak, N.E. 1995. Biogeographic patterns of fish species richness in Ontario lakes in relation to historical and environmental factors. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 1462-1474.
- 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.
- Marsh-Matthews, E., and W. J. Matthews. 2000. Spatial variation in relative abundance of a widespread, numerically-dominant fish species and its effect on fish assemblage structure. *Oecologia* 125:283-292.
- Matthews, W. J. 1987. Physicochemical tolerance and selectivity of stream fishes as related to their geographic ranges and local distributions. Pages 1111-120 *in* W.J. Matthews and D.C. Heins, editors. *Community and evolutionary ecology on North American stream fishes*. University of Oklahoma Press, Norman.

- Matthews, W. J. 1988. North American prairie streams as systems for ecological study. *Journal of the North American Benthological Society* 7:387-409.
- Matthews, W. J. 1998. *Patterns in freshwater fish ecology*. Chapman and Hall, New York.
- Matthews, W. J., and H.W. Robison. 1988. The distribution of fishes of Arkansas: a multivariate analysis. *Copeia* 1988:358-374.
- McCormick, F. H. and R. M. Hughes. 1998. Physical habitat characterization. Pages 161-182 *in* Lazorchak, J. M., D. J. Klemm, and D. V. Peck, editors. Environmental monitoring and assessment program-surface waters: field operations and methods for measuring the ecological condition of wadable streams. U. S. Environmental Protection Agency, 620/R-94/004F, Washington, D.C.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. Chapman and Hall, London.
- McMahon, T. E. 1982. Habitat suitability index models: creek chub. FWS/OBS-82/10.4. U.S. Fish and Wildlife Service, Washington, D.C.
- McPhail, J. D., and C. C. Lindsay. 1970. *Freshwater fishes of northwestern Canada and Alaska*. Fisheries Research Board Canada, Bulletin 173, Ottawa.
- Metcalf, A. L. 1966. Fishes of the Kansas River system in relation to zoogeography of the Great Plains. University of Kansas Museum of Natural History Publication 17:23-189.
- Microsoft Corporation. 2003. *Microsoft Office Access SP1*. Microsoft Office Professional Edition.
- Modde, T., R. C. Ford, and M. G. Parsons. 1991. Use of habitat-based stream classification system for categorizing trout biomass. *North American Journal of Fisheries Management* 11:305-311.
- Oakes, R. M., K. B. Gido, J. A. Falke, J. D. Olden, and B. L. Brock. 2005. Modelling of stream fishes in the Great Plains, USA. *Ecology of Freshwater Fish* 14:1-14.
- Olden, J. D. and D. A. Jackson. 2002. A comparison of statistical approaches for modeling fish species distributions. *Freshwater Biology* 47:1976-1995.
- Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77:118-125.

- Ostle, W. R., R. E. Schneider, J. M. Aldrich, T. M. Faust, R. L. B. McKim, and S. J. Chaplin. 1997. The status of biodiversity in the Great Plains. The Nature Conservancy, Arlington, Virginia.
- Ostrand, K. G., and G. R. Wilde. 2001. Temperature, dissolved oxygen, and salinity tolerances of five prairie stream fishes and their role in explaining fish assemblage patterns. *Transactions of the American Fisheries Society*, 130:742-749.
- Ostrand, K. G., and G. R. Wilde. 2002. Seasonal and spatial variation in a prairie stream-fish assemblage. *Ecology of Freshwater Fish* 11:137-149.
- Ostrand, K. G., and G. R. Wilde. 2004. Changes in prairie stream fish assemblages restricted to isolated streambed pools. *Transactions of the American Fisheries Society* 133:1329-1338.
- Patton, T. M., W. A. Hubert, F. J. Rahel, and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains of Wyoming. *North American Journal of Fisheries Management* 20:394-398.
- Pflieger, W. L. 1971. A distributional study of Missouri fishes. University of Kansas Museum of Natural History, Publication 20:225-570.
- Pflieger, W. L. 1997. The fishes of Missouri. Missouri Department of Conservation, Jefferson City.
- 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.
- Quist, M. C., F. J. Rahel, and W. A. Hubert. 2005. Hierarchical faunal filters: an approach to assessing effects of habitat and nonnative species on native fishes. *Ecology of Freshwater Fish* 14:24-39.
- R Foundation for Statistical Computing. 2004. R 2.0.1. Available: <http://www.r-project.org>. Accessed: 1/15/2005.
- Rabeni, C. F., and S. P. Sowa. 1996. Integrating biological realism into habitat restoration and conservation strategies for small streams. *Canadian Journal of Fisheries and Aquatic Sciences* 53:252-259.
- Rahel, F. J. and W. A. Hubert. 1991. Fish assemblages and habitat gradients in a Rocky Mountain-Great Plains stream: biotic zonation and additive patterns of community change. *Transactions of the American Fisheries Society* 12:319-332.

- Rashleigh, B. 2004. Relation of fish community structure to environmental characteristics in the upper French Broad River basin, North Carolina. *Environmental Monitoring and Assessment* 93:139-156.
- Ritter, D. F., R. C. Kochel, and J. R. Miller. 1995. *Process geomorphology*, 3rd edition. William C. Brown Publishers, Dubuque, Iowa.
- Roth, N. E. J. D. Allan, and D. L. Erickson. 1996. Landscape influence on stream biotic integrity assessed at multiple scales. *Landscape Ecology* 11:141-156.
- Samson, F. B., and F. L. Knopf. 1994. Prairie conservation in North America. *BioScience* 44:418-421.
- Schlosser, I. J. 1987. A conceptual framework for fish communities in small warmwater streams. Pages 17-24 *in* W. J. Matthews and D. C. Heins, editors. *Community and Evolutionary Ecology of North American Stream Fishes*. University of Oklahoma Press, Norman, Oklahoma.
- Schlosser, I. J. 1991. Stream fish ecology: a landscape perspective. *BioScience* 41:704-712.
- Sigler, W. F. and R. R. Miller. 1963. *Fishes of Utah*. Utah Department of Fish and Game, Salt Lake City.
- Slagle, S. E. 1984. Hydrology of area 45, northern Great Plains and Rocky Mountain coal provinces, Montana and North Dakota. U.S. Geological Survey, Helena.
- 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.
- Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysicist Union* 38:913-920.
- Tonn, W. M., and J. J. Magnuson. 1982. Patterns in the species composition and richness of fish assemblages in northern Wisconsin lakes. *Ecology* 63:1149-1166.
- Tonn, W. M., J. J. Magnuson, and A. M. Forbes. 1983. Community analysis in fishery management: an application with northern Wisconsin lakes. *Transactions of the American Fisheries Society* 112:368-377.
- Vannote R. L., G. W. Minshall, K.W. Cummins, K.R. Sedell, and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130-137.

- Vayssières, M. P., R. E. Plant, B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11:679-694.
- Wall, S. S., C. R. Berry Jr., C. M. Blausey, J. A. Jenks, and C. J. Kopplin. 2004. Fish-habitat modeling for gap analysis to conserve the endangered Topeka shiner (*Notropis topeka*). *Canadian Journal of Fisheries and Aquatic Sciences* 61:954-973.
- Wang, L., J. Lyons, and P. Kanehl. 2001. Impacts of urbanization on stream habitat and fish across multiple scales. *Environmental Management* 28:255-266.
- Welker, T. L., and D. L. Scarnecchia. 2004. Habitat use and population structure of four native minnows (family Cyprinidae) in the upper Missouri and lower Yellowstone Rivers, North Dakota (USA). *Ecology of Freshwater Fish* 13:8-22.
- Whittier, T. R., R. M. Hughes, and D. P. Larsen. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. *Canadian Journal of Fisheries and Aquatic Sciences* 45:1264-1278.
- Wildlife Spatial Analysis Lab (WSAL). 2003. Montana Land Cover Type (Map). University of Montana, Missoula.
- Zale, A. V., D. M. Leslie, W. L. Fisher, and S. G. Merrifield. 1989. The physiochemistry, flora, and fauna of intermittent prairie streams: a review of the literature. U. S. Fish and Wildlife Service, Biological Report 89(5), Washington, D.C.
- 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.

APPENDIX A
GIS METHODOLOGY AND DOCUMENTATION

I. Working with Stream Layers:

A. I chose a stream network [National Hydrography Dataset (NHD)] that had topology (e.g., direction and connection) and density sufficient for this project (e.g., all sample locations fell along the stream network).

B. Visual analysis showed varying densities of streams within one-degree quadrangles. I wanted all quadrangles to have similar densities of streams; thus, I had to censor smaller streams. To do so, I classified quadrangles of interest by a density classification in the attribute table of the quadrangle layer (1 = low density; 2 = medium/average density; 3 = high density).

C. I created a new shapefile called “Megawatersheds.” In this layer, I created watersheds upstream from where the confluence of major rivers or streams (i.e., Milk, Marias, Bighorn, etc.) flowed into either the Missouri or Yellowstone rivers. Watersheds were created by assigning fifth- or sixth-code hydrologic unit codes (HUCs) within a given megawatershed the same unique identifier (see directions on delineating watersheds). Borders between HUCs with the same identifier were dissolved, creating large watersheds. The NHD stream layer was clipped to these megawatersheds to make assigning stream order computationally efficient and faster.

D. Strahler (1957) stream orders were assigned using an extension provided by the Environmental Systems Research Institute (ESRI) website (www.esri.com). Looking at the distribution of stream lengths and orders within quadrangles classified as density “3”, I discovered that first order streams appeared to be

overly abundant. I decided to try eliminating first order streams of less than a given length until density “3” quads appeared to look more like density “2” quads. To perform this task, I first joined the classified quadrangle layer to the assigned stream network. Next, I created a dummy field in the attribute table of the joined quad classification-stream network layer called “Theme 1”; in this column, I assigned all density “3” quads with a “1” to select and change the attributes of only these quads. Within these quads, I chose a stream winnowing start point (i.e., first order streams of < 1000 m in length) and assigned them a blank value in the field “Theme1.” I themed the joined layer to show only the streams classified as “1” in “Theme1” and visually analyzed this theme to see if it met our density needs. I continued to eliminate other stream lengths of first order (e.g. < 5000 m, <2500 m, <3500m) until the density of streams in density “3” quads closely matched the density of streams in density “2” quads.

E. After these first order streams were eliminated from the NHD layer, a new shapefile was created that contained our kept streams. Stream orders needed to be reassigned for this new shapefile. However, I first examined the new stream layer for anomalies (i.e., misdirected flow, disconnected streams, irrigation channels). These anomalies were coded and fixed using the appropriate tools in order to assure that stream orders were assigned correctly.

II. Delineating Watersheds:

A. I decided to use fifth- and sixth-code HUC's rather than digital elevation models (DEMs) to delineate watersheds upstream of our sample sites for three reasons:

- 1) Using DEMs would require a large database and, thus, very precise data management.
- 2) Little relief exists in eastern Montana to allow for accurate watershed delineation.
- 3) Using DEMs is computationally challenging overall.

B. The smallest HUCs that exist for Montana are sixth-order. However, at the time of analysis, the sixth-code HUCs were not complete. Blank areas existed near the Musselshell and Missouri rivers. For these missing HUCs, we used the fifth-code HUCs.

C. I utilized the polygon tool to assist in determining the shape of the watershed upstream of the sample site and to select which HUCs would be included in the watershed.

D. The selected HUCs were assigned a unique identifier in a created attribute column (i.e., "WaterNumb") that would match the site number to which they drained.

E. I dissolved the borders between the polygons that had the same unique identifier to result in a single polygon that represented the watershed upstream of the sampling site.

F. The HUCs often included some area upstream of the sample site. In these cases, I would use the “Cut Polygons Features” editing tool to eliminate any areas of our dissolved watersheds that fell upstream of the sampling site.

G. When a smaller watershed drained within a larger watershed, the polygon was given two or more unique identifiers. The first ID indicated the largest watershed into which the HUC drained. The second (or third or fourth) ID indicated decreasingly smaller watersheds until the HUC was assigned the ID of the smallest watershed.

III. Analyses Involving Calculations of Point Layer Values Within a Watershed:

A. For analyses involving point layers (e.g., the number of dams and the number of groundwater wells within the watershed), I first began by joining the watershed layers to the point layer of interest.

B. Once the layers were joined, I right-clicked on the unique watershed identifier field and chose the command *Summarize*. To obtain basic counts, I could have selected any variable and any summary of that variable. For this project, I consistently used the minimum of longitude of the points within the watershed. A count summary always followed.

IV. Analyses Involving Calculations of Polyline Layer Values Within a Watershed:

A. For analyses involving polylines (e.g., the total length of road or irrigated channels within a watershed), I followed the join command as described above for point calculations.

B. Once the layers were joined, I right-clicked on the unique watershed identifier field and chose the command *Summarize*. To obtain total lengths, I would choose to sum on the line segment length field. To calculate densities, I would simply divide the total length of the given parameter within a watershed by the total area of the watershed.

V. Analyses Involving Calculations of Polygon Layer Values Within a Watershed:

A. For analyses involving polygons (i.e., volume of potential precipitation falling within the watershed, percent land use within the watershed), I would perform a union between the watershed layer to the polygon feature of interest.

B. Once the union was completed, I would create a new field within the unionized shapefile according to the calculation at hand (e.g., area, volume).

C. When calculating an area, I would perform a search in the Help Menu under “making field calculations.” The result (usually the eighth one on the list) provided the code to calculate individual unionized polygon areas. This code was implemented by right-clicking on the newly created field, choosing *Calculate Values*, and copying and pasting the code in the appropriate spaces.

D. If a volume was to be calculated, I had to ensure the multipliers were the appropriate units and converted if needed. Volumes were calculated in a new field by using the *Calculate Values* feature described above.